

# Gene Flow

**James Mallet**

*Galton Laboratory, Department of Biology, University College  
London, 4 Stephenson Way, London NW1 2HE, UK*

---

## What is Gene Flow?

'Gene flow' means the movement of genes. In some cases, small fragments of DNA may pass from one individual directly into the germline of another, perhaps transduced by a pathogenic virus or other vector, or deliberately via a human transgenic manipulation. However, this kind of gene flow, known as horizontal gene transfer, is rare. Most of the time, gene flow is caused by the movement or dispersal of whole organisms or genomes from one population to another. After entering a new population, immigrant genomes may become incorporated due to sexual reproduction or hybridization, and will be gradually broken up by recombination. 'Genotype flow' would therefore be a more logical term to indicate that the whole genome is moving at one time. The term 'gene flow' is used probably because of an implicit belief in abundant recombination, and because most theory is still based on simple single locus models: it does not mean that genes are transferred one at a time. The fact that gene flow is usually caused by genotype flow has important consequences for its measurement, as we shall see.

## Two Meanings of 'Gene Flow'

We are often taught that 'dispersal does not necessarily lead to gene flow'. The term 'gene flow' is then being used in the sense of a final *state* of the population, i.e. successful establishment of moved genes. This disagrees somewhat with a more straightforward interpretation of gene flow as *actual*

*movement* of genes. The tension between actual movement and successful establishment is at the heart of many misunderstandings of the term 'gene flow' in studies of genetic variability at marker loci (Slatkin and Barton, 1989). Under some stringent assumptions (see below) we can measure genetic variability (a description of the *state* of populations) and obtain an estimate of 'gene flow' in the form of  $N_e m$  (the product of population size and fraction of migrants), but this tells us almost nothing about *actual* fraction of the population moving,  $m$ . Actual gene flow,  $m$ , could be very high or very low, depending on the value of  $N_e$ .

'Reproductive isolation', as employed in the biological species concept, can be viewed as a kind of inverse of gene flow, and has a similar ambiguity. 'Reproductive isolation' is a combination of some very different things: pre-mating isolation, or sexual behaviour which inhibits *actual* gene flow, and post-mating isolation, selection against hybrids which determines whether genes that do flow survive in their new genomic environment. Reproductive isolation is therefore most straightforwardly interpreted as a stable but strongly divergent balance between gene flow and selection, rather than as complete isolation from actual gene flow. As with 'gene flow', the term 'reproductive isolation' is used ambiguously in many publications, sometimes to mean pre-mating isolation, sometimes to mean post-mating isolation, and sometimes both. It is therefore advisable to distinguish hybrid inviability, sterility and mate choice rather than lumping all into a single term 'reproductive isolation' or 'isolating mechanisms' (Mallet, 2000).

Of course, there is nothing wrong with ambiguous terminology provided we know what we mean, but it is all too easy to become deluded into measuring 'gene flow' via  $N_e m$  and then to attempt to use this knowledge in ways which require knowledge of actual gene flow,  $m$ , between populations, as in the retardation of adaptation. Similarly, to say that sexual species are defined by their 'reproductive isolation' is fine provided one realizes at the same time that species may not *actually* be reproductively isolated: horizontal gene transfer rates can be non-zero.

### Which 'Gene Flow' do we Want to Measure?

Both types of gene flow are of course worth measuring. However, the balance version of gene flow simply measures the state of the population, rather than the process of gene flow leading to that state. Thus, it is normally more useful to measure actual gene flow because it is true of all loci, whether under selection or not, and is of interest in all models of evolution. For instance, supposing we were able to measure actual gene flow distance,  $\sigma_x$  (the standard deviation of parent-offspring distances), from studies of neutrally drifting genes (in practice, we usually cannot; see below), we could then apply this measure in ecological and population genetic problems unrelated to genetic drift. As an ecological example, the velocity of spread of

an invading species is expected to be  $v_x = \sqrt{2r\sigma_x^2}$ , where  $r$  is the intrinsic rate of increase (Andow *et al.*, 1993). In contrast, 'gene flow' in the sense of the balance of gene flow and drift given by  $N_e m$  is of use only in understanding population differentiation under drift.

The remainder of the chapter will examine the measurement of the balance state and of actual gene flow between local populations, geographic races and sympatric host races or species.

## The Balance Between Gene Flow and Genetic Drift

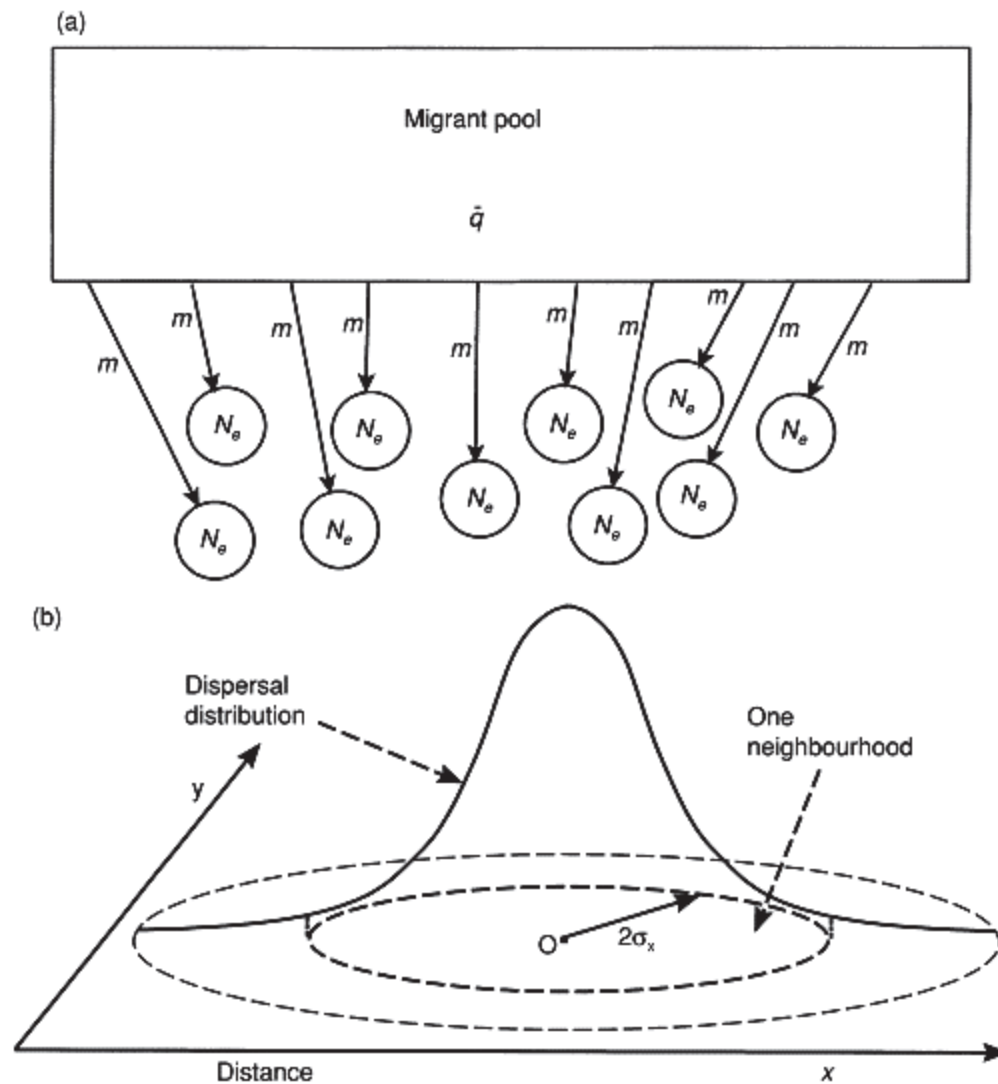
One of the simplest effects of gene flow is on genetic drift in interconnected populations. The probability of identity by descent increases in a population by a factor  $F_{ST} = 1/(2N_e)$  in each generation.  $N_e$  is the 'effective size' of an ideal diploid population that loses heterozygosity at the same rate as the actual population of size  $N$ . Because some individuals in a population can monopolize resources or reproduction,  $N_e$  is often less than, although of the same order as,  $N$  (Nunney, 1993). Frankham (1995) reviewed data which suggest that  $N_e$  may typically be as low as 5–10% of the value of  $N$ . In the absence of gene flow, drift would eventually cause fixation at neutral alleles in all subpopulations: if the initial allele frequency in all the populations were  $\bar{q}$ , subpopulations would diverge so that a fraction  $\bar{q}$  of the populations eventually become fixed for the allele, while the remaining  $1 - \bar{q}$  will be fixed for some alternative allele. Dispersal, however, reduces this effect by carrying alleles between populations, and a balance between drift and gene flow can result. The probability of identity by descent can be measured by the variance in gene frequency  $s_q^2$  between populations expressed as a fraction of its theoretical maximum  $\bar{q}(1 - \bar{q})$  when all populations are fixed, i.e.

$$F_{ST} = \frac{s_q^2}{\bar{q}(1 - \bar{q})}.$$

$F_{ST}$  can reach an equilibrium when drift, which causes divergence of neutral allele frequencies and local homozygosity, is balanced by the homogenizing influence of gene flow. A simple but spatially unrealistic model of metapopulation structure, the 'island model' (Fig. 16.1a), has an immigration fraction  $m$  per generation into each subpopulation, or 'island', of size  $N_e$ ; migrants can come from anywhere else in the metapopulation or from a 'mainland', and therefore have a constant gene frequency  $\bar{q}$ . Under these circumstances, it can be shown (Wright, 1969: 291) that

$$F_{ST} \approx \frac{1}{1 + 4N_e m}$$

will result at equilibrium between gene flow and drift. This relation can remain approximately valid even in more realistic population structures, such as the 'stepping stone' model, in which migrants are exchanged mainly with



**Fig. 16.1.** Models of population structure. (a) Wright's island model. A large number of 'islands' each with population size  $N_e$  are assumed. A fraction  $m$  of each subpopulation is replaced per generation from a universal pool of migrants having neutral gene frequency  $\bar{q}$  equivalent to the average over all subpopulations. Under this model, the potential for local drift then depends on the combined parameter  $N_e m$ . (b) Wright's isolation by distance model. In this model, population structure is more spatially realistic so that individuals are distributed continuously in a two-dimensional plane. If the dispersal is  $\sigma_x$  (the standard deviation of parent-offspring distances), the potential for local drift can be shown to depend on the 'neighbourhood population size', numerically the number of individuals within a circular neighbourhood of radius  $2\sigma_x$ . The neighbourhood size,  $N_b = 4\pi\rho_e\sigma_x^2$ , is the product of the area of this circle and the effective population density  $\rho_e$ .  $N_b$  is again a product of gene flow and population numbers, and it plays a role in isolation by distance equivalent to  $N_e m$  in the island model.

local subpopulations (Slatkin and Barton, 1989). If, in addition to local dispersal, a small fraction of migrants are exchanged over very long distances,  $F_{ST}$  may stabilize over the whole metapopulation structure as in the island model. If there is no long-distance dispersal, mutation or weak selection to restore polymorphisms,  $F_{ST}$  will increase with distance as drift within subpopulations becomes more and more independent (Wright, 1969; Slatkin, 1993).

If populations are small and/or migrants are rare so that  $N_e m \ll 1$ , drift on the islands outweighs gene flow, and gene frequencies in subpopulations become nearly fixed at  $q = 0$  or  $q = 1$ , so  $F_{ST} \rightarrow 1$ . Conversely, if populations are large and/or migrants are common so that  $N_e m \gg 1$ , migration outweighs drift and all gene frequencies converge on  $\bar{q}$ , so that  $F_{ST} \rightarrow 0$ . The value  $N_e m \approx 1$  is therefore a kind of cusp between these two regimes, where populations can have almost any gene frequency with equal probability (Wright, 1969).

Wright (1969: 292) points out that  $N_e m$ , as a product of population size and fraction of the population migrating, is numerically equivalent to the number of immigrants that arrive in any subpopulation in each generation.  $N_e m$  has therefore been interpreted as a kind of 'gene flow',  $\hat{M}$ , especially by recent authors (Slatkin, 1987, 1993). This 'gene flow' can be estimated by studying allele frequencies in natural populations, because

$$\hat{M} = N_e m \approx \frac{1}{4} \left( \frac{1}{F_{ST}} - 1 \right).$$

However, it is worth remembering that this 'gene flow' in fact consists of a ratio between two dimensionless probabilities (actual gene flow,  $m$ , and genetic drift,  $1/2N_e$ ); each consists of a fraction of INDIVIDUALS per total INDIVIDUALS). The contributions of drift and gene flow to gene frequency variation are both, of course, relative to a single generation, and therefore also have units  $\text{TIME}^{-1}$ .  $N_e m$  itself is formed from a ratio of these probabilities, and its units  $(\text{TIME}^{-1})^{-1} \times \text{TIME}^{-1}$  cancel to give a quantity that is dimensionless (Barton and Rouhani, 1991: 501), as it must be if it is to be proportional to  $1/F_{ST}$ . In a sense, then, the interpretation of  $N_e m$  as a number-based form of gene flow is illusory: although numerically correct, this parameter actually has no units at all. Quite apart from the obvious difficulties with assumptions under the unrealistic island model (Weir, 1996; Bossart and Pashley Prowell, 1998; Whitlock and McCauley, 1999),  $N_e m$  is not a very useful quantity because the equation merely transforms our data, in the form of  $F_{ST}$ , rather than solving for the gene flow parameter of interest,  $m$ , valid for use where drift is not involved. We might of course employ  $N_e m$  to 'predict' equilibrium levels of  $F_{ST}$  under neutral drift, but this is of very limited use because we already know what the  $F_{ST}$  is: it is the information from which we obtained our estimate of  $N_e m$ . In many cases, it will be more sensible, and perhaps more honest, simply to cite values of  $F_{ST}$  from which this 'gene flow' is derived rather than making the transformation to  $N_e m$  or  $\hat{M}$ , whether based

on  $F_{ST}$ , rare alleles (Slatkin, 1987) or gene genealogies (Slatkin, 1989, 1993; Slatkin and Maddison, 1989).

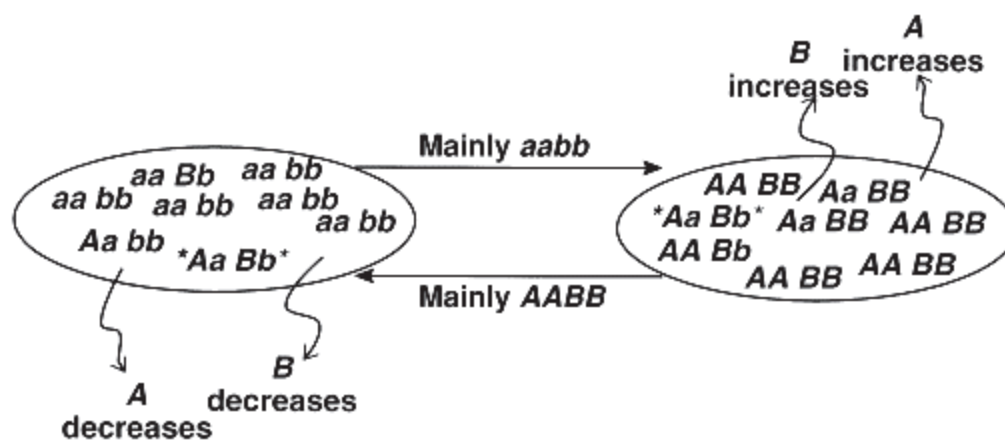
Wright (1969) also developed a more lifelike model of continuous population structure known as 'isolation by distance'. This model is far from tractable (Felsenstein, 1975), explaining its unpopularity in genetic marker studies. However, its spatial realism makes isolation by distance potentially far more useful than the island model. The metapopulation is assumed more or less continuous, rather than discrete, and migration is modelled as a diffusion process with a bivariate Gaussian distribution: any individual born at the origin  $O$  has a dispersal distribution that is symmetrical around the birth place, and its own offspring will be born along the  $x$  dimension with probability given by the variance of the dispersal distribution  $\sigma_x^2$  (Fig. 16.1b). This variance is related to the 'diffusion parameter',  $D$ , much used in ecology (e.g. Kareiva, 1982; Andow *et al.*, 1993): in fact,  $D = \sigma_x^2 / 2$ . Assuming that dispersal in  $x$  and  $y$  dimensions are both given by  $\sigma_x^2$ , 86.5% of offspring are born within a circular 'neighbourhood' of  $\pm 2\sigma_x$  around the birthplace of the parent (Fig. 16.1b – note that the familiar 95.4% probability integral of the Gaussian distribution at  $\pm 2\sigma_x$  applies only in one dimension). The equilibrium  $F_{ST}$  of neutral alleles under drift is controlled by the 'neighbourhood population size',  $N_b = 4\pi\sigma_x^2\rho_e$ , which is the area of the neighbourhood multiplied by the effective population density  $\rho_e$  (Wright, 1969). Drift enters as population density  $1/\rho_e$ , a probability term with units DISTANCE<sup>2</sup>, while gene flow,  $\sigma_x^2$ , represents the probability of movement also with units DISTANCE<sup>2</sup>; the two again cancel. Because both drift and gene flow effects are per generation, TIME<sup>-1</sup> also enters as a unit in each, and again cancels in the ratio. Thus 'neighbourhood population size',  $N_b$ , although numerically the effective number of individuals within a circle of radius  $2\sigma_x$ , is another dimensionless quantity, the continuous population equivalent of  $N_e m$  (Slatkin and Barton, 1989; Barton and Rouhani, 1991). Curiously,  $N_b$  has usually been seen as a kind of 'effective population size' (e.g. Hartl and Clark, 1989), rather than 'gene flow' as for  $N_e m$ , but both these incomplete interpretations are equally misleading.  $N_e m$  and  $N_b$  are neither measures of gene flow nor of effective population size. The truth is somewhere in between: both are dimensionless ratios which determine the relative strength of gene flow and drift. As with the stepping stone model, the distances across which  $F_{ST}$  is measured will determine its magnitude (Wright, 1969; Slatkin, 1993; Rousset, 1997) unless some additional 'restoring force' of mutation, long-distance migration or weak stabilizing selection controls the average gene frequency across the whole metapopulation of neighbourhoods (Slatkin and Barton, 1989).

In conclusion, while it is possible to estimate the dimensionless quantities  $N_e m$  and  $N_b$ , it is hard to infer anything very useful about actual gene flow from these estimates alone (see also McCauley and Eanes, 1987; Bossart and Pashley Prowell, 1998; Whitlock and McCauley, 1999). This may seem surprising, because this is exactly what a rather large number of studies have

apparently attempted (see the useful reviews by Roderick, 1996; Peterson and Denno, 1998). Of course, comparative studies of  $F_{ST}$  do give *some* idea of gene flow, but inferences are rather weak: a species with a high  $F_{ST}$  is likely to have lower levels of gene flow ( $m$  or  $\sigma_x^2$ ) than a different species with low  $F_{ST}$ , provided their population densities are comparable. But in many cases, we cannot make this assumption. In fact, it is rare that we can assume  $F_{ST}$  is even at equilibrium with respect to drift and gene flow. Somewhat embarrassingly, I myself have co-authored a paper that purports to estimate gene flow via  $N_e m$  (Korman *et al.*, 1993). I defend this paper on the grounds that we generated some information about the spatial scale of gene flow, as well as  $N_e m$  (see below).

### Multilocus Data and Interlocus Correlations

Because dispersal causes a flow of whole genotypes, or 'genotype flow', it will lead to a correlation between immigrant alleles at different loci, 'linkage disequilibria' (Fig. 16.2), as well as to eventual homogenization of gene frequencies. Thus, a balance between gene flow, drift and recombination will control linkage disequilibria within populations, just as a balance between drift and gene flow controls  $F_{ST}$ . In principle, therefore, combined data on

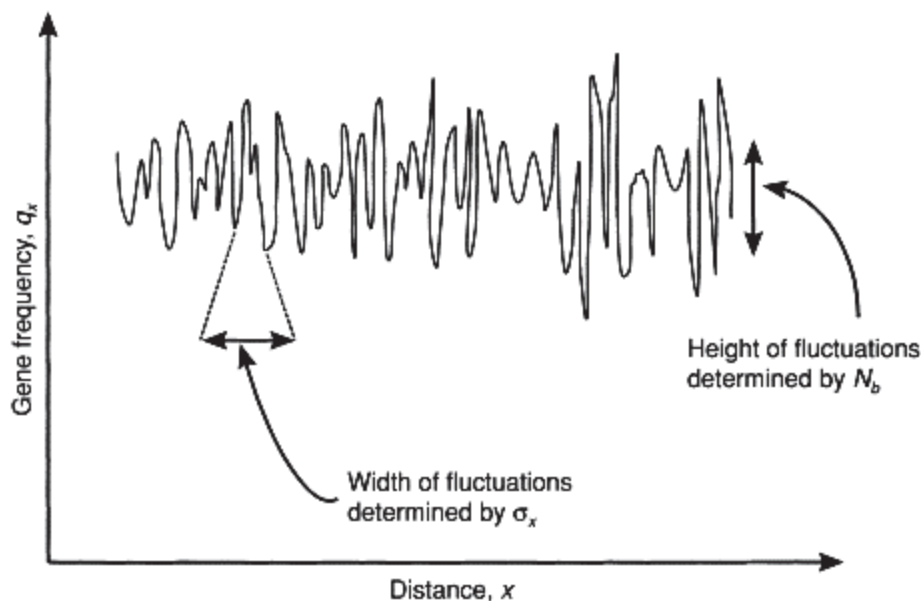


**Fig. 16.2.** Gene flow and gametic correlations. A pair of populations are assumed to exchange individuals. If the two populations come to differ in gene frequency at each of two loci ( $A$ ,  $B$ ), whether caused by selection or drift (bent arrows), exchange of individuals will ensure that there are gametic correlations, or linkage disequilibria, within each population. For example, the left-hand population consists mainly of  $ab$  gametes, but immigrants will bring in mainly  $AB$  gametes. If these immigrants mate, their offspring will be mainly  $AB/ab$  genotypes (starred in the diagram). Thus the presence of an immigrant allele  $A$  makes it more likely that another immigrant allele  $B$  will also be present in the same genome: the two loci will be correlated. This principle applies also to spatially separated areas within continuous populations, and is particularly important in clines and hybrid zones.

$F_{ST}$  and linkage disequilibria can be used separately to estimate the gene flow ( $m$  or  $\sigma_x^2$ ) required for the maintenance of disequilibria, as well as the neighbourhood size ( $N_e m$  or  $N_b$ ) at equilibrium: two independent effects are known, allowing us to solve for two unknowns (Vitalis and Couvet, 2000). In what is in effect the same method, multilocus information may be used to infer the source area from which migrants originated (Cornuet *et al.*, 1999). This genotypic identification technique may prove especially useful for determining sources of exotic pests such as the Mediterranean fruit fly, *Ceratitidis capitata* (Weidemann) (Davies *et al.*, 1999a,b). However, it is as yet unclear whether linkage disequilibria expected under drift (which are, after all, weak second-order effects) provide sufficient information with reasonable sample sizes.

### Spatial Information in Gene Flow/Drift Balance

Spatial information present in the measurement of gene frequencies under neutral drift may in principle be used to give spatial information about gene flow. In a continuous population, the magnitude of local gene frequency fluctuations ( $F_{ST}$ ) is controlled by  $N_b$ , while the spatial extent of these fluctuations is controlled by  $\sigma_x^2$  (Barton and Clark, 1990; see also Fig. 16.3). The



**Fig. 16.3.** Drift in continuous populations. In a continuous population, gene frequencies will fluctuate under genetic drift due to finite neighbourhood size  $N_b$ . Locally, these fluctuations will be correlated due to dispersal,  $\sigma_x$ . Thus the spatial scale of the decline in gene frequency correlation with distance (in other words, the width of the gene frequency fluctuations) will depend on  $\sigma_x$ , while the variability of gene frequencies will depend on  $N_b$ .

distance at which gene frequency fluctuations in local populations become more or less independent should therefore be useful for understanding the spatial extent of gene flow as measured by  $\sigma_x$ . However, because local fluctuations change between generations, one should carry out all local sampling as a 'snapshot' within a single generation to preserve this spatial information. Needless to say, it has rarely been recognized that spatial information is destroyed when sampling is done over multiple generations. There are many studies where hierarchical analyses of  $F_{ST}$  have been used to investigate the relative importance of gene flow at different spatial scales (Peterson and Denno, 1998), but few explicitly state that samples were obtained as a snapshot. In the important cotton pest *Heliothis virescens* (F.) we collected all samples within a single generation, and showed that allozyme frequency variation was significant at the lowest spatial scale (8 km); this implies that dispersal is restricted at this scale, even though  $F_{ST}$  was very low overall ( $\approx 0.002$ ) across the entire southeastern USA (Korman *et al.*, 1993). On the basis of this evidence for restricted gene flow ( $\sigma_x \leq 10$  km  $\text{gen.}^{-1/2}$ ), we argued that insecticide resistance in *Heliothis* is manageable on a local or countywide scale of  $\sim 10$  km. The relationship of  $F_{ST}$  (or its inverse,  $\hat{M} = N_e m$ ) with distance can similarly be used to give an idea of the degree of isolation by distance (Slatkin, 1993).

It is possible to use 'spatial autocorrelation' (Sokal *et al.*, 1989) to determine the distance at which gene frequency correlations between subpopulations (again sampled as a snapshot) drop to an insignificant level. Recently, simulations have shown how  $\sigma_x$  affects patterns of spatial genetic autocorrelation (Epperson and Li, 1997; Epperson *et al.*, 1999), but direct analytical methods are not yet available. This distance should give a rough idea of  $\sigma_x$  (see also Rousset, 1997). An interesting feature of drift in two-dimensional populations is that it will generally outweigh balancing selection locally. Drift around a mean gene frequency determined by selection could in principle be as informative as completely neutral drift (Barton and Clark, 1990). In fact, drift around any equilibrium gene frequency could be used: for example, variation around the gene frequency cline expected for selection against heterozygotes can also give estimates of  $N_b$  and  $\sigma_x$  (Barton and Clark, 1990; N.H. Barton, unpublished analyses).

### Direct Estimates of Dispersal and 'Slatkin's Paradox'

A puzzling phenomenon in empirical studies of population structure is known as 'Slatkin's paradox'. Direct estimates of effective population size and dispersal distance often suggest low values of  $N_e m$  or  $N_b$  ( $\sim 0.1$ – $10$ ; e.g. see Frankham, 1995) which predict high expected equilibrium levels of  $F_{ST}$ . In contrast, actual measures of  $F_{ST}$  at marker loci such as allozymes, especially in flying insects, are often very low ( $\ll 0.1$ ), implying large neighbourhood sizes ( $N_b, N_e m \gg 10$ ; for reviews of insect cases see McCauley and Eanes,

1987; Roderick, 1996; Peterson and Denno, 1998). For example, the butterfly *Euphydryas editha* (Boisduval) differs little in allozyme frequencies between the Rocky mountains and the Sierra Nevada in the western USA, even though it is almost inconceivable that this weakly flying, colony-forming butterfly ever disperses across the vast Great Basin desert (Slatkin, 1987).

Possible explanations of Slatkin's paradox include three major candidates:

1. The lack of allelic fixation might be due to balancing selection on allozyme loci (e.g. Avise, 1994). However, in cases where  $F_{ST}$  at different loci are similar, it is not really tenable to expect that the strengths of balancing selection are also so similar at every locus.
2. Slatkin (1987) himself argued that these low  $F_{ST}$  values resulted from frequent colonization and extinction, and that founders from distant populations would cause greater spatial homogeneity than expected given local movement measurable in mark-recapture studies. In other words, gene frequencies are similar across large areas because drift and gene flow equilibrate on a time scale longer than that of population turnover. In northern latitudes, genetic homogeneity could be a consequence of events as long ago as  $10^4$  years. Recolonization occurred after the last retreat of the ice, and there have been high effective population numbers since then. Under these circumstances, an equilibrium between slow drift and low existing levels of gene flow will not yet have been reached over the entire range. On a more local scale, analysis of populations of the fungus beetle *Phalacrus substriatus* (Gyllenhal) shows clearly that long-distance colonizations cause a much lower  $F_{ST}$  than expected on the basis of the mainly local post-colonization gene flow and population sizes observed (Ingvarsson and Olsson, 1997). However, colonization/extinction cycles may also inflate  $F_{ST}$  compared with the island model, instead of reducing it, depending on the number and genetic similarity of original colonists in each subpopulation (Whitlock and McCauley, 1990). If populations go through frequent bottlenecks when colonizing new habitat, and if members of each founder group are from similar population sources,  $F_{ST}$  will be higher than that predicted from post-colonization gene exchange equilibrium, as found by Whitlock (1992) in another fungus-feeding beetle, *Bolitotherus cornutus* (Panzer). Thus, consideration of population turnover can explain Slatkin's paradox but does not necessarily do so.
3. Another highly probable reason for Slatkin's paradox is that long-range movement of marked individuals will usually be underestimated because study sites are finite. Small numbers of very long-range dispersers which have a large genetic effect on  $\sigma_x^2$  and lead to strongly lowered  $F_{ST}$  may escape the study site and so be missed in field studies. Thus Dobzhansky and Wright measured dispersal in *Drosophila pseudoobscura* Frolova as  $\sigma_x = 59\text{--}81$  m day<sup>-1/2</sup> depending on the study site, leading to an expected 7-day lifetime dispersal of only  $\sigma_x = 157\text{--}218$  m gen.<sup>-1/2</sup> (Wright, 1978). It therefore appeared

that local genetic drift could be quite extensive in this species. However, later studies showed that oases were recolonized every year by *D. pseudoobscura* females dispersing many tens of kilometres across desert habitat (Jones *et al.*, 1981; Coyne *et al.*, 1982). Of course, dispersal over desert may be enhanced compared with the habitats in which earlier studies were done. None the less, the existence of such long-range movements casts grave doubt on the original estimates.

Explanations (2) and (3) are in fact related, because long-distance colonists are also those that fall outside the normal dispersal range of the majority. The distribution of dispersal found in many mark-recapture studies is not Gaussian but markedly leptokurtic (i.e. most individuals disperse near the parent, but there is a long-distance 'tail' containing significantly more individuals than expected under the Gaussian model). It can easily be shown that the strength of leptokurtosis (measured by  $\sigma_x^4$ ) has little effect on overall  $F_{ST}$  at neutral loci; use of  $\sigma_x^2$  alone in the Gaussian model usually suffices (Wright, 1969: 303–307). Instead, and more importantly, the practical effect of leptokurtosis is to cause the value of  $\sigma_x^2$  itself to be greatly underestimated in mark-recapture studies, when a long tail escapes detection beyond the edge of the study site. Thus leptokurtosis will homogenize distant populations, leading to much stronger reduction of equilibrium  $F_{ST}$  over large areas than expected from local observations of dispersal (Rousset, 1997).

All in all, therefore, it remains extremely difficult to obtain useful measures of gene flow either directly, via mark-recapture, or indirectly, using allele frequency variation due to genetic drift.

## Gene Flow – Selection Balance

One reason why inferring gene flow indirectly is so difficult is that, in reasonably abundant mobile species (such as flying insects),  $N_e$  and  $m$  are large, so equilibrium levels of drift measured by  $F_{ST}$  or disequilibria are expected to be very small, leading to large errors in its measurement. Selection, sampling or even laboratory scoring errors can dominate the weak signal due to drift. We found clear examples of laboratory errors by previous workers, almost certainly due to scoring problems, when we showed  $F_{ST}$  to be an order of magnitude less than previously recorded in *Heliothis* (Korman *et al.*, 1993). In contrast to drift, selection in natural populations is often a much stronger force: inferences about gene flow from the balance between gene flow and selection will be more robust than those from the balance between gene flow and drift (Lenormand *et al.*, 1998).

The frequency  $\hat{q}$  of an allele at equilibrium between immigration  $m$  and counterselection  $s$  ( $s$  is the fractional increase of mortality of genotypes bearing the allele) within a population is given by  $\hat{q} \approx m/s$ , providing that  $m < s$  (Haldane, 1930). This has potential applications for the possibility of adaptation, for example in insecticide resistance: gene flow will prevent

the evolution of resistance in a treated population until the selection for resistance reaches a critical level, whereupon the treated population flips to a new state of high resistance caused both by selection and also by the attainment of a higher population density more resistant to genetic swamping (Comins, 1977). The flip corresponds to a 'cusp catastrophe' which occurs at a critical value of  $m/s$  (Comins, 1977).

### Minimum Size of an Area for Adaptation

The balance between selection and gene flow is also tractable when we consider the problem spatially. The extent to which an insect is a 'slave of the environment' depends to a large extent on its rate of dispersal (Loxdale and Lushai, 1999). Thus, insecticide resistance (or any other adaptation) will evolve locally if the area treated has a radius much greater than about  $\sigma_x/\sqrt{s}$ . The reason for this, as originally shown by Haldane and Fisher, is that the equilibrium width of a cline in allele frequency, where selection on an allele with fitnesses  $(1+s)$  and  $(1-s)$  in two adjacent areas is given by  $w \approx \sqrt{(3\sigma_x^2/s)}$  (Fig. 16.4; Slatkin, 1973; Endler, 1977; Roughgarden, 1979). This analytical theory of clines depends on a diffusion approximation, and so requires weak selection, but the results remain approximately valid for  $s \leq 0.2$ . For insecticide resistance to evolve, the critical diameter  $d_c$  of an area for adaptation requires at least a small multiple of cline width

$$d_c > \pi \sqrt{\frac{\sigma_x^2}{8s}}$$

(Nagylaki, 1975; Roughgarden, 1979; Barton and Clark, 1990), so that insecticide resistance can attain a high frequency inside the treated area (Fig. 16.4; see also Lenormand and Raymond, 1998).

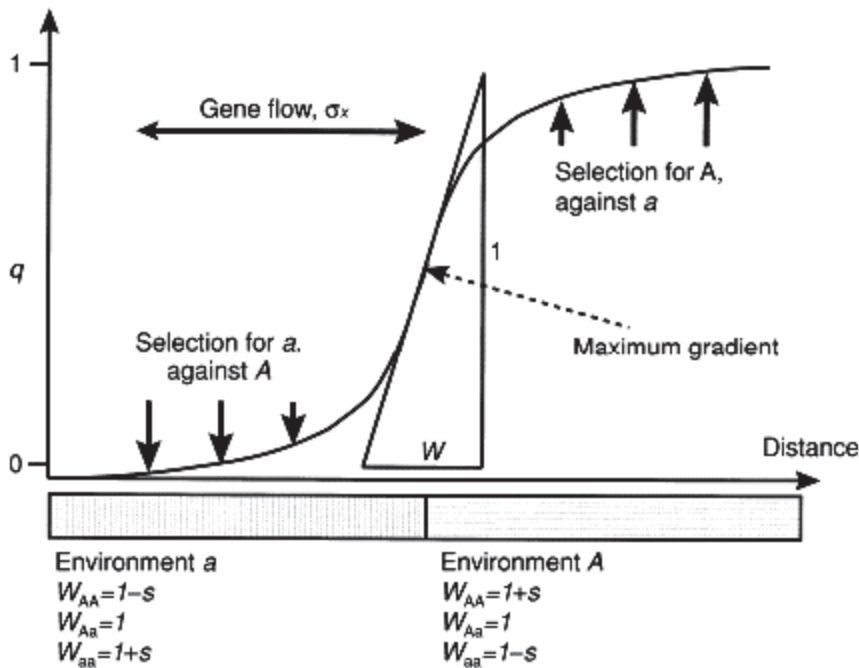
### Indirect Measurement of Gene Flow in Clines

Although theory relates cline width to gene flow  $\sigma_x$  and selection  $s$  in a potentially constructive way, it is again hard to reverse the equation. We will often be able to measure cline widths, but would prefer to estimate  $\sigma_x$  and  $s$  separately rather than their ratio. An estimate of  $\sigma_x/\sqrt{s}$ , being merely a reformulation of cline width, is as useless, in its way, as the combined drift/gene flow quantity  $N_e m$ . If we had direct measures of either  $\sigma_x$  or  $s$ , we could of course solve for the other. Endler (1977) used field estimates of dispersal  $\sigma_x$  to estimate selection pressures acting in clines using this method. Many of the selection pressures Endler estimated were so low ( $10^{-5}$ – $10^{-9}$ ) as to be incredible, for instance in the case of chromosomal or butterfly warning colour hybrid zones. These low estimates of selection are probably due to Slatkin's paradox: a tenfold underestimate of  $\sigma_x$  would generate a 100-fold

underestimate of  $s$ , which enters into the cline equation as a square root. Even under ideal circumstances, there will often be situations where  $\sigma_x$ , and  $s$  are both poorly known.

A possible exception occurs when a cline is moving as a wave of advance. Turelli and Hoffmann (1991) estimated dispersal in *Drosophila simulans* Sturtevant from data on a moving cline of maternally inherited *Wolbachia* parasites. The *Wolbachia* infection, which causes unidirectional sexual incompatibility, was spreading north in California at a rate of about  $175 \text{ km year}^{-1}$ . The gene flow of  $\sigma_x \approx 45\text{--}60 \text{ km gen.}^{-1/2}$  was estimated using cline theory, the measured speed of the wave of advance of *Wolbachia*, and laboratory estimates of selection pressures. Again in accordance with Slatkin's paradox, the observed measurements were nearly two orders of magnitude larger than those estimated in Dobzhansky and Wright's experiments on *D. pseudoobscura* (Wright, 1969).

One might be tempted to employ pairwise measures of  $N_e m$  based on marker alleles across clines or hybrid zones as 'gene flow' estimates, as advocated by Porter (1990). This should be avoided. Estimating  $N_e m$  from  $F_{ST}$  assumes that drift and gene flow have reached equilibrium. However, the



**Fig. 16.4.** Selection in continuous populations. Selection,  $s$ , in a continuous population may favour allele  $a$  on the left of the diagram, and allele  $A$  on the right. At equilibrium, the gene frequencies will form a sigmoid cline over the boundary between the environments, whose width,  $w$ , is given by a ratio between gene flow and selection ( $\sigma_x/s$ ). The case of different environments is shown here; however, similar clines are formed in the case of intrinsic or frequency-dependent selection, for example in contact zones between races differing in an underdominant chromosomal rearrangement, or in warning colour pattern.

hiatus in gene frequency across a cline implies either a temporary situation where gene flow and drift have not equilibrated as required, or, more probably, that selection is maintaining allelic divergence (albeit possibly indirectly), so that a selection/gene flow balance rather than a drift/gene flow balance is the correct model.

## Gene Flow in Multiple Locus Clines

Often, multiple locus clines occur together in hybrid zones. For example, in the butterfly *Heliconius erato* (L.), three loci determine colour pattern differences across a hybrid zone near Tarapoto, Peru. *Heliconius* are warningly coloured, which causes purifying frequency-dependent selection against rare morphs. Analytical solutions for this type of selection suggest that cline widths at equilibrium should stabilize at  $w \approx \sqrt{(8\sigma_x^2 / s)}$  to  $w \approx \sqrt{(12\sigma_x^2 / s)}$ , depending on dominance (Mallet and Barton, 1989a).

Dispersal across a multiple locus hybrid zone causes the immigration of whole genotypes, rather than flow of single genes. This genotype flow between two populations with distinct gene frequencies should cause a non-independence in the form of linkage disequilibria or correlations between genes within genotypes sampled in these populations (Fig. 16.2). As first shown by Barton (1982), genotype flow across such clines of allele frequency will lead to a stable maximum of correlations  $R_{AB}$  between a pair of genes,  $A$  and  $B$  of approximately

$$R_{AB} \approx \frac{4\sigma_x^2}{c_{AB}w_Aw_B}$$

near the centre of the hybrid zone where gene frequencies are 50%. Here,  $R_{AB}$  is the interlocus correlation coefficient,  $w_A$  and  $w_B$  are the widths of clines at loci  $A$  and  $B$ , and  $c_{AB}$  is the recombination rate. The correlation equilibrates due to a balance between immigration, which increases disequilibrium, and recombination, which reduces it. By sampling genotypes, we can measure  $w_A$  and  $w_B$  as well as  $R_{AB}$  directly. Gametic correlations can also, of course, be affected directly by selection for particular gene combinations, or epistasis. In *Heliconius*, fitnesses at the different loci were assumed multiplicative, i.e. non-epistatic. Additional epistasis is indeed likely in colour pattern combinations. However, provided that selection is not too strong, this second order effect on disequilibria will be weak in comparison to gene flow (Mallet and Barton, 1989a; Barton and Shpak, 2000). Given our equations for equilibrium width of a cline and disequilibria between loci, we now have two equations and two unknowns,  $\sigma_x$  and  $s$ , which we can therefore solve. Analysis of three unlinked colour pattern genes in a hybrid zone in *H. erato* led to estimates of  $\sigma_x \approx 2.6 \text{ km gen.}^{-1/2}$ , and  $s \approx 0.15\text{--}0.33 \text{ gen.}^{-1}$  per locus (Mallet *et al.*, 1990, 1998a). An estimate of

selection based on mark-recapture agreed approximately with the indirect measure from cline theory (Mallet and Barton, 1989b), but again in accordance with Slatkin's paradox, dispersal estimated via mark-recapture (Mallet, 1986,  $\sigma_x \approx 0.296 \text{ km gen.}^{-1/2}$ ) was an order of magnitude less than that estimated in clines. The discrepancy was presumably due to the difficulty of tracing long-distance movers in the field dispersal study.

Recently, work with insecticide resistance clines in the mosquito *Culex pipiens* L. has used similar techniques to obtain robust estimates of selection and gene flow from genotypic data. This mosquito has been treated for many years with insecticides during the summer near Montpellier in southern France. The coastal strip where insecticides are applied is approximately 20 km wide. This has led to high summer coastal frequencies of resistance alleles at two major loci: target site insensitivity *Ace-1*, and amplified esterase *Ester*. These loci are linked with recombination rate  $c \approx 14.5\%$ . Resistance is selected against in untreated areas inland, leading to the formation of clines in the two resistance loci. Analysis of cline widths and disequilibria of the two loci gave estimates of resistance allele selective advantages of 30 and 16% in the presence of insecticides, and disadvantages of 12 and 5.5% in their absence, for *Ace-1* and *Ester* respectively. Gene flow was estimated as  $\sigma_x \approx 6.6 \text{ km gen.}^{-1/2}$  (Lenormand *et al.*, 1998).

Further analyses of *Culex* were able to show variation in selection and dispersal through the season. In the autumn, migration was inland, towards overwintering sites. In the winter, no spraying is performed and the adults roost in caves, dispersing little, and there is strong selection against resistant alleles in the females. However, there is little or no effective selection against resistance alleles in the males during winter, since the females store sperm after mating in the autumn. Breeding in spring therefore causes a rise in resistance frequency as these unselected sperm unite with female gametes to form offspring (Lenormand *et al.*, 1999; Lenormand and Raymond, 2000). In accordance with Slatkin's paradox, gene flow was generally higher than expected on the basis of mark-recapture estimates, probably because long-distance dispersers were missed in field studies; selection estimates were consistent with laboratory data (Lenormand and Raymond, 1998).

The practical value of these kinds of estimates of gene flow and selection can be profound. Lenormand *et al.* (1999) recommend that if the coastal insecticide treatment area was halved to about 10 km wide, gene flow would swamp the selection for insecticide resistance, and there would be little problem controlling the mosquito. This year (2000), health authorities in the south of France are attempting to test this prediction for public benefit. The *Culex* story is perhaps the most complex and informative analysis of selection and gene flow ever achieved in the field, and the results are of potential economic and public health importance. This work shows the power that can be achieved through indirect measures of gene flow via cline theory and genetic data.

## Gene Flow Between Host Races or Species

As we move up the evolutionary continuum from populations through to species, hybridization rates tend to decrease. For pairs of taxa near the species boundary, we can use mass-action population genetic theory similar to that discussed above (Barton and Gale, 1987), rather than investigating the results of occasional historical events to study gene flow. For example, in the apple maggot *Rhagoletis pomonella* (Walsh), disequilibrium measurements suggest that gene flow between genetically distinct apple and haw host races was about  $m \approx 20\%$  per generation (Barton *et al.*, 1988; Feder *et al.*, 1988, 1990). Direct measurements of habitat and mate choice between apple and haw host races suggest a slightly lower level of  $m \approx 6\%$  (Feder *et al.*, 1994). However, these results are not too dissimilar, given that gene flow must be highly contingent on the particular spatial and temporal arrangements of hawthorn and apple fruiting trees.

Hybridization and gene flow between species is something that, under the biological species concept, apparently ought not to happen. There is now increasing realization that hybridization, while rare, does occur between clearly marked taxonomic species in the wild. Among insects, hybridization seems especially prevalent in Lepidoptera, perhaps because strong colour pattern divergence makes hybrids easily identifiable, as in birds. Among European butterflies, 12% of all species are known to hybridize with at least one other species (Guillaumin and Descimon, 1976), but the rate is probably similar elsewhere, such as in North American swallowtails (6%, Sperling, 1990) and heliconiines (25%, Mallet *et al.*, 1998a). Few *Drosophila* species are known to hybridize, but it is notable that those that do are within the *melanogaster* Meigen, *pseudoobscura* Frolova, Hawaiian picture-wing, and other groups that are genetically well studied (Gupta *et al.*, 1980). It seems not unlikely that interspecific hybridization among related *Drosophila* is as frequent, though less evident, as in Lepidoptera. However, while species that hybridize are fairly common, the rates at which hybrids occur are very low, almost always, in insects, less than 1/1000 individuals per sympatric species pair.

We should avoid the fallacy that  $N_e m$  measures 'gene flow' between species, host races or parapatric populations separated by hybrid zones (see also above), as is sometimes recommended (e.g. Porter, 1990). Supposing a newly arisen pair of species never hybridizes ( $m = 0$ ), a very long time indeed is required before all marker loci drift to diagnostic fixation so that  $F_{ST} = 1$  and  $N_e m = 0$  is inferred.

To be identifiably different, host races or sibling species must have multiple genetic differences and, in sexual taxa, this implies little gene flow. Therefore, individual hybrids may be identifiable at multiple loci, and mass-action models become less practical for estimating rates of gene flow. For example, in the narrow contact zone between *Heliconius erato* and

*Heliconius himera* Hewitson, hybrids identified via colour pattern were clearly also identifiable using molecular data (Jiggins *et al.*, 1997). In a direct analysis of hybridization in nature, backed up by laboratory studies of mate choice, it was shown that the production of hybrids was similar to the approximately 5% frequency of  $F_1$  hybrids in the centre of the contact zone (Jiggins *et al.*, 1997; McMillan *et al.*, 1997; Mallet *et al.*, 1998b). In another pair of species, *Heliconius charithonia* (L.) and *Heliconius peruvianus* Felder & Felder, hybrids were identified via allozyme genotypic signatures (Jiggins and Davies, 1998). As already mentioned, this genotypic identification technique also has potential for determining the geographic sources of conspecific immigrants (Cornuet *et al.*, 1999; Davies *et al.*, 1999b). We have recently used genotypic identification to estimate hybridization rates of ~2% between sympatric larch and pine host races of the larch budmoth *Zeiraphera diniana* Guenée (Emelianov *et al.*, 1995). These estimates are similar to those based on direct analyses of cross-attraction of host races in the field (Emelianov *et al.*, 2000).

If hybrids between species are rare, and when formed often sterile or inviable, can 'gene flow' between species ever occur? Hybridization itself is a form of actual gene flow in the sense I have advocated here, but a valid question is: can genes be transferred via backcrossing between populations normally considered as separate species? This is difficult to answer, because the rarity of hybridization itself makes large-sample direct observations almost impossible. Instead, accurate typing and identification of alleles using sensitive DNA-based techniques will usually be required. Phenotypic manifestations alone, as in colour pattern or allozyme banding patterns, will usually be insufficient evidence. In several cases, evidence strongly supports gene transfer. Molecular homologies suggest that transposable P-elements in some strains of *D. melanogaster* originated recently from elements common in the unrelated *Drosophila willistoni* group, and were transferred, perhaps by parasitic mite vectors (Houck *et al.*, 1991). Some individuals of *Drosophila mauritiana* Tsacas & Davis, a species in the *simulans* group endemic to the island of Mauritius, share mtDNA haplotypes with the worldwide species *D. simulans* Sturtevant, while other *D. mauritiana* have a more divergent mtDNA. This suggests recent introgression from *D. simulans* into *D. mauritiana*, presumably via hybridization (Solignac and Monnerot, 1986). Very different gene genealogies for *Adh* compared with other genes in the *D. pseudoobscura* group suggest transfer of *Adh* genes between closely related species (Wang *et al.*, 1997). If these interpretations are correct, gene transfer may have the utmost importance for systematics: estimated molecular phylogenies, or indeed any phylogenies of closely related species, may depend on which characters are used in their construction, not only because of errors in phylogeny estimation and ancestral polymorphisms, but also because different genes may have genuinely different histories within a single group of species.

## Prospects for Estimating Gene Flow

The use of selection/gene flow balance and disequilibrium across clines to estimate gene flow has been achieved in only a handful of cases. Apart from the insect examples cited here, only three other taxa, all vertebrates, have been studied in this way (Barton, 1982; Szymura and Barton, 1991; Sites *et al.*, 1994). In every case, development of new population genetic theory and new multidimensional likelihood analyses has been required. There may be some general models that can be applied via standardized computer programs, for example in multilocus clines where the precise nature of selection on each locus is less important than aggregate polygenic effects on cline width and disequilibria (Szymura and Barton, 1991; Stuart Baird *et al.*, unpublished). Every other example studied so far had a unique mode of selection or spatial situation, which caused difficulties for standardized methods. Similarly, analyses of hybridization and gene flow between species require a good understanding of the theory of gene flow/selection balance between a pair of demes (Barton and Gale, 1987) or of coalescence theory (Wang *et al.*, 1997). Practitioners of these newer gene flow methods will have to develop theoretical skills and gain a good understanding of likelihood fitting or analysis of deviance. Gene flow/selection balance is complex. To study it, we must all become more numerate.

Various readers of this chapter have commented on the grim picture I paint for the estimation of gene flow from gene frequency data using the drift/gene flow equilibrium assumption. Problems range from the error-proneness of estimating very small values of  $F_{ST}$ , the virtual uselessness of the resultant  $N_e m$  combined parameter, the slow attainment of equilibrium between gene flow and selection over large areas, the possibility of selection, and laboratory scoring errors. What if no selection/dispersal balance with multiple locus clines exists in the species? Is it hopeless to attempt to estimate gene flow indirectly? I fear that the answer is often 'yes'. At minimum, anyone seeking to calculate  $N_e m$  from their data on  $F_{ST}$  must avoid unrealistic expectations that the 'gene flow' they measure thereby will be very useful. Such expectations have existed since the 1930s, when Dobzhansky's hopes for estimating gene flow from the allelism of lethals in *D. pseudoobscura* were dashed by Sewall Wright (Provine, 1986: 367). It is especially unwise to estimate 'gene flow' using  $N_e m$  when a cline or other situation clearly indicates that a drift/gene flow equilibrium is unlikely. Only in well justified comparative analyses between species or populations, and where spatial and temporal information about gene frequency samples are preserved, may some inferences about gene flow be made. Anyone attempting to estimate gene flow from marker data should consider carefully whether their inferences will be as useful as they would like. Many previous studies have used an  $N_e m$  route to estimate 'gene flow', but this popularity does not ensure utility.

## Conclusions

Estimating gene flow, or its converse, reproductive isolation, is essential for understanding evolution and diversification. In this chapter, I have reviewed the somewhat confused current state of the field. In particular, an impression of limited dispersal is often based on field observations of local insect movement; this conflicts with genetic evidence for high rates of gene flow. I attempt to provide some hope that we may in future comprehend these areas of evolutionary biology more clearly. Molecular advances have enabled considerable improvements on 'gene flow' measured as combined parameters such as  $N_e m$ , or 'reproductive isolation' in the sense of pre- and post-mating isolation. Incorrect application of population genetic theory to these new data has hitherto held back our understanding. In the first few decades of this century, improvements in theory and data analysis will allow us to topple gene flow from its current almost mystical status as a dimensionless quantity to being an observable, measurable parameter in everyday use. Unless better methods for analysing drift/gene flow models are developed, the information may have to come from situations where selection/gene flow balance can be observed. Because of the complexity of selection/dispersal balance in continuous populations, estimating gene flow will rarely be as easy as grinding up a few bugs, running a few gels, and then analysing the data using an off-the-shelf computer program. However, these new methods will ultimately provide far more insight into evolution, particularly the important case of evolution under natural selection.

## Summary

Insect populations have provided many of the best studies of gene flow, a parameter which plays a central role in topics as diverse as population structure, adaptation (e.g. insecticide resistance), conservation genetics and speciation. In the literature, 'gene flow' often refers to  $N_e m$  (effective population size  $\times$  migration rate) estimated from frequencies of genetic markers, typically allozymes, by assuming an equilibrium between gene flow and neutral genetic drift.  $N_e m$  and its continuous population analogue, 'neighbourhood population size',  $N_b$ , are actually both dimensionless parameters that measure the relative strength of gene flow and drift, rather than being simple measures of 'gene flow' or 'population size' alone. Inferring gene flow using this method therefore has many pitfalls. A more robust procedure is to exploit equilibria between gene flow and selection when suitable geographic variation in selection is present. New population genetic theory makes selection-based methods particularly useful, and selection is often a stronger force than genetic drift in many insect populations, leading to estimates with a lower margin of error. Finally, gene flow may also occur between sympatric

host races or closely related species that hybridize, albeit rarely. I give examples from ongoing work with insects, especially fungus-eating beetles, *Heliconius* butterflies (Nymphalidae) and other Lepidoptera, *Culex* mosquitoes (Culicidae), and *Drosophila* (Drosophilidae). These newer methods provide the power to estimate actual gene flow ( $m, \sigma_x^2$ ), rather than merely dimensionless gene flow measures such as  $N_e m$  and  $N_b$ .

## Acknowledgements

This work was supported by NERC and BBSRC grants. I am very grateful for useful comments and discussions on this manuscript by Nick Barton, Roger Butlin, Chris Jiggins, Thomas Lenormand, Hugh Loxdale, Russ Naisbit, Graham Stone, Chris Thomas and Mike Whitlock. I am particularly grateful to Nick Barton for pointing out the dimensionlessness of  $N_e m$  and  $N_b$ , and for the prototype of Fig. 16.3 which traces to an unpublished grant proposal, c. 1984.

## References

- Andow, D.A., Kareiva, P.M., Levin, S.A. and Okubo, A. (1993) Spread of invading organisms: patterns of spread. In: Kim, K.C. and McPherson, B.A. (eds) *Evolution of Insect Pests: Patterns of Variation*. John Wiley & Sons, New York, pp. 219–242.
- Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, London.
- Barton, N.H. (1982) The structure of the hybrid zone in *Uroderma bilobatum* (Chiroptera: Pyllostomatidae). *Evolution* 36, 863–866.
- Barton, N.H. and Clark, A.G. (1990) Population structure and processes in evolution. In: Wohrman, K. and Jain, S. (eds) *Population Genetics and Evolution*. Springer-Verlag, Berlin, pp. 115–174.
- Barton, N.H. and Gale, K.S. (1993) Genetic analysis of hybrid zones. In: Harrison, R.G. (ed.) *Hybrid Zones and the Evolutionary Process*. Oxford University Press, New York, pp. 13–45.
- Barton, N.H. and Rouhani, S. (1991) The probability of fixation of a new karyotype in a continuous population. *Evolution* 45, 499–517.
- Barton, N.H. and Shpak, M. (2000) The effect of epistasis on the structure of hybrid zones. *Genetical Research, Cambridge* 75, 179–198.
- Barton, N.H., Jones, J.S. and Mallet, J. (1988) No barriers to speciation. *Nature* 336, 13–14.
- Bossart, J.L. and Pashley Prowell, D. (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution* 13, 202–206.
- Comins, H.N. (1977) The development of insecticide resistance in the presence of migration. *Journal of Theoretical Biology* 64, 177–197.

- Cornuet, J.-M., Piry, S., Luikhart, G., Estoup, A. and Solignac, M. (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153, 1989–2000.
- Coyne, J.A., Boussy, I.A., Prout, T., Bryant, S.H., Jones, J.S. and Moore, J.A. (1982) Long-distance migration of *Drosophila*. *American Naturalist* 119, 589–595.
- Davies, N., Villablanca, F.X. and Roderick, G.K. (1999a) Bioinvasions of the medfly *Ceratitidis capitata*: source estimation using DNA sequences at multiple intron loci. *Genetics* 153, 351–360.
- Davies, N., Villablanca, F.X. and Roderick, G.K. (1999b) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology and Evolution* 14, 17–21.
- Emelianov, I., Mallet, J. and Baltensweiler, W. (1995) Genetic differentiation in the larch budmoth *Zeiraphera diniana* (Lepidoptera: Tortricidae): polymorphism, host races or sibling species? *Heredity* 75, 416–424.
- Emelianov, I., Drès, M., Baltensweiler, W. and Mallet, J. (2001) Host plant effects on assortative pheromone attraction by host races of the larch budmoth. *Evolution* (in press).
- Endler, J.A. (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.
- Epperson, B.K. and Li, T.Q. (1997) Gene dispersal and spatial genetic structure. *Evolution* 51, 672–681.
- Epperson, B.K., Huang, Z. and Li, T.Q. (1999) Measures of spatial structure in samples of genotypes for multiallelic loci. *Genetical Research* 73, 251–261.
- Feder, J.L., Chilcote, C.A. and Bush, G.L. (1988) Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* 336, 61–64.
- Feder, J.L., Chilcote, C.A. and Bush, G.L. (1990) The geographic pattern of genetic differentiation between host associated populations of *Rhagoletis pomonella* (Diptera: Tephritidae) in the Eastern United States and Canada. *Evolution* 44, 570–594.
- Feder, J.L., Opp, S.B., Wlazlo, B., Reynolds, K., Go, W. and Spisak, S. (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences USA* 91, 7990–7994.
- Felsenstein, J. (1975) A pain in the torus: some difficulties with the model of isolation by distance. *American Naturalist* 109, 359–368.
- Frankham, R. (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research, Cambridge* 66, 95–107.
- Guillaumin, M. and Descimon, H. (1976) La notion d'espèce chez les lépidoptères. In: Bocquet, C., Générmont, J. and Lamotte, M. (eds) *Les Problèmes de l'Espèce dans le Règne Animal*, Vol. I. Société zoologique de France, Paris, pp. 129–201.
- Gupta, J.P., Dwivedi, Y.N. and Singh, B.K. (1980) Natural hybridization in *Drosophila*. *Experientia* 36, 290.
- Haldane, J.B.S. (1930) A mathematical theory of natural and artificial selection. 6. Isolation. *Proceedings of the Cambridge Philosophical Society* 26, 220–230.
- Hartl, D.L. and Clark, A.G. (1989) *Principles of Population Genetics*, 2nd edn. Sinauer, Sunderland, Massachusetts.
- Houck, M.A., Clark, J.B., Peterson, K.R. and Kidwell, M.G. (1991) Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* 253, 1125–1129.

- Ingvarsson, P.K. and Olsson, K. (1997) Hierarchical genetic structure and effective population sizes in *Phalacroscus substriatus*. *Heredity* 79, 153–161.
- Jiggins, C.D. and Davies, N. (1998) Genetic evidence for a sibling species of *Heliconius charithonia* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* 64, 57–67.
- Jiggins, C.D., McMillan, W.O., King, P. and Mallet, J. (1997) The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity* 79, 495–505.
- 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.
- Karciva, P.M. (1982) Experimental and mathematical analysis of herbivore movement: quantifying the influence of plant spacing and quality on foraging discrimination. *Ecological Monographs* 52, 261–282.
- Korman, A., Mallet, J., Goodenough, J.L., Graves, J.B., Hayes, J.L., Hendricks, D.E., Luttrell, R.G., Pair, S.D. and Wall, M. (1993) Population structure in *Heliothis virescens* (F.) (Lepidoptera: Noctuidae): an estimate of gene flow. *Annals of the Entomological Society of America* 86, 182–188.
- Lenormand, T. and Raymond, M. (1998) Resistance management: the stable zone strategy. *Proceedings of the Royal Society of London Series B* 265, 1985–1990.
- Lenormand, T. and Raymond, M. (2000) Analysis of clines with variable selection and variable migration. *American Naturalist* 155, 70–82.
- Lenormand, T., Guillemaud, T., Bourguet, D. and Raymond, M. (1998) Evaluating gene flow using selected markers: a case study. *Genetics* 149, 1383–1392.
- Lenormand, T., Bourguet, D., Guillemaud, T. and Raymond, M. (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* 400, 861–864.
- Loxdale, H.D. and Lushai, G. (1999) Slaves of the environment: the movement of herbivorous insects in relation to their ecology and genotype. *Philosophical Transactions of the Royal Society of London B* 354, 1–18.
- Mallet, J. (1986) Dispersal and gene flow in a butterfly with home range behaviour: *Heliconius erato* (Lepidoptera: Nymphalidae). *Oecologia* 68, 210–217.
- Mallet, J. (2001) Species, concepts of. In: Levin, S. (ed.) *Encyclopedia of Biodiversity*, Vol. 5. Academic Press, New York, pp. 427–440.
- Mallet, J. and Barton, N.H. (1989a) Inference from clines stabilized by frequency-dependent selection. *Genetics* 122, 967–976.
- Mallet, J. and Barton, N.H. (1989b) Strong natural selection in a warning color hybrid zone. *Evolution* 43, 421–431.
- Mallet, J., Barton, N.H., Lamas, G., Santisteban, J., Muedas, M. and Eeley, H. (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* 124, 921–936.
- Mallet, J., McMillan, W.O. and Jiggins, C.D. (1998a) Mimicry and warning color at the boundary between races and species. In: Howard, D.J. and Berlocher, S.H. (eds) *Endless Forms: Species and Speciation*. Oxford University Press, New York, pp. 390–403.
- Mallet, J., McMillan, W.O. and Jiggins, C.D. (1998b) Estimating the mating behavior of a pair of hybridizing *Heliconius* species in the wild. *Evolution* 52, 503–510.
- McCauley, D.E. and Eanes, W.F. (1987) Hierarchical population structure analysis of the milkweed beetle *Tetraopes tetraophthalmus* (Forster). *Heredity* 58, 193–201.

- McMillan, W.O., Jiggins, C.D. and Mallet, J. (1997) What initiates speciation in passion-vine butterflies? *Proceedings of the National Academy of Sciences USA* 94, 8628–8633.
- Nagylaki, T. (1975) Conditions for the existence of clines. *Genetics* 80, 595–615.
- Nunney, L. (1993) The influence of mating system and overlapping generations on effective population size. *Evolution* 47, 1329–1341.
- Peterson, M.A. and Denno, R.F. (1998) Life-history strategies and the genetic structure of phytophagous insect populations. In: Mopper, S. and Strauss, S.Y. (eds) *Genetic Structure and Local Adaptation in Natural Insect Populations*. Chapman & Hall, New York, pp. 263–322.
- Porter, A.H. (1990) Testing nominal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (*Limenitis*: Nymphalidae). *Systematic Zoology* 39, 131–148.
- Provine, W. (1986) *Sewall Wright and Evolutionary Biology*. University of Chicago Press, Chicago, Illinois.
- Roderick, G.K. (1996) Geographic structure insect populations: gene flow, phylogeography, and their uses. *Annual Review of Entomology* 41, 325–352.
- Roughgarden, J. (1979) *Theory of Population Genetics and Evolutionary Ecology: an Introduction*. Macmillan, New York.
- Rousset, F. (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* 145, 1219–1228.
- Sites, J.W., Barton, N.H. and Reed, K.M. (1994) The genetic structure of a hybrid zone between two chromosome races of the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae) in central Mexico. *Evolution* 49, 9–36.
- Slatkin, M. (1973) Gene flow and selection in a cline. *Genetics* 75, 733–756.
- Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science* 236, 787–792.
- Slatkin, M. (1989) Detecting small amounts of gene flow from phylogenies of alleles. *Genetics* 121, 609–612.
- Slatkin, M. (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47, 264–279.
- Slatkin, M. and Barton, N.H. (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43, 1349–1368.
- Slatkin, M. and Maddison, W.P. (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123, 603–614.
- Sokal, R.R., Jacquez, G.M. and Wooten, M.C. (1989) Spatial autocorrelation analysis of migration and selection. *Genetics* 121, 845–855.
- Solignac, M. and Monnerot, M. (1986) Race formation, speciation and introgression within *Drosophila simulans*, *D. mauritiana*, and *D. sechellia* inferred from mtDNA analysis. *Evolution* 40, 531–539.
- Sperling, F.A.H. (1990) Natural hybrids of *Papilio* (Insecta: Lepidoptera): poor taxonomy or interesting evolutionary problem? *Canadian Journal of Zoology* 68, 1790–1799.
- 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.
- Turelli, M. and Hoffmann, A.A. (1991) Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature* 353, 440–442.

- Vitalis, R. and Couvet, D. (2001) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics* (in press).
- Wang, R.L., Wakeley, J. and Hey, J. (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* 147, 1091–1106.
- Weir, B.S. (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, Massachusetts.
- Whitlock, M.C. (1992) Nonequilibrium population structure in forked fungus beetles: extinction, colonization, and the genetic variance among populations. *American Naturalist* 139, 952–970.
- Whitlock, M.C. and McCauley, D.E. (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* 44, 1717–1724.
- Whitlock, M.C. and McCauley, D.E. (1999) Indirect measures of gene flow and migration:  $F_{ST} \approx 1/(4Nm+1)$ . *Heredity* 82, 117–125.
- Wright, S. (1969) *Evolution and the Genetics of Populations*, Vol. 2. *The Theory of Gene Frequencies*. University of Chicago Press, Chicago, Illinois.
- Wright, S. (1978) *Evolution and the Genetics of Populations*, Vol. 4. *Variability Between and Among Natural Populations*. University of Chicago Press, Chicago, Illinois.