

# Spatial models for hybrid zones

RICHARD DURRETT<sup>†</sup>, LINDA BUTTEL<sup>‡</sup> & RICHARD HARRISON\*<sup>§</sup>

<sup>†</sup>Department of Mathematics, Cornell University, Ithaca, NY 14853, U.S.A., <sup>‡</sup>Theory Center, Cornell University, Ithaca, NY 14853, U.S.A. and <sup>§</sup>Section of Ecology and Systematics, Cornell University, Ithaca, NY 14853, U.S.A.

We introduce a spatially explicit model of natural hybrid zones that allows us to consider how patterns of allele frequencies and linkage disequilibria change over time. We examine the influence of hybrid zone origins on patterns of variation at two loci, a locus under selection in a two-patch environment, and a linked neutral locus. We consider several possible starting conditions that represent explicit realizations of two alternative scenarios for hybrid zone origins: primary intergradation and secondary contact. Our results indicate that in some circumstances, differences in hybrid zone origins will result in substantially different patterns of variation that may persist for thousands of generations. Our conclusions are generally similar to those previously derived from partial differential equations, but there are also some important differences.

**Keywords:** clinal variation, hybrid zone, linkage disequilibrium, secondary contact, stochastic spatial models.

## Introduction

Hybrid zones occur when genetically distinct groups of individuals meet and mate, resulting in at least some offspring of mixed ancestry (Barton & Hewitt, 1985, 1989; Harrison, 1990). Hybrid zone origins have been a puzzle to evolutionary biologists for more than a century (Harrison, 1993). The fundamental problem is to infer the relative importance of history and current ecology as explanations for observed distributions of genotypes and phenotypes. Two alternative scenarios have been proposed to explain patterns of variation in hybrid zones (Mayr, 1942, 1963; Harrison, 1990). One suggests that hybrid zones are the result of *secondary contact* between populations that were previously geographically isolated. Although Mayr imagined that initial genetic differentiation occurred in allopatry, models of secondary contact simply require that the current hybrid zone is a result of the coming together of already differentiated forms. The origin of observed genetic differentiation may substantially predate the origin of the current hybrid zone.

The alternative explanation, *primary intergradation*, is that hybrid zones have arisen *in situ* in response to spatially varying selection, e.g. existing environmental gradients or patchy environments. Although the secondary contact explanation has traditionally been favoured (consistent with the widely held view that genetic divergence requires geographical isolation),

Endler (1977) argued that to infer origins from current patterns of variation is difficult. Using analysis of single-locus clines (see his fig. 6.1), he showed that primary intergradation and secondary contact produce patterns that are indistinguishable unless 'we observe a zone of intergradation within a few hundred generations of secondary contact'.

If hybrid zones are the result of secondary contact, then initially they will be characterized by a set of coincident clines (for all traits that have diverged during the period of isolation). In contrast, coincident clines are not necessarily expected in primary intergradation, unless many independent traits respond in similar fashion to a single environmental gradient or mosaic. Therefore, Barton & Hewitt (1981) argued that concordant clines are evidence of secondary contact. Such evidence will be particularly convincing if some of the clines reflect changes in characters that are known to be 'neutral'.

Here we use a simulation model to examine the influence of hybrid zone origins on observed patterns of variation at two loci (one under selection, the other neutral). We consider several possible starting conditions, defined in terms of (i) initial gametic disequilibrium and (ii) initial spatial distribution of the possible two-locus genotypes. Using different intensities of selection and different amounts of recombination, we document the nature of the hybrid zones formed from each of the starting conditions by assaying allele frequencies at the two loci and patterns of disequilibrium between them. Our models suggest that in some

\*Correspondence. E-mail: rgh4@cornell.edu

circumstances primary intergradation and secondary contact will result in substantially different patterns of variation, and that observations of divergence of neutral loci across hybrid zones and significant linkage disequilibrium in hybrid zones may often be evidence of secondary contact.

### Description of the model

Following the approach of Cox & Durrett (1995), we represent our spatially distributed population by a grid of sites, each of which is always occupied by exactly one individual. Mathematically, the set of sites is  $Z^2$ , the points in the plane with integer coordinates. In words, we have an infinite checkerboard with one individual on each square. Insisting on having exactly one individual at each site is the spatial analogue of having a constant population size.

As mentioned in the introduction, we will be mainly concerned with the situation in which there is one locus under selection, with alleles  $A$  and  $a$ , and a second neutral locus with alleles  $B$  and  $b$ . To describe the state of the system at time  $t$  we have to give the relevant genetic information about the diploid individual at each site. We represent the state of a site  $x$  at time  $t$  by a string with the first two letters from  $\{A,a\}$  and the second two from  $\{B,b\}$ . For example  $AabB$  indicates that the first chromosome has alleles  $A$  and  $b$  and the second has  $a$  and  $B$ .

In our system, time is discrete:  $t = 0, 1, 2, \dots$ . Think of  $t$  as being the generation number or think of the successive states as representing snapshots of the system taken at a fixed time interval. The state of the individual at  $x$  at time  $t + 1$  is determined by picking two parents from nearby sites, choosing twice with replacement according to the following rule. Independently pick displacements in the two coordinates  $l_1$  and  $l_2$  according to the following distribution:

-4	-3	-2	-1	0	1	2	3	4
0.04	0.08	0.12	0.16	0.20	0.16	0.12	0.08	0.04

and let the parent be the individual at  $(x_1 + l_1, x_2 + l_2)$ .

It may sound unreasonable to have an unborn individual pick its parents. To see that this is essentially equivalent to the more reasonable scheme in which the two neighbouring parents mate and their offspring disperse from the place at which mating occurs, consider the genealogies of a given individual in the two processes. In each case the distances from parent to child are independent with some fixed distribution.

Once parents are selected, they produce gametes in the usual way. For example, if the genetic state of a parent is  $AabB$  then the gamete contributed will be  $Ab$  or  $aB$  with

probability  $(1 - r)/2$  each, whereas recombinants  $AB$  and  $ab$  have probability  $r/2$ . To incorporate selection into the model we assign fitnesses

$$\phi(A,A,x), \phi(A,a,x) = \phi(a,A,x), \phi(a,a,x)$$

to the possible genotypes. In the model, if the  $A$  alleles of the proposed new individual are  $a_1$  and  $a_2$ , then with probability  $\phi(a_1, a_2, x)$ , the new replaces the old at  $x$  at time  $t + 1$ . With probability  $1 - \phi(a_1, a_2, x)$ , the proposed new individual is discarded and we allow the individual currently at  $x$  to hold on to the site in the next time period. Note that here selection acts by the mechanism: less fit types have lower probabilities of survival.

In the scheme described above it would also be possible to sample repeatedly until an acceptable individual is generated. We chose the ‘sample once’ approach to streamline the computer program. As all the fitnesses we study are at least 0.7 and most are at least 0.9, this will make little difference in the behaviour of the model. In this paper we focus on the ‘half-plane’ environment. We imagine the plane being divided in two by a vertical line, with the  $A$  allele having a selective advantage on the right, and the  $a$  allele having an advantage on the left. To describe this precisely, we let  $x_1$  be the first coordinate of the point  $x$  and define the fitnesses to be

$$\begin{array}{l} \phi(A,A,x) \quad \phi(A,a,x) = \phi(a,A,x) \quad \phi(a,a,x) \\ \text{if } x_1 > 0 \quad 1 \quad 1 - s/2 \quad 1 - s \\ \text{if } x_1 \leq 0 \quad 1 - s \quad 1 - s/2 \quad 1 \end{array}$$

The ‘half-plane environment’ is a simple mathematical idealization of a heterogeneous environment in which the two types have different fitnesses.

In addition to introducing the hybrid zone model, the primary focus of this paper is to consider several initial conditions that represent different hypotheses about hybrid zone origins. The first is our representation of primary intergradation or differentiation *in situ*. In this case colonists are drawn from a single polymorphic, randomly mating source population and arrive in large numbers to populate our environment.

*Random16.* We assign to each site one of the 16 possible states chosen at random. Here we are imagining the possibility that initial colonization of patches was independent of patch type.

The second and third initial conditions are examples of secondary contact, like those presumed to have occurred following postglacial range expansions in many temperate organisms which have ranges divided

by hybrid zones (Hewitt, 1989, 1993). To represent secondary contact, colonists are drawn from two genetically distinct source populations.

*Random2.* We randomly assign to each site either the state *AABB* or *aabb*. This is one version of secondary contact; for example, imagine two genotypes colonizing via long-distance dispersal so that initially nearby sites are independent.

*Front.* Our second version of secondary contact has individuals on the right half (i.e.  $x_1 > 0$ ) start in state *AABB* whereas those on the left are in state *aabb*. Here, we envisage two types that have expanded slowly into an unoccupied region.

Up to this point we have described the model on the two-dimensional integer lattice  $Z^2$ . The formulation of this model on an infinite grid is necessary, for example, to have a nontrivial equilibrium state for the hybrid zone involving the selected locus. Of course, when we simulate the system on the computer we have to restrict our attention to a finite set. With the use of arrays in the computer program in mind, we represent space as the set of pairs  $(m, n)$  where  $m$  and  $n$  are integers, say between 0 and 239. In this case, the left half becomes the sites with  $m < 120$  and the right those with  $m \geq 120$ . To avoid edge effects on the top and bottom we assume that sites on the top row  $n = 239$  are adjacent to those on the bottom  $n = 0$ . This has the effect of turning a flat plane into a tube but it has the advantage that the world looks the same seen from any site in a given column.

We do not want the sites on the left and right edges to be neighbours because that would give us two hybrid zones. On these edges we use the analogue of Neumann or no flux boundary conditions from partial differential equations. These are perhaps best described by example. If site (2, 37) decided to use the nonexistent individual at  $(-1, 35)$  as a parent, the request bounces off an imaginary wall at  $m = 0$  and lands at  $(1, 35)$ . This may sound odd but is calculated to avoid one trouble we ran into in early simulations. At first we used ‘fixed’ boundary conditions in which the finite grid is imagined to be surrounded by sites that are always *aabb* if  $m < 0$  and *AABB* if  $m > 239$ . In this case if we use the front start then the limiting frequency of *Bs* as time tends to infinity is not flat at 0.5, as one would expect, but instead is an approximate straight line that connects frequency 0 at  $m = -1$  to frequency 1 at  $m = 240$ . These values are caused by the unchanging individuals at the two ends.

## Simulation results

In this section we report the results of our simulations, organizing our findings by the statistic being investigated and the dependent variables being considered. In

each subsection we will compare our results with those in earlier work.

### Allele frequencies at the selected locus

When the *A* and *a* alleles each have a selective advantage  $s = 0.1$  in different halves of the half-plane environment, a steep cline across the boundary between the two environmental patches appears and reaches equilibrium within 100 generations. Figure 1 summarises the results of simulations from Random2, Random16 and Front starts and shows that the shape of the resulting cline at the selected locus is independent of the initial conditions. In this figure and those that follow, position  $j$  refers to a column  $j$  of the grid. Our estimate of the allele frequency for *A* (or *B*) at any position  $j$  is the fraction of alleles of that type in the five columns centred at  $j$  (i.e. columns  $j - 2, j - 1, j, j + 1$  and  $j + 2$ ).

The final panel in Fig. 1 shows the equilibrium prediction for the associated partial differential equation (PDE) model introduced by Haldane (1948). In the notation of eqn 11 on p. 739 of Slatkin (1973) this may be written as

$$A(\xi) = -2 + 3 \tanh^2\left(\frac{\xi}{2} + \tanh^{-1}\left(\sqrt{\frac{2}{3}}\right)\right),$$

where  $\xi = s^{1/2}x/\sigma$ .

A more careful comparison between the PDE prediction and the spatial model is given in Fig. 2. There the analytical result is overlaid on the column averages for a simulation on a  $480 \times 480$  grid when  $s = 0.01$ . As one can see the agreement is quite good.

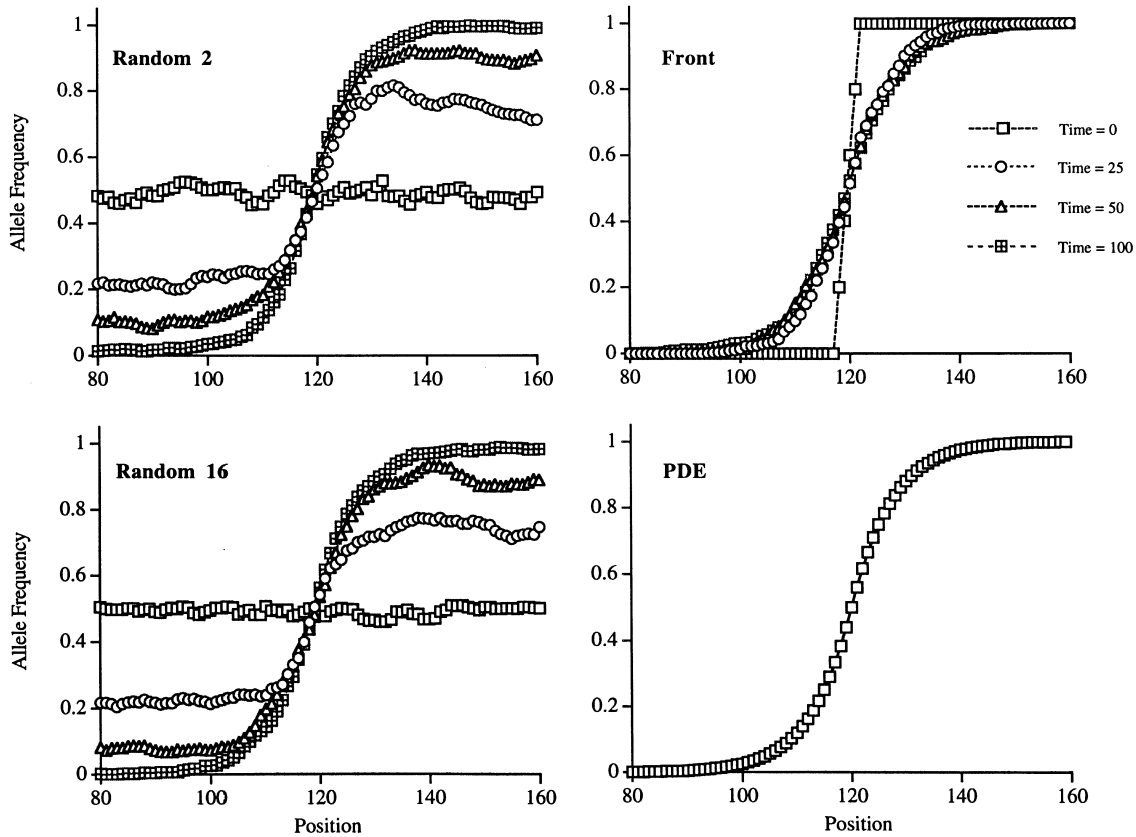
### Disequilibrium vs. initial condition

We investigated the state of the hybrid zone 500 generations after colonization for the three different initial conditions, when  $s = 0.1$  and recombination probability between neutral and selected loci is  $r = 0.01$  per generation. Linkage disequilibrium and its normalized version for column  $j$  were again calculated from the statistics of columns  $j - 2 \leq k \leq j + 2$  by using the formulae

$$D = p_{AB} - p_A p_B; \quad \hat{D} = \frac{D}{[p_A(1 - p_A) \cdot p_B(1 - p_B)]^{1/2}},$$

where  $p_A$  and  $p_B$  are the frequencies of the *A* and *B* alleles and  $p_{AB}$  is the frequency of the *AB* combination. In subsequent simulations, measurements of allele frequency and linkage disequilibrium are always made in this way.

As Fig. 3 shows, patterns of variation at the neutral locus and the magnitude of linkage disequilibrium

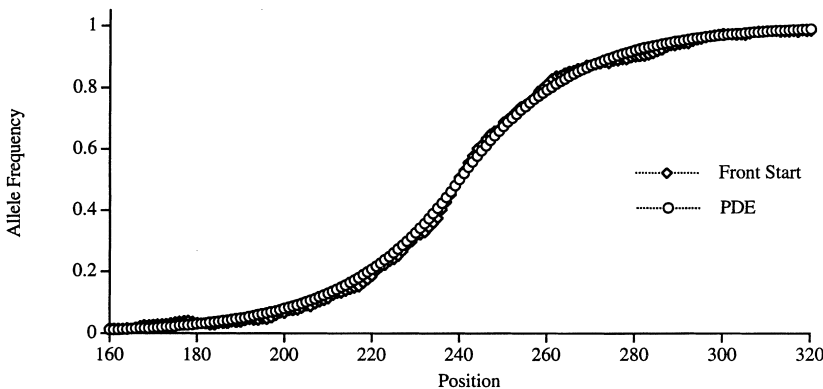


**Fig. 1** Clines in allele frequency for three different initial conditions. In these simulations, within 50–100 generations selection ( $s = 0.1$ ) results in the same cline regardless of initial conditions. Further, the limiting curve is close to the predictions of Haldane (1948) based on partial differential equations.

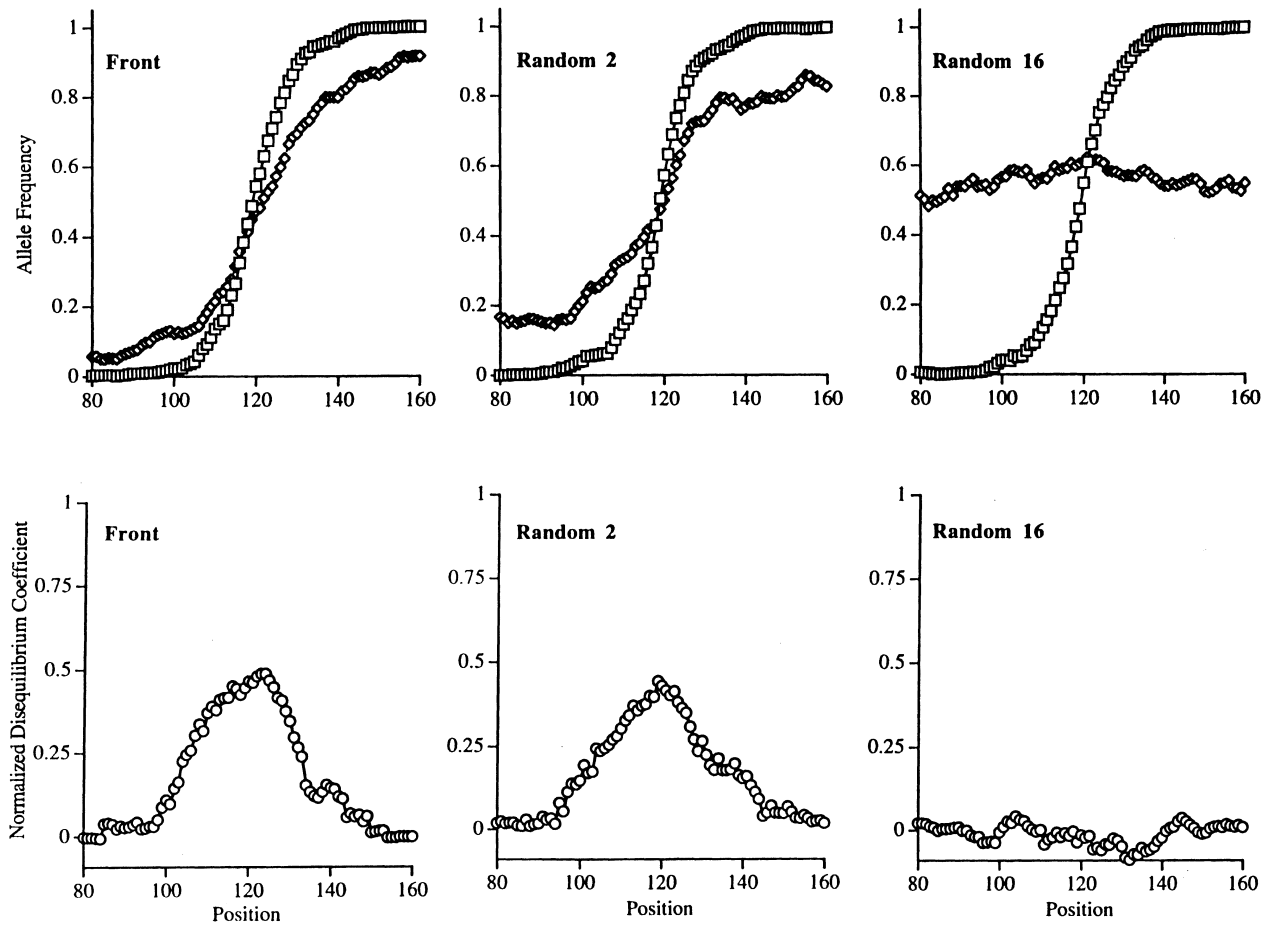
between the selected and neutral loci do depend on initial conditions. Following a Front or Random2 start (both of which involve colonization of the two half-plane environments by the two homozygous genotypes  $AABB$  and  $aabb$ ), a steep cline is established at the neutral locus and substantial disequilibrium (measured by the normalized disequilibrium coefficient) occurs at the boundary between the two patch types. In contrast, when the initial colonization event involves all

possible two-locus genotypes with equal probability (Random16), allele frequencies at the neutral locus are always uncorrelated with the state of the locus under selection and no cline develops (Fig. 3).

As we saw in Figs 1 and 3, the outcomes of Random2 and Front starts are very similar when selection is strong relative to recombination. However, when  $s$  and  $r$  are of similar magnitudes (e.g.  $r = 0.01$ ,  $s = 0.02$ ), the Random2 start results in a shallow cline at the neutral



**Fig. 2** A comparison of the PDE prediction and the spatial model equilibrium distribution on a  $480 \times 480$  grid when the selective advantage  $s = 0.01$ .



**Fig. 3** Allele frequencies after 500 generations for the locus under selection (squares) and a linked neutral locus (diamonds). The bottom graphs show values of the normalized linkage disequilibrium (circles). For all simulations,  $s=0.1$  and  $r=0.01$ . Note that there is no neutral cline or disequilibrium for the Random16 case.

locus and very little linkage disequilibrium even at the centre of the hybrid zone (Fig. 4). Under these conditions, recombination decouples alleles at the selected and neutral loci before a steep cline is established at the selected locus. In contrast, with a Front start, a steep cline exists for both loci immediately after contact, and linkage disequilibrium has not disappeared after 500 generations (Fig. 4).

It is important to note that others have approached these problems using a ‘continuum model’ of dispersal and selection introduced by Nagylaki (1975). For example, Barton (1979b) wrote a pair of partial differential equations for  $u(x)$ , the frequency of  $B$  at the neutral locus, and the disequilibrium  $D$ :

$$\frac{du}{dt} = \frac{m}{2}u'' + s(x)D$$

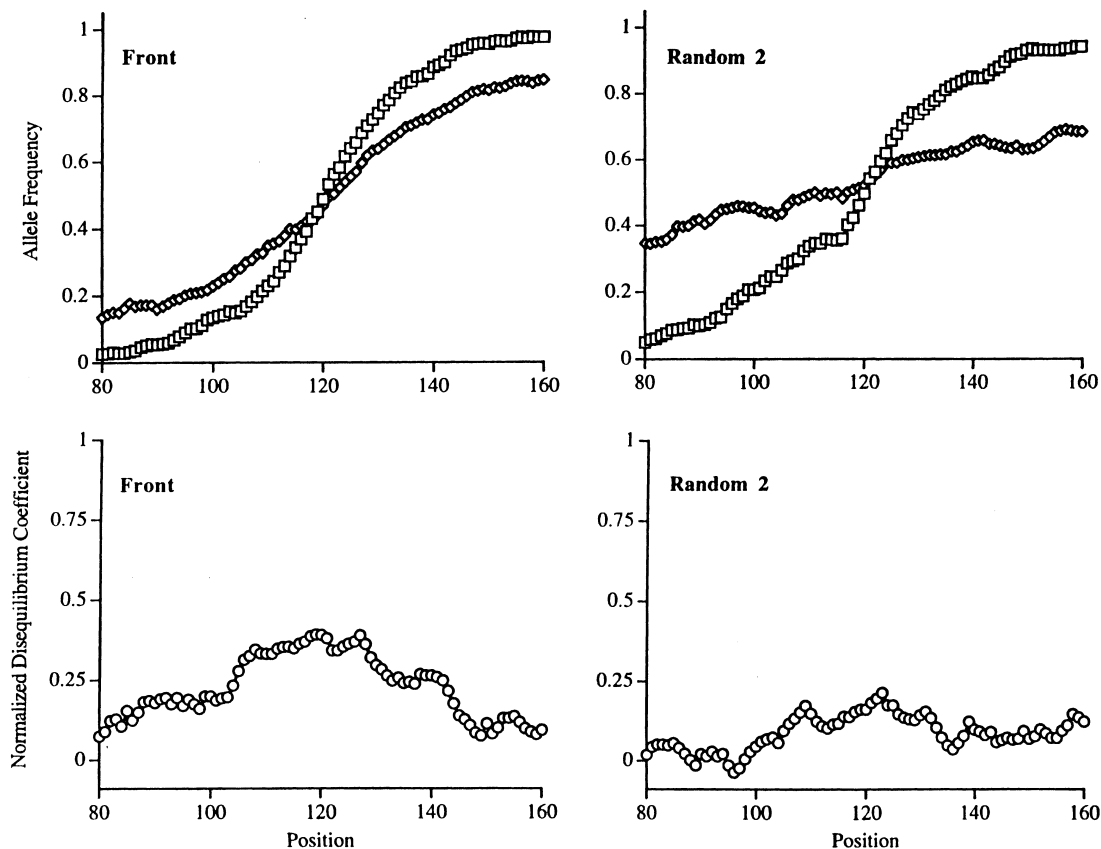
$$\frac{dD}{dt} = -rD + \frac{m}{2}D'' - s(x)D(p - q) + mp'u'$$

where  $p$  and  $q$  are the frequencies of the  $A$  and  $a$  alleles, and  $m$  is the migration rate.

Although these equations are easy to write, they are difficult if not impossible to solve explicitly. Barton (1979b) extracted some information about the solutions by considering the extreme cases in which one of  $r$  and  $s$  is much larger than the other. An advantage of our approach is that one can directly treat any combination of  $r$  and  $s$  rather than extrapolating from extreme cases. Of course, one can also do this by solving the PDE numerically.

#### *Disequilibrium vs. recombination and selection*

Patterns of variation across the hybrid zone depend on the relative magnitudes of selection and recombination. Based on the discussion in the previous subsection, we expect similar results from a Front or a Random2 start, so, for simplicity, we will only investigate the latter possibility. When  $s$  is large and  $r$  is small a steep neutral



**Fig. 4** Allele frequencies (squares = selected locus, diamonds = neutral locus) and normalized linkage disequilibria (circles) after 500 generations when recombination ( $r = 0.01$ ) and selection ( $s = 0.02$ ) are both weak. The Front start begins with a steep cline and large disequilibrium, but these never develop in the Random2 start.

cline, similar in slope to that of the selected locus, is produced (Fig. 5). Only when the cline at the neutral locus is steep is there detectable linkage disequilibrium near the centre of the zone (Fig. 6). When the neutral and selected loci are not tightly linked (e.g.  $r = 0.1$ ), even selection coefficients of 0.3 will not produce a steep neutral cline or maintain disequilibrium 500 generations after a Random2 start.

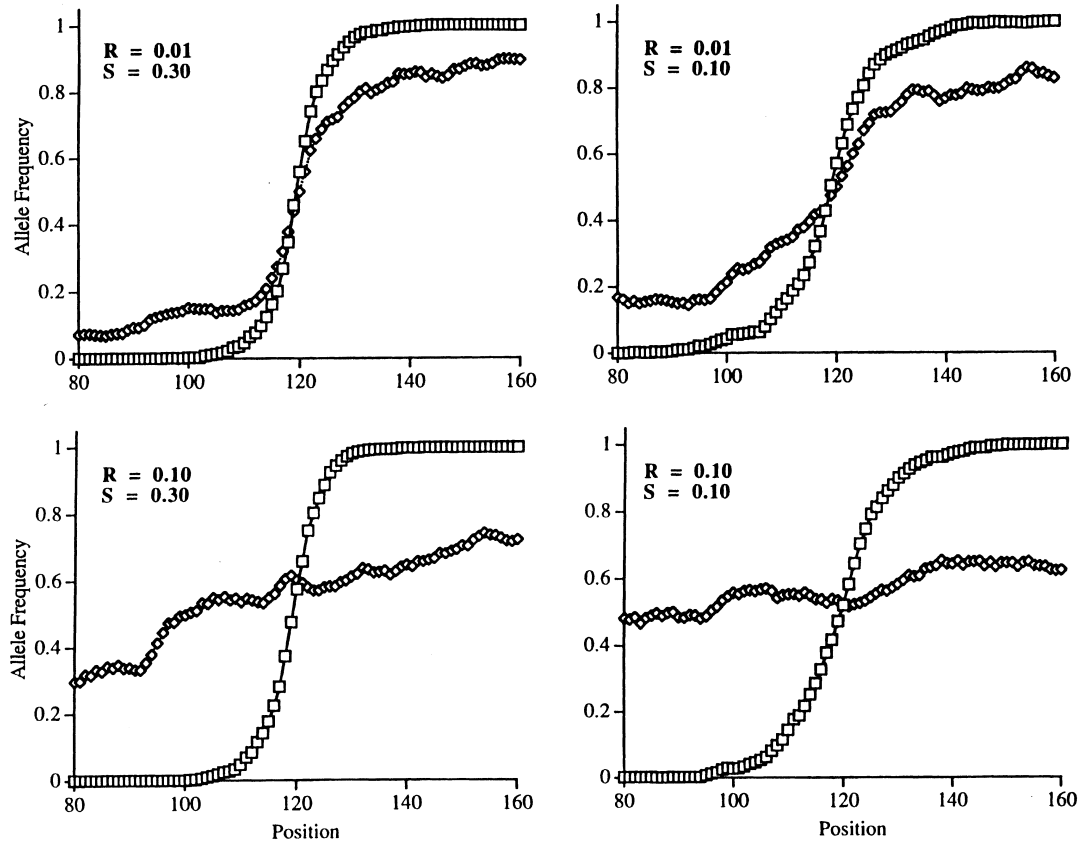
For any given strength of selection, the amount of linkage disequilibrium depends on the rate of recombination and the time since the two-patch environment was colonized. If the neutral and selected loci are not very tightly linked, then even with moderately strong selection ( $s = 0.1$ ) disequilibrium decays rapidly. In such cases, we would expect to detect disequilibrium only if we sampled a hybrid zone soon after it became established by secondary contact. However, when  $r$  is much smaller than  $s$  (e.g.  $r = 0.005$ ,  $s = 0.1$ ), substantial disequilibrium can persist for thousands of generations.

Because the number of generations to have an average of one recombination event is  $1/r$ , this should give the time scale on which disequilibrium disappears.

To make this point, Fig. 7 plots normalized disequilibrium vs.  $rN$ , where  $N$  is number of generations. From this graph we see that when  $rN$  is about 50, the normalized disequilibrium has dropped to about 0.1. For  $r = 0.02$  this is 2500 generations, for  $r = 0.005$  this is 10 000 generations.

It is interesting to compare the last simulation of our spatial model with the theoretical result that in a homogeneously mixing infinite population with a Random2 initial condition the disequilibrium decays like  $(1 - r)^n$  (see, e.g. Maynard Smith & Haigh (1974)). To compare this prediction with our simulation, we can return to Fig. 7 and note that the disequilibrium at time  $60/r$  is about 0.1. Using  $(1 - r)^{1/r} \approx e^{-1}$  and solving  $e^{-60c} = 0.1$ , we see that the disequilibrium in generation  $n$  is approximately  $(1 - r)^{cn}$ , where  $c = (\ln 10)/60 = 0.0383$ . Thus, the disequilibrium decays more than 25 times slower in our spatial model than in a homogeneously mixing one.

As in the case of the previous two subsections, one can also approach these problems using partial differential equations. Calculations on pages 417–419 of Barton (1986) (see also page 334 of Barton (1979a) and



**Fig. 5** Allele frequencies after 500 generations for the locus under selection (squares) and a linked neutral locus (diamonds). When  $r$  is small (0.01) a steep cline is observed at the neutral locus, but not when  $r$  is large (0.1), even under strong selection.

(1) on page 20 of Barton & Gale (1993)), show that if selection is much weaker than recombination then the strength of linkage disequilibrium should be

$$D = \frac{\sigma^2}{r} \cdot \frac{dp_A}{dx} \cdot \frac{dp_B}{dx},$$

where  $\sigma$  is the dispersal range and  $p_A$ ,  $p_B$  are the frequencies of  $A$  and  $B$  alleles.

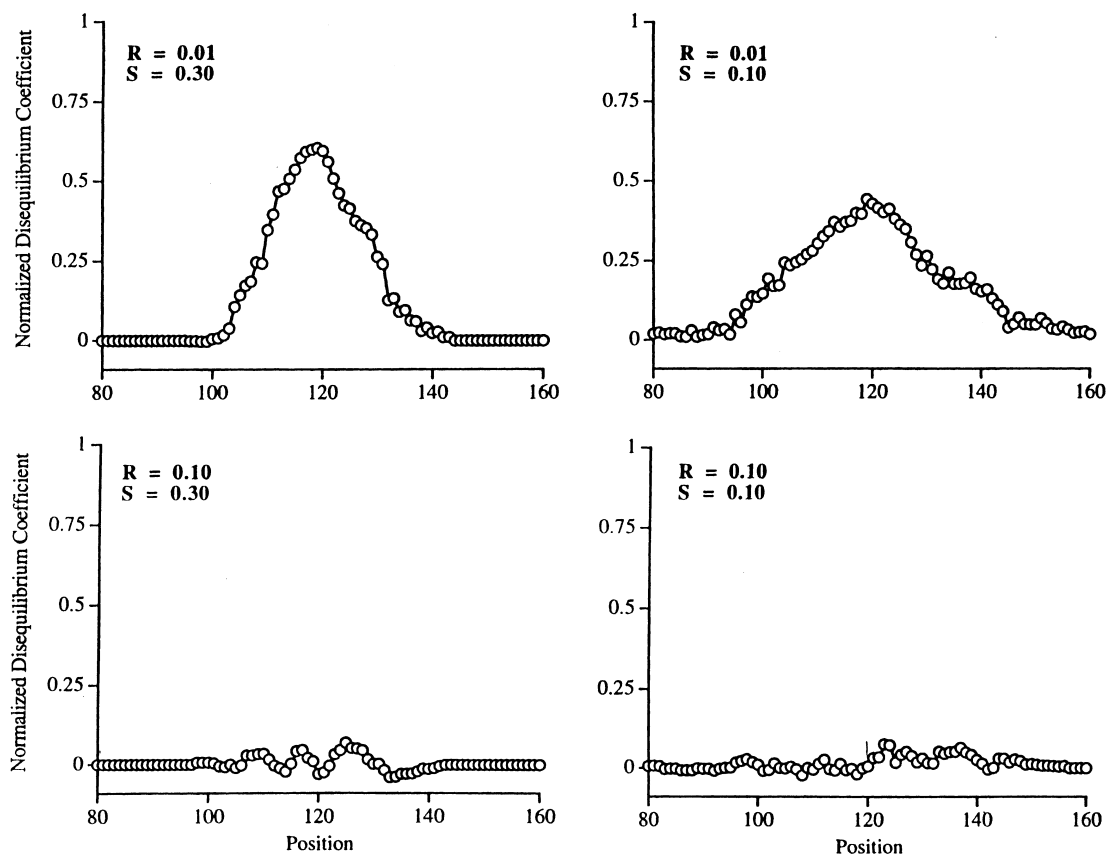
To compare this prediction with our results we carried out a number of simulations. When  $r = s = 0.01$  we found very little disequilibrium at any time and we had difficulty obtaining good estimates of the two quantities to be compared. Taking  $s = 0.04$  and using 20 simulation runs to calculate the normalized disequilibrium after 200 generations gives the result shown in Fig. 8. The extra runs were needed to reduce the variance in our estimates so that we could estimate the derivatives of the allele frequencies by taking differences of the measurements in adjacent columns. Figure 8 shows (i) the direct estimate from simulation of the disequilibrium, and compares it with (ii) the curve that results from using simulation estimates of  $dp_A/dx$  and  $dp_B/dx$  in the approximate

formula for  $D$  given above. Note that the curve derived from the formula has a higher peak and drops more quickly to 0 as we move away from the centre of the hybrid zone.

Our results do not contradict Barton's. We have  $r = 0.01$  and  $s = 0.04$ , so the assumption that  $s$  is much smaller than  $r$  does not apply. However, these observations do not support Barton's (1986) speculation on page 418 that 'the assumption that selection is much weaker than recombination may be unduly restrictive' for the approximation to be good. Indeed returning to the derivation of the approximation given on page 334 of Barton (1979a), one sees that he drops two terms from the PDE by setting  $s = 0$  and these terms will be important when  $r$  and  $s$  are of comparable size. On the other hand, Barton's formula is useful when  $r$  is much larger than  $s$  and the low level of disequilibrium present in the system is difficult to estimate from simulations.

## Discussion

We have presented a simple, spatially explicit model of a hybrid zone that allows us to examine patterns of



**Fig. 6** Patterns of normalized linkage disequilibria for the four simulations in Fig. 5. Only if  $r$  is small (0.01) is there significant disequilibrium in the centre of the hybrid zone.

variation over thousands of generations, given defined levels of selection, recombination and dispersal. We can use such a model to investigate the possible causes of patterns of variation observed in natural hybrid zones, and to predict their ongoing dynamics. In this paper we have focused primarily on a simple case in which selection is a consequence of differential survivorship in two different environments which are adjacent rectangular 'half-planes.'

We have defined initial conditions for hybrid zones in terms of the distribution of two-locus genotypes and the associated patterns of linkage disequilibrium. Our three initial conditions are explicit realizations of classical models of hybrid zone origin from Mayr (1942). The Random16 start corresponds to primary intergradation or differentiation *in situ*, whereas Front and Random2 correspond to secondary contact. Our results show that unless selection and recombination are both weak, after a few hundred generations there will be very little difference in outcome between the two scenarios for secondary contact. In both cases, the system settles rapidly into almost identical clines at the selected and neutral loci; there is significant linkage disequilibrium

early, which slowly decays with a rate determined by the recombination rate.

Endler's argument (Endler, 1977, p. 155) that one can distinguish primary intergradation and secondary contact only in recently established hybrid zones is based on the observation that in single-locus cline models with selection and gene flow, the approach to equilibrium is rapid. Therefore, given the same values for gene flow and selection, identical patterns of variation will be established quickly regardless of starting conditions. Our simulations (Fig. 1) support this conclusion for a single locus under selection. However, the behaviour of neutral loci is quite different, with patterns of variation depending both on the initial conditions and on linkage to a locus under selection. When selection is strong and the neutral and selected loci are very tightly linked, the signature of origin via secondary contact (clinal variation at neutral loci and significant disequilibrium at the centre of the hybrid zone) can persist for thousands of generations (see Fig. 7). In contrast, our model of primary intergradation (Random16) predicts that allele frequencies at neutral loci will not vary clinally across the hybrid zone and that disequilibrium between a



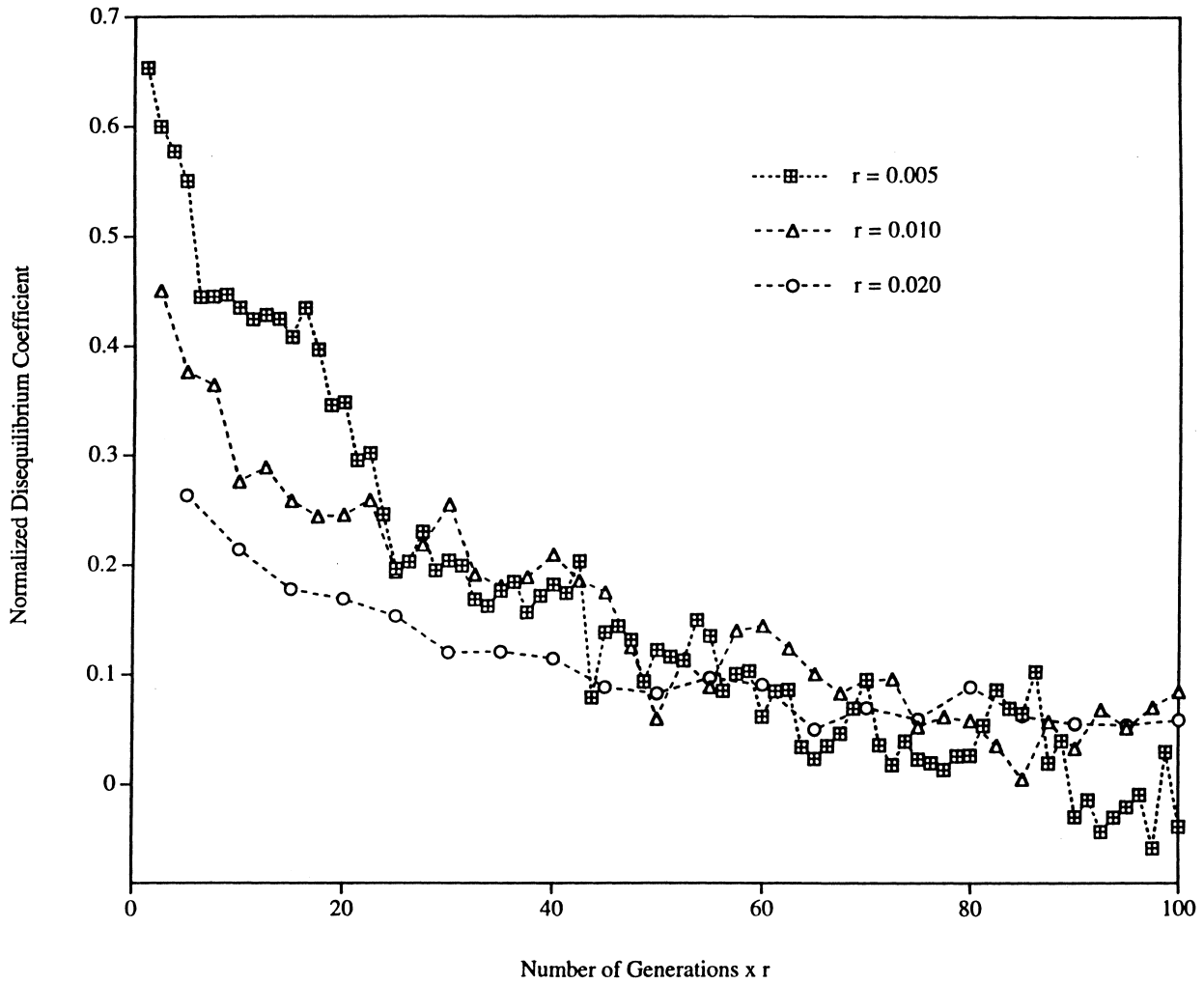


Fig. 7 Normalized linkage disequilibrium for the Random2 initial condition plotted against recombination times number of generations ( $rN$ ). Values are the average of 10 runs. It takes  $\approx 60/r$  generations for the normalized disequilibrium to drop to 0.1.

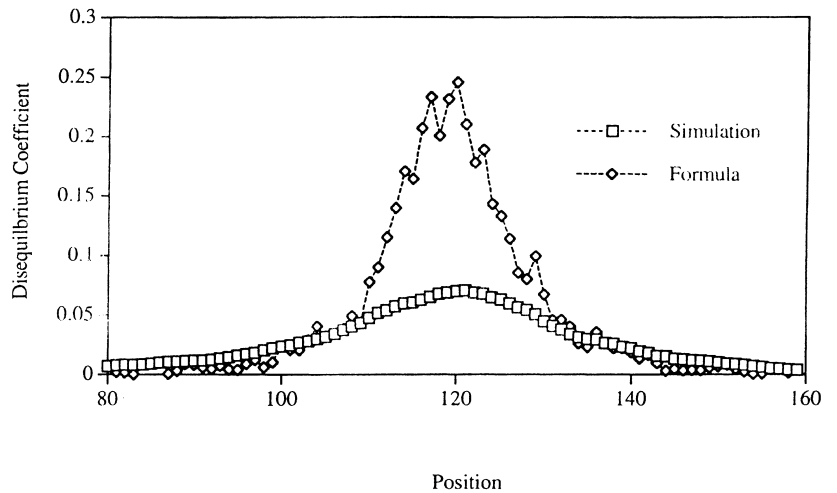


Fig. 8 Plot of Barton & Gale's (1993) formula for disequilibrium in terms of allele frequencies vs. a direct measurement from simulation at time 200 when  $r = 0.01$  and  $s = 0.04$ . Note that the analytical result has a sharper peak and more rapid decay than in the simulation.

neutral locus and a closely linked locus under selection will decay to 0 quickly (Figs 3 and 4).

In many natural hybrid zones, transects across the zone reveal parallel clinal variation of allele or haplotype frequencies (e.g. Hunt & Selander, 1973; Patton *et al.*, 1979; Sage & Selander, 1979; Harrison, 1986; Howard, 1986; Szymura & Barton, 1986, 1991; Good, 1989; Harrison & Bogdanowicz, 1997). The data often are derived from electrophoretic surveys of protein (allozyme) variation or from DNA restriction fragment length polymorphisms. These polymorphic molecular markers are often assumed to be effectively neutral, both with respect to the sort of environmental heterogeneity that we model in our half-plane environment and with respect to hybrid fitness. When markers are neutral, patterns of clinal variation must be a consequence of very recent secondary contact or of tight linkage to loci under selection (either spatially varying selection or hybrid unfitness). The age of most hybrid zones is difficult to ascertain directly, although some north temperate hybrid zones appear to have been established five to 15 thousand years ago, soon after retreat of the glaciers (Hewitt, 1989). If molecular markers are neutral and contact occurred more than a few thousand years before the present, persistence of clines for polymorphic allozyme or DNA markers argues for very tight linkage to loci under selection. The observation that many nuclear gene markers vary in parallel across hybrid zones would then seem to imply that a substantial fraction of the genome experiences the effects of selection.

## Acknowledgements

Preliminary work on the model was supported in part by the Systematic Biology Program at NSF (grant BSR-8906707 to R.G.H.). R.T.D. was supported by NSF grants DMS 93-01070 and DMS 96-26201. L.A.B. was supported by NSF grant BIR 94-23339. Simulations were carried out on computers at the Cornell National Supercomputer Facility. We would like to express our appreciation to an anonymous referee for providing us with a number of valuable suggestions and references.

## References

- BARTON, N. H. 1979a. Gene flow past a cline. *Heredity*, **43**, 333–339.
- BARTON, N. H. 1979b. The dynamics of hybrid zones. *Heredity*, **43**, 341–359.
- BARTON, N. H. 1986. The effects of linkage and density dependent regulation on gene flow. *Heredity*, **57**, 415–426.
- BARTON, N. H. AND GALE, K. S. 1993. Genetic analysis of hybrid zones. In: Harrison, R. G. (ed.) *Hybrid Zones and the Evolutionary Process*, pp. 13–45. Oxford University Press, Oxford.
- BARTON, N. H. AND HEWITT, G. M. 1981. Hybrid zones and speciation. In: Atchley, W. R. and Woodruff, D. S. (eds) *Evolution and Speciation: Essays in Honour of M. J. D. White*, pp. 109–145. Cambridge University Press, Cambridge.
- BARTON, N. H. AND HEWITT, G. M. 1985. Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.*, **16**, 113–148.
- BARTON, N. H. AND HEWITT, G. M. 1989. Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.
- COX, J. T. AND DURRETT, R. 1995. Hybrid zones and voter model interfaces. *Bernoulli*, **1**, 343–370.
- ENDLER, J. A. 1977. *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, NJ.
- GOOD, D. A. 1989. Hybridization and cryptic species in *Dicamptodon* (Caudata: Dicamptodontidae). *Evolution*, **43**, 728–744.
- HALDANE, J. B. S. 1948. The theory of a cline. *Genetics*, **48**, 277–284.
- HARRISON, R. G. 1986. Pattern and process in a narrow hybrid zone. *Heredity*, **56**, 337–349.
- HARRISON, R. G. 1990. Hybrid zones: windows on evolutionary process. In: Futuyma, D. and Antonovics, J. (eds) *Oxford Surveys in Evolutionary Biology*, vol. 7, pp. 69–128. Oxford University Press, Oxford.
- HARRISON, R. G. 1993. Hybrids and hybrid zones: historical perspective. In: Harrison, R. G. (ed.) *Hybrid Zones and the Evolutionary Process*, pp. 3–12. Oxford University Press, Oxford.
- HARRISON, R. G. AND BOGDANOWICZ, S. M. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution*, **51**, 493–505.
- HEWITT, G. M. 1989. The subdivision of species by hybrid zones. In: Otte, D. and Endler, J. A., (eds) *Speciation and its Consequences*, pp. 85–110. Sinauer Associates, Sunderland, MA.
- HEWITT, G. M. 1993. After the ice: *parallelus* meets *erythropus* in the Pyrenees. In: Harrison, R. G., (ed.) *Hybrid Zones and the Evolutionary Process*, pp. 140–164. Oxford University Press, New York.
- HOWARD, D. J. 1986. A zone of overlap and hybridization between two ground cricket species. *Evolution*, **40**, 34–43.
- HUNT, W. G. AND SELANDER, R. K. 1973. Biochemical genetics of hybridization in European house mice. *Heredity*, **31**, 11–33.
- MAYNARD SMITH, J. AND HAIGH, J. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.*, **23**, 23–35.
- MAYR, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York.
- MAYR, E. 1963. *Animal Species and Evolution*. Belknap Press, Cambridge, MA.
- NAGYLAKI, T. 1975. Conditions for the existence of clines. *Genetics*, **80**, 595–615.
- PATTON, J. L., HAFNER, J. C., HAFNER, M. S. AND SMITH, M. F. 1979. Hybrid zones in *Thomomys bottae* pocket gophers: genetic, phenetic and ecologic concordance patterns. *Evolution*, **33**, 860–876.

- SAGE, R. D. AND SELANDER, R. K. 1979. Hybridization between species of the *Rana pipiens* complex in central Texas. *Evolution*, **33**, 1069–1088.
- SLATKIN, M. 1973. Gene flow and selection in a cline. *Genetics*, **75**, 733–756.
- SZYMURA, J. M. AND BARTON, N. H. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *Bombina variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141–1159.
- SZYMURA, J. M. AND BARTON, N. H. 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**, 237–261.