

# The genetic consequences of long distance dispersal during colonization

RICHARD A. NICHOLS\* & GODFREY M. HEWITT†

School of Biological Sciences, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS

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

Rare long distance dispersal may have little impact on gene frequencies in established populations but it can dramatically increase gene flow during episodes of range expansion. We model the invasion of new territory by genetically distinct populations of the same species to investigate the dynamics of such episodes. If long distance dispersal is sufficiently frequent, the populations do not spread as a wave of advance but instead found intermingled isolates. We argue that this process can explain many otherwise puzzling patterns in the geographical distribution of alleles.

**Keywords:** colonization, dispersal, founder events, gene flow, hybrid zone, travelling wave.

## Introduction

In most species, populations from different parts of the species range are genetically distinct. It has become increasingly apparent that in order to interpret this geographical variation we need to understand the history of a species' distribution (Hewitt, 1989; Avise *et al.*, 1987) and episodes of range expansion have particularly dramatic genetic consequences. The process of range expansion can produce genetic patterns that persist for many hundreds or thousands of generations; for example the present day distribution of human blood groups can be traced back to the consequences of the Neolithic agricultural revolution (Ammerman & Cavalli-Sforza, 1984), and hybrid zones in many plants and animals seem to have formed soon after the last ice age (Barton & Hewitt, 1985).

Fisher's (1937) influential model characterized the spread of an advantageous gene through a population as an advancing wave. Subsequently population expansion has also been modelled as an advancing wave (Skellam, 1951; van den Bosch *et al.*, 1988). In such models the dispersal is characterized by the variance ( $\sigma^2$ ) in parent-offspring distance (Hengeveld, 1989). However, natural examples of range expansion often do not involve a simple wave of advance; instead long distance colonizations set up new populations distant from the parent population. These new populations then act as foci for local spread. Long distance coloni-

zations have been implicated in the spread of organisms ranging from cholera and potato blight (*Pytophophthora infestans*) to cheat grass (*Bromus tectorum*) and oak trees (*Quercus* spp.) (Pyle, 1969; Cliff *et al.*, 1981; van der Plank, 1967; Mack, 1981; Hengeveld, 1989). Indeed, the post-glacial advance of many organisms seems to have been very rapid, with clear evidence from fossil pollen and beetles (Huntley & Birks, 1983; Bennett, 1988; Coope, 1990).

The theory of the dynamics of range expansion has been extensively studied by Mollinson (1977). He identifies an important transition in the behaviour of an expanding population depending on the form of the function relating dispersal to distance:  $V(x)$ . If  $V$  has exponentially bounded tails then the population spread tends to proceed as a wave of advance. In stochastic simulations where  $V$  has thicker tails then the population tends to spread in leaps and bounds.

In this paper we investigate the genetic consequences of this difference in dispersal behaviour when two genetically distinct populations meet. The work was stimulated by surveys which revealed genetic mixing between races of the grasshopper *Chorthippus parallelus* in the Pyrenees. The measured rate of dispersal ( $\sigma = 30$  m) is insufficient to account simply for the penetration of allozyme markers, morphological and behavioural characters some 20 km into the range of the other race (Butlin & Hewitt, 1985a,b; Hewitt, 1989; Butlin *et al.*, 1991). We therefore developed a simulation that could emulate the two types of range expansion.

\*Correspondence.

## A model of range expansion

The model incorporates migration and population growth with selection and/or genetic drift acting on two alleles at a single locus. It consists of a rectangular array of demes, 79 by 40. The direction of the shorter axis is designated North.

### Migration

The number of migrants from each deme is drawn from the binomial distribution with parameters  $m$  (migration rate) and  $N_t$  (the number of adults in the deme). Migrants are chosen at random from the deme and moved to their new population. A dispersal function (see below) specifies the distribution of displacements for the new population from the old. For each migrant a random direction was chosen uniformly from 0 to 360° and a displacement drawn from the dispersal distribution. Migrants crossing the eastern and western boundaries were discarded. The northern and southern boundaries were connected, so that a migrant crossing to the north appeared in the south and *vice versa*.

### Population growth

The population in a deme grows until it reaches the carrying capacity according to the equation:

$$N_{t+1} = N'_t + rN'_t(k - N'_t)/k, \quad (1)$$

where  $N'_t$  is the population size after migration,  $N_{t+1}$  is the adult population size in the next generation,  $r$  is the intrinsic rate of increase and  $k$  is the carrying capacity (20 individuals in these simulations).

### Genetic drift and selection

The two alleles in the population can be designated  $A$  and  $B$ . The frequency of  $A$  in adults after migration ( $p_t$ ) determines the frequency in the gamete pool. Each allele in the following generation is of type  $A$  with probability:

$$p_{t+1} = \frac{p_t^2 w_{AA} + p_t(1 - p_t)w_{AB}}{p_t^2 w_{AA} + 2p_t(1 - p_t)w_{AB} + (1 - p_t)^2 w_{BB}}, \quad (2)$$

where  $w_{AB}$  is the fitness of the  $AB$  genotype.

Hence the number of  $A$  alleles in generation  $t + 1$  is drawn from the binomial distribution with parameters  $p_{t+1}$  and  $N_{t+1}$ .

### Modelling colonization

In the runs reported here, the model was set up to compare the genetical consequences of dispersal patterns

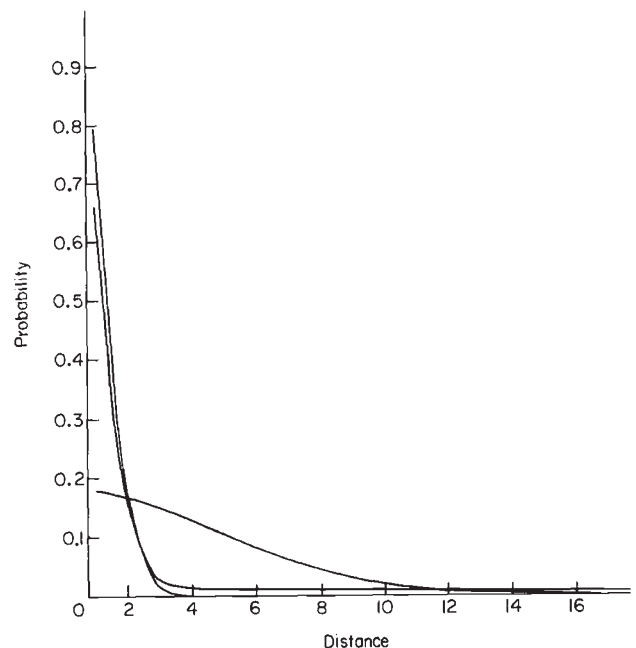
during colonization. In the light of Mollinson's (1977) results, a dispersal function was chosen for which the tails of the distribution could be modified but the variance kept constant. This was achieved using the weighted sum of two normal distributions:

$$(1 - a)N[0,1] + aN[0,b].$$

$N[x,\sigma]$  represents the normal distribution with mean  $x$  and variance  $\sigma^2$ . The unit of distance was the spacing between demes. The series of distributions used here was chosen to have a range of values for  $a$ , and a value of  $b$  such that the variance remained 25. The proportion of individuals that migrate ( $m$ ) was set to 0.3 and the rate of population growth ( $r$ ) was set to 0.9 so that there was some lag (approximately seven generations) between colonization and a deme growing to full size. These combinations lead to colonization of the array of demes in around 25 generations which was sufficiently short for repeated simulation. The range of values of interest makes  $\ln(a)$  (hereafter  $\alpha$ ) a convenient measure (Table 1). Figure 1 illustrates the function for the two extreme values and one intermediate:  $\alpha = 0, 2$  and 4.5. The intermediate and larger values of  $\alpha$  pro-

**Table 1** The set of parameter values for the dispersal function

$\alpha$	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
$b$	5	5.7	6.9	8.5	10.7	13.5	17.1	21.9	28.9	46.3



**Fig. 1** Dispersal functions. The distribution of distances moved by migrants. Curves are plotted for  $\alpha = 0, 2$  and 4.5 (lowest to highest at distance = 1).

duce distributions which are very similar over short distances but differ in the tails.

This range of functions produces patterns of colonization which encompass both a conventional steady Fisherian wave of advance ( $\alpha = 0$ ) and jerky progress by a sequence of long distance colonizations ( $\alpha = 4.5$ ). The two extremes are shown in Fig. 2. Where  $\alpha = 4.5$ , a mottled pattern of colonies are set up ahead of the front and these act as foci which expand and coalesce. Single demes colonized ahead of the front in generation 3 have become large patches by generation 10.

### Colonization dynamics

Three replicate simulations were run for each parameter combination. They were initiated with the easternmost demes populated by AA individuals and the westernmost with BB. In one series there was selection of intensity 0.5 against heterozygotes ( $W_{AA} = W_{BB} = 1$ ;  $W_{AB} = 0.5$ ), in the other the alleles were neutral ( $W_{AA} = W_{BB} = W_{AB} = 1$ ).

### Results

Once the intervening territory had been colonized, there was a cline in the frequency of the A allele from west to east. The width of this cline was evaluated as a measure of genetic introgression after different times. The width parameter  $w$  was estimated by fitting the following equation to the average gene frequency ( $p$ ) at each distance ( $x$ ) from the western margin:

$$p = 0.5 + 0.5(e^{\delta} - e^{-\delta}) / (e^{\delta} + e^{-\delta}),$$

where  $\delta = (x - c)/w$  and  $c$  is the centre of the cline fitted for each replicate and  $w$  is fitted across replicates.

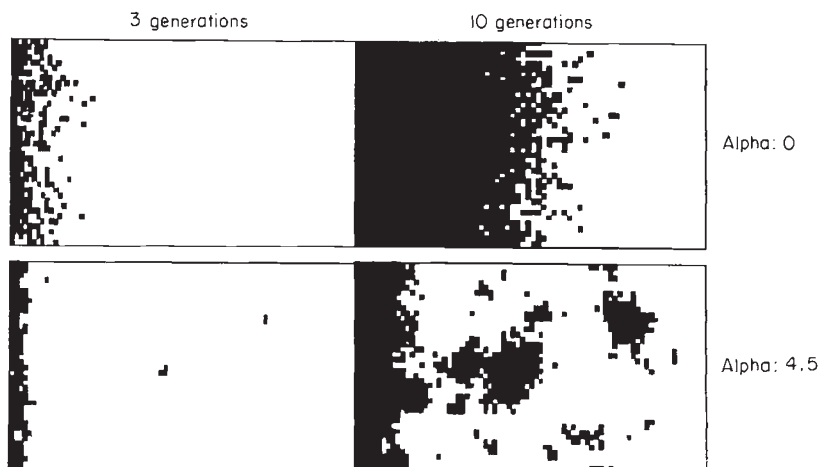
The introgression of neutral alleles increased markedly with  $\alpha$  (Fig. 3). This pattern has a simple

explanation. When  $\alpha = 0$  the eastern and western populations spread as expanding fronts. The fronts met forming a sharp transition in A frequency. With increasing  $\alpha$ , some dispersing individuals jumped ahead of the fronts and set up new expanding colonies of a few demes. The predominantly AA colonies were established intermingled with those that were predominantly BB in a broad band between the two main fronts. As the colonies spread, met and exchanged migrants, a wide cline was formed. In all cases, the clines continued to get wider with time.

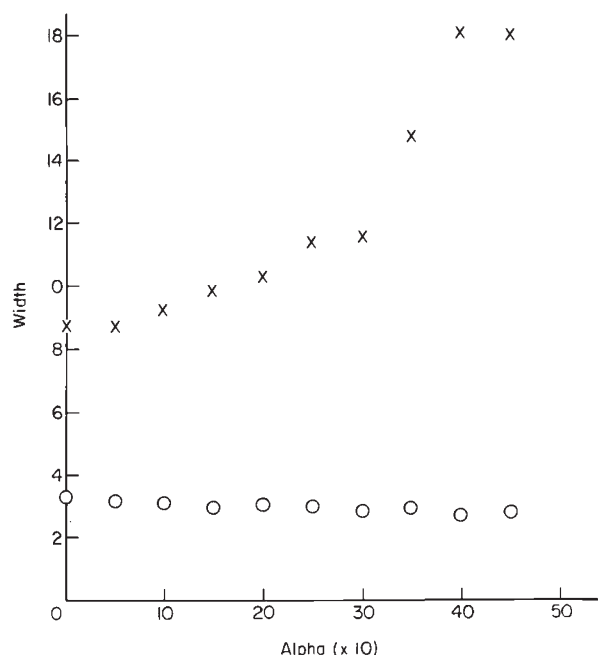
The case of underdominance (selection against heterozygotes) was dramatically different. The clines were initially broader when  $\alpha$  was large but within a few hundred generations they actually narrowed to an equilibrium width (Fig. 4). The equilibrium is produced by a balance between dispersal and selection and to a first approximation is expected to be a function of the dispersal variance (Barton, 1979). Indeed the equilibrium width does not vary dramatically with  $\alpha$  (Fig. 3). There is a very small but highly significant trend but in this case width decreases with  $\alpha$  ( $F_{[1,8]} = 51.3$ ;  $P \leq 0.001$ ). This can be explained by the ineffectiveness of the long distance migrants as agents of gene flow once the demes have become established. If they arrive in foreign territory, their alleles will be at low frequency and hence immediately selected against (eqn 2). They therefore have negligible impact compared with the period when they arrived in virgin territory and could establish local colonies.

### Discussion

The simulation results indicate that rare, long distance dispersal could produce genetic introgression over long distances during those rare episodes when the same tract of new territory is colonized by diverse genomes. Such dynamics could account for clines seen



**Fig. 2** Maps of the demes with extant populations (shaded squares) after three and ten generations for the two extreme dispersal functions. Only the most westerly demes were inhabited initially.

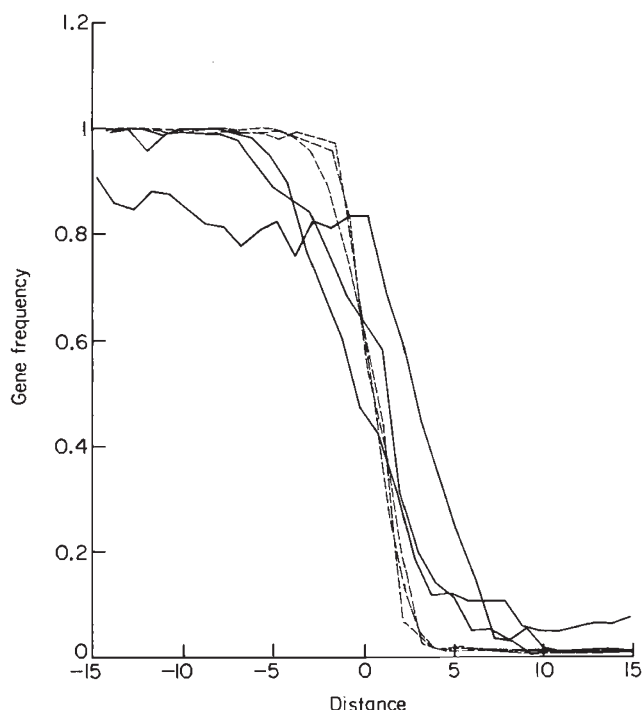


**Fig. 3** The effect of the shape of the dispersal function on cline width. The width parameter ( $w$ ) is plotted against  $\alpha$  for neutral alleles after 25 generations (crosses) and underdominant alleles after 500 generations (circles). Each width was estimated over three replicates.

in a great variety of species which are wide relative to the dispersal variance but stable. Well studied examples include *Peromyscus*, *Thomomys* and *Lepomis*; the widths of these and many other transitions are illustrated in Barton & Hewitt (1985). In order to assess the relevance of the model, we need to address two broad questions: do many species have the attributes that were represented in the model and can we use the model to interpret specific examples of genetic introgression?

A wide range of species does appear to have the type of dispersal and the history of range expansion that could interact to produce introgression over long distances. The distribution of distances moved by propagules is frequently found to have thick tails necessary for colonization to proceed in jumps. For example, the distribution of rust spores (*Uromyces appendiculatus*) around an affected leaf is of the form  $a/(1 + bx^2)$  (Aylor & Ferrandino, 1990); similar distributions are commonly found in plant seeds (Harper, 1977). Comparable long distance animal dispersal is notable during times of invasion and may occur particularly when individuals find themselves in a hostile environment (Jones *et al.*, 1981).

Climatic changes have affected most species' ranges. As well as the major changes associated with the ice ages, there have been smaller more frequent climatic



**Fig. 4** An example of clines that get narrower with time. The frequency of an underdominant allele is plotted against distance from the centre of the cline. The clines were generated by the dispersal function with  $\alpha = 4.5$ . Solid lines show the clines after 25 generations; broken lines show the narrower equilibrium clines at 500 generations.

fluctuations, with periods as short as 78 and 180 years (Dansgaard *et al.*, 1971). These were associated with range changes most noticeable as tree line fluctuations of 100–200 m (Wardle, 1974). In locations where genetically distinct populations met, the two genomes will have been mixed, reassorted and then fragmented again with the next climatic cycle. It is worth noting that colonization of vacant habitats is not only restricted to periods of range expansion. Thus Atkinson & Shorrocks (1984) have pointed out that there can be stochastic loss of species from a locality. They argue that these vacant localities can provide refuges from interspecific competition (see Shorrocks, 1991) but they can equally well provide vacant sites for colonization by foreign alleles and genomes.

Given the broad patchy mixing predicted by the pioneer colonization model, a resultant reassorted genome may establish itself and subsequently spread far from its initial origins. This type of dynamics would account for the broad and asymmetrical distributions of recombinant genomes seen between races in *Viatica*, *Mus*, *Caledia* and *Scilla* (Hewitt, 1979; Gyllenstein & Wilson, 1987; Shaw *et al.*, 1988; Parker *et al.*, 1991).

Although the history of range expansion can account for such broad patterns, there are competing



explanations for variation in cline width and the non-random distribution of recombinant genomes. These include small effective population size which can lead to broader clines (Nichols, 1989), and selection which may cause the spread of some genes or combinations (Shaw *et al.*, 1988) or restrict them to certain habitat types (Harrison & Rand, 1989). It is possible to collect information to distinguish between these alternative explanations of patterns of genetic introgression. This history of colonization may be apparent in phylogenies of alleles (Avise *et al.*, 1987), small population size can be detected by the effects on other loci (Nichols *et al.*, 1990) and selection inferred from similar outcomes in different localities or correlations with environmental variables (Harrison & Rand, 1989; Shaw *et al.*, 1985).

The most likely example of which we have direct experience is the hybrid zone in the Pyrenees between subspecies of the grasshopper (*Chorthippus parallelus* (Butlin & Hewitt, 1985a,b; Butlin *et al.*, 1991). This zone shows genetic patterns indicative of exceptional gene flow. The *C.p.parallelus* and *C.p.erythropus* ranges would have met after the last ice age some 9000 years ago and the intermingling of the two genomes appears to have generated allozyme, morphological and behavioural clines over 20 km wide since then, yet the dispersal distance has been measured as 30 m (Virdee & Hewitt, 1990). At that rate a neutral gene would only be expected to have produced a cline of 2–3 km (Endler 1977). The mixing process seen in the model could explain these unexpectedly broad clines in *Chorthippus*. It would nevertheless be compatible with the narrow clines seen at other loci in this hybrid zone being under selection, notably nucleolar organizer regions (NORs). The two *Chorthippus* races differ in the number and location of NORs (Gosalvez *et al.*, 1988), so it seems likely that the offspring of hybrids would be at a disadvantage. The NOR cline is less than 1 km wide (Hewitt, 1989). In the model proposed here, selection against hybrids would have caused a narrower cline to crystalize out of the initially broad distribution in a few hundred generations.

The narrowing of clines maintained by selection against hybrids is a consequence of alleles being eliminated from those localities where they are rare and the fixation of the common allele. If the habitat is broken up into isolates by barriers to gene flow, then an allele may be locally abundant in a region where it is otherwise rare. The allele can therefore be maintained in the isolate and the cline would then be less susceptible to narrowing. It is notable that in *Chorthippus* the NOR transition and other narrow (selected) character clines are wider in some places than others (Butlin *et al.*, 1991; Ferris *et al.*, 1993) and that this seems to be

associated with different patterns of habitat fragmentation. This could reflect the existence of isolates established by long distance colonization. The centres of NOR clines and those at other loci do not coincide (Ferris *et al.*, 1993). This non-coincident pattern would arise if isolates were established by recombinant individuals, the NOR frequencies would then be affected independently by other loci. Work is currently in progress to collect more detailed information about habitat fragmentation and concurrently to map molecular genetic markers to evaluate our ideas about the history of past colonization and present gene flow in the region.

### Acknowledgements

We would like to thank Michael Shaw for valuable discussion. The work was supported by an NERC grant to G.M.H. and an SERC grant to R.A.N.

### References

- AMMERMAN, A. J. AND CAVALLI-SFORZA, L. L. 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press, Princeton.
- ATKINSON, W. D. AND SHORROCKS, B. 1984. Aggregation of larval Diptera over discrete and ephemeral breeding sites: the implications for co-existence. *Am. Nat.*, **124**, 336–351.
- AVISE, J. L., ARNOLD, J., BALL, R. M., BERMINGHAM, E., LAMB, T., NEIGEL, J. E., REEB, C. A. AND SAUNDERS, N. C. 1987. Intra-specific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.*, **18**, 489–522.
- AYLOR, D. E. AND FERRANDINO, F. J. 1990. Initial spread of bean rust close to an inoculated bean leaf. *Phytopathology*, **80**, 1469–1476.
- BARTON, N. H. 1979. The dynamics of hybrid zones. *Heredity*, **43**, 341–359.
- BARTON, N. H. AND HEWITT, G. M. 1985. Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.*, **16**, 113–148.
- BENNET, K. D. 1988. Post-glacial vegetation history: ecological considerations. In: Huntley, B. and Webb, T. III (eds), *Vegetation History*, Kluwer, Dordrecht, pp. 699–724.
- BUTLIN, R. K. AND HEWITT, G. M. 1985a. A hybrid zone between *Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus* (Orthoptera: Acrididae): morphological and electrophoretic characters. *Biol. J. Linn. Soc.*, **26**, 269–285.
- BUTLIN, R. K. AND HEWITT, G. M. 1985b. A hybrid zone between *Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus* (Orthoptera: Acrididae): behavioural characters. *Biol. J. Linn. Soc.*, **26**, 287–299.
- BUTLIN, R. K., RITCHIE, M. G. AND HEWITT, G. M. 1991. Comparisons among morphological characters and between localities in the *Chorthippus parallelus* hybrid zone (Orthoptera: Acrididae). *Phil. Trans. R. Soc. B*, **334**, 297–308.

- CLIFF, A. D., HAGGETT, P., ORD, J. D. AND VERSEY, G. R. 1981. *Spatial Diffusion*, p. 27. Cambridge University Press, Cambridge.
- COOPE, G. R. 1990. The invasion of Northern Europe during the Pleistocene by Mediterranean species of Coleoptera. In: di Castri, F., Hansen, A. J. and de Bussohe, M. (eds) *Biological Invasions in Europe and the Mediterranean Basin*, pp. 203–215. Kluwer, Dordrecht.
- DANSGAARD, W., JOHNSON, S. J., CLAUSEN, H. B. AND LANGWAY, C. C. 1971. Climatic records revealed by the Camp Century ice core. In: Turekian, K. K. (ed.) *Late Cenozoic Glacial Ice Ages*, pp. 36–37. Yale University Press, CT.
- ENDLER, J. 1977. *Geographic Variation, Speciation and Clines*. Princeton University Press, Princetown.
- FERRIS, C., RUBIO, J. M., SERRANO, L., GOSALVEZ, J. AND HEWITT, G. M. 1993. One way introgression of a subspecific sex chromosome marker in a hybrid zone. *Heredity*, **71**, 119–129.
- FISHER, R. A. 1937. The wave of advance of advantageous genes. *Ann. Eugen.*, **7**, 355–369.
- GOSALVEZ, J., LOPEZ-FERNANDEZ, C., BELLA, J. L., BUTLIN, R. K. AND HEWITT, G. M. 1988. A hybrid zone between *Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus* (Orthoptera: Acrididae): chromosomal differentiation. *Genome*, **30**, 656–663.
- GYLLENSTEN, V. AND WILSON, A. E. 1987. Interspecific mitochondrial DNA transfer and the colonization of Scandinavia by mice. *Genet. Res.*, **49**, 25–29.
- HARPER, J. L. 1977. *Population Biology of Plants*. Academic Press, London.
- HARRISON, R. C. AND RAND, D. M. 1989. Mosaic hybrid zones and the nature of species boundaries. In: Otte, D. and Endler, J. A. (eds) *Speciation and its Consequences*, pp. 111–133. Sinauer Associates, Sunderland.
- HEWITT, G. M. 1975. A sex-chromosome hybrid zone in the grasshopper *Podisma pedestris* (Orthoptera: Acrididae). *Heredity*, **35**, 375–385.
- HEWITT, G. M. 1979. *Animal Cytogenetics III, Orthoptera*. Gebruder-Borntraeger, Stuttgart.
- 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.
- HENGVELD, R. 1989. *Dynamics of Biological Invasions*, pp. 60–62. Chapman and Hall, London.
- HUNTLEY, B. J. AND BIRKS, H. J. B. 1983. *An Atlas of Past and Present Pollen Maps for Europe: 0–13,000 Years Ago*. Cambridge University Press, Cambridge.
- JONES, J. S., BRYANT, S. H., LEWONTIN, R. C., MOORE, J. A. AND PROUT, T. 1981. Gene flow and the geographical distribution of a molecular polymorphism in *Drosophila pseudoobscura*. *Genetics*, **98**, 157–178.
- MACK, R. N. 1981. Invasion in *Bromus tectorum* L. into western North America: an ecological chronicle. *Agro-Ecosystems*, **7**, 145–165.
- MOLLINSON, D. 1977. Spatial contact models for ecological and epidemic spread. *J. R. Statist. Soc. B*, **39**, 283–326.
- NICHOLS, R. A. 1989. The fragmentation of tension zones in sparsely populated areas. *Am. Nat.*, **134**, 969–977.
- NICHOLS, R. A., HUMPAGE, E. A. AND HEWITT, G. M. 1990. Gene flow and the distribution of karyotypes in the alpine grasshopper *Podisma pedestris* L. (Orthoptera, Acrididae). *Biol. San. Veg. Plagas (Fuera de serie)*, **20**, 373–379.
- PARKER, J. S., LOZANO, R., TAYLOR, M. AND RUIZ REJON, M. 1991. Chromosomal structure of populations of *Scilla autumnalis* in the Iberian Peninsula. *Heredity*, **67**, 287–297.
- PYLE, G. F. 1969. Diffusion of cholera in the United States. *Geog. Anal.*, **1**, 59–75.
- SHAW, D. D., COATES, D. J., ARNOLD, M. L. AND WILKINSON, P. 1985. Temporal variation in the chromosomal structure of a hybrid zone and its relationship to karyotypic repatterning. *Heredity*, **55**, 293–306.
- SHAW, D. D., MARCHANT, A. D., ARNOLD, M. L. AND CONTRERAS, N. 1988. Chromosomal rearrangements, ribosomal genes and mitochondrial DNA: contrasting patterns of introgression across a narrow hybrid zone. In: Brandham, P. E. (ed.) *Kew Chromosome Conference IV*, HMSO, London, pp. 121–129.
- SHORROCKS, B. 1991. Competition on a divided and ephemeral resource: a cage experiment. *Biol. J. Linn. Soc.*, **43**, 211–220.
- SKELLAM, T. G. 1951. Random dispersal in theoretical populations. *Biometrika*, **38**, 196–218.
- VAN DEN BOSCH, F., ZADOCKS, J. C. AND METZ, J. A. J. 1988. Focus expansion in plant disease. II. Realistic parameter sparse models. *Phytopathology*, **78**, 59–64.
- VAN DER PLANK, J. E. 1967. Spread of plant pathogens in space and time. In: Gregory, P. H. and Monteith, J. L. (eds) *Airborne Microbes*, pp. 227–247. Cambridge University Press, Cambridge.
- VIRDEE, S. R. AND HEWITT, G. M. 1990. Ecological components of a hybrid zone in the grasshopper *Chorthippus parallelus* (Zetterstedt) (Orthoptera: Acrididae). *Bol. San. Veg. Plagas (Fuera de Serie)*, **20**, 299–309.
- WARDLE, J. 1974. Alpine timberlines. In: Ives, J. D. and Barry, R. G. (eds) *Arctic and Alpine Environments*, pp. 371–402. Methuen, London.