

The effects of linkage and density-dependent regulation on gene flow

N. H. Barton

Department of Genetics and Biometry, University
College London, 4 Stephenson Way, London NW1
2HE, U.K.

The rate of gene flow across a hybrid zone may be reduced by the presence of a physical barrier, by a reduction of population density caused by reduced fitness of hybrids (the "hybrid sink" effect), and by linkage. If the reduction in hybrid fitness is not extreme, the strength of the barrier to gene flow caused by these effects is $B = w(\rho^*/\rho_0^*)^2(\bar{W}^*/\bar{W}_0^*)^{(2/R+1/\bar{r})}$. Here, w is the width of the cline; ρ^* is the carrying capacity; \bar{W}^* is the mean fitness of the population, excluding effects of density; R is the strength of density-dependent regulation; and \bar{r} is the harmonic mean recombination rate between the locus whose flow is being calculated, and loci under selection. +, 0 denote populations outside the hybrid zone, and at its centre, respectively. This relation is illustrated using data from hybrid zones in *Bombina* and *Podisma*, and its implications for interpretation of data from nature are discussed.

INTRODUCTION

Genetically distinct populations may often be separated by narrow hybrid zones. Whether these populations already constitute separate biological species, or will eventually diverge into separate species, depends, in part, on the rate of gene exchange between them. Hybrid zones can impede gene flow in several ways. They may coincide with physical barriers (Key, 1968; Barton, 1979a); the reduced fitness of hybrids may cause a reduction in population size, which will act as a physical barrier (the "hybrid sink" effect; Hall, 1973; Barton, 1980), and linkage will impede flow between different genetic backgrounds (Barton, 1979b; Bengtsson, 1985; Barton and Bengtsson, 1986; Spirito *et al.*, 1983; Petry, 1983). Here, I will show that these three effects can be combined into a simple and general expression. The strength of a barrier to gene flow depends primarily on the reduction in mean fitness of the hybrid population. This allows mean fitness to be inferred from clinal patterns in neutral markers. Conversely, some general restrictions can be put on the rate of evolution of reproductive isolation by random drift (Barton and Charlesworth, 1984).

DEFINITIONS

The meaning of the term "gene flow", and the definition of "barrier strength", are discussed by

Barton and Bengtsson (1986), who consider both demic and continuous population structures. This paper is concerned solely with gene flow across a continuous habitat; in this section, the terms relevant to the continuous case will be briefly introduced.

Selection is assumed throughout to be sufficiently weak that allele frequencies do not change significantly between generations or (in the case of a stepping-stone model) between adjacent demes. With this assumption, gene flow can be approximated by diffusion (Nagylaki, 1975). Consider an allele which is moving across a continuous, one dimensional habitat. The flux of alleles diffusing past any point is proportional to the gradient of allele frequency at that point: the steeper the gradient, the greater the flux. A local barrier will cause a sharp step in the frequency (u), in just the same way as an obstruction in a stream will cause a sharp step in the water level. The size of the step (Δu) is proportional to the flux, and hence to the gradient (u') (fig. 1). We can therefore define the strength of a local barrier to gene flow in a continuous habitat as the ratio between the size of the step and the gradient ($B = \Delta u/u'$; Nagylaki, 1976). If the barrier is not quite local, but has some appreciable spatial extent, we must assume that the allele whose flow we are considering is sufficiently weakly selected that selection has negligible effect over the width of the barrier. If the barrier is asymmetric, allowing more alleles to penetrate in one direction than the other,

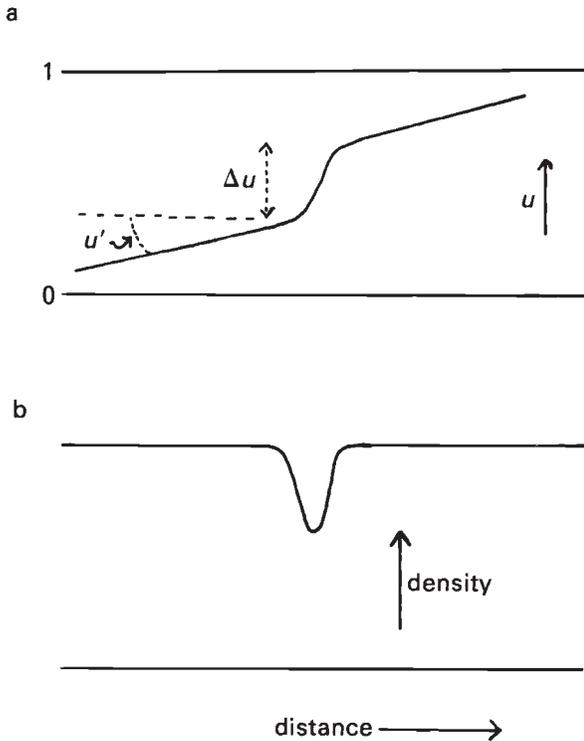


Figure 1 (a) The stepped cline produced by a barrier of strength $B = \Delta u / u'$. (b) The pattern of population density, ρ , which may produce a barrier to gene flow: $B = (\rho_+ / \rho_0)^2$; equation (4).

we must define two barrier strengths ($B_{\pm} = \Delta u / u'_{\pm}$), corresponding to the different fluxes of genes into the populations on either side.

This measure of barrier strength has the dimensions of distance. It can be thought of as the length of open habitat which would present an equivalent obstacle to the spread of a neutral allele: such an allele will be delayed for about $(B/\sigma)^2$ generations, where σ is the dispersal range. However, it is important to note that a local barrier will impede a slightly advantageous allele much less than a length B of open habitat. A barrier must be much greater than the dispersal range (σ) to have much effect on a neutral allele, and must be virtually complete to prevent the spread of a slightly advantageous allele (Barton, 1979b).

PHYSICAL BARRIERS

We will now find the strength of various kinds of barrier, beginning with the simplest case: a local reduction in population density. Assume that the distribution of distance between parent and

offspring has a mean of zero, and a variance of σ^2 . This ensures that dispersal causes no directed flux of genes when density is uniform. We also assume that the parent-offspring distance has a variance which is independent of position, and of density. Then, the effects of gene flow in a habitat with varying density, $\rho(x)$, is approximated by (Nagylaki, 1976):

$$\dot{u} = \frac{\sigma^2}{2} u'' + \sigma^2 u' (\log \rho)' \quad (1)$$

(where $\dot{u} = \partial u / \partial t$, and $u' = \partial u / \partial x$).

Suppose that u is held at equilibrium by a flux of genes from outside the region of interest. Equation (1) can then be rewritten as:

$$0 = \frac{\sigma^2}{2\rho^2} (\rho^2 u')' \quad (2)$$

Integrating, we see that $\rho^2 u'$ is constant. If the densities on either side of the barrier are equal ($\rho_+ = \rho_-$), and if density decreases towards a minimum at the centre (ρ_0), and then increases again, the allele frequency will change in a stepped pattern (fig. 1). The size of the step, and hence the barrier strength, can be found by integrating equation 2 again:

$$u(x) = u'_+ \left(\frac{\rho_+}{\rho(x)} \right)^2$$

$$\Delta u = \left\{ \int_{-\infty}^{\infty} u'_+ \left(\frac{\rho_+}{\rho(x)} \right)^2 dx \right\} - u'_+ x \quad (3)$$

$$B = \frac{\Delta u}{u'_+} = \int_{-\infty}^{\infty} \left\{ \left(\frac{\rho_+}{\rho(x)} \right)^2 - 1 \right\} dx.$$

This expression can be simplified by introducing the width of the barrier, w . This is defined in the usual way (Endler, 1977; fig. 1) as the total change in gene frequency, divided by the maximum gradient: $w = \Delta u / u'_0$. Now, from equation (2):

$$B = \frac{\Delta u}{u'_+} = \frac{\Delta u}{u'_0} \left(\frac{\rho_+}{\rho_0} \right)^2 = w \left(\frac{\rho_+}{\rho_0} \right)^2 \quad (4)$$

The barrier due to a local reduction in density depends primarily on the relative densities at the edge and the centre (ρ_+ / ρ_0). Since the barrier strength has the dimensions of distance, it is also proportional to the width of the barrier, w (which can be found by taking the ratio of equations (3) and (4)). Thus, the ratio of densities could be inferred from the spatial pattern of allele frequencies, $u(x)$, through the two measures, B and w .

Of course, dispersal rates, σ , may also vary from place to place: a local reduction in dispersal can also act as a barrier. Such variations can be

dealt with in a similar way, by extending equation (1) to include the effects of gradients in σ (Nagy-laki, 1976). When this is done, the density, ρ , is replaced by the product $\rho\sigma^2$. Although we have used one dimensional equations, the results apply to straight clines in two dimensions; most relevant data is from such two dimensional populations. In this case, $\rho\sigma^2$ is proportional to Wright's (1943) "neighbourhood size", $N = 4\pi\rho\sigma^2$. The barrier strength is thus proportional to the square of the ratio between the neighbourhood sizes at the centre and at the edge: $(B/\sigma_+) = (w/\sigma_0)(N_+/N_0)^2$. This approach will be developed more fully in a later paper; for the moment, it is sufficient to note that the above relations give measures of neighbourhood size which are independent of those obtained from random fluctuations in gene frequencies (Wright *et al.*, 1942).

So far, the densities on either side of the barrier have been assumed equal. If they differ, then there will be a net flux from the dense population into the less dense population, and a corresponding difference in the gradients on either side. The ratio of gradients, and of rates of gene flow, is $(u'_+/u'_-) = (\rho_-/\rho_+)^2$.

THE "HYBRID SINK"

The effects of a simple physical barrier can now be extended to include genetic effects. Suppose that selection acts on a number of loci, distinct from the marker locus whose flow we are considering. This selection maintains a set of narrow clines, either because different alleles are favoured on different sides of the clines, or because heterozygous and/or recombinant genotypes are less fit. In either case, populations within the cline will be less fit than those outside, and so there will be some reduction in population density. This will, in turn, act as a barrier to gene flow at loci which are themselves not directly selected. This has been described as the "hybrid sink" effect (Hall, 1973; Barton, 1980), because genes flow in from either side, and are lost in the unfit population at the centre.

In order to quantify the relation between the genetic constitution of the population, and its density, we must consider the mean absolute fitness, \bar{W} . This is the average number of offspring per parent, and is a function of the density, ρ , the distribution of genotypes in the population, g , and the spatial position, x . Assume that, for any given position and distribution of genotypes, the population reaches a definite equilibrium density, rather

than cycling or becoming chaotic. At equilibrium, the mean fitness must (in a sexually reproducing population) equal 2. The density is then given by $\bar{W}(\rho, g, x) = 2$, and is an implicit function of the location and the genotypic composition (x, g). Following the same approach as Barton (1980) we expand the mean fitness in a Taylor series about some arbitrary density, ρ^* . It is convenient to work with the log mean fitness and log density:

$$\log(\bar{W}) \cong \log(\bar{W}^*) - R \log(\rho/\rho^*). \quad (5)$$

Here, $R = -\partial \log(\bar{W})/\partial \rho$ at ρ^* , and is a measure of the strength of density-dependent regulation: when selection is "hard", R is small, and when it is "soft", R is large (Wallace, 1975). Under the logistic equation, the intrinsic rate of increase is $1+R$. R may depend on the composition of the population, and on position.

We now choose ρ^* to be the equilibrium density which a population would have if it contained the genotypes found outside the cline; for convenience, assume that the system is symmetric, so that $\rho^* = \rho_- = \rho_+$. This is a function of position, and measures the carrying capacity of the population at each location. \bar{W}^* is now the mean fitness which the population would have at a density ρ^* ; it depends on genotype proportions as well as on position. By definition, $\bar{W}^*(g_+, \infty) = 2$. Solving equation (5), we find:

$$\rho(x, g) \cong \rho_+^*(x) \left(\frac{\bar{W}^*(g, x)}{2} \right)^{(1/R(g, x))} \quad (6)$$

This is only accurate when density does not vary much from $\rho^*(x)$: extending it to a wider range of densities is equivalent to assuming a particular form of density-dependence, $\bar{W} \propto \rho^{-R}$. The analysis could, of course, be repeated using other arbitrary functions, such as the logistic equation: the results would not differ much unless density varies greatly with genotype.

Combining equation (6) with equation (4):

$$B = w \left(\frac{\rho_+^*}{\rho_0^*} \right)^2 \left(\frac{\bar{W}_+^*(2/R_+)}{\bar{W}_0^*(2/R_0)} \right). \quad (7)$$

The barrier strength depends on the ratio of carrying capacities (ρ_+^*/ρ_0^*), expressing the effect of any physical barrier, and on the ratio of mean fitness (\bar{W}_+^*/\bar{W}_0^*), expressing the "hybrid sink" effect.

LINKAGE

In order to pass across a hybrid zone, a gene must recombine into the foreign genetic background

before it is eliminated by selection against the alleles with which it is associated. To put this another way, linkage disequilibrium will be generated as individuals carrying parental combination of alleles disperse into the centre of the zone (Li and Nei, 1974; Feldman and Christiansen, 1975). This disequilibrium allows some fraction of the total selection maintaining the hybrid zone to act on each individual locus, even if that locus does not itself affect fitness. A set of parallel clines may be greatly steepened by these effects (Slatkin, 1975; Barton, 1983).

Suppose that each genotype has constant fitness relative to other genotypes: although absolute fitness may depend on density and position, relative fitnesses are assumed to be independent of density, position, and genotype proportions. Then, the frequency of a neutral allele (u) is given, at equilibrium, by:

$$0 = \frac{\sigma^2 u''}{2} + \sum_i \left\{ \frac{D_{Ni}}{2} \frac{\partial \log \bar{W}^*}{\partial p_i} \right\} + \sum_{i \neq j} \left\{ \frac{D_{Nij}}{2} \frac{\partial \log \bar{W}^*}{\partial D_{ij}} \right\} \dots \quad (8a)$$

This equation follows from the general equations for selection and gene flow, which are derived in the Appendix. D_{Ni} is the two-way disequilibrium between the neutral locus (N) and the i th selected locus. D_{Nij} is the three-way disequilibrium, and so on. p_i is the frequency at the i th locus. (It is assumed that there are two alleles at each locus; however, extension to several alleles is straightforward.) \bar{W}^* is the mean fitness which the population would have at a density ρ^* . At equilibrium, the mean absolute fitness, \bar{W} , must equal 2: the tendency for natural selection to increase mean fitness is counterbalanced by an increase in density (Fisher, 1930). Hence, equation (8) involves the mean fitness at a fixed density, excluding the counterbalancing effects of density-dependence. (In this section, we do not include the effect of varying carrying capacity ($\rho^*(x)$): this was treated above.)

The effects of selected loci on the neutral cline depend on the strength of the various disequilibria, $D_{Nijk} \dots$. These are given by:

$$0 = \frac{\sigma^2 D_{Ni}''}{2} + \sigma^2 u' p_i' + \sum_{j \neq i} \left\{ \frac{D_{Nij}}{2} \frac{\partial \log \bar{W}^*}{\partial p_j} \right\} + \sum_{\substack{j \neq k \\ \neq i}} \left\{ \frac{D_{Nijk}}{2} \frac{\partial \log \bar{W}^*}{\partial D_{jk}} \right\} \dots - r_{Ni} D_{Ni} \quad (8b)$$

$$0 = \frac{\sigma^2 D_{Nij}''}{2} + \sigma^2 (u' D_{ij}' + p_i' D_{Nj}' + p_j' D_{Ni}') + \frac{\sigma^2}{2} (u'' D_{ij} + p_i'' D_{Nj} + p_j'' D_{Ni}) + \sum_{\substack{k \neq i \\ \neq j}} \left\{ \frac{D_{Nijk}}{2} \frac{\partial \log \bar{W}^*}{\partial p_k} \right\} \dots - r_{Nij} D_{Nij}. \quad (8c)$$

Here, r_{Ni} is the probability of recombination between N and i , and r_{Nij} is the probability of a recombination anywhere between loci N , i , or j (Bennett, 1954). The differentials of log mean fitness with respect to allele frequency ($\partial \log \bar{W}^* / \partial p_i$) give the selection coefficients on each locus, whilst the differentials with respect to disequilibria ($\partial \log \bar{W}^* / \partial D_{ij}$, etc.) give the coefficients of epistasis. We assume that all these coefficients are of the same order of magnitude, and are proportional to some measure $s - a$ reasonable assumption if there are no special constraints on the selection scheme. If s is much smaller than r (the rate of recombination between adjacent loci; $r \cong r_{Nij} \dots$) then equation (8b) is dominated by two terms: the first, $\sigma^2 u' p_i'$, representing the generation of disequilibrium by dispersal into the centre of the hybrid zone, and the second, $-r_{Ni} D_{Ni}$, representing the decay of disequilibrium through recombination. This can be shown explicitly by rescaling distance relative to a characteristic distance $\sigma / \sqrt{2s}$, and expanding in powers of s/r (Barton, 1983). (The assumption that selection is much weaker than recombination may be unduly restrictive: Barton (1983) showed that, when selection acts against heterozygotes, the above two terms dominate even when selection is strong relative to recombination, provided that a sufficiently large number of loci is involved ($\log(n) \gg s/r$.)

With this approximation, the pairwise disequilibrium is $D_{Ni} \cong \sigma^2 u' p_i' / r_{Ni}$. Similarly, the three-way disequilibrium, D_{Nij} , is of order $(s/r)^2$, and can be ignored. Substituting into equation (8a):

$$0 = \frac{\sigma^2 u''}{2} + \frac{\sigma^2 u'}{2} \sum_i \left\{ \frac{p_i'}{r_{Ni}} \frac{\partial \log \bar{W}^*}{\partial p_i} \right\} + O((s/r)^2). \quad (9)$$

Now, assume that selection is spread evenly over the genome, so that there is no tendency for the neutral locus to be more closely linked to more (or less) strongly selected sections of chromosome. The sum over terms inversely proportional to r_{Ni} can then be replaced by a term inversely proportional to \bar{r} , the harmonic mean recombination rate

between the neutral locus and the selected loci:

$$0 \equiv \frac{\sigma^2 u''}{2} + \frac{\sigma^2 u'}{2\bar{r}} \sum_i \left\{ p'_i \frac{\partial \log \bar{W}^*}{\partial p_i} \right\}. \quad (10)$$

In general, the mean fitness is a function of the allele frequencies (p_i), disequilibria ($D_{ij}...$), and spatial position (x). The gradient of log mean fitness is therefore:

$$\begin{aligned} (\log \bar{W}^*)' &\equiv \frac{d \log \bar{W}^*}{dx} = \sum_i \left\{ p'_i \frac{\partial \log \bar{W}^*}{\partial p_i} \right\} \\ &+ \sum_{i \neq j} \left\{ D'_{ij} \frac{\partial \log \bar{W}^*}{\partial D_{ij}} \right\} \dots \\ &+ \frac{\partial \log \bar{W}^*}{\partial x}. \end{aligned} \quad (11)$$

Since we assume that disequilibria are weak ($O(s/r)$), and that relative fitnesses do not depend directly on location, only the first term in equation (11) need be considered. This allows equation (10) to be integrated:

$$\begin{aligned} 0 &= \frac{\sigma^2 u''}{2} + \frac{\sigma^2 u' (\log \bar{W}^*)'}{2\bar{r}} \\ \therefore 0 &= \frac{\sigma^2 \bar{W}^{*(-1/\bar{r})}}{2} \left\{ u' \bar{W}^{*(1/\bar{r})} \right\}' \end{aligned} \quad (12)$$

$u' \bar{W}^{*(1/\bar{r})}$ is constant, and so:

$$B = w \left\{ \frac{\bar{W}_+^*}{\bar{W}_0^*} \right\}^{(1/\bar{r})} \quad (13)$$

(Note that the exponent in the formula given by Barton and Jones (1983), $(2/\bar{r})$, is too large by a factor of two.)

In deriving equation (13), several assumptions have been made: namely, that selection is weak enough that gene flow can be approximated by diffusion ($s \ll 1$); that selection is weaker than recombination, so that linkage disequilibria are small ($s \ll r$, $D \ll 1$); that selection is spread evenly over the genetic map; that relative fitnesses are not frequency- or density-dependent; and that relative fitnesses do not depend directly on location. Only two of these assumptions are seriously restrictive. The requirement that disequilibria be small means that equation (13) cannot be applied in cases where the distribution of genotypes is bimodal, being clustered around the two parental forms. A method for finding the rate of gene flow in such cases is derived by Barton and Bengtsson (1986). The second serious restriction is that fitnesses be spatially invariant. It is hard to judge the relative contributions of selection against hybrids, and

selection favouring different alleles in different places, in maintaining hybrid zones (Barton and Hewitt, 1985). If spatial variation is important, one would expect it to contribute to the spatial pattern of mean fitness in the same way as does spatial variation in allele frequencies. It will introduce a term in equation (11) which may be of similar magnitude to the first term ($\sum_i p'_i \partial \log \bar{W}^* / \partial p_i$), and which would be expected to be negative to the left, and positive to the right. Equation (13) will then overestimate the barrier strength, since only part of the gradient in mean fitness (equation (11)) will contribute to linkage disequilibria (equation (10)).

The above arguments are closely related to those on "genetic load", in that they relate the combined effects of many loci (in this case, on gene flow at a neutral locus) to the mean fitness of the population. However, some of the arguments which have been made against the use of "load" in interpreting polymorphism and substitution (e.g., Sved *et al.*, 1967, Wallace, 1975) do not apply here. Epistatic gene interactions are accounted for in the derivation, whilst the effect of gene frequencies on population size is introduced explicitly through the function $\bar{W}(\rho, g, x)$.

DISCUSSION

Equations (1), (7), and (13) can be combined to give a single expression for the barrier to gene flow caused by physical barriers, the "hybrid sink", and linkage. Since equation (13) depends on the assumption that relative fitnesses of genotypes are constant, the degree of density-dependent regulation, R , must also be assumed independent of genotype. Thus:

$$B = w \left(\frac{\rho_+^*}{\rho_0^*} \right)^2 \left(\frac{\bar{W}_+^*}{\bar{W}_0^*} \right)^{\{(2/R)+(1/\bar{r})\}} \quad (14)$$

In order for there to be a substantial reduction in gene flow, the barrier should be wide (w large), the carrying capacity and mean fitness at the centre should be low ($\rho_+^* \gg \rho_0^*$, $\bar{W}_+^* \gg \bar{W}_0^*$), selection should be hard (R small), and linkage should be tight (\bar{r} small).

This relationship can be illustrated by applying it to data from two hybrid zones: between the fire-bellied toads, *Bombina bombina* and *B. variegata* (where linkage disequilibrium is strong), and between two chromosomal races of the grasshopper, *Podisma pedestris* (where there is evidence on the effects of both the external environment and hybridisation, on population density).

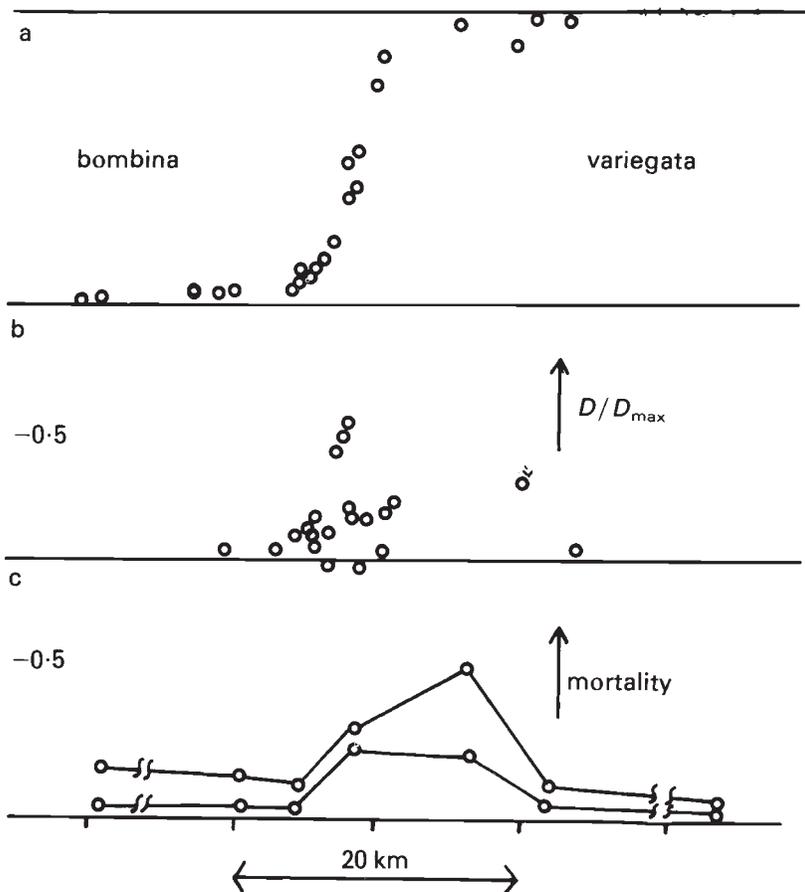


Figure 2 (from Szymura and Barton, 1986, fig. 4) (a) Allele frequency, averaged over 5 enzyme loci, plotted against distance. (b) The standardised linkage disequilibrium, D/D_{max} , averaged over all pairs of loci. (c) Mortality from cleavage to gastrulation (lower curve), and from gastrulation to feeding (upper curve), across the hybrid zone in *Bombina*.

Bombina bombina and *B. variegata* meet in a narrow hybrid zone which runs for over 1000 Km through eastern and central Europe. Szymura (1976, and Szymura and Barton, 1986) has made a detailed study of the zone near Cracow, in southern Poland, including an electrophoretic survey of 5 enzyme loci which are fixed for different alleles in the two taxa. The clines at these loci are parallel, and change sharply at the same place as the zone defined by morphological and other characters. However, alleles introgress at low frequencies up to 30 Km in either direction, giving the stepped pattern characteristic of a barrier to gene flow (fig. 2). The close similarity of the clines at these loci, and their extensive introgression, implies that selection on each individual enzyme locus is weak or absent, and that the shape of the clines is determined primarily by the barrier to gene flow. The strength of this barrier, and the

width of the cline, can be estimated from the data of fig. 2(a). The most likely estimates, and support limits, are $B = 220$ (48–415) Km, and $w = 6.15$ (5.45–6.45) Km. (Introgression is in fact significantly asymmetric, with more from *bombina* into *variegata* than vice versa; the above estimate of B is an average over both directions).

The sharp step in allele frequency is probably caused by linkage disequilibria: significant disequilibria are observed between each pair of enzyme loci, even though these are on different chromosomes *Bombina* has 12 pairs of metacentric chromosomes, each with ≈ 1 chiasma per arm (Morescalchi, 1965); if selection is spread over 300 evenly spaced loci, the harmonic mean recombination rate is $\bar{r} \approx 0.12$. (This figure of 300 loci is estimated from the *Bombina* data; \bar{r} is not sensitive to this number, however). In order to account for

the observed cline shape ($B/w \cong 36$), hybrid populations must have mean fitness $(B/w)^{-\bar{r}} = 0.65(0.60-0.77)$ of that of pure populations. This estimate may be in error for three reasons. First, although the observed disequilibria are likely to give the main contribution, there may also be some reduction in density across the zone. Second, equation (13) is only accurate when disequilibria are weak: this is not a serious problem here, since average pairwise disequilibria are 13 per cent of the maximum possible value at the centre (Szymura and Barton, 1986; see fig. 2(b)). Finally, selection may favour different genotypes in different places, as well as acting against hybrids: *bombina* and *variegata* show many adaptations to their characteristic habitats. Spatial variation in selection would cause underestimation of the reduction in mean fitness (see equation (11)). Despite these potential difficulties, the above interpretation is supported by an increased frequency of morphological abnormalities in hybrids, and by increased mortality in hybrid crosses (fig. 2(c)).

The relation between population density and barrier strength can be illustrated using data from the hybrid zone in *Podisma pedestris*. This involves two chromosomal races which differ by a Robertsonian fusion between the *X* and an autosome; there is also substantial hybrid inviability which is not associated with the chromosomal difference (Barton and Hewitt, 1981). The two types meet in a hybrid zone which runs for ~ 150 Km through the Alpes Maritimes in south-eastern France. In most of the transects which have been studied, the chromosomal cline is trapped by physical barriers, and is sharply stepped (Barton and Hewitt, 1981). At the western end of the cline, near Seyne, however, there is an area where, although the cline is pinned at either end by inhospitable scree slopes, it runs across a region of high density (Nichols, 1984; fig. 3(a)). The position and shape of the chromosomal cline can be explained by the observed density: the cline is smooth and sigmoid (Nichols, 1984; fig. 2(b)). In general, it is hard to disentangle the effects of physical environment and genetic composition on population density. In this case, however, Nichols (1984) has shown that although most of the variation in density is related to differences in vegetation, the number of nymphs hatching in hybrid populations is half that in pure populations with matched vegetation. However, the adult densities do not differ significantly. Thus, although mean fitness (\bar{W}^*) is reduced by hybridisation, density-dependent regulation is strong enough to bring the population back to its original adult density. Since dispersal is primarily by adults,

one would not expect, and does not observe, any step in gene frequency as a result of the "hybrid sink" effect. (It is puzzling, however, that there is no evidence of any effect of linkage disequilibrium, despite strong selection against hybrids in field and laboratory; see Barton, 1983.)

In most of the above derivations, the populations on either side of the barrier were assumed to have the same densities, and the same mean fitnesses: the rate of gene flow would then be the same in either direction. In fact, hybrid zones are often markedly asymmetrical, and this asymmetry is sometimes consistent across loci (e.g., *Mus musculus/domesticus*, *Caledia captiva*, *Bombina bombina/variegata*, *Uroderma bilobatum*, *Ranidella insignifera/pseudinsignifera*; Barton and Hewitt, 1985). Asymmetric introgression does not necessarily imply an asymmetric barrier: the introgressing alleles could be subject to different selection pressures after penetrating the barrier, or could disperse at different rates on either side. However, if this were the case, it would be hard to explain a consistent direction of asymmetry. In contrast, if a barrier to gene exchange is asymmetric, that asymmetry will be in the same direction at all loci, since much of the contribution to linkage disequilibria is from relatively loosely linked loci (see Barton, 1983, p. 465). An asymmetrical barrier would seem to reflect an asymmetry in either mean fitness or in carrying capacity ($\bar{W}_+^* \neq \bar{W}_-^*$, or $\rho_+^* \neq \rho_-^*$). However, this interpretation is not necessarily correct. If carrying capacity (ρ^*) were uniform, then the genotypes which give greater mean fitness will tend to advance; as they move forward, alleles from the less fit race will be pushed into the fitter race (Moran, 1981). This tendency will counteract the tendency of alleles to move from the fitter race into the less fit, caused by linkage and by the "hybrid sink" effect. Similarly, if carrying capacity varies from place to place, the hybrid zone will tend to move into an equilibrium position in which the tendency of the fitter race to advance is counterbalanced by a gradient in density (Barton, 1979a). In either case, it is not at all obvious which effect will predominate: I can find no simple relation between the relative fitnesses of the two populations, and the density gradient needed to counterbalance this difference. However, one can argue that although carrying capacity may fluctuate greatly from place to place, the average capacity may not differ much between the habitats occupied by the two taxa. Thus, although a difference in mean fitness may be counterbalanced by a local density gradient, the net direction of introgression will still be from the fitter population into the less

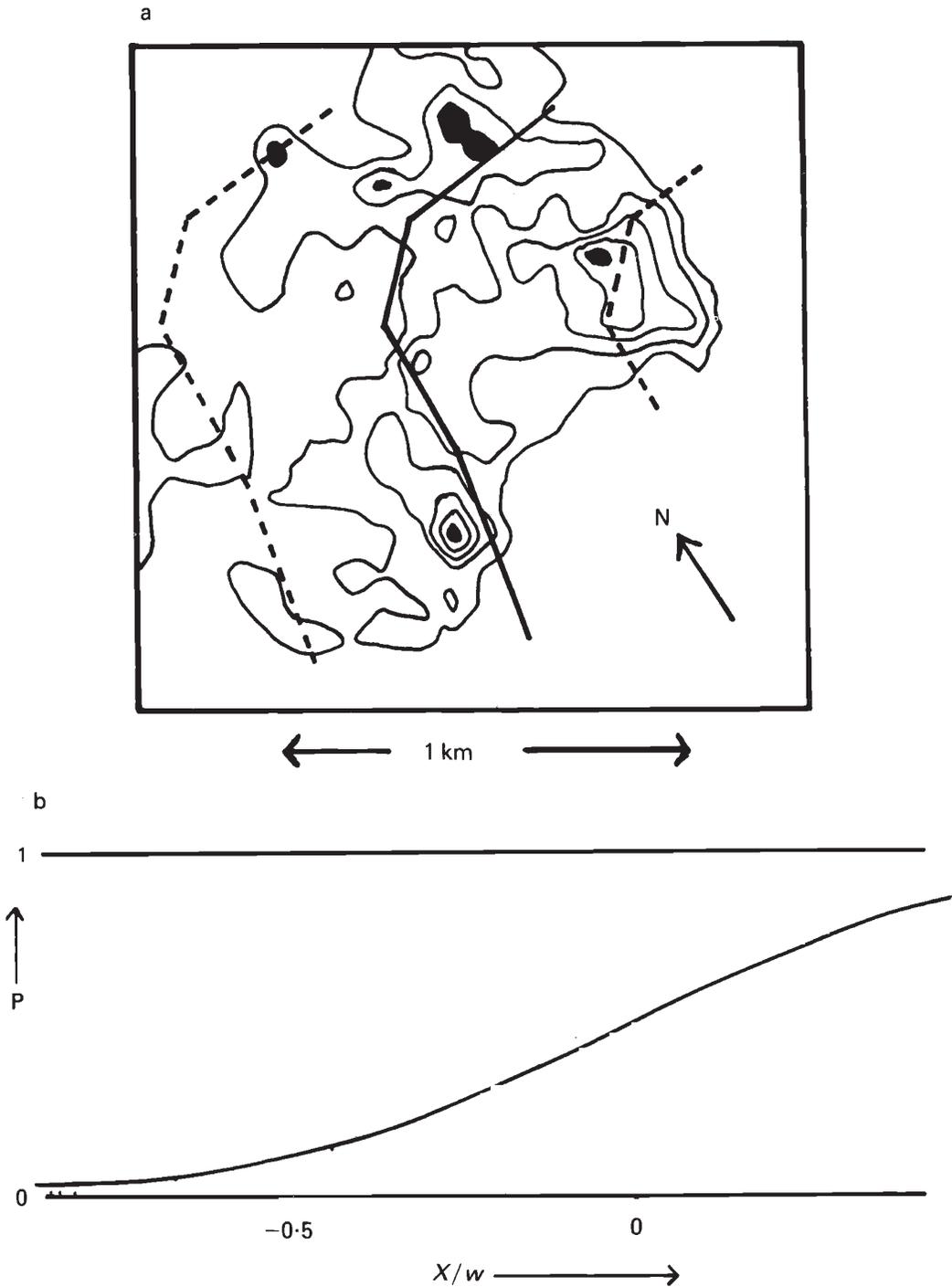


Figure 3 (from Nichols, 1984, and unpublished data) (a) The position of the chromosomal cline in *Podisma pedestris*, superimposed on population density, for the area near Seyne. Density contours are plotted at intervals of $0.5 \text{ adults m}^{-2}$; solid shading indicates maximum density. The left and right dotted lines give the 10 per cent and 90 per cent frequencies of the chromosomal fusion, and the central solid line gives the 50 per cent frequency; these are estimated by maximum likelihood. (b) The shape of the chromosomal cline: distance is measured from the central contour shown in (a), and is given as a ratio of the estimated cline width. The smooth curve plots the theoretical frequency of the fusion, in the absence of any effects of linkage or density variation. The points give the observed frequency of the fusion in 105 samples, averaging 16.5 males.

fit. An additional argument, which supports this view, is that the exponent $(2/R + 1/\bar{r})$ must be large if the barrier is to be biologically significant; it will also tend to increase as the number of loci involved increases, since the harmonic mean recombination rate (\bar{r}) will decrease. In contrast, the density gradient needed to counterbalance a given difference in mean fitness will decrease as the number of loci increases, since the density gradient need only balance the selection pressure at each individual locus. The difference in mean fitness (\bar{W}^*) is therefore likely to dominate over any tendency for the cline to move: introgression should therefore be from the fitter race into the less fit one.

Acknowledgements I would like to thank Shahin Rouhani and Richard Nichols for their helpful comments on the manuscript, and the latter for providing the data used in fig. 3. This work was supported by a grant from the Science and Engineering Research Council (GR/C/91529).

REFERENCES

- ANTONELLI, P. L. AND STROBECK, C. 1977. The geometry of random drift. I. Stochastic distance and diffusion, *Adv. Appl. Prob.*, **9**, 238–249.
- BARTON, N. H. 1979a. The dynamics of hybrid zones. *Heredity*, **43**, 341–359.
- BARTON, N. H. 1979b. Gene flow past a cline. *Heredity*, **43**, 333–339.
- BARTON, N. H. 1980. The hybrid sink effect. *Heredity*, **44**, 277–278.
- BARTON, N. H. 1983. Multilocus clines. *Evolution*, **37**, 454–471.
- BARTON, N. H. AND BENGTSOON, B. O. 1986. The barrier to gene exchange between hybridising populations. *Heredity*, in press.
- BARTON, N. H. AND CHARLESWORTH, B. 1984. Genetic revolutions, founder effects, and speciation. *Ann. Rev. Ecol. Syst.*, **15**, 133–164.
- BARTON, N. H. AND HEWITT, G. M. 1981. A chromosomal cline in the grasshopper *Podisma pedestris*. *Evolution*, **35**, 1008–1018.
- BARTON, N. H. AND HEWITT, G. M. 1985. Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.*, **16**, 113–148.
- BARTON, N. H. AND JONES, J. S. 1983. Mitochondrial DNA: new clues about evolution. *Nature*, **306**, 317–318.
- BENGTSOON, B. O. 1985. The flow of genes through a genetic barrier. In Greenwood, P. J., Harvey, P. and Slatkin, M. (eds.) *Evolution: Essays in Honour of John Maynard Smith*, Cambridge Univ. Press, Cambridge.
- BENNETT, J. H. 1954. On the theory of random mating. *Ann. Eugenics*, **18**, 311–317.
- ENDLER, J. A. 1977. *Geographic Variation, Speciation, and Clines*. Princeton Univ. Press, Princeton.
- EWENS, W. 1979. *Mathematical Population Genetics*. Springer Verlag, Berlin.
- FELDMAN, M. W. AND CHRISTIANSEN, F. B. 1975. The effect of population subdivision on two loci without selection. *Genet. Res.*, **24**, 151–162.
- FISHER, R. A. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- HALL, W. 1973. *Comparative Population Cytogenetics, Speciation, and Evolution of the Iguanid Lizard Genus, Sceloporus*. Ph.D. Thesis, Harvard Univ.
- HASTINGS, A. 1986. Multilocus population genetics with weak epistasis. II. Equilibrium properties of multilocus selection: what is the unit of selection? *Genetics*, **112**, 157–171.
- KEY, K. H. L. 1968. The concept of stasipatric speciation. *Syst. Zool.*, **17**, 14–22.
- LI, W. H. AND NEI, M. 1974. Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Pop. Biol.*, **6**, 173–183.
- MORAN, C. 1981. Genetic demarcation of geographical distribution by hybrid zones. *Proc. Roy. Soc. Aust.*, **11**, 67–73.
- MORESCALCHI, A. 1965. Osservazioni sulla carilogia di *Bombina*. (In Italian, with English summary.) *Boll. Zool.*, **32**, 207–218.
- NAGYLAKEI, T. 1975. Conditions for the existence of clines. *Genetics*, **80**, 595–615.
- NAGYLAKEI, T. 1976. Clines with variable migration. *Genetics*, **83**, 867–886.
- NAGYLAKEI, T. AND CROW, J. F. 1974. Continuous selective models. *Theor. Pop. Biol.*, **5**, 257–283.
- NICHOLS, R. A. 1984. *The Ecological Genetics of a Hybrid Zone in the Alpine Grasshopper Podisma pedestris*. Ph.D. Thesis, Univ. of East Anglia.
- PETRY, D. 1983. The effect on neutral gene flow of selection at a linked locus. *Theor. Pop. Biol.*, **23**, 300–313.
- SLATKIN, M. 1972. On treating the chromosome as the unit of selection. *Genetics*, **72**, 157–168.
- SLATKIN, M. 1975. Gene flow and selection in a two-locus system. *Genetics*, **75**, 733–756.
- SPIRITO, F., ROSSI, C. AND RIZZONI, M. 1983. Reduction of gene flow due to the partial sterility of heterozygotes for a chromosome mutation. I. Studies on a “neutral” gene not linked to the chromosome mutation. *Evolution*, **37**, 785–797.
- SVED, J. A., REED, T. C. AND BODMER, W. F. 1967. The number of balanced polymorphisms that can be maintained by natural selection. *Genetics*, **55**, 469–481.
- SZYMURA, J. M. 1976. Hybridisation between discoglossid toads *Bombina bombina* and *B. variegata* in Southern Poland as revealed by the electrophoretic technique. *Z. Zool. Syst. Zooforsch.*, **14**, 227–236.
- SZYMURA, J. M. AND BARTON, N. H. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads. *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution* (in press).
- WALLACE, B. 1975. Hard and soft selection revisited. *Evolution*, **29**, 465–473.
- WRIGHT, S. 1937. The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. USA*, **23**, 307–320.
- WRIGHT, S. 1943. Isolation by distance. *Genetics*, **28**, 114–138.
- WRIGHT, S., DOBZHANSKY, T. AND HOVANITZ, W. 1942. Genetics of natural populations. VII. The allelism of lethals in the third chromosome of *Drosophila pseudoobscura*. *Genetics*, **27**, 363–394.

APPENDIX

This section is a generalisation of Appendix A in Barton (1983) to describe arbitrary selection pressures on a multilocus system. Suppose that two alleles segregate at each of n loci: the state of a gamete is written as $l = (l_1, l_2, \dots, l_n)$, where each

of the l 's can have value 0 or 1. The treatment is thus restricted to biallelic loci. However, it could readily be extended to the multiallelic case either by regarding a multiallelic locus as being made up of several tightly linked sites, or by letting the l 's become vectors, consisting of one element for each allele. This element would take the value 1 when the locus carried the corresponding allele.

The state of the population could be described by the frequency of each of the 2^n possible gametes, $g(l)$. Alternatively, the description could be in terms of the allele frequencies at each locus, $p_i = E(l_i)$, the pairwise disequilibria, $D_{2;ij} = E((l_i - p_i)(l_j - p_j))$, the three-way disequilibria, $D_{3;ijk} = E((l_i - p_i)(l_j - p_j)(l_k - p_k))$, and so on. ($E(f)$ denotes the expectation of f over the population; this definition follows that of Slatkin (1972), and differs slightly from that of Bennett (1954); see Hastings, 1986). The deviation of the actual state of a gene from its expectation, $l_i - p_i$, will be abbreviated to ζ_i .

Selection

The effect of selection can be found by first finding the change in gamete frequencies, $g(l)$, and then transforming the equations to give the change in allele frequencies and disequilibria. Write the fitness of a diploid individual composed of the two gametes l, l^* as $W(l, l^*)$. Then, the change in gamete frequencies is (assuming random mating):

$$\Delta g(l) = \frac{(\sum_{l^*} g(l^*) W(l, l^*) - \bar{W}) g(l)}{\bar{W}}. \quad (\text{A1})$$

Here, $\bar{W} = E(W)$ is the mean fitness:

$$\bar{W} = \sum_l \sum_{l^*} g(l) g(l^*) W(l, l^*). \quad (\text{A2})$$

Now, the differential of log mean fitness with respect to gamete frequency is:

$$\frac{\partial \log \bar{W}}{\partial g(l)} = 2 \sum_{l^*} \frac{g(l^*) W(l, l^*)}{\bar{W}}. \quad (\text{A3})$$

Equation (A1) can therefore be rewritten:

$$\Delta g(l) = \sum_m \left\{ g(l) (\delta(l, m) - g(m)) \frac{1}{2} \frac{\partial \log \bar{W}}{\partial g(m)} \right\} \quad (\text{A4})$$

(where $\delta(l, m)$ is the Kronecker delta).

This equivalent to the formula derived by Wright (1937). If fitnesses are frequency-dependent, the differential on the right-hand side of equation (A4) must be replaced by the right-hand term in equation (A3), which gives the marginal effect on fitness of each type of gamete.

We now wish to transform equation (A4) into an equation for allele frequencies and linkage disequilibria, or in other words, to change the coordinates by which we describe the state of the population from $g(l)$ to the set $(p, D_2, D_3, \dots, D_n)$. Since the linkage disequilibria are a non-linear function of the gamete frequencies, this transformation is extremely complicated. However, if we assume that selection is weak enough that changes in gamete frequencies over one generation are small, then $\Delta g(l)$ in equation (A4) can be approximated by the continuous differential $\partial g(l)/\partial t$ (Nagylaki and Crow, 1974). The matrix $g(l)(\delta(l, m) - g(m))$, which connects the change between generations with the gradient in log mean fitness, can now simply be rotated to the new set of coordinates. Writing this matrix as M :

$$M(p_i, p_j) = \sum_l \sum_m \left\{ \frac{\partial p_i}{\partial g(l)} M(g(l), g(m)) \frac{\partial p_j}{\partial g(m)} \right\} \quad (\text{A5})$$

(and similarly for elements involving $D_{2;ij}$). The differentials required for this transformation are:

$$\frac{\partial p_i}{\partial g(l)} = l_i$$

$$\frac{\partial D_{N;ij\dots}}{\partial g(l)} = (\zeta_i \zeta_j \dots) - (l_i D_{N-1;j\dots} + l_j D_{N-1;i\dots}) \quad (\text{A6})$$

The extra terms in the equation for $\partial D/\partial g$ arise because the disequilibria are defined as $E(\zeta_i \zeta_j \dots)$; the ζ_i themselves depend on gamete frequencies through p_i . Substituting equation (A6) into (A5):

$$M(p_i, p_j) = D_{2;ij} \quad (\text{A7a})$$

$$M(p_i, D_{N;jk\dots}) = D_{N+1;ijk\dots} - (D_{2;ij} D_{N-1;k\dots} + D_{2;ik} D_{N-1;j\dots}) \quad (\text{A7b})$$

$$M(D_{N;ij\dots}, D_{M;kl\dots}) = D_{N+M;ij\dots,kl\dots} - (D_{N+1;kij\dots} D_{M-1;l\dots} + D_{N+1;l ij\dots} D_{M-1;k\dots}) - (D_{N-1;j\dots} D_{M+1;ikl\dots} + D_{N-1;i\dots} D_{M+1;jkl\dots}) + (D_{2;ik} D_{N-1;j\dots} D_{M-1;l\dots} + D_{2;il} D_{N-1;j\dots} D_{M-1;k\dots}) - D_{N;ij\dots} D_{M;kl\dots} \quad (\text{A7c})$$

In many cases, two or more indices may coincide. These cases can be dealt with using the same method, and their solution is still given by equation (A7). However, the "singular" disequilibria must

be defined as (for example) $D_{ijk\dots} = E(\zeta_i \zeta_j \zeta_k \dots)$. This gives:

$$D_{N;ijk\dots} = p_i q_i D_{N-2;jk\dots} - (p_i - q_i) D_{N-1;ijk\dots} \quad (\text{A8})$$

Coincidence of three or more indices can be dealt with by repeated application of equation (A8). $D_0 = 1$, and $D_{1;i} = 0$. In the special case of selection on two loci, equations (A7), (A8) give the same equations as those derived by a direct algebraic transformation of equation (A1), for an arbitrary pattern of (weak) selection. The second set of terms in equations (A7b) was omitted from equation (A7b) of Barton (1983); however, this error only affects third and higher order disequilibria, and does not affect the arguments in the body of that paper.

The matrix M plays the key role in the equations for the effects of selection on a multi-locus system. The same matrix determines the effects of sampling drift: $M(g(l), g(m))/2N_e$ is the covariance in gamete frequencies produced by sampling from a population of effective size N_e . By the same argument as was used above, the matrix can be rotated to give the effect of sampling drift on allele frequencies and disequilibria. This depends only on the assumption that changes in gamete frequencies due to drift are small. For two loci, this gives the same coefficients as in Ewens' (1979) equation (6.75). The matrix M can be thought of as setting a natural measure of genetic distance, or metric (Antonelli and Strobeck, 1977). The coordinate transformations made above rely on the fact that the quantities $\partial g(l)/\partial t$, $M(g(l), g(m))$ and $\partial \log \bar{W}/\partial g(l)$ are tensors, allowing equation (A4) to be expressed in an arbitrary coordinate system.

Dispersal

The effects of gene flow are given by equation (A11) of Barton (1983). In order that gene flow can be approximated by diffusion, selection must be weak.

Recombination

Following Hastings' (1986) notation, let $r_{S_N|S_M}$ be the probability of a recombination event which separates the set of N loci, S_N , from the set of M loci, S_M . This event may involve multiple cross-overs and/or segregation of separate chromosomes. It may involve any number of loci ($N + M \leq n$). The recombination event destroys a gamete carrying alleles in state $(l_i l_j \dots l_k l_l \dots)$, and

creates a gamete carrying two sets of alleles, derived from different parents, and in state $(l_i l_j \dots)(l_k^* l_l^* \dots)$. (Here, $i, j \in S_N$, and $k, l \in S_M$). The change in disequilibrium is therefore (assuming random mating):

$$\begin{aligned} \Delta D_{N+M;ij\dots kl\dots} &= -r_{S_N|S_M} (E(\zeta_i \zeta_j \dots \zeta_k \zeta_l \dots) \\ &\quad - E(\zeta_i \zeta_j \dots) E(\zeta_k^* \zeta_l^* \dots)) \\ &= -r_{S_N|S_M} (D_{N+M;ij\dots kl\dots} \\ &\quad - D_{N;ij\dots} D_{M;kl\dots}). \end{aligned} \quad (\text{A9})$$

The total change in $D_{N+M;ij\dots kl\dots}$ caused by recombination is the sum of equation (A9) over all possible partitions, $S_N | S_M$. This expression differs from that of Bennett (1954) for fourth and higher order disequilibria: Bennett's definition of these higher order disequilibria differs from that used here by terms involving products of pairs of lower order disequilibria, in such a way that the products in equation (A9) (e.g., $D_{N;ij\dots} D_{M;kl\dots}$) disappear. Then, the rate of decay of the L 'th order disequilibrium is $(1 - r_{S_L}) D_L$, where $(1 - r_{S_L})$ is the probability of any recombination which disrupts the set of L loci. When disequilibria are weak, the effect of the product terms becomes negligible, so that this rate of decay applies to the definition used here. (Note that in the main part of this paper, the more compact notation r_{ijk} is used to denote the probability of a recombination anywhere between the three loci i, j, k ; it is equal to the term $(1 - r_{S_3})$ used in this Appendix, and the r_{ijk} used by Hastings (1986).)

Selection, dispersal, and recombination

In order that the effects of these three processes can be added, selection must be weak; this assumption has in any case already been made. Recombination may be frequent, however, provided that the population is measured immediately after recombination takes place.

Hastings (1986) conjectures that when selection is weak, higher-order disequilibria "directly reflect the presence of additive epistasis at the loci involved"; he shows that this is true for second and third-order disequilibria (Result 5). We can see from the above equations that this is indeed so. When disequilibria are weak (either because epistatic components are much weaker than additive selection and recombination, or because all components of selection are much weaker than recombination), the leading terms in equation (A7) are those on the diagonal:

$$M(D_{N;ij\dots}, D_{N;ij\dots}) = p_i q_i p_j q_j \dots + O(D). \quad (\text{A10})$$

The equation for disequilibrium is therefore, to leading order in D :

$$\Delta D_{N;ij\dots} = (p_i q_i p_j q_j \dots) \left\{ \frac{1}{2} \frac{\partial \log \bar{W}}{\partial D_{N;ij\dots}} \right\} - (1 - r_{S_N}) D_{N;ij\dots} \quad (\text{A11})$$

This confirms that the N th-order disequilibrium depends only on the corresponding N th-order coefficient of epistasis, and the probability of a recombination somewhere between the loci involved.