# The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations

BRIAN CHARLESWORTH*†, MAGNUS NORDBORG‡
AND DEBORAH CHARLESWORTH†

*Department of Ecology and Evolution, University of Chicago, 1101 E 57th Street, Chicago, IL 60637, USA*

(*Received 3 February 1997 and in revised form 17 June 1997*)

**Summary**

Levels of neutral genetic diversity in populations subdivided into two demes were studied by multi-locus stochastic simulations. The model includes deleterious mutations at loci throughout the genome, causing 'background selection', as well as a single locus at which a polymorphism is maintained, either by frequency-dependent selection or by local selective differences. These balanced polymorphisms induce long coalescence times at linked neutral loci, so that sequence diversity at these loci is enhanced at statistical equilibrium. We study how equilibrium neutral diversity levels are affected by the degree of population subdivision, the presence or absence of background selection, and the level of inbreeding of the population. The simulation results are compared with approximate analytical formulae, assuming the infinite sites neutral model. We discuss how balancing selection can be distinguished from local selection, by determining whether peaks of diversity in the region of the polymorphic locus are seen within or between demes. The width of such diversity peaks is shown to depend on the total species population size, rather than local deme sizes. We show that, with population subdivision, local selection enhances between-deme diversity even at neutral sites distant from the polymorphic locus, producing higher $F_{ST}$ values than with no selection; very high values can be generated at sites close to a selected locus. Background selection also increases $F_{ST}$, mainly because of decreased diversity within populations, which implies that its effects may be distinguishable from those of local selection. Both effects are stronger in selfing than outcrossing populations. Linkage disequilibrium between neutral sites is generated by both balancing and local selection, especially in selfing populations, because of linkage disequilibrium between the neutral sites and the selectively maintained alleles. We discuss how these theoretical results can be related to data on genetic diversity within and between local populations of a species.

## 1. Introduction

The levels of genetic diversity in natural populations, and the forces maintaining this diversity, have been debated by population geneticists for many years. It is well known that diversity may sometimes be present because of the effects of genetic drift on selectively neutral mutations in finite populations (Kimura, 1983). It is also well known that various kinds of selection can maintain diversity (Crow & Kimura, 1970; Gillespie, 1991). These two processes have been

studied together in theoretical models of the effects of balanced polymorphism on genetic diversity at neutral sites close to selectively maintained sites. If selection maintains two or more alleles at a site, neutral sites closely linked to the site under selection behave as though they were in different, partially isolated populations, and an accumulation of neutral mutations in the surrounding region is expected (Strobeck, 1983; Hudson & Kaplan, 1988).

The theory of this process has been well studied for randomly mating populations. At statistical equilibrium between genetic drift and mutation, the degree of increase in variability at a neutral site over the standard neutral expectation caused by linkage to a long-established balanced polymorphism is inversely related to the product of the population size and the frequency of recombination between the neutral site

* Corresponding author.
† Present address: ICAPB, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JT, UK.
  Tel: +44(0) 131-650-5750. Fax: +44(0) 131-650-6564.
  e-mail: brian.charlesworth@.ed.ac.uk
‡ Present address: Department of Genetics, Lund University, Sölvegatan 29, 223 62 Lund, Sweden.

and the polymorphic locus (Strobeck, 1983; Hudson & Kaplan, 1988; Nordborg, 1997). In addition, models that include the joint effects on neutral variability of population subdivision and selection at linked loci show that transient population differentiation can sometimes be greatly prolonged over standard neutral expectations when immigrant alleles at the selected loci are subject to counter-selection (Barton, 1979; Bengtsson 1985; Barton & Bengtsson, 1986; Barton & Gale, unpublished results). Petry (1983) has examined the equilibrium variance in allele frequencies at a neutral locus among local populations when there is this type of selection at a linked locus, but no comparable study of more general neutral models has been done other than that of Nordborg (1997).

Another type of study combining neutral sites with loci under selection is the recent work on the effect of neutral diversity of selection against deleterious mutations at linked loci, known as 'background selection' (Charlesworth *et al.*, 1993; Hudson & Kaplan, 1994, 1995; Nordborg *et al.*, 1996 *a*, *b*). In inbred populations, the effects of background selection are expected to be particularly strong, because the low heterozygosity makes recombination ineffective as far as its populational consequences are concerned (Narain, 1966). Selfing populations should therefore show considerably reduced neutral or nearly-neutral variability (Charlesworth *et al.*, 1993; Nordborg *et al.*, 1996 *b*; Nordborg, 1997). Alternatively, selection may take the form of 'selective sweeps' (Maynard Smith & Haigh, 1974; Kaplan *et al.*, 1989; Stephan *et al.*, 1992; Braverman *et al.*, 1995; Simonsen *et al.*, 1995), where diversity is reduced by repeated fixations of selectively favoured alleles. Again, the reduced effective rate of genetic recombination makes this process particularly important in highly inbred populations (Hedrick, 1980), although its effects on neutral diversity have not been explicitly studied in this case. Recurrent, weak directional or fluctuating selection pressures may also reduce variability at linked sites (Gillespie, 1994).

In real populations, the situation is likely to be much more complicated than in such simple models, since it is very unlikely that only one form of selection will be acting, and species are often divided into numerous partially isolated populations. The situation when background selection and balancing selection both operate has been previously investigated by Monte Carlo simulation (Nordborg *et al.*, 1996 *b*). The reduction in genetic diversity due to background selection when recombination is restricted, or the population is highly inbreeding, still holds true for most neutral sites in the presence of a balanced polymorphism, but diversity shows a pronounced peak close to the locus at which the polymorphism is maintained.

Expressions for expected nucleotide site diversities have now been derived under models that allow for the simultaneous operation of several of these factors,

using the coalescent method (Nordborg, 1997). This approach has generated approximations for the expected coalescence time for a pair of alleles sampled at a neutral locus, and hence for equilibrium neutral diversity levels under the infinite sites model (Kimura, 1971; Hudson, 1990), as a function of the product of population size and recombination distance from a locus with a selectively maintained polymorphism. The model incorporates both population subdivision and background selection. The selected locus which generates the peak of diversity at neutral sites can be subject to either balancing selection (which is uniform across local populations), or local selective differences (whereby different alleles are favoured in different populations). Partial self-fertilization can also be included.

The present paper describes the results of simulations of these various scenarios for populations with different levels of self-fertilization, and compares them with the analytical results. The analytical and simulation results enable two important questions to be investigated in relation to population subdivision. The first is whether the widths of peaks of diversity depend on local deme sizes, or on the population size of the species as a whole. If the widths depend on total species population size, this implies that our previous simulations using limited population sizes would generate much wider peaks than are likely in real situations, so that loci with selectively maintained variation may be detectable empirically by very sharply defined diversity peaks (Nordborg *et al.*, 1996 *b*). The second question is whether data on diversity at neutral loci can be used to distinguish between situations with local selective differences and ones where a balanced polymorphism is maintained within local populations. We also relate the results to empirical data on genetic diversity within and between natural populations. Finally, our findings shed light on the interpretation and utility of different statistics for describing population differentiation, when neutral loci are linked to loci under selection.

## 2. Methods

### (i) *General description of the model*

The simulations of systems with many neutral sites and background selection caused by deleterious mutations at many loci, together with selection at a single further polymorphic locus, were performed as described previously for models of the simultaneous effects on neutral diversity of background selection and balancing selection (Nordborg *et al.*, 1996 *a*, *b*). A total population of $N$ diploid hermaphrodites with selfing rate $S$ was assumed. In simulations with background selection, each chromosome was assumed to have 1000 loci subject to deleterious mutations uniformly distributed along its length, with a total rate $U = 0.8$ per diploid genome. (This unrealistically high value was used in order for background selection

to produce large effects in outcrossing, as well as inbred, populations). Deleterious mutations at all loci were assumed to have the same selection coefficient (0·1) against homozygous mutations, in all the runs involving background selection. The dominance coefficient was 0·2 (except for a few runs with a value of 0·5: see below), so that the fitness of heterozygotes was 0·98 relative to homozygotes for wild-type. To model populations without background selection, $U$ was simply set to zero.

Two demes, of equal and constant size $N/2$, were assumed in these simulations. Migration between the demes was assumed to occur with rate $m$ (the probability that an individual is an immigrant from the other deme). Two alternative selection regimes were used, in addition to the selection just described against deleterious mutations. In both models, all fitness effects interacted multiplicatively to determine genotypic fitness. One selection model assumed the same regime of balancing selection in both demes as in our previous simulations (i.e. there was a locus subject to frequency-dependent selection in the centre of the chromosome). In order to maintain an average allele frequency of 0·5, the fitness of homozygotes for the most common allele in this model was assumed to be $2(1-p)$, and the fitness of the heterozygote was $1·5-p$, where $p$ is the frequency of the common allele (Nordborg *et al.*, 1996*b*). As discussed by Nordborg *et al.* (1996*b*), the exact model of balancing selection is not important for the results; all that matters is that the alleles at the locus in question are maintained indefinitely in the population, close to the deterministic equilibrium frequency of one half.

Alternatively, local selection (without balancing selection within demes) was modelled by assuming directional selection acting in opposite directions in the two subpopulations. The gene involved will be referred to as the locus under local selection. The heterozygous and homozygous fitnesses at this locus are denoted by $1-hs$ and $1-s$, respectively (see Appendix, section (i)). The selection model assumed $h = 0·5$, and $s = 0·5$ for 'strong', or $s = 0·1$ for 'moderate', local selection. With free migration ($m = 0·5$), the model is equivalent to Levene's (1953) model for the maintenance of polymorphism in a randomly mating population with spatial variation in genotypic fitnesses, and the condition $h \leqslant 0·5$ is sufficient for Levene's conditions for maintenance of polymorphism in an infinite population to be met. It is well known that restriction of migration reduces the stringency of these conditions (Karlin, 1982). With the population sizes and selection coefficients used in the simulations, variation is maintained almost indefinitely. The simulations were intended to show the qualitative effects of the various forces studied, and their relative magnitudes, so we did not choose biologically plausible values of the selection coefficients at the locus responsible for selection for local adaptation. Rather, we used values that would produce clear-cut effects.

We refer to the selected locus in both cases as the polymorphic locus.

### (ii) *Simulation methods*

In each generation, the sequence of events was as follows. To model migration between the two demes, a random number is picked for each new individual being generated in a given sub-population in the next generation, to decide whether it is to be a migrant or not. If it is a migrant, it is created from a parent or parents in the other subpopulation, instead of coming from the same subpopulation. If the individual is to be a product of self-fertilization, a single parent is chosen; otherwise two parents are chosen. Parents are chosen according to their relative fitnesses within each deme, in a similar manner to the method used in our previous simulations (Nordborg *et al.*, 1996*a, b*). Each haplotype of the new individual is then created from the genome of the parent by the recombination and mutation processes. No interference between crossovers was assumed in the recombination process. The whole procedure corresponds to fertility selection among adults, followed by gamete production (involving recombination and mutation), followed by migration of zygotes or juveniles.

To obtain the simulation data on neutral diversity under an infinite-sites model (Kimura, 1971), 100 neutral sites were modelled. In most of the runs, individuals had one chromosome, each with a map length of 1 morgan, and with 50 neutral sites on each side of the locus subject to balancing or local selection. In what follows, we refer to the frequency of recombination between the polymorphic locus and a neutral site as $r$. Some runs were done with three chromosomes, with 50 neutral sites on the chromosome with the polymorphic locus in the centre, as just described, and 25 on each of the other two chromosomes, i.e. with a total of 50 loci that were unlinked to the polymorphic locus.

Nucleotide diversities ($\pi$) at each neutral site were estimated from the mean of $2\Sigma z_t (1-z_t)$, over replicated introductions at the site of single variants, where $z_t$ is the frequency of the neutral variant at time $t$, and summation is over all times until either fixation or loss occurs (Charlesworth *et al.*, 1993; Hudson & Kaplan, 1995; Nordborg *et al.*, 1996*a, b*). Each data point for a neutral site represents the mean of $2·4 \times 10^4$ introductions of a variant at the site in question. The total genetic diversity at the neutral sites ($\pi_T$) was also decomposed into that within and between allelic classes at the polymorphic locus. Diversity within allelic classes, which will be written here as $\pi_A$, was estimated from the mean of $2\Sigma \{x_t(1-x_t)+y_t(1-y_t)\}$, where $x_t$ and $y_t$ are the frequencies of the neutral variant within the first and second allelic classes, respectively. Diversity between allelic classes with respect to the polymorphic locus was calculated as the difference between the total diversity values and $\pi_A$,

and is denoted by $\pi_{T-A}$. The within-deme ($\pi_S$) diversity was calculated in an analogous manner, using frequencies of the variants in the two demes, rather than two allelic classes. The component of diversity between subpopulations is denoted by $\pi_{T-S}$. We thus have

$$\pi_{T-A} = \pi_T - \pi_A \tag{1a}$$

and

$$\pi_{T-S} = \pi_T - \pi_S \tag{1b}$$

All the results on neutral site diversity, including those for non-zero selfing rates, are expressed relative to neutral expectation for the case of $S = 0$, i.e. no corrections were made for the reduction in effective population size due to selfing (Pollak, 1987).

$F_{ST}$ values were calculated as the ratio of the between-deme component of diversity to the total (Nei, 1987, p. 162), i.e.

$$F_{ST} = 1 - \frac{\pi_S}{\pi_T} \tag{1c}$$

This is equivalent to $K_{ST}$ of Hudson *et al.* (1992), which is derived from measures of the numbers of nucleotide differences between sequences. A similar quantity, $F_{AT}$, can be defined for variation between allelic classes at the polymorphic locus.

## 3. Results

### (i) *Analytical theory for subdivided populations*

Our simulations were guided by approximate analytical results based on coalescent theory which are presented elsewhere (Nordborg, 1997). In that paper, however, the case of low migration rates between subpopulations subject to local selection was not considered. Since this case is biologically important, it is treated in the Appendix, Section (i). To facilitate comparisons of the simulation results with the

analytical equations for equilibrium diversity levels in the different situations studied, the relevant formulae are summarized in Table 1. As will be shown, there is generally good agreement between the formulae and the simulation results, and qualitatively the theory predicts most of the effects observed. We here describe the effects of the different factors (polymorphism maintained by local selection versus balancing selection within demes, population subdivision versus free migration, presence or absence of background selection, and inbreeding versus outcrossing), making clear where there are discrepancies between the theoretical and simulation results.

The formulae predict neutral diversity values as functions of the migration rates between the subpopulations, the allele frequency $q$ of the counter-selected alleles in each subpopulation under local selection, the equilibrium inbreeding coefficient for a neutral locus in an infinite population with selfing rate $S, \hat{F} = S/(2-S)$ (Crow & Kimura, 1970, p. 94), and the frequencies of recombination between the neutral sites and the selectively maintained locus. In the formulae in Table 1, $\alpha$ is 1 in outcrossing populations without background selection, and is reduced by inbreeding or background selection (see Nordborg *et al.*, 1996 a, b; Nordborg, 1997). As explained in Section 2(ii), diversity is partitioned into biologically meaningful and observable components: within-population diversity, diversity within allelic classes with respect to the polymorphic locus, diversity between allelic classes with respect to the polymorphic locus, and diversity between demes.

### (ii) *No population subdivision*

We start with the simplest case of random mating and no background selection. Because of the high migration rate, the allele frequencies at the polymorphic

Table 1. *Expected neutral nucleotide site diversities relative to the standard value for a randomly mating population of size* N

|  | Partitioned with respect to allelic class | | Partitioned with respect to subpopulation | |
|---|---|---|---|---|
|  | Within | Between | Within | Between |
| Free migration | $\alpha$ | $\dfrac{1}{4N\tilde{r}}$ | $\alpha + \dfrac{1}{4N\tilde{r}}$ | $0$ |
| Subdivision with balancing selection | $\alpha + \dfrac{1}{8Nm} + \dfrac{1}{4N(2m+\tilde{r})}$ | $\dfrac{1}{4N\tilde{r}}$ | $\alpha + \dfrac{1}{4N\tilde{r}} + \dfrac{1}{4N(2m+\tilde{r})}$ | $\dfrac{1}{8Nm}$ |
| Subdivision with local selection | $(1-q)\alpha + \dfrac{1}{2N\tilde{s}}$ | $\dfrac{1}{8N\tilde{m}}$ | $(1-q)\alpha + \dfrac{1}{2N\tilde{r}}$ | $\dfrac{1}{8N\tilde{m}}$ |

$\alpha$ represents the effects on diversity of the frequency of self-fertilization and background selection ($\alpha = 1$ in the absence of these factors, and is otherwise $< 1$). Expressions for $\alpha$ are given by Nordborg (1997). $\tilde{s}$ is the effective selection coefficient against a rare allele with local selection and partial self-fertilization; $\tilde{r}$ is the effective recombination frequency with partial self-fertilization (Nordborg, 1997); $\tilde{m}$ is the effective migration rate with local selection and partial self-fertilization. See the text and Section (i) of the Appendix for further details.
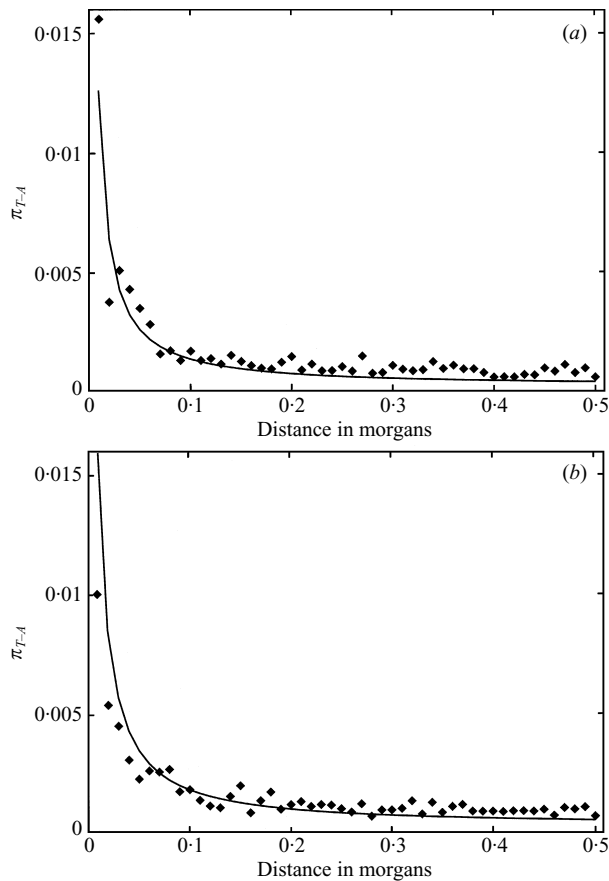
Fig. 1. Neutral genetic diversities from simulations of outcrossing populations of size $N = 2000$, with free migration between subpopulations. Diversity values between allelic classes ($\pi_{T-A}$) with respect to a polymorphic locus (located at map position zero) are shown as a function of map distance. Each data point from the simulation results (filled diamonds) is the average of two results, from equivalent neutral sites on each side of the selected locus. Values predicted from the analytical theory are shown as continuous lines. (*a*) Results of simulations assuming a balanced polymorphism maintained by frequency-dependent selection (see text). (*b*) Results with a polymorphism maintained by strong local selection. There is no background selection.

locus will be very similar in the two demes. With free migration (Table 1 with $Nm \to \infty$), terms involving $1/Nm$ will be very small and we thus expect the two models of selection maintaining polymorphism to yield similar results, with a major fraction of the total diversity being within subpopulations, and a small between-population component of diversity ($\pi_{T-S}$). Diversity within allelic classes, $\pi_A$, should show no dependence on proximity to the polymorphic locus. Our simulation results for a single, non-subdivided ($m = 0.5$), completely outcrossing population with a total $N$ value of 2000 agree with these predictions, for both types of selection. $\pi_S$ values are close to 1 for outcrossing populations with no background selection, and are similar at sites throughout the entire chromosome. $\pi_{T-S}$ is low and also bears no relation to map position.

The component of diversity between allelic classes with respect to the polymorphic locus, $\pi_{T-A}$, is also low for most sites, but there is a peak at sites close to the polymorphic locus, again as expected (see Table 1). The simulation results for this component of diversity for the two selection models are shown in Fig. 1. The values of $\pi_{T-A}$ from the simulations are plotted with the chromosome folded at the polymorphic locus, so that each point shown represents the mean of data from two sites, one from each chromosome arm. Note that, because the simulation models assume no interference, $r \approx 0.32$ for the most distal loci on the chromosome (which are 50 cM from the selected locus), according to Haldane's (1919) mapping function. The predicted diversity values, using the equations in Table 1, are also plotted. Since the analytical theory assumes small $Nr$ (Nordborg, 1997), these predictions are expected to work only for diversity at neutral sites which are rather close to the polymorphic locus. In practice, however, they are often surprisingly similar to the observed simulation results, even for quite loose linkage. For the purposes of making the general nature of the results clear, and since qualitative agreement is excellent, we therefore show the results of plotting the equations for the entire chromosome.

Without population subdivision, genetic diversity under local selection is, as expected, indistinguishable from that when there is balancing selection within each population. Both forms of selection produce peaks in $\pi_{T-A}$ at sites close to the polymorphic locus (whereas there is no such relationship within allelic classes). This is true in populations with high selfing (cf. Nordborg *et al.*, 1996*b*), as well as with random mating (results not shown).

Since the diversity within allelic classes, $\pi_A$ (which is the major component in the models without population subdivision), shows essentially no relationship to the distance from the polymorphic locus, we can simplify the description of the results, and make it easier to compare the effects of selfing and other factors, by comparing the predicted $\pi_A$ with the average from simulations with the same parameter values. The results are summarized in Fig. 2. The first two pairs of columns for each selfing rate set correspond to the two selection models whose results have just been described, illustrating the good quantitative agreement for both outcrossing and selfing populations. The analytical theory accurately predicts the effects of selection on $\pi_A$ for both the balancing selection and local selection models without background selection, even with high levels of selfing (Fig. 2*a*). The effect of inbreeding on $\pi_A$ almost reaches its full extent once populations are 90 % selfing, and a further increase in the selfing rate has little further effect on this diversity measure. Populations with 50 % selfing have diversities intermediate between the fully outcrossing and 90 % selfing values (results not shown). The quantitative predictions for diversity
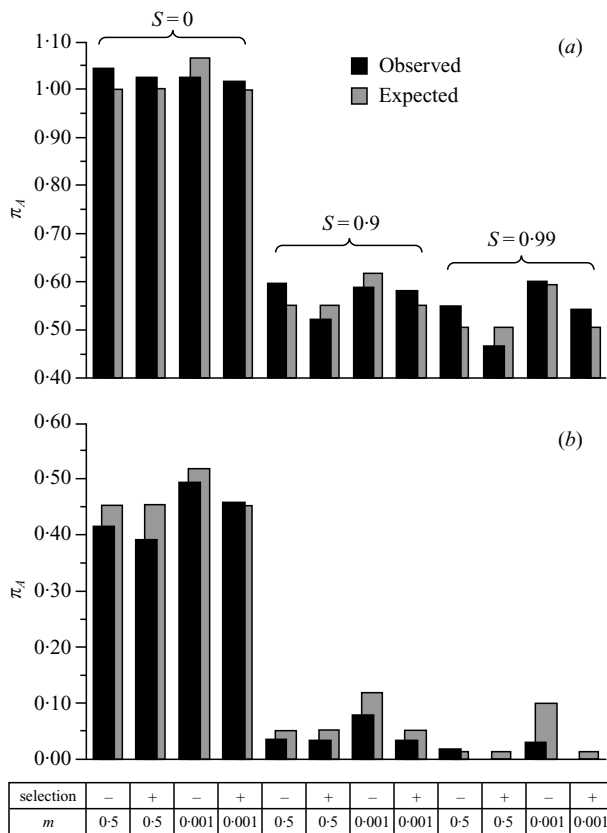
Fig. 2. Observed and expected diversities within allelic classes with respect to a polymorphic locus, for populations of total size 2000 with a range of selfing rates. The results for the two selection models are displayed, with or without subdivision into two demes with restricted migration. (a) Results without background selection; (b) with background selection. The − and + signs in the boxes below (b) indicate the operation of balancing selection or strong local selection, respectively; the values of the migration rate $m$ are indicated in the lower set of boxes.
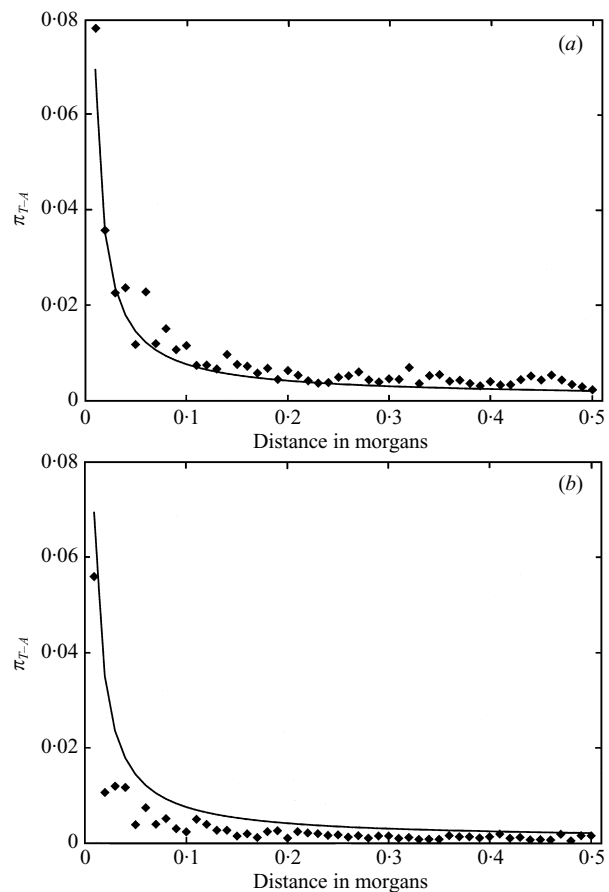


Fig. 3. Observed and expected diversities between allelic classes with respect to a locus under balancing selection (with no local selection or population subdivision), for populations of total size 2000 with selfing rate $S = 0.9$. (a) Results without background selection; (b) with background selection.

between allelic classes with respect to the polymorphic locus are not in complete agreement with the simulation results. For instance, selfing increases $\pi_{T-A}$ much as predicted, but the values are slightly higher than expected at sites loosely linked to a locus under balancing or local selection (Figs. 1, 3a).

With background selection also operating, diversity within populations, and within allelic classes with respect to the polymorphic locus, is reduced. Fig. 2b shows that, for both selection models with free migration, the observed $\pi_A$ values are similar to the expected ones, although the reduction due to background selection is generally somewhat greater than predicted by the analytical theory, in both inbreeding and outcrossing populations. Simulation results are missing for the cases with 99 % selfing with strong local selection and background selection, because the polymorphism was not maintained under these conditions, presumably because the low effective population size induced by background selection with this selfing rate allowed fixation of one allele or the other. Despite these minor discrepancies, however, overall

agreement with the analytical formulae is good, and the simulations agree with the theory in showing a pronounced peak of diversity when chromosomes with different alleles at the polymorphic locus are compared, but show no such localized effect for the other diversity measures. A more important disagreement with the theoretical results is that $\pi_{T-A}$ values are also rather lower than predicted, at least in populations with $S = 0.9$ with a polymorphism maintained by frequency-dependent balancing selection (Fig. 3b), i.e. background selection decreases diversity between allelic classes as well as within them.

(iii) *Subdivided populations with balancing selection*

The equivalence between the effects of balancing selection and local selection on genetic diversity at neutral sites no longer holds when there is population subdivision, i.e. with restricted migration (Table 1). When $m$ is reduced to 0·001 ($Nm = 2$) the theoretical results again give good predictions of diversity within allelic classes, for both inbreeding and outcrossing populations without background selection (right-hand pairs of results within each set in Fig. 2a). For both selection models, unless there is both high selfing and
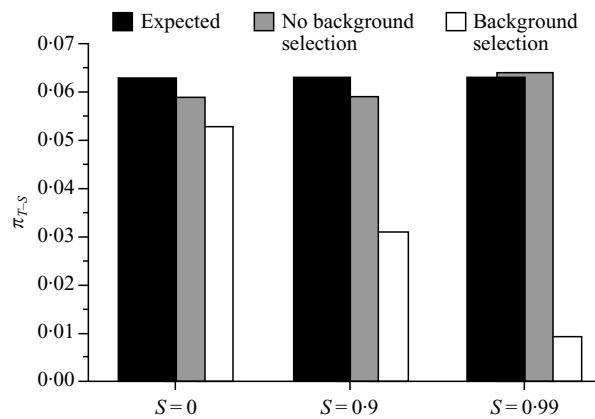
Fig. 4. Observed and expected diversities between two subpopulations with no local selection, for populations of total size 2000, migration rate $m = 0.001$, and selfing rates from 0 to 0.99. Simulation results with and without background selection are shown.

background selection (see below), diversity within allelic classes is the major component, except at sites close to the polymorphic locus. Its magnitude is similar for both models, and $\pi_S$ is roughly the same as $\pi_A$ for most sites.

When there is a balanced polymorphism maintained by frequency-dependent selection within demes, most of the diversity is, as expected, within the demes, and there is a peak in the between-allele component of diversity close to the polymorphic locus. As predicted by the analytical results, the effect of the balanced polymorphism on neutral diversity is proportional to $1/4N\tilde{r}$ (where $\tilde{r}$ is the effective recombination rate; see Appendix, Section (i)), just as in the case of a single population of size $N$. The division into two smaller populations slightly reduces $\pi_S$. $\pi_{T-S}$ is now non-zero, due to the fact that migration between populations is restricted; it bears no relationship to the map position of the neutral sites with respect to the polymorphic locus. From the formula in Table 1, we expect $\pi_{T-S} = 1/(8Nm)$ for any selfing rate, i.e. a value of 0.0625 with our parameters, but the means from the simulations are slightly lower than this value (Fig. 4). The effect of background selection on $\pi_S$ is similar to that in a single population, but $\pi_{T-S}$ is also somewhat reduced (Fig. 4), which is not predicted by the analytical theory. As with the reduction in $\pi_{T-A}$ (see above), the effect is stronger, the higher the selfing rate.
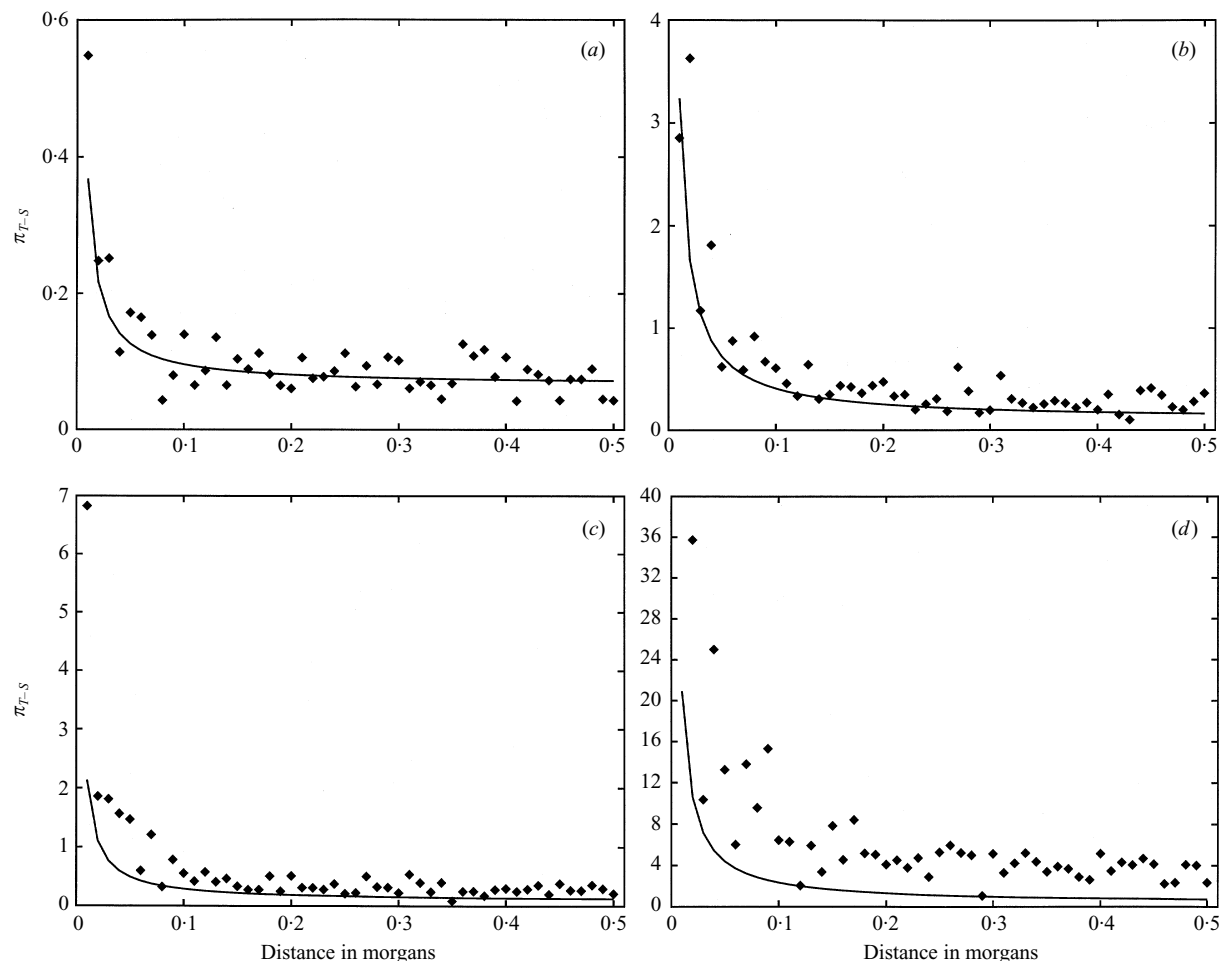


Fig. 5. Observed and expected diversities between two subpopulations of a population of total size 2000, and migration rate 0.001. There is no background selection. Results with moderate local selection are shown for the case of outcrossing (a), and with a selfing rate 0.9 (b); (c) and (d) show the same two cases with strong local selection. Diversity values are displayed as functions of map distance from the locus subject to local selection.
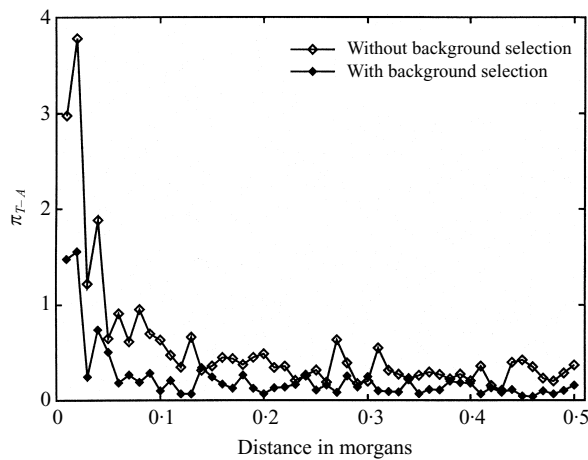
Fig. 6. Simulation results with and without background selection for diversity between allelic classes with respect to a locus subject to moderate local selection, for a subdivided population of total size 2000 ($m = 0.001$) and selfing rate 0.9. Diversity values are displayed as functions of map distance from the locus subject to local selection.
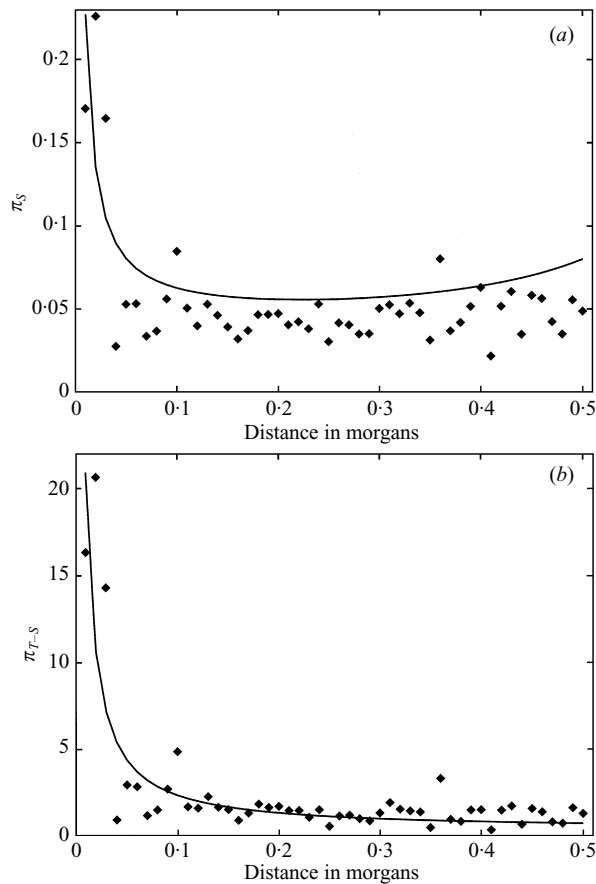


Fig. 7. Observed and expected diversities within (*a*) and between (*b*) subpopulations of a population of total size 2000, $m = 0.001$ and $S = 0.9$, with strong local selection and background selection. The continuous lines are the analytical predictions. Diversity values are displayed as functions of map distance from the locus subject to local selection.

### (iv) *Subdivided populations with local selection*

With local selection, the equilibrium frequency of the favoured allele at the polymorphic locus in each deme

is usually close to 1. When $s \gg m$ the frequency of the rare allele, which is present because of migration from the other deme, is given by

$$q \approx \frac{m}{\tilde{s}}, \qquad (2)$$

where $\tilde{s}$ is the effective selection coefficient against the minority allele in a population defined in section (i) of the Appendix (Haldane, 1930; Nordborg, 1997). (When the above condition does not hold, the disfavoured allele will no longer be rare, and (2) is not exact.)

Application of this expression to the diversity formulae in Table 1 shows that the $\pi_{T-A}$ component of diversity at the neutral sites in a subdivided population is expected to be increased by local selection, and this is indeed found in the simulations. Examples with moderate and strong local selection, and without background selection, are shown in Fig. 5. Although the increased diversity is especially marked at sites close to the polymorphic locus, it extends throughout the chromosome, as the equations predict, and does not disappear with large $\tilde{r}$. With moderate local selection, the equilibrium frequencies of the rare allele in each deme are about 0.02 and 0.01 for populations with $S = 0$ and $S = 0.9$, respectively, and the corresponding frequencies for strong selection are 0.004 and 0.002, but the actual frequencies in the simulations fluctuate widely since $Nm$ is small. Applying (A3) to the results in Table 1, $\pi_{T-S}$ should fall to 0.073 and 0.0128 for loci at the distal end of the chromosome in outcrossing populations with moderate and local selection, respectively, and to 0.170 and 0.956 for populations with $S = 0.9$. The observed values are often higher than these values (Fig. 5). For outcrossing populations, agreement was good for moderate local selection, but with strong selection the mean value from the simulations for the most distal 25 neutral loci was about 0.30 (Fig. 5*C*), and 0.19 for unlinked loci (the expectation for the latter is 0.104). With 90% selfing, even moderate local selection produced greater differentiation than the equations predict, with mean $\pi_{T-S}$ values of 0.29 and 3.87 for distal loci with the two strengths of selection (Fig. 5*b, d*), and 0.15 and 1.55 for unlinked loci (expectations of 0.131 and 0.635, respectively). As expected, the increased diversity between demes is paralleled by a similarly high diversity between allelic classes at the polymorphic locus, which, of course, has very different allele frequencies in the two subpopulations.

With background selection also acting, both $\pi_{T-S}$ and $\pi_{T-A}$ values are somewhat reduced, as in the case without local selection described above. These results are again contrary to the theoretical predictions. With outcrossing, the mean values of $\pi_{T-A}$ for loci loosely linked to the polymorphic locus are similar to those given above within background selection. With 90% selfing, background selection reduces between-allele diversity at loci loosely linked to the polymorphic
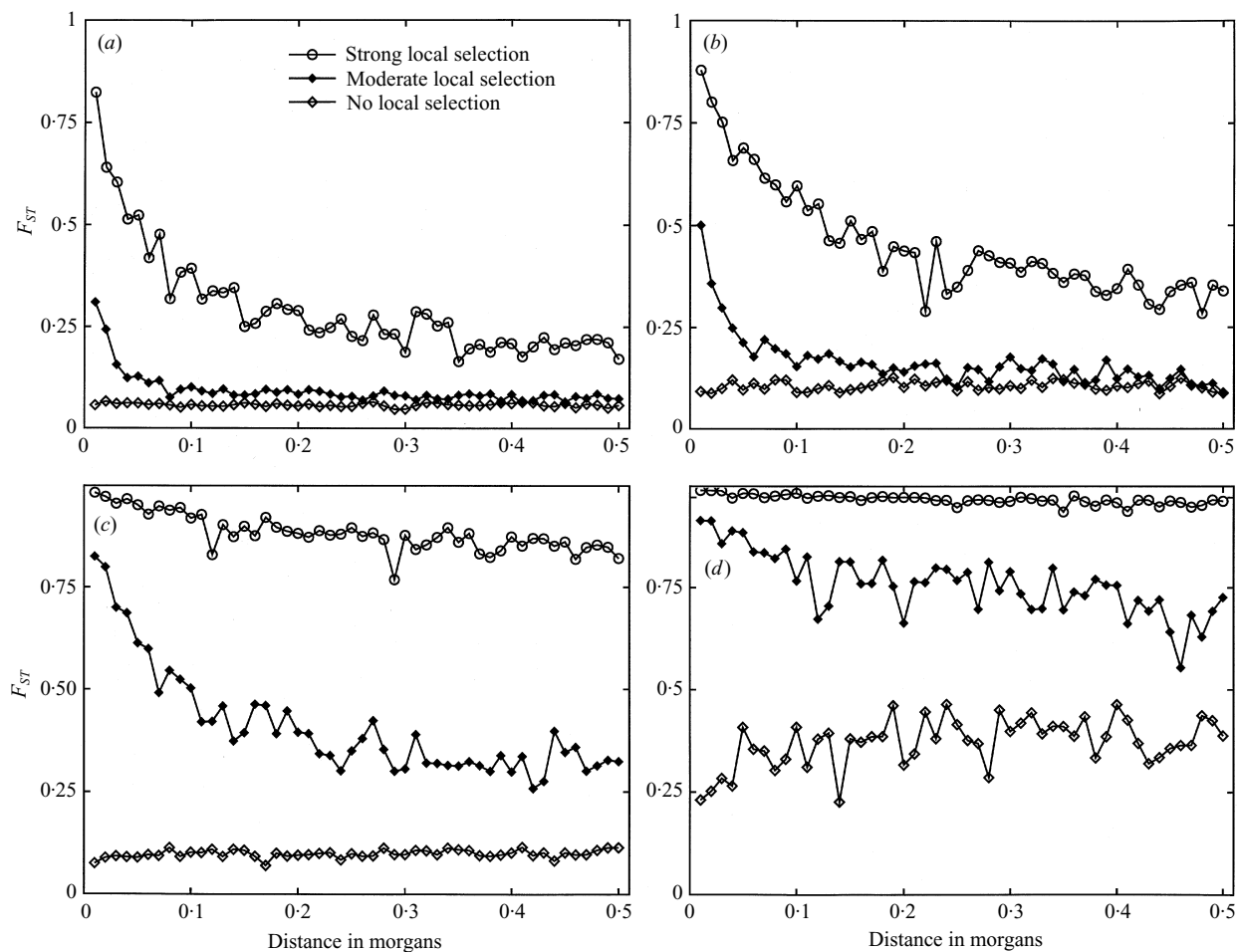
Fig. 8. Observed $F_{ST}$ values in simulations of subdivided outcrossing and inbreeding populations of total size 2000 ($m = 0.001$). (*a*) $S = 0$, no background selection; (*b*) $S = 0$, with background selection; (*c*) $S = 0.9$, no background selection; (*d*) $S = 0.9$, with background selection. Diversity values are displayed as functions of map distance from the locus subject to local or balancing selection.

locus, to about one half of the values given above, for moderate local selection (Fig. 6), and one third for strong selection.

In inbreeding populations, small peaks of within-population neutral diversity were seen at sites close to the locus under local selection, as well as in $\pi_{T-S}$ and $\pi_{T-A}$ (Fig. 7). Such peaks of diversity might give the appearance that a balanced polymorphism is maintained within such populations, when the selection is, in fact, between localities. The existence of such peaks might seem unexpected, but can be explained by the fact that migrant genomes rarely recombine with genotypes in the recipient populations, and so the association between alleles at the polymorphic locus and those at neutral sites is persistent. Because of the rarity of the selectively disfavoured allele in a given subpopulation, the peaks should be small, relative to those between demes, as is indeed observed (Fig. 7). On this interpretation, peaks within populations should be between, rather than within allelic classes, which is consistent with the absence of any detectable peak of diversity within alleles with respect to the polymorphic locus, and with the finding that $\pi_S$

exceeds $\pi_A$. These peaks are predicted by the analytical theory, and should be noticeable when $\alpha$ is small, which will occur when there is selfing and background selection. In theory they are not restricted to selfing populations (Table 1), but in outcrossing populations are too narrow to be seen in our simulations, where the closest neutral site is 1 cM from the polymorphic locus. With background selection, $\pi_S$ in the simulations is also consistently lower than predicted (Fig. 7*a*).

Due to the nearly complete conservation of $\pi_S$ while $\pi_{T-S}$ is increased, total diversity is increased when there is restricted migration as well as local selection and thus, not surprisingly, the proportion of diversity that is between populations is increased by local selection, compared with the situation when a balanced polymorphism is maintained within each local population. In outcrossing populations without background selection, the $\pi_{T-S}$ values in the simulations remain considerably lower than $\pi_S$ (compare, for instance, the results on diversity values between demes or alleles for moderate local selection in Fig. 5*a* with Fig. 2*a* for within-allele values). This is true even

when local selection is very strong; for the outcrossing case corresponding to that just described, the mean $\pi_{T-S}$ over all sites is 0·60 and, of course, most sites distant from the polymorphic locus had lower values than this.

### (v) *Effects of local selection and background selection on $F_{ST}$*

The effects on $\pi_{T-S}$ relative to $\pi_T$ in the simulations are illustrated for different levels of local selection in Fig. 8, in terms of $F_{ST}$ values (see (1c)). The values agree well with the predicted ones for outcrossing populations when there is no local selection or background selection. For the parameters of Fig. 8a without local selection $F_{ST}$ should be unrelated to map distance from the polymorphic locus, and its expected value is 0·059, compared with a mean of 0·059 from the simulations. With local selection, however, agreement with the predictions is poorer. Since between-deme diversity values are higher in the simulations than expected (see above), $F_{ST}$ values are higher than expected on strict neutrality. Background selection does not alter these qualitative conclusions, but it strongly decreases $\pi_S$. Although $\pi_{T-S}$ values also decrease, these effects are relatively minor, so the effect is to increase $F_{ST}$ considerably (Fig. 8a, b). With the parameter values of our simulations, $\pi_{T-S}$ for outcrossing populations nevertheless never exceeds $\pi_S$ (i.e. $F_{ST}$ is always less than 0·5), even with very strong local selection, except at sites close to the polymorphic locus, where $\pi_{T-S}$ becomes extremely high in all cases where local selective differences exist (Fig. 8a, b).

With selfing and local selection, $\pi_{T-S}$ and $\pi_{T-A}$ are again similar in magnitude. The quantitative effects of local selection on $\pi_{T-S}$, and on total diversity, can be much larger than under outcrossing, even comparing cases with only moderate selection, although $\pi_{T-S}$ is still lower than $\pi_S$ unless local selection is very strong (Fig. 8c). When there is background selection as well, the effect can be extremely pronounced, because $\pi_S$ is strongly reduced. Even with moderate local selection, most of the diversity can be between demes, even for loci which are loosely linked to the polymorphic locus (Fig. 8d).

### 4. Discussion

### (i) *The joint effects of background selection and balancing selection on neutral diversity*

We have previously shown that background selection reduces genetic diversity in highly inbreeding populations at most neutral sites, even in the presence of a balanced polymorphism that is maintained for an indefinite period of evolutionary time (Nordborg *et al.*, 1996a, b). In outcrossing populations, unusually high diversity is found only at sites very closely linked to a locus subject to balancing selection. This might

not have been expected to be true in highly selfing populations, because of the reduced effective rate of recombination implied by high homozygosity (Narain, 1966). In agreement with the analytical results, simulations have shown that background selection is effective in reducing diversity, even in very highly selfing populations, except at sites quite closely linked to the polymorphic locus, although the region of enhanced diversity linked to this locus is wider than for outcrossing populations (Nordborg *et al.*, 1996b; Nordborg, 1997).

The results presented here show that the major effect of background selection is still to reduce diversity within populations, even with subdivision of a population and local selection, as in the less complex models studied before (Charlesworth *et al.*, 1993; Nordborg *et al.*, 1996b), i.e. the effects of background selection within populations can largely be considered without regard to events in other demes, even when there is migration between them. There is, however, a tendency for background selection to reduce diversity between demes to some extent (see Section 3(iii)), which is not predicted by the analytical results (Table 1).

We previously suggested that peaks of diversity might offer a means of empirically detecting loci with balanced polymorphisms, and that the reduced diversity caused by background selection in inbred populations should make such populations particularly useful for such studies (Nordborg *et al.*, 1996b). The practicality of this approach relies on the peaks being very local. Since the width of the peak of diversity around a locus with a polymorphism maintained by balancing selection size is proportional to $1/N\tilde{r}$, it will decrease with total population size. Our previous simulations used population sizes much smaller than those of most species, and so must over-predict peak widths associated with polymorphic loci. Given that real populations are often patchily distributed, with small local population sizes, it is important to know whether the peak width is determined by the sizes of the local populations or by the total species population size. In the case of two demes with symmetrical migration studied here, both the analytical and simulation results show that the peak width around a locus with a polymorphism maintained by balancing selection depends on the total species population size, not the local ones. Thus, if the species $N$ value is $10^6$, only loci with $\tilde{r}$ of the order of $10^{-6}$ are expected to see a large effect of balancing selection, so that peaks of high diversity will be very narrow, even in largely selfing populations.

The intuitive reason for this is that the decline in diversity with $\tilde{r}$ reflects the reduced coalescent time for the neutral sites that are far away from the selected locus (Hudson & Kaplan, 1988). With conservative migration in the sense of Nagylaki (1982) (i.e. when migration does not alter deme size), the coalescence time for a pair of neutral alleles sampled from the

same populations is determined by the metapopulation size (Maruyama, 1971; Nagylaki, 1982; Slatkin, 1987; Strobeck, 1987; Hudson, 1990). The effect of linkage to a locus maintained by balancing selection thus reflects the overall rate at which recombination events occur in the species as a whole, which is dependent on $N\tilde{r}$. It remains to be tested whether this holds more generally, but there seems little reason to doubt the generality of the conclusion, given that an effective population size can always be defined for a meta-population with an arbitrary population structure (Nagylaki, 1982; Whitlock & Barton, 1997). It can be conjectured that this effective metapopulation size would replace $N$ in the equations in Table 1, but this remains to be rigorously established.

### (ii) *Effects of population subdivision and local selection on neutral diversity*

Diversity differences under local selection show a different relationship with $N$. Between-population and between-allelic class variability depends on $1/N\tilde{m}$, where $\tilde{m}$ is an effective migration rate that depends on the relation between $\tilde{r}$ and the intensity of local selection (Table 1). With population subdivision, such that $Nm$ is of order 1, local selection can increase diversity between allelic classes with respect to the polymorphic locus, especially at sites closely linked to this locus (these diversity values are very similar to those shown in Fig. 5 for between-deme diversity). Even with moderate local selection and outcrossing, $\pi_{T-S}$ in our simulations rises to about 10 % of the total diversity (averaged over the whole chromosome), compared with an insignificant fraction when there is no local selection. A similar, but more extreme, effect occurs in selfing populations.

Because of the difference between allele frequencies at the selected locus in the two sub-populations, this enhanced diversity is mainly between demes. Local selection lowers the effective amount of migration, because migrants will tend to carry the disadvantageous allele for the populations into which they move, and so are selected against (Barton, 1979; Petry, 1983; Bengtsson, 1985; Barton & Bengtsson, 1986; Barton & Gale, unpublished results). The net effect is to increase $\pi_{T-S}$ (and thus $F_{ST}$) over the neutral expectation. The analytical theory (see Table 1 and the Appendix, Section (i)) predicts substantial effects in many situations but under-predicts the effects found in the simulations, except for outcrossing populations without strong local selection (Section 3(iii)). For sites close to a locus under local selection, extremely high $F_{ST}$ values can be generated, especially when $\pi_S$ is reduced by background selection (Fig. 8).

While the selection coefficients that we have used many seem unrealistically high, it should be borne in mind that larger effects are to be expected from the same total intensity of local selection, the larger the number of selected loci. This is because it is more difficult for a migrant neutral allele to disentangle itself from a block of linked loci subject to negative selection than from a single locus (Barton & Bengtsson, 1986; Barton & Gale, unpublished results). This effect will be very important in highly selfing populations, where even loci on different chromosomes behave as though closely linked (Narain, 1966; Nordborg, 1997). We thus conclude that large effects of local selection are quite possible, even for sites loosely linked to polymorphic loci, especially in inbreeding populations.

### (iii) *Distinguishing between local selection and balancing selection*

In principle, if populations are somewhat isolated from one another, it should be possible to detect local selection, and distinguish it empirically from polymorphism maintained by balancing selection within populations, because only the former will show peaks in diversity between populations (i.e. in $\pi_{T-S}$) at particular locations in the genome, as well as between alleles. Peaks of $\pi_S$ are also generated at neutral sites linked to the locus under local selection, which makes it appear as though there is balancing selection within populations (Fig. 7), but the much larger diversity peaks between populations should still make it possible to distinguish the causes of the peaks. Such tests to determine the kind of selection that is acting can, of course, be applied only if one is aware of the population boundaries, so that one can make comparisons between and within populations. If the population subdivision is not apparent, unexpectedly high $\pi_S$ values may be found, as is sometimes observed in inbred plant populations (Bonnin *et al.*, 1996). This might be incorrectly attributed to balancing selection at a locus closely linked to the one studied, but is distinguishable because diversity should be high for loci throughout the genome in this case.

### (iv) *The effects of local or balancing selection on linkage disequilibria among neutral loci*

With data on diversity at a set of loci in different populations, peaks of diversity may thus indicate the presence of selection and suggest its nature. Such peaks can also be detected as linkage disequilibrium between neutral loci, because selection at a locus produces long coalescence times for alleles at linked loci. Thus alleles linked to one allele at the polymorphic locus should tend to form a set different from those linked to the other allele. If we were to compare alleles at two of these linked loci, they would not be in linkage equilibrium. In the Appendix, Section (ii), we provide some analytical results that relate $\pi_{T-A}$ and $F_{AT}$ (the analogue of $F_{ST}$ for between-allele diversity) values for a pair of loci to their expected squared

linkage disequilibrium and the Ohta–Kimura standardized linkage disequilibrium measure (Ohta & Kimura, 1969), respectively. High levels of $F_{AT}$ for a set of neutral loci correspond to high levels of linkage disequilibrium among them, so that a haplotype structure would be perceived if multi-locus genotype data were collected.

If evidence were obtained for such a situation, it would be helpful to ask whether the different haplotypes tend to be found in different populations, suggesting local selection. If they coexist within populations, some form of balancing selection that can maintain variability within demes is indicated. Studies of inbreeding populations have often found that multi-locus allozyme genotypes tend to be associated with local environments (e.g. Hamrick & Allard, 1972; Nevo *et al.*, 1994), and this has been thought to be evidence for locally co-adapted genotypes (reviewed in Hedrick, 1979). Such data do not, however, necessarily provide convincing evidence for selection on the loci scored, nor for epistatic fitness interactions (Hedrick *et al.*, 1978; Hedrick, 1979). Hedrick (1979) suggested that the correlations between different loci that are to be expected in inbreeding populations will promote genetic hitch-hiking, such that selection for advantageous alleles at one locus will cause allele frequency differences at neutral loci, and he showed that effects of the magnitudes observed could readily be generated. The hitch-hiking model does not account for the maintenance of diversity, however, since advantageous alleles will presumably eventually spread throughout all populations if there is some migration. Data of this kind are more plausibly explained by local selective differences of the kind modelled here.

Our results on the effects of local selective differences focus on steady-state values of diversity measures that will be produced over evolutionary time, rather than on transient alterations that will occur if a genetically variable population experiences a changed breeding system. The results can account for persistent local allele frequency differences at neutral loci, even in the face of some migration, provided that gene flow is sufficiently restricted for local adaptation to occur. Since migration between habitat patches subject to the same selection regimes is not impeded, this process should also eventually lead to similar genotypes at neutral loci in similar environments, as is observed in inbreeding plant populations. Our results do not, however, deal with newly established alleles. If an allele at a polymorphic locus has recently risen to a high frequency under local or balancing selection, it will be expected to show greatly reduced variability with respect to closely linked neutral sites, which will also exhibit strong linkage disequilibrium, but these sites will not be unusually divergent from their counterparts linked to the alternative allele. Asymmetries in the amounts of diversity within allelic classes or between populations would suggest the existence of this kind of transient state (see, for example, Hudson *et al.*, 1994).

## (v) *Discrepancies between simulations and analytical theory*

As discussed above, we have useful approximate analytical expectations for the effects on the expected levels of diversity at neutral loci of population subdivision, selection pressures maintaining variation at other loci in the genome, and background selection. In general, the theoretical results fit our simulation results well, although there are some discrepancies, particularly in cases where there is background selection and high selfing. This good agreement suggests that we have a reasonable understanding of the major features of these rather complex situations, subject to the assumptions of the model. For the most part, the different factors act nearly additively (see Table 1).

Some of the discrepancies between the theory and the simulation results may be due to our having simulated rather small populations ($N = 2000$). The theory assumes constant allele frequencies at the selected loci, which is clearly untrue in practice. In addition, the approximations used to derive the expected diversity values in the case of frequency-dependent balancing selection assume small $Nr$. We thus expect the best agreement in our simulations either for diversity measures where proximity to a locus subject to such selection has no effect, or for sites close to such a locus. We have few data for sites tightly linked to the balanced polymorphism, because of the long coalescence times at such sites, but it is clear that there is in general no obvious pattern of this kind. Large population sizes would reduce fluctuations in allele frequencies at the selected locus and should thus improve agreement, but this remains to be rigorously tested.

With background selection, diversity between allelic classes with respect to a polymorphic locus was lower than predicted, with both local and balancing selection. This occurred at neutral sites close to the polymorphic locus at least as much as at distant sites, so that high $Nr$ is not the explanation. Background selection was also associated with reduced diversity between demes, as well as within them. An assumption of the analytical theory that may be relevant to these discrepancies is the absence of linkage disequilibrium between the loci subject to background selection. With population subdivision, especially with local selection, this is likely to be violated in the simulations. Subdivision will cause between-deme heterosis, since genotypes from within each deme are more likely to share deleterious alleles at any given locus than hybrid individuals formed from crosses between demes (or between allelic classes, in the case of a balanced polymorphism maintained within a deme). Migration and recombination rates will therefore be somewhat

greater in the simulations than expected if we ignore this effect, particularly in inbred populations. The presence of large numbers of deleterious mutations may therefore reduce diversity between allelic classes and between demes.

To test this, some runs were done with additive effects of the deleterious mutations. In this case, one would predict that effects of this kind would be much smaller than those observed with our standard dominance coefficient of 0·2 (although some heterosis is still expected because mutations at different loci affect fitness multiplicatively rather than additively in our simulations, so that hybrids will always have higher fitness than the average genotype). In runs with subdivided populations, 90 % selfing, and no local selection, however, changing the dominance coefficient to 0·5 gave almost no reduction in the effect of background selection on $\pi_{T-S}$, though it somewhat reduced the effect on $\pi_{T-A}$. Heterosis thus cannot account for most of this discrepancy. Further theoretical and simulation work is needed to solve this problem.

### (vi) *Measurement of between-population diversity*

In the next section, we briefly review empirical findings on patterns of genetic diversity, in order to ask how the models here help us to understand them, but before doing so it is worthwhile to discuss how genetic diversity is quantified. There is no single standard notation, so we here review some of the terminology. Total genetic diversity can be partitioned into within- and between-population values. For allozyme data, the components are usually denoted by $H_T$, $H_S$ and $D_{ST} = H_T - H_S$ (Nei, 1973, 1987; Brown, 1979), where the within-deme diversity ($H_S$) and the excess between populations ($D_{ST}$) are the allozyme analogues of the nucleotide-site diversity measures $\pi_S$ and $\pi_{T-S}$ which we have used above. Diversity between sets of populations is often measured by quantities such as $F_{ST}$, which for allozymes can be defined as $(H_T - H_S)/H_T$ (i.e. the ratio of the excess between-deme diversity to that within demes), which is identical to Nei's $G_{ST}$ (Nei, 1973). Such measures have the advantage that data using different types of genetic markers, such as allozymes and DNA sequence diversity, can be compared, since they measure the relative magnitude of between- versus within-population diversity. Other measures, which differ slightly from these, have been proposed by Weir & Cockerham (1984) and Slatkin (1993).

A serious problem with $F_{ST}$, however, is that it depends inversely on within-deme diversity (Nei, 1973, 1987, p. 190). As explained in Section 1, it is now known that several processes (bottlenecks, background selection, selective sweeps and fluctuating selection) can reduce within-deme diversity. Effects of selection at linked loci will be more important in inbreeding than outcrossing populations, and in genomic regions where recombination is infrequent (Charlesworth *et al.*, 1993). Measures such as $F_{ST}$, and measures derived from it (see Brown, 1979; van Dijk *et al.*, 1988), can be very useful for descriptive purposes, but the fact that they depend on both genetic divergence and genetic diversity within populations causes problems when comparisons are being made between species with different mating systems or between regions of the genome with different rates of crossing over. This is especially true if interpretations of the causes of different patterns of between-population variation are to be attempted. It may often be better for comparative purposes to use $D_{ST}$ values for describing allozyme data (analogous to our $\pi_{T-S}$). These are not as often reported as $F_{ST}$, however, and are also less used in theoretical studies. In reviewing the published data (see below), it is thus necessary to consider $F_{ST}$.

There will also be problems if data from loci linked to loci subject to local selection are used, as both $\pi_{T-S}$ and $F_{ST}$ can be much higher than for unlinked loci (see Section 4(iii)). In selfing populations, where loci are not independent even if they are not close in the genetic map, this is a real possibility. The effects of local selection, inflating $F_{ST}$ for neutral loci elsewhere in the genome, imply that estimates of $Nm$ derived from $F_{ST}$ (e.g. Slatkin, 1993) could seriously underestimate migration if local selection is occurring. Estimates based on loci closely linked to a locus subject to local selection might be seriously in error, but even neutral loci distant from any selected loci will yield biased estimates (Fig. 8).

The difficulties in using $F_{ST}$ to interpret differences in patterns of diversity are exemplified by data on sequence divergence between isolated *Drosophila* populations. Such data might offer a test of our prediction that, even in outcrossing populations, local selection on certain loci in the genome should sometimes produce large differences in allele frequency at other loci, if gene flow is restricted (see Fig. 5*a*). Other things being equal, these effects should be most pronounced in genomic regions with low recombination, if loci involved in local adaptation are present in all regions. Results of this kind have been reported for sequence comparisons of loci in regions of low versus high recombination regions of *D. anannassae* (Stephan, 1994) and *D. melanogaster* (Begun & Aquadro, 1993, 1995). However, $\pi_S$ is much lower at the loci in the regions with high $F_{ST}$ than at loci with low $F_{ST}$ in these studies. Whatever the causes of these difference in $\pi_S$, the above discussion suggests that comparisons of divergence between populations should be made on the basis of the extent of $\pi_{T-S}$, not $F_{ST}$. When such a comparison is made, there is no evidence for a significant effect of chromosomal region (B. Charlesworth, unpublished results). Thus, these apparent differences in the extent of divergence are explicable solely in terms of differences in levels of within-population variability.

Table 2. *Summary of studies of comparisons of allozyme or microsatellite diversity in related populations with different selfing rates* (S)

| Genus | Species or population | $S$ | $H_S$ | $H_T$ | $D_{ST}$ | $F_{ST}$ | Reference |
|---|---|---|---|---|---|---|---|
| *Phlox* | *drummondii* | 0 | 0·176 | 0·246 | 0·070 | 0·283 | Levin (1978) |
| | *roemeriana* | 0 | 0·277 | 0·356 | 0·079 | 0·223 | |
| | *cuspidata* | High | 0·046 | 0·074 | 0·028 | 0·373 | |
| *Oenothera* | *grandis* | 0 | 0·19 | 0·22 | 0·03 | 0·136 | Ellstrand & Levin (1980) |
| | *laciniata* | High | 0·14 | 0·18 | 0·04 | 0·222 | |
| *Plectritis* | *congesta* | 0·3 | 0·226 | 0·266 | 0·040 | 0·150 | Layton & Ganders (1984) |
| | *brachystemon* | 0·98 | 0·060 | 0·166 | 0·106 | 0·639 | |
| *Gilia* | *achilleifolia* (4 pops.) | 0·21 | 0·209 | 0·272 | 0·063 | 0·232[a] | Schoen (1982) |
| | *achilleifolia* (3 pops.) | 0·71 | 0·139 | 0·300 | 0·161 | 0·537[a] | |
| *Lolium* | 3 spp. | 0·08[b] | 0·341 | 0·384 | 0·043 | 0·099 | Loos (1993) |
| | *remotum* | High | 0·004 | 0·442 | 0·438 | 0·750 | |
| *Plantago* | *lanceolata* | 0 | 0·127 | 0·133 | 0·006 | 0·045 | Wolff (1991) |
| | *major* (total) | High | 0·047 | 0·056 | 0·018 | 0·321 | |
| *Mimulus* | *guttatus*[c] | 0·49 | 0·506 | 0·657 | 0·151 | 0·230 | Awadalla (1996) |
| | *lacinatus*[c] | 0·88 | 0·311 | 0·642 | 0·321 | 0·515 | |

[a] Re-calculated from the published $D_{ST}$ values.
[b] Based on the published $F_{IS}$ value.
[c] Based on microsatellites.

(vii) *Patterns of neutral diversity in inbreeding and outbreeding populations*

There are few good data to compare with the results of our models, apart from the *Drosophila* results which we have just discussed. The ideal data would be DNA sequence diversity, preferably using neutral differences, since these relate most directly to co-alescence times and to the infinite sites model used in our simulations. Data on several loci are needed because, if there are local selection pressures, between-population differences are not independent of map position, and very high values can be produced at loci in the region of a locus subject to local selection (see Section 4(iv) above). Allozyme data are much less useful because, even assuming selective neutrality, differences between genotypes probably do not in-crease linearly with coalescence times (Marshall & Brown, 1975). Moreover, allozyme allelic differences may not be neutral. Several kinds of evidence suggest some selection on such variants (e.g. Karl & Avise, 1992; Pogson & Zouros, 1994; Kreitman, 1996). Since nearly all currently available data are from allozyme surveys, however, we will discuss them here.

It has been estimated that between-population allozyme diversity values of selfing plants are up to 5 times greater than those of outcrossers. Hamrick & Godt (1996) find that breeding system is a major factor influencing $F_{ST}$ values of plant species, being involved in all the two-trait combinations that explained large proportions of the variance in this quantity. They estimate average $F_{ST}$ values for inbreeding and outcrossing dicotyledonous plants to be 0·587 and 0·184, yielding a ratio of 3·2. Much of this difference is caused by the approximately twofold higher within-population diversity in outbreeding species (0·165 *v.* 0·091 for selfers). The few available within-genus comparisons available give similar results (Table 2). In addition, a survey of microsatellite diversity in the highly inbreeding plant *Arabidopsis thaliana* found very high diversity between populations and uniformity within populations (Todokoro *et al.*, 1995); in the inbreeding plant *Medicago truncatula*, 67% of the variance in gene frequency for randomly amplified DNA (RAPD) markers was found to be between subpopulations (Bonnin *et al.*, 1996). A similar result has been found for restriction fragment length polymorphism (RFLP) variation in *A. thaliana* (M. Kreitman, personal communication). It would be interesting to have comparable surveys of related outbreeding species. Such a comparison in *Mimulus* species, using microsatellites, displays a pattern consistent with the other data just cited, even though the species compared differed only moderately in their selfing rates (Awadalla & Ritland, 1997); in this dataset, much of the difference in $F_{ST}$ is clearly caused by lower $H_S$ in the more inbreeding populations.

The existence of consistent differences in patterns of diversity between inbreeding and outcrossing species, with inbreeders having low within-population di-versity and often low total diversity, but relatively or even absolutely high divergence between populations, is well known in plant population biology. The results summarized here agree well with those reviewed in the past for morphological and quantitative genetic diversity, as well as diversity for genetic markers (Baker, 1953; Brown, 1979; Mitchell-Olds & Bergelson, 1990; Heywood, 1991).

## (viii) *Interpretation of the observed patterns*

A full explanation does not yet exist for the patterns just reviewed. Previous discussions have mostly focussed on the effect of changing from an outbreeding to an inbreeding mating system, assuming the prior existence of genetic diversity. For instance, it is often stated that inbreeding alone should not lead to loss of genetic variation, but merely changes how it will be distributed (Hamrick & Godt, 1990). This ignores the effect of inbreeding on effective population size (Pollak, 1987), which must affect the long-term equilibrium diversity. Furthermore, the breeding system influences the maintenance of diversity under selection. For instance, balanced polymorphisms maintained by overdominance are maintained under more restrictive conditions in inbred than outbred populations (Kimura & Ohta, 1971), and the breeding system also influences expected levels of quantitative genetic variability (Charlesworth & Charlesworth, 1995). With spatial differences in selection, inbreeding can relax the conditions for different alleles to be maintained in different subpopulations, presumably because it reduces gene flow via pollen (Hedrick, 1990). Background selection and selective sweeps will have much more profound effects on within-population diversity in inbreeders than outbreeders (Charlesworth *et al.*, 1993).

To understand the possible causes of breeding system differences in diversity patterns, it is helpful to review briefly the effects of different factors on individual diversity measures in such populations at statistical equilibrium under mutation and genetic drift. Isolation alone enhances $\pi_{T-S}$, but within-population diversity can be predicted from the effective metapopulation size, which is calculated by weighting the contributions of each deme appropriately (Maruyama, 1971, 1972; Nagylaki, 1982; Slatkin, 1987; Strobeck, 1987; Tajima, 1990, 1992; Whitlock & Barton, 1997). Large asymmetries in deme size or highly asymmetric patterns of migration that lead to a few populations being the source of most migrants can greatly reduce the effective metapopulation size below the total species number (Tajima, 1990, 1992; Nordborg, 1997). Local extinction and recolonization can also lead to greatly reduced effective metapopulation size (Maruyama & Kimura, 1980; Whitlock & Barton, 1997). We have not included these complications in the present study, but they may obviously be of considerable importance in many cases in nature.

As intuitively expected, subdivision without local selection increases the fraction of diversity between, as opposed to within, populations, regardless of proximity to a locus under balancing selection. Selfing reduces the effective sizes of the subpopulations, which will tend to increase $\pi_{T-S}$ and decrease $\pi_S$. Thus a higher proportion of the total variability is between demes, compared with outcrossing populations. As-

suming the infinite allele neutral model and the island model of migration, Maruyama & Tachida (1992) obtained an expression for $F_{ST}$ as a function of the selfing rate. Nordborg (1997, eq. 55) has also derived such a result for the case of two populations with migration, using the coalescence approach, and showed that $F_{ST} = 1/(1 + 4Nm/[1 + \hat{F}])$, where $\hat{F}$ is the equilibrium inbreeding coefficient in an infinite population with a given rate of selfing. These results for neutral models show that selfing alone can increase $F_{ST}$ by at most a factor of 2, as expected if the effect is entirely due to reduction in $\pi_S$. Our simulations with no local selection (but with a polymorphism maintained within populations by balancing selection) agree with this conclusion (compare the cases with no local selection in Fig. 8 *a, c*). Very high observed values of $F_{ST}$ in selfing versus comparable outcrossing populations must therefore be attributed to additional effects of the kinds discussed above.

Can we explain the observed differences in diversity by local selection pressures combined with population subdivision, together with the differences in breeding systems, or do we need background selection as well, in order to account for the data? Although our simulations use parameter values that are not biologically very realistic, it is nevertheless worth asking whether effects of magnitudes comparable to those actually observed can be produced under plausible assumptions with restricted migration. The simulations with no local selection produced $F_{ST}$ values around 0·1 for populations with 90 % selfing, and lower with outcrossing (Fig. 8). These are similar to those estimated using allozymes in outcrossing plants, as just discussed. With background selection, the values increase to about 0·4 for inbred, and 0·1 for outbred populations, respectively. Background selection alone could thus potentially account for increased $F_{ST}$ under selfing, although our simulations assumed a high per chromosome mutation rate to deleterious mutations (Section 2(i)), which makes the results unrealistic for outcrossing populations. In highly inbreeding populations, however, the lack of independence of loci means that the mutation rate used is not unreasonable as a per-genome value (Charlesworth *et al.*, 1990; Johnston & Schoen, 1995).

Local selection could also explain the observed differences in $F_{ST}$ between inbreeders and outbreeders. Moderate local selection, without background selection, increased $F_{ST}$ in our simulations to a similar extent to background selection, except for loci close to the selected locus, where very high values occurred (Fig. 8). Our local selection models thus yield values that are reasonable for both outcrossing and inbreeding populations. Moreover, the selection coefficient assumed in our model of moderate local selection is similar to the lower range of values estimated from reciprocal transplant experiments on plants between different environments; in some cases, there is evidence for selection as intense as that

assumed in our strong local selection model (Bell, 1996; Linhart & Grant, 1996). These estimates presumably represent somewhat stronger selection than typically occurs, as there is probably some bias towards studying and reporting situations with detectable selection. However, they measure selection on entire genotypes, whereas our model assumes a single genetic difference. Since even a single locus affects a wide region of the genome, it is likely that the effects of several selected loci produce larger effects than we observe, for the same total selection intensity (see Section 4(ii)). Models of local selection involving several loci require future study, but it appears that our selection parameters may not be unreasonable for plant populations. Simulations of selfing populations yielded $F_{ST}$ values for loci loosely linked to the selected locus that also correspond with those often observed, with values of 0·4–0·5 under moderate local selection. When there was also background selection the values were higher than most observed ones. Greater migration would, of course, reduce the values. On the other hand, in our models we compared inbreeders and outbreeders with the same migration rates, but in inbreeding plant populations migration via pollen is effectively lowered because of the rarity of outcrossing (Hedrick, 1990). This will only cause increased between-population differentiation, and hence cannot in itself explain the greatly reduced within-population variability associated with inbreeding (Table 2).

The evidence that inbreeding populations generally have lower within-population diversity levels than outbreeders suggests a possible role for background selection or other forms of selection at linked loci. However, other processes such as population bottlenecks associated with colonization events could lead to reduced within-population diversity and increased $F_{ST}$ values. As pointed out by Ingvarsson *et al.* (1997), increases in $F_{ST}$ due to extinction and recolonization (Slatkin, 1977; Wade & McCauley, 1988; Whitlock & McCauley, 1990; Ingvarsson, 1997) may occur primarily through reductions in within-population diversity rather than by increases in absolute between-population diversity, since the effective size of a metapopulation is reduced by local extinction and recolonization events (Whitlock & Barton, 1997), causing a reduction in total diversity in the metapopulation (Maruyama & Kimura, 1980). Absolute measures of between-population differentiation may thus be reduced by extinction and recolonization, in contrast to the effects of local selection. Since our simulation results for highly selfing populations indicate that background selection also reduces between-population differentiation (Fig. 4), the role of selfing in enhancing the effects of selection at linked sites may be difficult to distinguish from the effects of local extinction and recolonization events by this criterion alone, if such events are more common in inbreeders.

In has in fact been suggested that bottlenecks are more important in inbreeding than outbreeding populations, because levels of diversity (and $N_e$ values estimated from them) vary more from one population to another in inbreeders, consistent with recent bottlenecks in the ancestry of populations with low diversity values (Schoen & Brown, 1991). However, this evidence for bottlenecks is not conclusive, and the data could as well be interpreted in terms of low average within-population diversity in inbreeders caused by background selection or other factors. The occasional populations with high values (Schoen & Brown, 1991) could arise by population mixture, by local selection when subdivision is not apparent (see above), or even by their being the product of the recent evolution of selfing. Since it is likely that selfing often evolves, but selfing taxa do not persist for very long evolutionary time periods (Stebbins, 1957; Schoen *et al.*, 1996), this last possibility cannot be ignored. Unless direct evidence for bottlenecks is obtained, or the alternative interpretations ruled out, this issue is unresolved.

## Appendix

### (i) *The effective rate of migration with population subdivision and local selection*

This section extends the approach of Bengtsson (1985) and Barton & Bengtsson (1986) to the case of a partially selfing population, using a heuristic argument. We assume that there are two subpopulations of size $N/2$, connected by migration at rate $m$. As in the text, let the frequency of the favoured allele at the locus subject to local selection in a given subpopulation be $p$, and the frequency of the disfavoured allele be $q$ ($q$ is assumed to be so small that second-order terms in $q$ are negligible). The fitnesses of the heterozygotes and homozygotes for the disfavoured allele are $1-hs$ and $1-s$, respectively. If $q$ is small, the effect of selection on genotype frequencies in each generation is of order $sq$. The deviation of Wright's fixation index, $F$, from its neutral value must therefore also be $O(sq)$ (cf. Nagylaki, 1997). With a frequency of self-fertilization $S$, the equilibrium $F$ value for a neutral locus is $\hat{F} = S/(2-S)$ (Crow & Kimura, 1970, p. 94) so that $F$ in the present case is approximately equal to $\hat{F} + O(sq)$. Neglecting terms in $O(q^2)$, the standard selection equation with non-random mating (Wright 1969, p. 244) implies that the change in frequency of the disfavoured allele due to selection is

$$\Delta q_s \approx -qs\{\hat{F} + h(1-\hat{F})\}. \tag{A 1 a}$$

The effective selection coefficient against the locally disfavoured allele is therefore given by

$$\tilde{s} \approx s\{\hat{F} + h(1-\hat{F})\}. \tag{A 2 b}$$

If we follow the fate of a set of disfavoured alleles introduced into a particular subpopulation in a given generation, it follows that their frequency will be reduced by a factor of $(1-\tilde{s})$ each generation. If migration precedes selection in each generation (as assumed in our simulations), the frequency of this cohort of disfavoured alleles will be reduced by $(1-\tilde{s})^i$ after $i$ generations. If selection precedes migration, the factor is only $(1-\tilde{s})^{i-1}$. Clearly, only the survivors of selection contribute to the gene pool of the subpopulation in any given generation.

Next, consider a neutral locus that is linked to the selected locus. With probability $p$, the migrants into a given subpopulation will carry a neutral allele associated with the allele at the locally selected locus that is now disfavoured by selection. As shown by Nordborg (1997), the effective rate at which the neutral allele recombines with the selected allele in an inbred population is

$$\tilde{r} \approx r(1-\hat{F}). \tag{A 2}$$

If migration precedes selection, the probability that the neutral allele persists for $i-1$ generations, and then recombines to become associated with the locally favoured allele is thus $(1-\tilde{s})^i (1-\tilde{r})^{i-1}\tilde{r}$. Summing over all generations, the net fraction of neutral alleles which are initially associated with the locally disfavoured allele, but escape selective elimination, is the sum of terms of this form over all generations, i.e. $(1-\tilde{s})\tilde{r}/(1-[1-\tilde{r}][1-\tilde{s}])$. We can define an effective migration rate, $\tilde{m}$, for the neutral locus as the product of the migration rate and the probability that the neutral allele survives selection (Bengtsson, 1985). If the case in which the allele at the neutral locus is initially associated with the locally favoured allele at the selected locus (which has probability $q$) is taken into account, we have

$$\tilde{m} \approx m\left\{q + \frac{p(1-\tilde{s})\tilde{r}}{(1-[1-\tilde{s}][1-\tilde{r}])}\right\}. \tag{A 3}$$

If selection precedes migration, the factor $(1-\tilde{s})$ in the numerator of the second term in braces is omitted.

The degree of differentiation between subpopulations at a neutral locus is then given by using this effective migration rate in the standard equation for migration and drift in the case of two subpopulations (Hudson, 1990; Slatkin, 1991), i.e.

$$\pi_{T-S} = \frac{1}{8N\tilde{m}}. \tag{A 4}$$

A similar result was obtained by Petry (1983), for the case of an island model of migration with local selection, such that immigrants into a local population are disfavoured by selection at a particular locus. He showed that, for weak selection and low recombination (such that $r$ and $hs$ are of the order of the reciprocal of population size), the effective migration rate for a neutral locus linked to the selected locus is $\tilde{m} = r/(r+hs)$, and that this accurately predicts the variance in allele frequencies between subpopulations when substituted into the standard neutral formulae. This expression for effective migration rate is close to (A 3) for the random-mating case if selection is weak, $q$ is small, and recombination is rare.

### (ii) *Linkage disequilibria with two allelic classes or two subpopulations*

Assume that there are two alleles maintained by selection, with frequencies $p$ and $q$. Consider a neutral site $i$ which has recombination frequency $r_i$ with the polymorphic locus, and that is segregating for two nucleotide variants. Let $\delta_i$ be the difference in the frequency of one of these variants between the two allelic classes at the polymorphic locus. The coefficient of linkage disequilibrium between the polymorphic locus and site $i$ is (cf. Nordborg *et al.*, 1996*a*)

$$D_i = pq\delta_i$$

and the corresponding correlation coefficient is

$$\rho_i = \frac{pq\delta_i}{\sqrt{[pqx_i(1-x_i)]}}, \tag{A 5b}$$

where $x_i$ and $1-x_i$ are the frequencies of the two alternative variants at site $i$.

It follows from equation (5) of Nei & Li (1973) that the components of the linkage disequilibrium and correlation coefficients between two segregating neutral sites $i$ and $j$ due to the differences in their variant frequencies between the two allelic classes are

$$D_{ijT-A} = (pq)^2\delta_i\delta_j$$
$$= D_i D_j \tag{A 6a}$$

and

$$\rho_{ijT-A} = \rho_i\rho_j. \tag{A 6b}$$

If we take expectations of the squares of the right-hand side of (A 5*a*) over the statistical equilibrium, the expected squared linkage disequilibrium between the polymorphic locus and the $i$th segregating neutral site is thus

$$E\{D_i^2\} = (pq)^2 E\{\delta_i^2\} = \tfrac{1}{2}(pq)\,\pi_{iT-A}, \tag{A 7a}$$

where $\pi_{iT-A}$ is the between-allele component of diversity under the infinite sites model for site $i$. This follows from the relation $\pi_{iT-A} = 2pqE\{\delta_i^2\}$.

Similarly, if we take expectations of the squares of the denominator and numerator of the right-hand side of (A 5*b*) and substitute into (A 6*b*), we obtain the Ohta–Kimura (Ohta & Kimura, 1969) standardized linkage disequilibrium measure for the polymorphic locus and the $i$th neutral site

$$\sigma_{id}^2 = \frac{\pi_{iT-A}}{2E\{x_i(1-x_i)\}} = F_{iAT}, \tag{A 7b}$$

where $F_{iAT}$ is analogue of $F_{ST}$ for the between-allele diversity relative to within-allele diversity at site $i$.

A similar argument applied to the component of linkage disequilibrium between two neutral sites due to differences in variant frequencies between the two allelic classes yields the expressions

$$E\{D^2_{ijT-A}\} = \tfrac{1}{4}(pq)^2 E\{\delta_i \delta_j\}, \tag{A 8\,a}$$

$$\sigma^2_{ijT-A} = \frac{(pq)^2 E\{\delta_i^2 \delta_j^2\}}{4E\{x_i(1-x_i)\,x_j(1-x_j)\}} \tag{A 8\,b}$$

Correlations between the genealogical histories of the two neutral sites must cause positive covariances between $\delta_i^2$ and $\delta_j^2$, so that

$$E\{D^2_{ijT-A}\} \geqslant \tfrac{1}{4}(pq)^2 \pi_{iT-A}\, \pi_{jT-A}. \tag{A 9\,a}$$

The expectation of the denominator of (A 8$b$) exceeds the product of the expectations of $x_i(1-x_i)$ and $x_j(1-x_j)$, by the covariance between $x_i(1-x_i)$ and $x_j(1-x_j)$ due to shared geneaological history. This is likely to be very small in large populations, even for closely linked sites, since the expectations are taken over repeated introductions of new variants at the sites in question, which involve independent mutational events. For example, in the case of a random-mating population of size $N$ (and with no polymorphic locus), Pluzhnikov & Donnelly (1996, Appendix, final equation) show that the covariance in diversities between two sites is always very small compared with the product of the expected diversities, with a maximum of about 1/9 for absolutely linked sites. A similar result can be derived from equations (7) of Ohta & Kimura (1971).

Evaluating the denominator in (A 8$b$), and ignoring the terms due to shared genealogies, we thus have

$$\sigma^2_{ijT-A} \geqslant F_{iAT}\, F_{jAT}. \tag{A 9\,b}$$

This shows that the extent of differentiation between the allelic classes at the polymorphic locus with respect to linked neutral sites is directly related to the amount of linkage disequilibrium between neutral sites.

## References

Awadalla, P. (1996). Microsatellite variation and evolution in mixed mating species of *Mimulus* (Scrophulariaceae). MSc thesis, University of Toronto.

Awadalla, P. & Ritland, K. (1997). Microsatellite variation in *Mimulus* species of contrasting mating systems. *Molecular Biology and Evolution* (in press).

Baker, H. G. (1953). Race formation and reproductive method in flowering plants. *Society for Experimental Biology Symposia* **7**, 114–145.

Barton, N. H. (1979). Gene flow past a cline. *Heredity* **43**, 333–339.

Barton, N. H. & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity* **56**, 357–376.

Begun, D. J. & Aquadro, C. F. (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.

Begun, D. J. & Aquadro, C. F. (1995). Evolution at the tip and base of the X chromosome in an African population of *Drosophila melanogaster*. *Molecular Biology and Evolution* **12**, 382–390.

Bell, G. (1996). *Selection. The Mechanism of Evolution.* New York: Chapman and Hall.

Bengtsson, B. O. (1985). The flow of genes through a genetic barrier. In *Evolution. Essays in Honour of John Maynard Smith* (ed. P. J. Greenwood, P. H. Harvey & M. Slatkin), pp. 31–42. Cambridge: Cambridge University Press.

Bonnin, I., Huguet, T., Gherardi, M., Prosperi, J.-M. & Olivieri, I. (1996). High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *American Journal of Botany* **83**, 843–855.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796.

Brown, A. H. D. (1979). Enzyme polymorphism in plant populations. *Theoretical and Applied Genetics* **15**, 1–42.

Charlesworth, D. & Charlesworth, B. (1995). Quantitative genetics in plants: the effect of breeding system on genetic variability. *Evolution* **49**, 911–920.

Charlesworth, B., Charlesworth, D. & Morgan, M. T. (1990). Genetic loads and estimates of mutation rates in very inbred plant populations. *Nature* **347**, 380–382.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory.* New York: Harper and Row.

Ellstrand, N. C. & Levin, D. A. (1980). Recombination system and population structure in *Oenothera*. *Evolution* **32**, 245–263.

Gillespie, J. H. (1991). *The Causes of Molecular Evolution.* Oxford: Oxford University Press.

Gillespie, J. H. (1994). Alternatives to the neutral theory. In *Non-Neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 1–17. London: Chapman and Hall.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distance between loci of linked factors. *Journal of Genetics* **8**, 299–309.

Haldane, J. B. S. (1930). A mathematical theory of natural and artificial selection. VI. Isolation. *Proceedings of the Cambridge Philosophical Society* **26**, 220–230.

Hamrick, J. L. & Allard, R. W. (1972). Microgeographical variation in allozyme frequencies in *Avena barbata*. *Proceedings of the National Academy of Sciences of the USA* **69**, 2100–2104.

Hamrick, J. L. & Godt, M. J. (1990). Allozyme diversity in plant species. In *Plant Population Genetics, Breeding, and Genetic Resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler & B. S. Weir), pp. 43–63. Sunderland, Mass.: Sinauer.

Hamrick, J. L. & Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B* **351**, 1291–1298.

Hedrick, P. W. (1979). Hitch-hiking: an alternative to coadaptation for the barley and slender wild oat examples. *Heredity* **43**, 79–86.

Hedrick, P. W. (1980). Hitch-hiking: a comparison of linkage and partial selfing. *Genetics* **94**, 791–808.

Hedrick, P. W. (1990). Mating systems and evolutionary genetics. In *Population Biology: Ecological and Evol-*

*utionary Viewpoints* (ed. K. Wöhrmann & S. Jain), pp. 83–118. New York: Springer-Verlag.

Hedrick, P. W., Jain, S. K. & Holden, L. (1978). Multilocus systems in evolution. *Evolutionary Biology* **2**, 104–184.

Heywood, J. S. (1991). Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics* **22**, 335–355.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1–45.

Hudson, R. R. & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics* **120**, 831–840.

Hudson, R. R. & Kaplan, N. L. (1994). Gene trees with background selection. In *Non-Neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 140–153. London: Chapman and Hall.

Hudson, R. R. & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics* **141**, 1605–1617.

Hudson, R. R., Boos, D. D. & Kaplan, N. L. (1992). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**, 138–151.

Hudson, R. R., Bailey, K., Skerecky, D., Kwiatowski, J. & Ayala, F. J. (1994). Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* **136**, 1329–1340.

Ingvarsson, P. K. (1997). The effect of delayed population growth on the genetic differentiation of local populations subject to frequent extinctions and colonizations. *Evolution* **51**, 29–35.

Ingvarsson, P. K., Olsson, K. & Ericson, L. (1997). Extinction–recolonisation dynamics in the mycophagous beetle *Phalacrus substriatus*. *Evolution* **51**, 187–195.

Johnston, M. O. & Schoen, D. J. (1995). Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. *Science* **267**, 226–229.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitchhiking effect' revisited. *Genetics* **123**, 887–899.

Karl, S. A. & Avise, J. C. (1992). Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* **256**, 100–102.

Karlin, S. (1982). Classifications of selection–migration structures and conditions for a protected polymorphism. *Evolutionary Biology* **14**, 61–204.

Kimura, M. (1971). Theoretical foundations of population genetics at the molecular level. *Theoretical Population Biology* **2**, 174–208.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kimura, M. & Ohta, T. (1971). *Theoretical Topics in Population Genetics*, Princeton, NJ: Princeton University Press.

Kreitman, M. (1996). The neutral theory is dead. Long live the neutral theory. *BioEssays* **18**, 678–683.

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *American Naturalist* **87**, 331–333.

Levin, D. A. (1978). Genetic variation in annual *Phlox*: self-compatible vs self-incompatible species. *Evolution* **32**, 245–263.

Linhart, J. & Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics* **27**, 237–277.

Loos, B. P. (1993). Allozyme variation within and between populations in *Lolium* (Poaceae). *Plant Systematics and Evolution* **188**, 101–113.

Marshall, D. R. & Brown, A. H. D. (1975). The charge state model of protein polymorphisms in natural populations. *Journal of Molecular Evolution* **6**, 149–163.

Maruyama, T. (1970). Rate of decrease of genetic variability in a subdivided population. *Biometrika* **57**, 299–311.

Maruyama, T. (1971). An invariant property of a structured population. *Genetical Research* **18**, 81–84.

Maruyama, T. (1972). Some invariant properties of a geographically structured finite population: distribution of heterozygotes under irreversible mutation. *Genetical Research* **20**, 141–149.

Maruyama, T. & Kimura, M. (1980). Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Sciences of the USA* **77**, 6710–6714.

Maruyama, K. & Tachida, H. (1992). Genetic variability and geographic structure in partially selfing populations. *Japanese Journal of Genetics* **67**, 39–51.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genetical Research* **23**, 23–35.

Mitchell-Olds, T. & Bergelson, J. (1990). Statistical genetics of an annual plant *Impatiens capensis*. I. Genetic basis of quantitative variation. *Genetics* **124**, 407–415.

Nagylaki, T. (1982). Geographical invariance in population genetics. *Journal of Theoretical Biology* **99**, 159–172.

Nagylaki, T. (1997). The diffusion model for migration and selection in a plant population. *Journal of Mathematical Biology* **35**, 391–431.

Narain, P. (1966). Effect of linkage on homozygosity of a population under mixed selfing and random mating. *Genetics* **54**, 303–314.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* **70**, 3321–3323.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M. & Li, W.-H. (1973). Linkage disequilibrium in subdivided populations. *Genetics* **75**, 213–219.

Nevo, E., Krugman, T. & Beiles, A. (1994). Edaphic natural selection of allozyme polymorphisms in *Aegilops peregrina* at a Galilee microsite in Israel. *Heredity* **72**, 109–112.

Nordborg, M. (1997). Structured coalescent processes on different time scales. *Genetics*, in press.

Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996*a*). The effect of recombination on background selection. *Genetical Research* **67**, 159–174.

Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996*b*). Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proceedings of the Royal Society of London, Series B* **163**, 1033–1039.

Ohta, T. & Kimura, M. (1969). Linkage disequilibrium due to random genetic drift. *Genetical Research* **13**, 47–55.

Ohta, T. & Kimura, M. (1971). Linkage disequilibrium between two segregating nucleotide sites under steady flux of mutations in a finite population. *Genetics* **68**, 571–580.

Petry, D. (1983). The effect on neutral gene flow of selection at a linked locus. *Theoretical Population Biology* **23**, 300–313.

Pluzhnikov, A. & Donnelly, P. (1996). Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* **144**, 1321–1328.

Pogson, G. H. & Zouros, E. (1994). Allozyme and RFLP heterozygosities as correlates of growth rate in the scallop *Placopecten magellanicus*: a test of the associative overdominance hypothesis. *Genetics* **137**, 221–231.

Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117**, 353–360.

Schoen, D. J. (1982). Genetic variation and the breeding system of *Gilia achilleifolia*. *Evolution* **36**, 361–370.

Schoen, D. J. & Brown, A. H. D. (1991). Intraspecific

variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences of the USA* **88**, 4494–4497.

Schoen, D. J., Morgan, M. T. & Bataillon, T. (1996). How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Philosophical Transactions of the Royal Society of London, Series B* **351**, 1281–1290.

Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

Slatkin, M. (1977). Gene flow and genetic drift in a population subject to frequent local extinctions. *Theoretical Population Biology* **12**, 253–262.

Slatkin, M. (1987). The average number of sites separating DNA sequences drawn from a subdivided population. *Theoretical Population Biology* **32**, 42–49.

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. *Genetical Research* **58**, 167–175.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279.

Stebbins, G. L. (1957). Self fertilization and population variation in the higher plants. *American Naturalist* **91**, 337–354.

Stephan, W. (1994). Effects of genetic recombination and population subdivision on nucleotide sequence variation in *Drosophila ananassae*. In *Non-neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 57–66. London: Chapman and Hall.

Stephan W., Wiehe, T. H. E. & Lenz, M. W. (1992). The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theoretical Population Biology* **41**, 237–254.

Strobeck, C. (1983). Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* **103**, 545–555.

Strobeck, C. (1987). Average number of nucleotide differences in a sample from a single population: a test for population subdivision. *Genetics* **117**, 149–153.

Tajima, F. (1990). Relationship between migration and DNA polymorphism in a local population. *Genetics* **126**, 231–234.

Tajima, F. (1992). Measurement of DNA polymorphism. In *Mechanisms of Molecular Evolution* (ed. N. Takahata & A. G. Clark), pp. 37–59. Sunderland, Mass.: Sinauer.

Todokoro, S., Terauchi, R. & Kawano, S. (1995). Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. *Japanese Journal of Genetics* **70**, 543–554.

van Dijk, H. V., Wolff, K. & Fries, A. D. (1988). Genetic variability in *Plantago* species in relation to their ecology. 3. Genetic structure of populations of *Plantago major*, *P. lanceolata* and *P. coronopus*. *Theoretical and Applied Genetics* **75**, 518–528.

Wade, M. J. & McCauley, D. E. (1988). Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**, 995–1005.

Weir, B. S. & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Whitlock, M. C. & Barton, N. H. (1997). The effective size of a subdivided population. *Genetics* **146**, 427–441.

Whitlock, M. & McCauley, D. E. (1990). Some population genetic consequences of colony formation and extinction. *Evolution* **44**, 1717–1724.

Wolff, K. (1991). Analysis of allozyme variability in three *Plantago* species and a comparison to morphological variability. *Theoretical and Applied Genetics* **81**, 119–126.

Wolff, K., Rogstad, S. H. & Schaal, B. A. (1994). Population and species variation of minisatellite DNA in *Plantago*. *Theoretical and Applied Genetics* **87**, 733–740.

Wright, S. (1969). *Evolution and the Genetics of Populations*, vol. 2, *The Theory of Gene Frequencies*. Chicago: University of Chicago Press.