

# A Comparison of Multilocus Clines Maintained by Environmental Adaptation or by Selection Against Hybrids

L. E. B. Kruuk,\* S. J. E. Baird,<sup>†</sup> K. S. Gale\* and N. H. Barton\*

\**Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom* and <sup>†</sup>*Department of Biology, Indiana University, Bloomington, Indiana 47405*

Manuscript received April 13, 1999  
Accepted for publication August 4, 1999

## ABSTRACT

There has recently been considerable debate over the relative importance of selection against hybrids ("endogenous" selection) vs. adaptation to different environments ("exogenous") in maintaining stable hybrid zones and hence in speciation. Single-locus models of endogenous and exogenous viability selection generate clines of similar shape, but the comparison has not been extended to multilocus systems, which are both quantitatively and qualitatively very different from the single-locus case. Here we develop an analytical multilocus model of differential adaptation across an environmental transition and compare it to previous heterozygote disadvantage models. We show that the shape of clines generated by exogenous selection is indistinguishable from that generated by endogenous selection. A stochastic simulation model is used to test the robustness of the analytical description to the effects of drift and strong selection, and confirms the prediction that pairwise linkage disequilibria are predominantly generated by migration. However, although analytical predictions for the width of clines maintained by heterozygote disadvantage fit well with the simulation results, those for environmental adaptation are consistently too narrow; reasons for the discrepancy are discussed. There is a smooth transition between a system in which a set of loci effectively act independently of each other and one in which they act as a single nonrecombining unit.

**S**PECIATION occurs when selection generates barriers to gene flow between two populations. The structure of these barriers to gene flow can be investigated in cases where they are not entirely effective: that is, in hybrid zones where the effect of natural selection in maintaining the integrity of the parental genomes is undermined by the homogenizing effects of migration and recombination, generating a set of clines (Barton and Hewitt 1985). Such intermediate stages—systems that endure despite having neither achieved nor abandoned reproductive isolation—provide indirect means of observing the mechanisms preventing gene flow and hence of elucidating the components of speciation. Hybrid zones therefore offer a "window on the evolutionary process" (Harrison 1990). In this article, we use theoretical models of hybrid zones to describe and compare the effects of alternative types of selection, specifically adaptation to different environments or hybrid unfitness, in maintaining divergence.

Hybrid zones have been described in a wide range of organisms (for reviews, see Barton and Hewitt 1981b, 1985; Hewitt 1988; Harrison 1990, 1993; Arnold 1997). They usually involve divergence at many loci and can often be described as a set of parallel clines in gene

frequency and morphological traits, which are maintained by a balance between dispersal and selection. Earlier theoretical descriptions of such clines concentrated on single- (or at most two-) locus models, with continuous approximations for time and space, and gene flow described by diffusion (Fisher 1937, 1950; Haldane 1948; Bazykin 1969, 1973; Slatkin 1973, 1975; Nagylaki 1975, 1976a; Endler 1977). However, if clines at multiple loci coincide, as is the case in many hybrid zones, associations between loci, or linkage disequilibria, will be generated by the mixing of populations with different allele frequencies (Li and Nei 1974). Selection on one gene spills over onto any genes that are in linkage disequilibria with it, increasing the effective selection on each locus and generating a sharp step in gene frequency. Single-locus models cannot adequately represent such a system: consider, for example, the dynamics of a neutral locus linked to a set of selected loci, the effect of which would be lost in a single-locus model. The extension of the above models to clines at multiple loci offers quantitative estimates of the magnitude of natural selection (Barton and Hewitt 1983; Barton 1986; Barton and Bengtsson 1986). For example, measurements of the shape of clines in genetic markers and the statistical associations between them can be used to estimate the strength of the barrier to gene flow, and even the number of genes under selection, as well as dispersal rates. These methods have been applied to a variety of hybrid zones [Bombina (Szymura and Barton 1986, 1991); *Heliconius* (Mallet *et al.*

*Corresponding author:* Loeske Kruuk, Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Bldgs., W. Mains Rd., Edinburgh EH9 3JT, United Kingdom.  
E-mail: loeske.kruuk@ed.ac.uk

1990); Podisma (Barton 1980; Barton and Hewitt 1981a; Jackson 1992); Pontia (Porter *et al.* 1998); Sceloporus (Sites *et al.* 1995); and Uroderma (Barton 1982)].

Various different forms of clines can be distinguished. Dispersal may be negligible, in which case selection maintains a stable equilibrium at each locality (p. 116, Barton and Hewitt 1985). More often, as considered in the diffusion models cited above, clines are maintained in a balance between dispersal and some form of selection. The selection that acts to maintain ancestral divergence can be classified according to whether it maintains different alleles in different environments or in different genetic backgrounds. After Moore and Price (1993), these scenarios can be termed, respectively, (i) “exogenous” selection, with fitness defined in relation to the environment and a geographic selection gradient determining the relative fitness of genotypes, and (ii) “endogenous” selection, with fitness defined by within-genome interactions such as heterozygote disadvantage or epistasis, and being independent of environment. With purely exogenous selection, each allele has an effect on fitness that depends on location and not on the other alleles with which it is combined (*i.e.*, no dominance or epistasis). With purely endogenous selection, fitness is independent of location, though fitness differences may of course still be mediated by interactions with the environment (for example, by different abilities to escape predators).

In reality, natural hybrid zones may well include both kinds of selection. Nevertheless, the relative predominance of either type has been the subject of considerable debate in the hybrid zone literature. Reviews of studies of hybrid zones in >100 taxa (Barton and Hewitt 1981b, 1985) concluded that most are maintained by selection against hybrids independent of the environment, *i.e.*, they are “tension zones” (Key 1968). With this justification, Barton’s (1983, 1986) multilocus models invoked endogenous selection, specifically heterozygote disadvantage; simulations showed that the effects of epistatic endogenous selection were qualitatively similar (Barton and Gale 1993; Barton and Shpak 1999). However, there is also substantial evidence of an important role for exogenous selection in hybrid zones (Arnold 1997; Harrison 1990). In single-locus models, the shape of the clines generated by exogenous *vs.* endogenous viability selection is effectively indistinguishable (Barton and Hewitt 1989; Barton and Gale 1993), but it is not clear whether the similarity extends to multiple-locus systems. Here, we develop an analogous treatment to Barton’s earlier model under the alternative selection regime: exogenous selection. We consider the simplest case of two parapatric populations adapted to different environments (Haldane 1948). The shape of clines generated by the two forms of selection, their effects on linkage disequilibria, and the transition from a situation in which multiple loci act independently of

each other to that in which the linkage disequilibria are strong enough to bind alleles of shared ancestry together are compared. For simplicity, we consider throughout selection pressures defined for individual genotypes, as would be the case for viability selection—see Gavrillets (1997b) for an illustration of the extra complexity introduced by fertility selection in even a single-locus model.

In this analytical treatment of clines maintained in a dispersal-selection balance, we rely on a key assumption that selection is weak. The analysis then becomes tractable as higher-order terms in  $s$  (the selection coefficient) and linkage disequilibria of third and higher order can be treated as negligible. We develop a simulation model to test the robustness of the analytical descriptions and their assumptions as selection becomes stronger. Finite population sizes in the simulation incorporate the effects of drift, necessarily ignored in the deterministic analytical models but known to widen clines (Slatkin and Maruyama 1975). We also investigate the assumption that linkage disequilibria are generated primarily by migration (and hence can yield estimates of gene flow) and test whether the shapes of clines observed in simulations match those predicted by analytical models of both environmental selection and heterozygote disadvantage.

## SIMULATION METHODS

**Simulation model:** The simulation model is based on that used by Barton and Gale (1993; Jackson 1992), extended to allow selection on diploids rather than haploids. A one-dimensional stepping-stone model is used to simulate contact between two populations initially fixed for two alternative alleles at a set of loci. We refer to the two types of alleles as 0, from the left-hand population, or 1, from the right. The population is distributed along a chain of demes, each of which contains  $N$  diploid individuals. Migration occurs between adjacent demes, with a fixed proportion  $m$  of individuals moving per generation, half in either direction; migrants are chosen at random. With this model of nearest-neighbor migration, the migration rate  $m$  is identical to the variance of distance moved per generation along the cline axis,  $\sigma^2$  (distance being measured in deme spacings).

Migration is followed by reproduction. Recombination occurs at a rate  $r$  between adjacent loci, with chromosome positions determined randomly and independently (*i.e.*, no interference); unlinked loci are represented by setting  $r = 0.5$ . Within each deme, individuals then mate randomly, and each pair produces two offspring. The new adult population of  $N$  individuals is chosen by sampling with replacement from the offspring population; the probability of an individual being sampled is determined by its fitness as defined under the particular selection regime (see below).

End demes remain fixed in gene frequency, to simu-

late an infinite pure population on either side of the cline (after Feldman and Christiansen 1975), with sufficient numbers of demes being used to ensure that end effects do not disrupt the dynamics. Statistics are calculated after migration and reproduction at 200 generation intervals between generations 600 and 3000 of each run. The results were checked to ensure that (i) the system had reached equilibrium; (ii) these intervals were long enough to avoid autocorrelation, so values could be treated as independent data points. The 95% confidence intervals presented are twice the standard errors. The model was tested to check that it was performing as expected under the random processes, for example, with respect to the decay of linkage disequilibrium in the absence of selection or migration but incorporating the effects of drift (Hill and Robertson 1966), or with respect to the decay with time of a step cline in the frequency of a neutral allele.

**Selection regimes:** Let  $L$  be the number of loci,  $s$  the selection coefficient, and  $x$  the distance across the cline, with  $x = 0$  defining the center. For each individual, the number of heterozygous loci it carries is given by  $H$ , and the number of alleles originating from the population to the right of the environmental step (those of type 1) by a "hybrid index"  $Z$ . The individual fitnesses for the two possible selection regimes considered are then defined as

Environmental adaptation:

$$W = \begin{cases} e^{-Zs}, & x < 0, \\ e^{-(2L-Z)s}, & x \geq 0. \end{cases} \quad Z = 0 \dots 2L; \quad (1a)$$

Heterozygote disadvantage:

$$W = e^{-Hs}, \quad H = 0 \dots L. \quad (1b)$$

**Statistical methods:** *Maximum gradient and cline width:* We estimate the gradient of a cline by first transforming the gene frequency  $p$  to a logit scale,  $y = \ln(p/(1 - p))$ . On this scale, a single-locus cline generated by such viability selection should appear approximately linear with distance (Barton and Hewitt 1989). The gradient in allele frequency at the center (the maximum gradient) is then estimated by regression between the points at which  $y = -2$  and  $y = +2$ . The gradient on the logit scale ( $dy/dx$ ) is exactly four times the gradient of allele frequency ( $dp/dx$ ) at the center (where  $p = 0.5$ ), and the width of the cline is defined as the inverse of its maximum gradient (Slatkin 1973). In the simulation runs, the width of the cline at each locus is calculated individually, and all widths are then averaged to give a mean value.

*Linkage disequilibrium:* The average pairwise linkage disequilibrium across all loci  $D$  is estimated using the variance in a hybrid index (see Barton and Gale 1993). Define  $X_i$  as the number of type 1 alleles at the  $i$ th locus ( $X_i = 0, 1, \text{ or } 2$ ) and the hybrid index  $Z$ , as above, as the number of type 1 alleles an individual possesses out

of the total possible  $2L$ , so that  $Z = \sum_{i=1}^L X_i$ . The pairwise linkage disequilibrium  $D_{ij}$  between loci  $i$  and  $j$  is given by the covariance between  $X_i$  and  $X_j$  (Hartl and Clark 1989), so the variance in  $Z$  is

$$\begin{aligned} \text{Var}(Z) &= \sum_{i,j=1}^L \text{cov}(X_i, X_j) = 2 \left( \sum_{i=1}^L p_i q_i + \sum_{i \neq j} D_{ij} \right) \\ \Rightarrow \bar{D} &= \frac{1}{2L(L-1)} \left( \text{Var}(Z) - 2 \sum_{i=1}^L p_i q_i \right), \quad (2) \end{aligned}$$

where  $p_i$  is the frequency of type 1 alleles at locus  $i$  ( $q_i = 1 - p_i$ ) and  $\bar{D}$  is the pairwise linkage disequilibria averaged across all loci. In general, the variance in  $Z$  will be inflated by any heterozygote deficit. This is because the possibility of autozygosity causes correlations between homologous genes at one locus in addition to the correlations between loci measured by the linkage disequilibrium (so  $\text{Var}(X_i) > p_i q_i$ ; Bulmer 1985). Here, however, statistics are measured immediately after random mating, so that the genotype frequencies will be in Hardy-Weinberg proportions and values of linkage disequilibrium estimated using this method will not be inflated.

The maximum value of linkage disequilibrium should occur at the center of the cline, where  $p = 0.5$ . However, if the cline is steep, there may be no single deme at  $p = 0.5$ . Rather than using the maximum  $D$  or the value of  $D$  nearest the center observed in the simulations, we use a linear regression of  $D$  on  $(pq)^2$  to give the value at  $p = 0.5$ . This approach relies on the fact that single-locus clines follow a tanh curve, the gradient of which will be proportional to  $pq$ , so  $D$  at any point should be proportional to the square of the gradient at that point (see below). However, with this method it is possible for impossibly high values to be obtained ( $D > 0.25$ ), as cline shape departs from the simple tanh curve when selection is strong.

### SINGLE-LOCUS CLINES

Consider the simple case of a cline at a single locus that is segregating for two alleles. Two populations fixed for alternate alleles are in contact and interbreed at their interface. If selection is weak, the change in gene frequency in one generation due to selection will be small, allowing a continuous time approximation for the effect of selection. Neglecting the second-order interaction between dispersal and selection, the rate of change in gene frequency  $p$  with respect to time  $t$  will be determined by the sum of the effects of dispersal (from Fisher 1937) and selection (from Wright 1931),

$$\frac{\partial p}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} + \frac{1}{2} \frac{\partial \log(\bar{W}(p))}{\partial p} pq \quad (q = 1 - p), \quad (3)$$

where  $\sigma^2$  is the variance in distance moved per generation ( $\sigma^2 > 0$ ) and selection is defined by  $\bar{W}(p)$ , the mean fitness in a population at frequency  $p$ .

For a given selection function, Equation 3 can be solved at equilibrium ( $\partial p/\partial t = 0$ ) to give an explicit description of gene frequency as a function of distance across the cline (Haldane 1948; Bazykin 1969). The typical spatial scale of the resulting cline depends on the ratio of the dispersal distance and the strength of selection, the “characteristic length” of the cline (Slatkin 1973).

#### MULTILOCUS CLINES

**Selection acting on multiple loci:** Extending a single-locus system to multiple loci can have several effects. Epistasis will affect the marginal selection on each locus in a manner that changes across the cline. Second, statistical associations between loci (*i.e.*, linkage disequilibrium) will increase the effective selection on each locus. We concentrate here on this second component [see Gavrillets (1997a) and Barton and Shpak (1999) for treatment of epistasis in multilocus clines]. If linkage disequilibria are weak, the presence of other loci under selection will not affect the fate of a particular allele. At the other extreme, with very strong linkage disequilibria, a given allele will behave as if subject to the total selection pressure acting across the genome. To describe this variation, a measure of how “congealed” (Turner 1967) the genome is under the given selection pressures is required. Barton’s (1983) analytical results show that, for heterozygote disadvantage, the behavior of such a system will be defined by the relative strengths of the selection acting on each locus ( $s$ ), the rate of recombination ( $r$ ), and the total number of loci under selection ( $L$ ). For the sake of tractability, the analysis that follows makes the assumption that this selection is weak, and in particular that the ratio  $s/r$ , the “coupling coefficient,” is small (Barton 1983). We show that the dynamics of the system can still be profoundly different from those of single-locus clines and furthermore that the assumption of a low coupling coefficient need not be unduly restrictive.

Because of the statistical associations between loci, a list of the gene frequencies at each locus [derived from a series of equations like (3)] will not fully describe the system. Gametic frequencies would give the complete picture but soon become unwieldy if more than a few loci are considered as a system with  $L$  loci has  $2^L$  gametic frequencies to be tracked. An alternative method is to track individual gene frequencies and the statistical associations between alleles at different loci; the two approaches are interchangeable. Here, we make the assumption (which can be checked by comparison with the simulations) that under weak selection pressures the associations higher than pairwise become negligible. The latter method is then analytically considerably more tractable (Barton 1983).

In general, the frequency  $p$  of an allele at any point in the cline is determined by four factors: (i) migration;

(ii) selection acting directly on that locus; (iii) selection acting on gene combinations containing the locus in question (epistasis); and (iv) selection acting directly on other loci (or other combinations of loci) with which the locus is associated. The first two effects are incorporated in the single-locus model (Equation 3), and, as mentioned above, the effects of epistasis are not considered here. The fourth effect, that of the selection acting on associated genes, is determined by the degree of association between loci. For two loci, with alleles  $A/a$  at one locus and  $B/b$  at the other, the pairwise linkage disequilibrium  $D$  is defined as the excess of coupling gametes over that expected under random association, for example,

$$D = \text{frequency}(AB \text{ gametes}) - \text{freq}(A) \times \text{freq}(B).$$

Define  $p_1 = \text{freq}(A)$  and  $p_2 = \text{freq}(B)$ . As above, the effect of selection on the first locus on its gene frequency is determined by the variance in allelic states,  $p_i q_i$ , within the locus,

$$\frac{\partial p_1}{\partial t} = \frac{\partial \log \bar{W}}{\partial p_1} \frac{p_1 q_1}{2} \quad (4)$$

(Wright 1931). Similarly, the effect of selection on the first locus on the gene frequency at the *second* locus is defined by the covariance in allelic states between loci, namely the linkage disequilibrium,  $D$ ,

$$\frac{\partial p_2}{\partial t} = \frac{1}{2} \frac{\partial \log \bar{W}}{\partial p_1} D \quad (5)$$

(Lewontin and Kojima 1960; Barton and Turelli 1991).

The description can be extended to a system involving  $L > 2$  selected genes, where the allele frequency  $p_i$  at any one locus is also affected by selection on the other  $L - 1$  loci (each with frequencies  $p_j$ ). For the multiplicative model of selection used here, it can be assumed that higher-order disequilibria make a negligible contribution (Barton 1983). The equilibrium solution to first order in  $s$  is therefore given by

$$0 = \frac{\partial p_i}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 p_i}{\partial x^2} + \frac{p_i q_i}{2} \frac{\partial \log \bar{W}}{\partial p_i} + \frac{1}{2} \sum_{j \neq i}^L \frac{\partial \log \bar{W}}{\partial p_j} D_{ij} \quad (6)$$

**The magnitude of linkage disequilibria:** The dynamics of the pairwise linkage disequilibria can be described in the same way as those of gene frequencies, as the cumulative effects of diffusion, the mixing of populations at different gene frequencies (Li and Nei 1974), recombination, selection on the disequilibria themselves, and selection on other loci that are in higher-order associations with the pair in question. Under the assumption of weak selection and no epistasis, only the first three factors are relevant, giving

$$\frac{\partial D_{ij}}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 D}{\partial x^2} + \sigma^2 \frac{\partial p_i}{\partial x} \frac{\partial p_j}{\partial x} - r_{ij} D_{ij} \quad (7)$$

from Barton (1986). The first term, the change in  $D$  as it diffuses out from a maximum value, will be of order  $sD$  (Barton 1983). As demonstrated below,  $D$  is itself of order  $s$ , so for weak selection the first term is assumed to be negligible relative to the second term, the product of two gene frequency gradients each of order  $\sqrt{s}$ . Assuming that each  $D_{ij}$  can be represented by a mean value  $D$  and each  $r_{ij}$  by a harmonic mean  $r$ , (7) reduces to

$$\frac{\partial D}{\partial t} = -rD + \sigma^2 \frac{\partial p_i}{\partial x} \frac{\partial p_j}{\partial x}. \quad (8)$$

Finally, if the clines are concordant (of the same width  $w$ , defined as the inverse of the gradient at the center), the equilibrium solution for the maximum value of  $D$  is given by

$$D_{\max} = \frac{\sigma^2}{r w^2}. \quad (9)$$

The disequilibria are therefore determined by the relative magnitudes of recombination and migration and are not explicitly affected by the manner in which selection acts. Barton and Shpak (1999) show that even in the presence of epistasis, (9) is a good approximation to the magnitude of linkage disequilibrium.

The most questionable assumption in the derivation of (9) is that linkage disequilibrium changes continuously in time. If (as will often be the case) linkage is loose or absent ( $r = 0.5$ ), then linkage disequilibrium will change substantially between generations. To take account of this case, and to facilitate comparison with the simulation results, we describe in the appendix the discrete analogue of the diffusion model, in which genes are exchanged among three demes. This shows that Equation 9 is exactly equivalent to the three-deme discrete-time analogue at a quasi-equilibrium (when differences between demes in  $D$  are negligible relative to differences in  $p$ ; Nagylaki 1976b), measured after migration and reproduction. The magnitude of  $D$  in a three-deme system is also greater than the value predicted for exchange between two demes (Li and Nei 1974; Barton and Gale 1993).

The simulation model can now be used to test the analytical predictions. Figure 1 compares the maximum values of linkage disequilibria observed in simulations with the prediction in Equation 9. The results are presented for environmental selection (Figure 1, a-c) and heterozygote disadvantage (Figure 1, d-f), and for varying degrees of linkage. For unlinked and loosely linked loci, the simulation results fit very well with the predicted values for both fitness regimes. This is so even with strong selection, where fitness at the center is reduced by as much as 80%. However, the predictions break down for tightly linked loci, presumably because

the assumption that  $s/r$  is small is violated. Note also that the lower migration rates in the simulations presented in Figure 1, a and d, or the number of loci used, do not affect the correspondence. Thus, except where adjacent loci are tightly linked, linkage disequilibrium is determined primarily by a balance between dispersal and recombination. Barton and Gale (1993, Figures 2 and 3) found a similarly close agreement for models of endogenous selection that incorporated epistasis.

**The consequences of associations between genes:** With the description of linkage disequilibrium (Equation 9) confirmed by the simulation results, (6) can be simplified to give, at equilibrium,

$$0 = \frac{\sigma^2}{2} \frac{\partial^2 p_i}{\partial x^2} + \frac{p_i q_i}{2} \frac{\partial \log \bar{W}}{\partial p_i} + \frac{1}{2} \sum_{j \neq i}^L \frac{\partial \log \bar{W}}{\partial p_j} \left( \frac{\sigma^2}{r_{ij}} \frac{\partial p_i}{\partial x} \frac{\partial p_j}{\partial x} \right). \quad (10)$$

The extent to which the loci under selection behave as entirely independent units or, at the other extreme, as a single unit can now be considered. The analysis presented below considers environmental selection developed from the equivalent treatment of heterozygote disadvantage in Barton (1983) and Barton and Bengtsson (1986).

The mean fitness in a population under environmental selection is approximated as

$$\bar{W}(p) = \begin{cases} e^{-2\Sigma p_i s}, & x < 0, \\ e^{-2\Sigma(1-p_i)s}, & x \geq 0, \end{cases} \quad (11)$$

where  $p = (p_1, p_2, \dots, p_L)$ . The selection gradient  $\partial \log [\bar{W}] / \partial p_i$  is just  $-2s$  for  $x < 0$  and  $+2s$  for  $x \geq 0$ . Consider  $x < 0$  (for  $x \geq 0$ , replace  $p$  by  $q$  throughout). Assuming that clines in different loci coincide ( $p_i = p \forall i$ ) and using the harmonic mean recombination rate,  $r$ , Equation 10 becomes

$$0 = \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} - s p q - \sigma^2 \frac{(L-1)s}{r} \left( \frac{\partial p}{\partial x} \right)^2. \quad (12)$$

Define a "summed coupling coefficient,"  $\phi = (L-1)s/r$ . Using the equivalence  $\partial^2 p / \partial x^2 \equiv \frac{1}{2} \partial (\partial p / \partial x)^2 / \partial p$  gives a first-order differential equation in  $(\partial p / \partial x)^2$ , which can be integrated to give

$$\left( \frac{\partial p}{\partial x} \right)^2 = \frac{s}{8\sigma^2} \frac{(1 - e^{4\phi p} (1 - 2\phi)) + 2\phi(p - q) - 8\phi^2 p q}{\phi^3} \quad (13)$$

(given the boundary condition that the gradient is zero at the edges). The analogous relationship for heterozygote disadvantage with mean fitness given by  $e^{-2\Sigma p_i q_i}$ , derived by Barton and Bengtsson (1986; Equation A21, although the original article is missing a factor of  $4\phi^2$  in the denominator), is

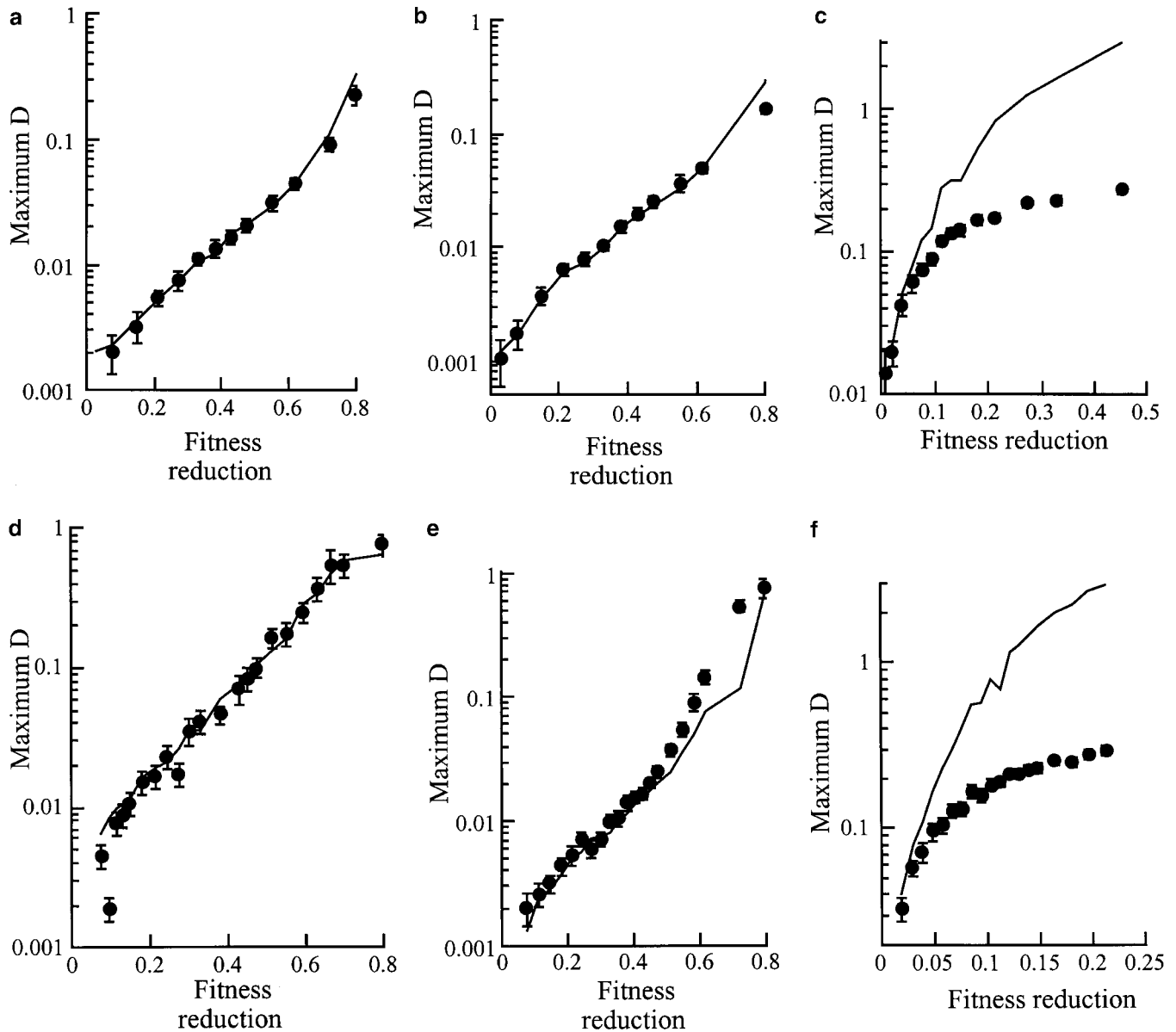


Figure 1.—Maximum linkage disequilibrium  $D$  against fitness reduction at  $p = 0.5$ . (Fitness reduction at  $p = 0.5$  is defined as  $1 - e^{-Ls}$  for environmental selection,  $1 - e^{-Ls/2}$  for heterozygote disadvantage.) Solid circles, simulation results with 95% confidence limits; line, prediction from observed width. (a) Environmental selection acting on 32 unlinked loci, with  $\sigma^2 = 0.2$ , recombination rate between adjacent loci  $r = 0.5$ ; (b) environmental selection on 32 loosely linked loci, with  $\sigma^2 = 0.5$ ,  $r = 0.2$ ; (c) environmental selection on 8 tightly linked loci, with  $\sigma^2 = 0.5$ ,  $r = 0.02$ ; (d) heterozygote disadvantage on 8 unlinked loci, with  $\sigma^2 = 0.2$ ,  $r = 0.5$ ; (e) heterozygote disadvantage acting on 32 loosely linked loci, with  $\sigma^2 = 0.5$ ,  $r = 0.2$ ; (f) heterozygote disadvantage acting on 8 tightly linked loci, with  $\sigma^2 = 0.5$ ,  $r = 0.01$ .

$$\left(\frac{\partial p}{\partial x}\right)^2 = \frac{s}{4\sigma^2} \frac{1}{\phi^2} (e^{4\phi pq} - 4\phi pq - 1). \quad (14)$$

The shapes of clines generated by these alternative selection regimes can be found by numerical integration of the formulas for  $\partial p/\partial x$  (Mathematica 3.0; Wolfram 1996). Ignoring a common scaling factor of  $\sqrt{\sigma^2}/s$  (the characteristic length; Slatkin 1973) gives the relationship between  $p$  and  $x$  in terms of  $\phi$  for either selection regime. These are plotted on both an untransformed

and a logit scale ( $= \text{Log}_e[p/q]$ ; Figures 2a and 2b, respectively) for a single-locus and a multilocus situation. Because the mean fitness for either selection regime is given by  $e^{-2Lps}$  and  $e^{-2Lpqs}$  for environmental selection and heterozygote disadvantage, respectively, values of  $s$  and  $\phi$  for heterozygote disadvantage are scaled by a factor of two so that the mean fitness in the center is equivalent under the two selection regimes.

On the logit scale (Figure 2b), heterozygote disadvantage in a single-locus model ( $\phi = 0$ ) yields a linear cline,

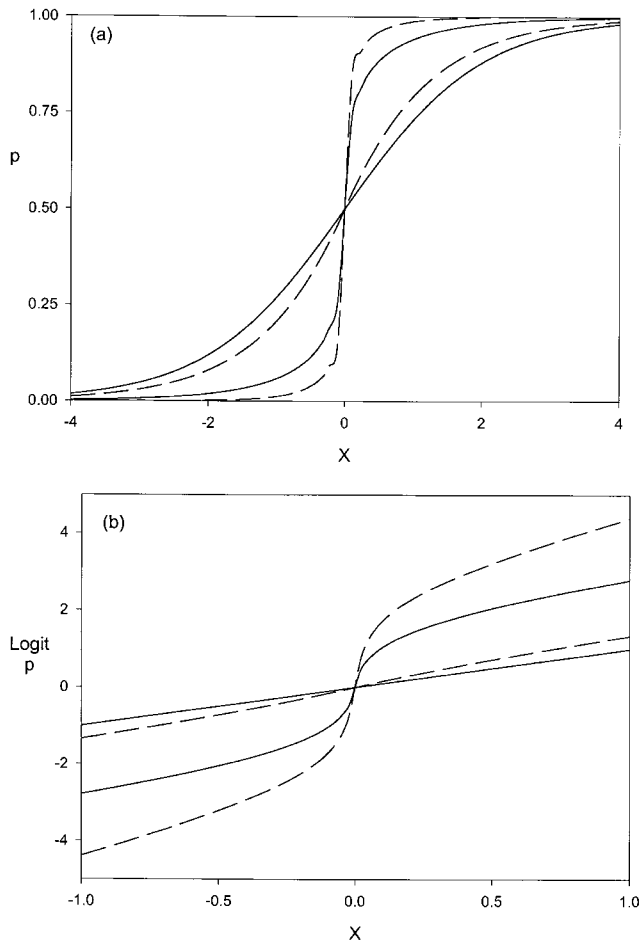


Figure 2.—Shape of clines under environmental selection and heterozygote disadvantage;  $X$  is the distance across cline. (a) Gene frequency  $p$ ; (b)  $\text{logit } p = \ln(p/(1-p))$  for environmental selection (solid lines) and heterozygote disadvantage (dashed lines). In a and b, the two shallow clines represent the single-locus case  $\phi = 0$ , the two steeper clines  $\phi = 5$ .

while environmental selection gives a slight step. This is because the selection gradient tends to zero at the center under heterozygote disadvantage but remains large under Haldane's (1948) model of a sharp ecotone. By contrast, the multilocus clines for both regimes show a stepped pattern. The step is due to the linkage disequilibria in the center increasing the effective selection pressures experienced by each locus, with positive feedback occurring as steeper gradients in turn generate more linkage disequilibria. The feedback is represented in the exponential terms in Equations 13 and 14. The clines under heterozygote disadvantage fall away more steeply in the tails than do those for environmental selection. The difference is predictable for the fitness functions used: the selection strength  $s$  was scaled up by a factor of two so as to give equivalent fitness in the center of the cline, but in the tails, where introgressing alleles are rare and therefore likely to be found only in heterozygous loci, the average selection *per locus* will inevitably be twice as great under heterozygote disadvan-

tage. However, there is no qualitative difference in the shape of the clines under the alternative selection regimes.

Once again, stochastic simulations give an indication of the robustness of Equations 13 and 14. Here, we compare the average cline width observed in simulations with that predicted from (13) and (14), defined as the inverse of the gradient at  $p = 0.5$ . Simulations fit the predictions for heterozygote disadvantage for unlinked loci (Figure 3c) and for weak to moderate selection when loci are linked (Figure 3d). The correspondence suggests that higher-order disequilibria, ignored in the analytical derivations but inevitably present in the simulations, are not generating further positive feedback, steepening the cline beyond what is expected from the pairwise disequilibria. However, with environmental selection the width is systematically greater than predicted (Figure 3, a and b), even for unlinked loci. Wider clines suggest that the maximum reduction in fitness is less than that expected, and we consider possible reasons for the discrepancy in the discussion.

**Cohesion of the genome:** With weak selection and loose linkage, the genome behaves as a set of effectively statistically independent loci. Then, the width of each cline is  $\sqrt{3\sigma^2/s}$  under environmental selection (Haldane 1948), and  $\sqrt{8\sigma^2/s}$  under heterozygote disadvantage (Bazykin 1969). At the other extreme of strong selection and tight linkage, all genes are bound together as a single unit under selection of magnitude  $Ls$ , and so the cline widths will be reduced by a factor  $\sim \sqrt{L}$  [Slatkin (1975) describes the effect for a two-locus system]. Figure 4 shows how clines maintained by selection on several loci transit between these two states. As selection (and hence the coupling coefficient) increases, the average width in a multilocus system changes smoothly from being equivalent to a single-locus cline to that expected if the whole genome acted together. The smooth change occurs in both the analytical predictions and the observed simulated values (although the two do not correspond exactly for environmental selection for the reasons outlined above). Note that as fitness decreases, the prediction itself becomes invalid as it extends outside the domain bounded by the two extreme cases.

The concept of the cohesion of the genome can be used to compare the analytical descriptions of environmental selection with heterozygote disadvantage. Define  $s^*$  as the *effective* selection pressure that would have to act on a single allele to give a cline of the observed width ( $s \leq s^* \leq Ls$ ). The ratio  $s^*/Ls$  gives a measure of the proportion of the selection on the whole genome acting on a single locus and increases from  $1/L$  to 1 as  $\phi$  increases. Figure 5 depicts the ratio for the two selection regimes, with  $\phi$  set so that the expected minimum fitness is equivalent. As before, the relationship becomes meaningless for high values of  $\phi$ , where the weak selection

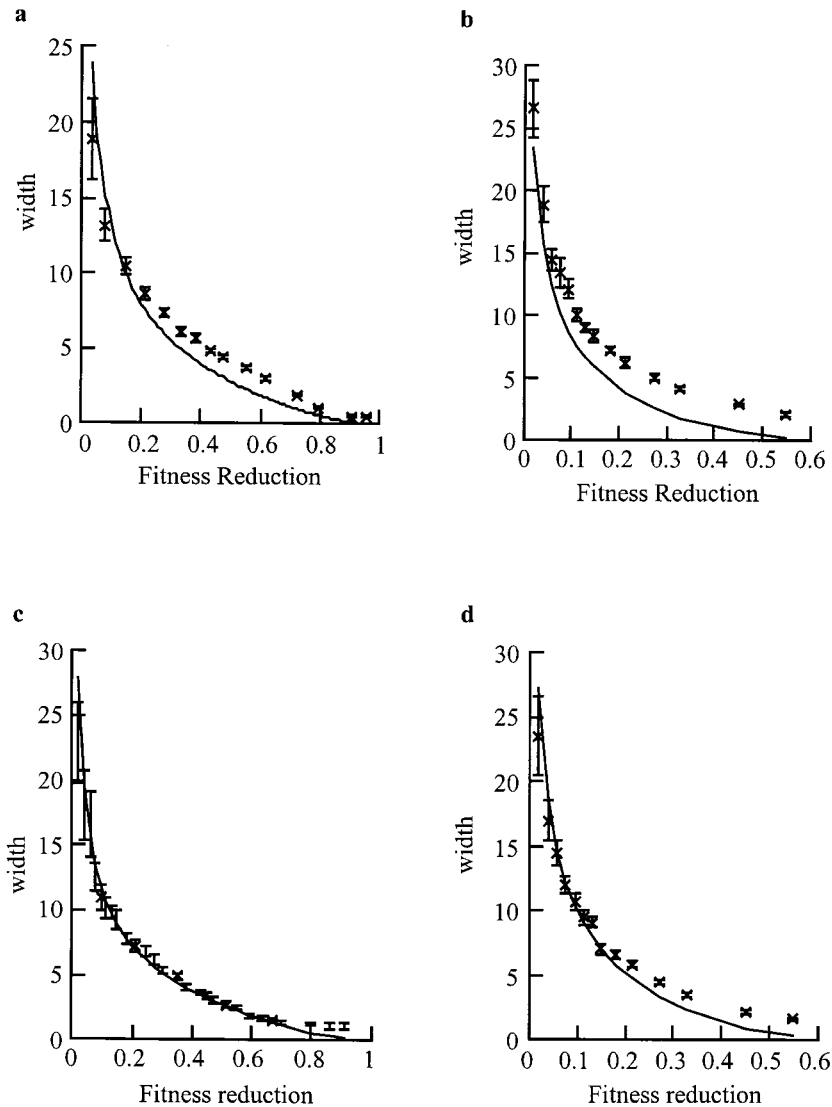


Figure 3.—Cline width against fitness reduction at  $p = 0.5$  (see Figure 1 legend). Crosses, simulation results with 95% confidence limits; line, predicted width (see text). (a) Environmental selection with 32 unlinked loci, with  $\sigma^2 = 0.2$ , recombination rate between adjacent loci  $r = 0.5$ ; (b) environmental selection with 8 loosely linked loci, with  $\sigma^2 = 0.5$ ,  $r = 0.1$ ; (c) heterozygote disadvantage with 8 unlinked loci, with  $\sigma^2 = 0.2$ ,  $r = 0.5$ ; (d) heterozygote disadvantage with 8 loosely linked loci, with  $\sigma^2 = 0.5$ ,  $r = 0.1$ .

assumptions are breaking down. More interestingly, the behavior of the system is remarkably similar for the two different regimes: the degree of interaction within the genome is analytically equivalent. However, the lack of correspondence between the stochastic simulations and the analytical predictions for environmental selection implies that the realized values of  $s^*$  will be less and that heterozygote disadvantage may actually keep the genome together more effectively than the environmental selection described here. This is confirmed in Figure 4, where greater selection strengths are required for environmental selection than for heterozygote disadvantage before the width of the observed cline corresponds to that for an entirely congealed genome.

#### DISCUSSION

Using the machinery set up in previous studies of clines maintained by hybrid dysfunction (Barton 1983, 1986; Barton and Bengtsson 1986), we have devel-

oped a multilocus model in which different alleles are favored on either side of a sharp environmental transition. An analytical description confirms the conclusion of single-locus models, that, at least with the simplest viability selection, the clines produced by either type of regime are indistinguishable in shape. The requirement that selection be relatively weak may seem unduly restrictive, particularly in the application of the theory to data from natural populations, but note that if, for example, a single allele crossing into the other environment type reduces an individual's fitness by only 5%, the width of the resulting gene frequency cline will be equal to less than eight times the average dispersal distance. Furthermore, because linkage disequilibrium increases the effective selection on each locus, the selection pressure acting on each locus increases disproportionately as more loci are considered. Comparisons of the analytical predictions with results from simulations also imply that the weak selection approximation holds for an apparently quite wide range of fitness values (*e.g.*, Figure 1,



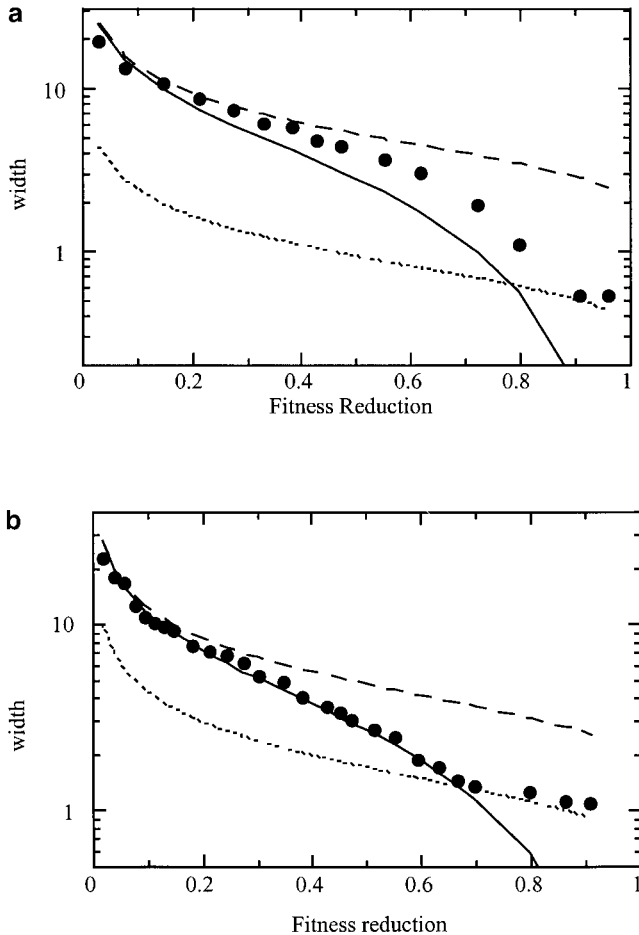


Figure 4.—Comparison of cline width from simulations and analytical model under increasing selection strength. (a) Environmental selection with parameters as for Figure 3a; (b) heterozygote disadvantage with parameters as for Figure 3c. (●) Simulation results; (—) analytical prediction from Equations 13 and 14 with width defined as the inverse of  $\partial p/\partial x$  at  $p = 0.5$ ; (---) analytical prediction for cline in a single locus; (· · ·) analytical prediction if whole genome acts as a single nonrecombining unit. Fitness reduction is defined at  $p = 0.5$  as in legend for Figure 1; width is on a logarithmic scale.

a and b)—including, for example, the estimated fitness reduction of 0.42 (support limits 0.32–0.46) in hybrid populations in the center of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata* (p. 255, Szymura and Barton 1991).

The comparison with stochastic simulations reveals an interesting distinction between the two systems: for environmental selection, the observed clines are wider than expected, *i.e.*, the gradient at the center is shallower, but with heterozygote disadvantage the observed cline width fits the prediction of Equation 14 well. The discrepancy is clearly not due to measurement error in the simulations as the observed width corresponds to that generating the linkage disequilibrium for both selection regimes (Figure 1). Furthermore, exact simulations (as used by Barton and Shpak 1999) show the same discrepancy from the analytical predictions (data

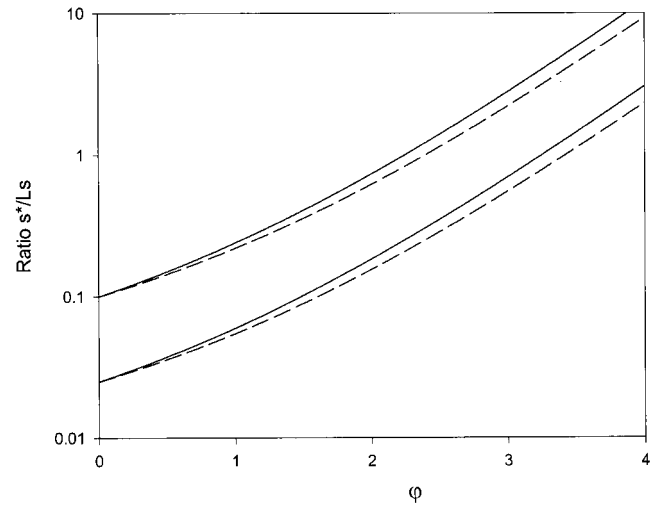


Figure 5.—Ratio of effective selection strength  $s^*$  to total possible selection acting,  $Ls$ , for both heterozygote disadvantage (---) and environmental selection (—), for 10 loci (top lines) and 40 loci (bottom lines), plotted against increasing values of  $\phi$ .

not shown): drift is not the culprit. We suggest two explanations for the discrepancy.

First, the expected gradient at the center is defined at the point where  $p = 0.5$ , which for environmental selection should fall at the point of the step transition in fitness ( $x = 0$ ). In a chain of discrete demes with a step selection function, there will be one deme immediately to the left of the transition and one to the right, but effectively no deme actually covering the center. If no population is at  $p = 0.5$ , then the minimum fitness will never be realized and the clines will be shallower.

The second reason involves the variance in  $z$ . Let the frequency of type 1 alleles carried by an individual be  $z$  ( $z = Z/2L$ ;  $0 \leq z \leq 1$ ), so that the mean value of  $z$  in a population is the gene frequency  $p$ :  $\bar{z} = p$ . Under environmental selection, fitness is defined by  $z$  (Equation 1a). For Equation 11, it was assumed that the mean fitness in a population at mean gene frequency  $p$  was equal to that of an individual with frequency  $p$ , *i.e.*, that  $\overline{W(\bar{z})} \cong W(\bar{z}) = W(p)$ . The assumption ignores the effect of any variance in  $z$ , but because fitness decays exponentially rather than linearly with gene frequency, any variance in  $z$  will increase the mean fitness. The variance in  $z$  has two components: one due to heterozygosity at individual loci and of order  $1/L$ , and the second due to associations between alleles at different loci and hence of order  $D$  (*cf.* Equation 2). In contrast, under heterozygote disadvantage fitness is defined by the number of heterozygous loci  $H$  (Equation 1b). The variance in  $H$  will also have one component due to heterozygosity at individual loci (Bulmer 1985) and one due to associations between loci. However, in this case, *two* sets of associations are required for the heterozygosity at one locus to be affected by heterozygosity at another: the

second component is therefore of order  $D^2$ . Because  $D \leq 0.25$ , this reasoning predicts a greater variance in  $z$  under environmental selection than in  $H$  under heterozygote disadvantage. As outlined above, enhanced variance will increase the mean fitness in the center, and weaker selection pressures result in wider clines.

A brief point should be made with respect to the assumption that the system reaches equilibrium. In an alternative approach to analyzing the consequences of interbreeding, Baird (1995) used Fisher's (1953) junction theory to track the fate of introgressing blocks of genetic material entering a population. Baird's simulation models show that the approach to an analytic equilibrium is slow, of the order of thousands of generations. The results from the simulations presented here fit more closely with equilibrium predictions within a shorter time span and were also checked to ensure that statistics were measured once clines had stabilized. The difference between these conclusions is probably due to the difference in the rates of recombination modeled. Baird considers the fate of a single chromosome of infinitely many loci, so recombination will be much lower than in our model, where recombination rates were chosen to represent either unlinked or loosely linked genes. Infinite deme sizes may also have slowed the approach to equilibrium relative to that in the finite demes considered here. Further comparison of the two systems would be interesting, particularly in the light of evidence that genes for reproductive isolation may be linked (Wu and Palapoli 1994).

We considered here the limit of weak selection. An alternative approach would have been to increase the number of loci  $L$ , ultimately reaching a system with infinitely many loci (*e.g.*, Baird 1995). As  $L$  increases, so do the number of higher-order disequilibria maintained by migration (Kirkpatrick and Servedio 1999). Their magnitude may be less, but whether this balances the concurrent increase in abundance is not clear—analysis in the limit of  $L$  rather than  $s$  would generate interesting comparisons. For the moment, our simulations afford a check on the magnitude of higher-order disequilibria, at least in the weak selection limit, as it is explicitly assumed that they are negligible in the analytical derivations (see appendix a, Barton 1983), but they are not excluded from the simulation models. If higher-order disequilibria *are* affecting the simulation dynamics, they will increase the effective selection pressure experienced at each locus in the same way that the pairwise disequilibria do. This would generate narrower clines than predicted by the analytical description, whatever the selection regime. We do not observe narrower clines either under environmental adaptation or heterozygote disadvantage, thus confirming the assumption that higher-order disequilibria are having a negligible effect.

The method presented here can theoretically be used

to represent any form of selection. Our simulations and analysis show that pairwise linkage disequilibria are closely approximated by a balance between dispersal and recombination, and Barton and Gale (1993) and Barton and Shpak (1999) confirm this for epistatic models. The epistatic selection coefficient  $\partial \text{Log}(\bar{W}) / \partial D_{ij}$  does not contribute significant linkage disequilibrium, while its effect on allele frequencies can be absorbed in a marginal selection coefficient. Genotype frequencies across a set of clines will be influenced primarily by the marginal selection gradient on each gene,  $\partial \text{Log}(\bar{W}) / \partial p_i$ , regardless of whether this is due to endogenous or exogenous selection. It would therefore be possible to construct an "exogenous" fitness function that would mimic the equilibrium maintained by any kind of endogenous selection: the effect of any change in genetic background can be simulated by a change in the environment. Note that even the distinction between viability and fertility selection is ultimately arbitrary. For example, the fertility coefficients defined in Gavrillets' (1997b) single-locus model are effectively equivalent to a two-locus "viability selection" model with epistatic interactions generating nonadditive selection pressures on the offspring produced by different pairings—again, it is the marginal selection gradient on each locus that determines the cline shape. The analysis here is restricted to a set of selected loci assumed to be each experiencing the same, weak selection pressures; but the extension to a system in which selection varies among loci, or some loci are even neutral, is straightforward (see Barton 1986; Barton and Bengtsson 1986; Piálék and Barton 1997; and Gavrillets and Cruzan 1998).

The exclusive action of either endogenous or exogenous selection is undoubtedly an unlikely simplification. Even if divergence were initially driven by differential adaptation to different environments, for a cline to be maintained entirely by exogenous selection requires that no epistatic interactions are ever involved and that selection acts independently on each locus (Barton and Hewitt 1985). Equally unlikely is a scenario in which selection is purely endogenous and fitness entirely independent of the environment. However, there is no reason for either type of selection to be exclusive, either theoretically [for example, Slatkin (1973) describes a single-locus model incorporating both types of selection] or in nature [for example, in the hybrid zone between two species of fire-bellied toads (*Bombina*), there is evidence of both adaptation to different environments (Nürnberg *et al.* 1995; Kruuk and Gilchrist 1997) and hybrid dysfunction (Szymura and Barton 1986; Kruuk *et al.* 1999)]. For evolutionary biology the distinction between the two selection regimes is important for our understanding of factors responsible for divergence: Is species diversity determined by rates of accumulation of genetic incompatibilities or by the diversity of alternative ecological niches?

However, the above results confirm that with the most straightforward of multilocus systems there is nothing intrinsically different in the population genetics of the two scenarios, and gene flow or selection pressures can be estimated in the same way for each.

We are grateful to J. Piálek, C. MacCallum, B. Nurnberger, S. Gavrilets, and M. Kirkpatrick for comments and discussion. This work was supported by Natural Environment Research Council (NERC) grants NERC GR3/11635 and GR3/9353, and the Darwin Trust of Edinburgh.

#### LITERATURE CITED

- Arnold, M. L., 1997 *Natural Hybridization and Evolution*. Oxford University Press, New York, Oxford.
- Asmussen, M. A., and J. Arnold, 1991 The effects of admixture and population subdivision on cytonuclear disequilibria. *Theor. Popul. Biol.* **39**: 273–300.
- Asmussen, M. A., J. Arnold and J. C. Avise, 1989 The effects of assortative mating and migration on cytonuclear associations in hybrid zones. *Genetics* **122**: 923–934.
- Baird, S. J. E., 1995 The mixing of genotypes in hybrid zones: a simulation study of multilocus clines. *Evolution* **49**: 1038–1045.
- Barton, N. H., 1980 The fitness of hybrids between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity* **45**: 47–59.
- Barton, N. H., 1982 The structure of the hybrid zone in *Uroderma bilopatum* (Chiroptera: Phyllostomatidae). *Evolution* **36**: 863–866.
- Barton, N. H., 1983 Multilocus clines. *Evolution* **37**: 454–471.
- Barton, N. H., 1986 The effects of linkage and density-dependent regulation on gene flow. *Heredity* **57**: 415–426.
- Barton, N. H., and B. O. Bengtsson, 1986 The barrier to genetic exchange between hybridising populations. *Heredity* **56**: 357–376.
- Barton, N. H., and K. S. Gale, 1993 Genetic analysis of hybrid zones., pp. 13–45 in *Hybrid Zones and the Evolutionary Process*, edited by R. G. Harrison. Oxford University Press, Oxford.
- Barton, N. H., and G. M. Hewitt, 1981a The genetic basis of hybrid inviability between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity* **47**: 367–383.
- Barton, N. H., and G. M. Hewitt, 1981b Hybrid zones and speciation, pp. 109–145 in *Evolution and Speciation*, edited by W. R. Atchley and D. S. Woodruff. Cambridge University Press, Cambridge, United Kingdom.
- Barton, N. H., and G. M. Hewitt, 1983 Hybrid zones as barriers to gene flow, pp. 341–359 in *Protein Polymorphism: Adaptive and Taxonomic Significance*, edited by G. S. Oxford and D. Rollinson. Academic Press, London.
- Barton, N. H., and G. M. Hewitt, 1985 Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**: 113–148.
- Barton, N. H., and G. M. Hewitt, 1989 Adaptation, speciation and hybrid zones. *Nature* **341**: 497–503.
- Barton, N. H., and M. Shpak, 1999 The effect of epistasis on the structure of hybrid zones. *Genet. Res.* (in press).
- Barton, N. H., and M. Turelli, 1991 Natural and sexual selection on many loci. *Genetics* **127**: 229–255.
- Bazykin, A. D., 1969 Hypothetical mechanism of speciation. *Evolution* **23**: 685–687.
- Bazykin, A. D., 1973 Population genetics of stabilising and disruptive selection. II. Systems of adjacent populations and populations within a continuous area. *Genetika* **9**: 156–166.
- Bulmer, M. G., 1985 *The Mathematical Theory of Quantitative Genetics*. Oxford University Press, Oxford.
- Endler, J. A., 1977 *Geographic Variation, Speciation and Clines*. Princeton University Press, Princeton, NJ.
- Feldman, M. W., and F. B. Christiansen, 1975 The effect of population subdivision on two loci without selection. *Genet. Res.* **24**: 151–162.
- Fisher, R. A., 1937 The wave of advance of advantageous genes. *Ann. Eugen.* **7**: 355–369.
- Fisher, R. A., 1950 Gene frequencies in a cline determined by selection and diffusion. *Biometrics* **6**: 353–361.
- Fisher, R. A., 1953 A fuller theory of “junctions” in inbreeding. *Heredity* **8**: 187–197.
- Gavrilets, S., 1997a Hybrid zones with Dobzhansky-type epistatic selection. *Evolution* **51**: 1027–1035.
- Gavrilets, S., 1997b Single locus clines. *Evolution* **51**: 979–983.
- Gavrilets, S., and M. Cruzan, 1998 Neutral gene flow across single locus clines. *Evolution* **52**: 1277–1284.
- Hal dane, J. B. S., 1948 The theory of a cline. *J. Genet.* **48**: 277–284.
- Harrison, R. G., 1990 Hybrid zones: windows on evolutionary process. *Oxf. Surv. Evol. Biol.* **7**: 69–128.
- Harrison, R. G. (Editor), 1993 *Hybrid Zones and the Evolutionary Process*. Oxford University Press, Oxford.
- Hartl, D. L., and A. G. Clark, 1989 *Principles of Population Genetics*, Ed. 2. Sinauer, Sunderland, MA.
- Hewitt, G. M., 1988 Hybrid zones—natural laboratories for evolutionary studies. *Trends Ecol. Evol.* **3**: 158–167.
- Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. *Genet. Res.* **8**: 269–294.
- Jackson, K. S., 1992 Population dynamics of a hybrid zone in the alpine grasshopper *Podisma pedestris*: an ecological and genetical investigation. PhD thesis, University College, London.
- Key, K. H., 1968 The concept of stasipatric speciation. *Syst. Zool.* **17**: 14–22.
- Kirkpatrick, M., and M. R. Servedio, 1999 The reinforcement of mating preferences on an island. *Genetics* **151**: 865–884.
- Kruuk, L. E. B., and J. S. Gilchrist, 1997 Mechanisms maintaining species differentiation: predator-mediated selection in a *Bombina* hybrid zone. *Proc. R. Soc. Lond. B* **264**: 105–110.
- Kruuk, L. E. B., J. S. Gilchrist and N. H. Barton, 1999 Hybrid dysfunction in fire-bellied toads (*Bombina*). *Evolution* **53**: 1611–1616.
- Lewontin, R. C., and K.-I. Kojima, 1960 The evolutionary dynamics of complex polymorphisms. *Evolution* **14**: 458–472.
- Li, W.-H., and M. Nei, 1974 Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Popul. Biol.* **6**: 173–183.
- Mallet, J. L. B., N. Barton, G. M. Lamas, J. C. Santisteban, M. M. Muedas *et al.*, 1990 Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* **124**: 921–936.
- Moore, W. S., and J. T. Price, 1993 Nature of selection in the Northern Flicker Hybrid Zone and its implications for speciation theory, pp. 196–225 in *Hybrid Zones and the Evolutionary Process*, edited by R. G. Harrison. Oxford University Press, Oxford.
- Nagylaki, T., 1975 Conditions for the existence of clines. *Genetics* **80**: 595–615.
- Nagylaki, T., 1976a Clines with variable migration. *Genetics* **83**: 867–886.
- Nagylaki, T., 1976b The evolution of one- and two-locus systems. *Genetics* **83**: 583–600.
- Nurnberger, B., N. H. Barton, C. MacCallum, J. Gilchrist and M. Appleby, 1995 Natural selection on quantitative traits in the *Bombina* hybrid zone. *Evolution* **49**: 1224–1238.
- Piálek, J., and N. H. Barton, 1997 The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics* **145**: 493–504.
- Porter, A. H., R. Wenger, H. Geiger, A. Scholl and A. M. Shapiro, 1998 The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. *Evolution* **51**: 1561–1573.
- Sites, J. W., N. H. Barton and K. M. Reed, 1995 The genetic structure of a hybrid zone between two chromosome races of the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae) in Central Mexico. *Evolution* **49**: 9–36.
- Slatkin, M., 1973 Gene flow and selection in a cline. *Genetics* **75**: 733–756.
- Slatkin, M., 1975 Gene flow and selection on a two-locus system. *Genetics* **81**: 787–802.
- Slatkin, M., and T. Maruyama, 1975 Genetic drift in a cline. *Genetics* **81**: 209–222.
- Szymura, J. M., and N. H. Barton, 1986 Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* **40**: 1141–1159.
- Szymura, J. M., and N. H. Barton, 1991 The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina*

- and *B. variegata*: comparisons between transects and between loci. *Evolution* **45**: 1141–1159.
- Turner, J. R. G., 1967 Why does the genotype not congeal? *Evolution* **21**: 645–656.
- Wolfram, S., 1996 *The Mathematica Book*, Ed. 3. Wolfram Media/Cambridge University Press, Cambridge, United Kingdom.
- Wright, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- Wu, C. I., and M. F. Palapoli, 1994 Genetics of postmating reproductive isolation in animals. *Annu. Rev. Genet.* **27**: 283–308.

Communicating editor: M. Slatkin

#### APPENDIX: EXCHANGE AMONG A CHAIN OF DEMES

We describe here the generation of pairwise linkage disequilibrium by migration between populations at different gene frequency. In a chain of demes, individuals in a population migrate to and from the two demes on either side of it; we therefore consider exchange between three demes. The model is an extension of the treatment of two demes considered by Li and Nei (1974) and Barton and Gale (1993). We show here that the three-deme scenario results in higher linkage disequilibrium than migration between two demes.

Suppose that gene frequency is lowest in deme 0, intermediate in deme 1, and highest in deme 2. Each generation, a proportion  $m$  of the diploid individuals in the central deme, are exchanged with individuals from deme 0 and deme 2, half in either direction. Let  $p_{1,x}$ ,  $p_{2,x}$ ,  $PP_x$ , and  $D_x$  be, respectively, the allele frequencies at the two loci, the frequency of the coupling gamete, and the linkage disequilibrium in deme  $x$ . After migration, the gene and gamete frequencies in deme 1 are given by

$$\begin{aligned} p'_{1,1} &= (1 - m)p_{1,1} + \frac{m}{2}p_{1,0} + \frac{m}{2}p_{1,2}; \\ p'_{2,1} &= (1 - m)p_{2,1} + \frac{m}{2}p_{2,0} + \frac{m}{2}p_{2,2}; \\ PP'_1 &= (1 - m)(p_{1,1}p_{2,1} + D_1) + \frac{m}{2}(p_{1,0}p_{2,0} + D_0) \\ &\quad + \frac{m}{2}(p_{1,2}p_{2,2} + D_2). \end{aligned} \quad (\text{A1})$$

In the production of gametes at meiosis, recombination can only break down the linkage disequilibrium within the pool of nonmigrants and the two pools of migrants separately. The frequency of the coupling gamete  $PP$  immediately after meiosis is therefore

$$\begin{aligned} PP'_1 &= (1 - m)(p_{1,1}p_{2,1} + (1 - r)D_1) \\ &\quad + \frac{m}{2}(p_{1,0}p_{2,0} + (1 - r)D_0) \\ &\quad + \frac{m}{2}(p_{1,2}p_{2,2} + (1 - r)D_2). \end{aligned} \quad (\text{A2})$$

Reproduction does not alter gene frequencies, and so the new linkage disequilibrium is just  $PP'_1 - p'_{1,1}p'_{2,1}$ . Substituting in Equations A1 and A2 leads to an expression for the linkage disequilibrium after one generation:

$$\begin{aligned} D'_1 &= (1 - r)D_1 + \frac{m}{2}(1 - r)(D_0 + D_2 - 2D_1) \\ &\quad + \frac{m}{2}(1 - m)(p_{1,0} - p_{1,1})(p_{2,0} - p_{2,1}) \\ &\quad + \frac{m}{2}(1 - m)(p_{1,2} - p_{1,1})(p_{2,2} - p_{2,1}) \\ &\quad + \frac{m^2}{4}(p_{1,2} - p_{1,0})(p_{2,2} - p_{2,0}). \end{aligned} \quad (\text{A3})$$

The first term describes the reduction in disequilibrium due to recombination, the next the averaging of disequilibria from the three demes due to migration; the last three terms are the linkage disequilibrium generated by the three pairwise combinations of demes. [If the two neighboring demes are identical, so  $p_{10} = p_{12}$  etc., and Equation A3 is identical to that for exchange between two demes (Li and Nei 1974; Barton and Gale 1993).]

Equation A3 can be solved for a “quasi-equilibrium” (Nagylaki 1976b), in which a balance has been reached between the loss due to recombination and the generation by migration. This assumes that differences between the linkage disequilibria in each deme change less than differences in the gametic frequencies, so that the diffusion of linkage disequilibrium can be neglected [second term in (A3)]. The system is taken to be symmetric, so that  $\Delta p_1 = p_{12} - p_{11} = p_{11} - p_{10}$  etc. Equation A3 then leads to

$$D_1 = \frac{m\Delta p_1 \Delta p_2}{r}. \quad (\text{A4})$$

Because  $m$  is equal to  $\sigma^2$ , the variance in parent-offspring distance (A4) is exactly equivalent to the prediction for the continuous case (Equation 2).

Note that Equation A4 is without the factor of  $(1 - m)$  in the equivalent formula for exchange between two demes (Li and Nei 1974; Barton and Gale 1993). This is because migrants enter a deme from two sides, so that the genetic make-up of the demes on either side of a central deme is important. If the neighboring demes are identical, the system collapses to the two-deme case, but in a chain of demes they will differ in opposite directions, and so the linkage disequilibria generated by exchange between three demes will be greater. Similarly, recursion equations developed by Asmussen and Arnold (1991) to describe the associations between cytoplasmic and nuclear loci generated by admixture imply that, when immigrants into a hybrid population come from two different source populations, the resulting associations will be inflated by differences be-

tween the latter; see also Asmussen *et al.* (1989). In the case considered here, the relationship applies to values measured after mating and recombination. If measurements are made after migration, mixing will have increased  $D$  by the amount  $m\Delta p_1\Delta p_2$ , and so the QLE approximation will be greater by a factor of  $(1 + r)$ .

In a system of  $L > 2$  loci, migration will also generate disequilibria of higher order (Kirkpatrick and Servedio 1999). Although the magnitude of each decreases

rapidly with order, so that three-way disequilibria are roughly three times smaller than pairwise and so forth, the number of associations increases with order, so their cumulative effect may be substantial. Here, we use the simulation model to test our explicit assumption (Equation 6) that selection coefficients involving higher-order disequilibria ( $\partial \text{Log}(\bar{W})/\partial D_{ijk}$  etc.) have a negligible effect on gene frequencies.