Effects of linkage on response to directional selection from new mutations

By P. D. Keightley and W. G. Hill

Institute of Animal Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN

(Received 22 June 1983 and in revised form 27 July 1983)

SUMMARY

The influence of linkage on the rates of response to continued directional selection of quantitative traits deriving from variation contributed by new mutations in finite populations is investigated. Mutant genes are assumed to have additive effects which are randomly sampled from a symmetric distribution, and to be randomly located on the chromosome. Results were obtained by Monte Carlo simulation.

The expected steady rate of response, when variability from new mutations is balanced by that lost by drift and selection, is reduced the tighter the linkage, but the reduction is small unless there are few, short chromosomes. For a given rate of new mutational variance per haploid chromosome set per generation, greater effects of linkage are obtained in large than in small populations, because more mutants segregate. The response and influence of linkage are essentially the same whether the new variance is due to many genes of small effect or few of large effect.

The variability of response between replicates and generations was investigated, and the contribution to this of new mutants or recombination of existing mutants compared. Usually most genetic variability was due to the occurrence of a new favourable mutant of large effect.

1. INTRODUCTION

Recent analyses of the results of selection experiments (Frankham, 1980) and theoretical studies (Hill, 1982a, b) have suggested that the mutations which have occurred after selection commenced can make a significant contribution to response in long-term experiments and breeding programmes. An important conclusion was that the asymptotic rate of response would be approximately proportional to population size, which has implications in design of practical improvement programmes. In the previous theoretical studies all loci were assumed to be unlinked. It is known, however, that linkage can affect the fixation probability of genes (Latter, 1966; Hill & Robertson, 1966; Robertson, 1970) and the results of several selection experiments have been interpreted in terms of release of variation by crossing-over (e.g. Mather, 1941; Thoday, Gibson & Spickett, 1964). It seemed necessary to investigate theoretically the effects of linkage on the contribution to selection response, and the present study is an attempt to do so, mainly using Monte Carlo simulation on a computer.
The range of parameters to be studied is limited; for example only additive gene action is considered. But the information on which the models can be based is itself very limited, most of it coming from studies of bristle number in Drosophila. This analysis is of the effects of directional selection for a quantitative trait in a finite population with continuous input of new mutations, in contrast to those of Lande (1976) and others who have discussed the effects of stabilizing selection, usually in infinite populations. To reduce computing time, the simulation was carried out using a model of differences in fertility.

2. MODEL

(i) Artificial selection by truncation

The genetic model which was used allowed genes to take any position on the chromosome and to have any value of effect on the trait under selection, and there were no restrictions imposed on the number of genes segregating in the population at any time.

In terms of the metric trait on which artificial selection was practised, the model was based on the following parameters.

Number and effects of mutants. The expected number of mutants per haploid chromosome set per generation is \( \lambda \). Mutants have additive effects both within and between loci on the metric trait under selection, and \( a \) is the difference between homozygotes in effect on the trait. Values of \( a \) are sampled from a distribution with variance \( \sigma_a^2 \).

Mutational variance. The expected increment in variance in the metric trait from new mutations each generation is \( \sigma_M^2 \), which is given by

\[
\sigma_M^2 = \frac{1}{2} \lambda \sigma_a^2
\]

for any population size (Hill, 1982a). Assuming the phenotypic variance is \( \sigma_P^2 \), the increment in heritability each generation is \( \sigma_M^2 / \sigma_P^2 \) for which estimates are around 0.001 for bristle number in Drosophila melanogaster (Hill, 1982b).

Selection. Truncation selection is assumed to be practised with intensity \( i \), so the selective value of a gene of effect \( a \) is given by

\[
s = i a / \sigma_P
\]

where \( s \) is the difference between homozygotes.

In the absence of linkage the fixation probability is, from the diffusion approximation (Kimura, 1957), given by

\[
u = 0 \quad \text{for } s < -1/N
\]

\[
u = 1/(2N) + s/2 \quad \text{for } |s| \leq 1/N
\]

\[
u = s \quad \text{for } s > 1/N
\]

approximately (Hill, 1982a). With an effective population size of \( N \), and constant symmetric distribution of mutant effects, the expected response from mutants occurring in one generation, which equals the asymptotic rate of response, is given by \( 2Ni\sigma_M^2 / \sigma_P \) approximately (Hill, 1982a).
Simulated model using fertility differences

If truncation selection is simulated directly, many more individuals have to be sampled and ranked than become parents. Therefore the computer simulation was undertaken by sampling selective values \( s \) rather than effects of mutants, and assigning fertility differences to the parents to reflect the probabilities that their offspring would be selected. The details of the simulated model are as follows:

**Population size.** \( N \) monoecious diploid individuals were bred by random mating but with selfing excluded.

**Distribution of mutant selective values.** The selective values of individual mutants were sampled independently from a 'double gamma' distribution (Hill, 1982b) with variance \( \sigma_s^2 \) given by

\[
f(s) = \frac{1}{2} \left[ \frac{(3)^{\frac{1}{2}}}{\sigma_s \pi |s|} \right] \exp \left\{ -\frac{|s| (3)^{\frac{1}{2}}}{\sigma_s} \right\}, \quad -\infty < s < \infty,
\]

which is illustrated in Fig. 1. It is a symmetric distribution, essentially a chi-square with 1 d.f. reflected about zero, and is intended to represent a model in which mutants of very small effect are much more common than those of large effect.

**Expected numbers of mutants.** Using a simulated model of fertility differences to represent artificial selection, the selective values \( s \) depend on the selection intensity \( i \) for specified gene effect (Equation 2). The relationship between the expected number of mutants \( \lambda \) and \( \sigma_s^2 \) therefore depends on the value of \( i \) which can be chosen to represent a conventional truncation selection scheme. Thus, from (2),

\[
\sigma_s^2 = i^2 \sigma_M^2 / \sigma_P^2
\]

and hence, from (1),

\[
\lambda \sigma_s^2 = 2i^2 \sigma_M^2 / \sigma_P^2.
\]
A convenient value is \( i = \sqrt{2} \sim 1.4 \), which corresponds to 20% selection. Taking

\[
\frac{\sigma_m^2}{\sigma_p^2} = 0.001 \quad \text{for Drosophila bristles, this gives}
\]

\[\lambda \sigma_s^2 = 0.004 \quad \text{(5)}\]

as a typical value. Thus combinations of \( \lambda \) and \( \sigma_s \) could be chosen to satisfy (5); for example with \( \sigma_s = 0.1 \), \( \lambda = 0.4 \) mutations are expected each generation.

**Method of selection.** The fitness of an individual was computed as

\[ g = 1 + \Sigma s_j - \mu \]

where \( \Sigma s_j \) is the sum of selective values over loci and chromosomes, assuming additivity, and \( \mu \) is the mean of the \( \Sigma s_j \) over the \( N \) individuals. To simulate progeny, parents were randomly sampled with probability proportional to \( g \). With this model, the expected response in fitness per generation equals the additive genetic variance in fitness \( (g) \), as predicted by Fisher’s (1930) ‘Fundamental Theorem of Natural Selection’, and the response in the metric trait is, from (2), equal to \( \sigma_p / i \) times the response in fitness.

**Linkage relationships.** One chromosome was present per haploid set, with map length \( (L) \) equalling the expected number of recombinants per generation. The number of crossovers contributing to each daughter chromosome was sampled from a Poisson distribution with parameter \( L \), and the position of each crossover sampled from a uniform distribution over the whole chromosome length.

**Number and location of mutants.** The number of new mutants on each haploid chromosome was sampled from a Poisson distribution with parameter \( \lambda \) (computed as in Equation (5)) and the position of each mutant on the chromosome sampled from a uniform distribution, i.e. from an infinite number of equally likely sites (loci). Each mutant was therefore identified by both an effect and a position. To speed up computation, every ten or so generations a ‘garbage collection’ routine was used to remove all fixed mutants, their value being added as a constant to the population mean.

(iii) **Other simulation programs**

Two other programs were used to deal with special limiting cases, and also to serve as checks on the main program for finite chromosome length.

**No recombination.** In this case, where \( L \to 0 \), no mutant site or recombination has to be considered, so mutant effects merely change the value of the chromosome and the chromosome is described solely by its value. In some runs, a model of several \( (k) \) chromosomes each with no recombination was simulated, with new mutants equally likely to fall on any chromosome.

**No linkage.** In this case, which can be regarded as either \( L \to \infty \) on one chromosome, or as infinitely many chromosomes, each new mutant was assumed to fall on a different chromosome or, equivalently, to be unlinked to all other mutants. Results from this model correspond to those for independent genes discussed previously (Hill, 1982a, b), except that in the earlier studies only single segregating genes were considered.

**Computation of results**

Most results are presented for the ‘asymptotic response rate’, the rate of change in the metric trait achieved when the increments in variance from mutation
balance the loss from selection and drift. In the simulations, a single population was started in complete homozygosity and run for very many — at least 1500 and typically 3000 — generations. The asymptotic rate was obtained by ignoring early, typically 50, generations. Because the responses in successive generations are auto-correlated, an exact standard error of the asymptotic rate is difficult to compute. In practice, response rates were computed in 10 generation averages, and the standard error of these was used as an estimate. Details of the standard errors are not given in the following results, but usually they were less than 5% of the rate of response.

In investigating the relative role of linkage and new mutants on patterns of response, the variability within the population was expressed in two ways: the gametic variance, $V_G$, the variance among the fitnesses ($\Sigma s_j$) of the individuals; and
the genic variance, $V_g$, computed as the variance which would be present if all genes were in linkage equilibrium, i.e. as the sum over loci, $l$, of $\frac{1}{2} p_l (1-p_l) s^2_l$, where $p_l$ is the frequency, and $s_l$ the effect of the gene at locus $l$. With mean negative disequilibrium among genes of positive value, for example, $V_G$ will be less than $V_g$. 

$$N = 20$$

![Graph](image)

**Fig. 3.** As Fig. 2, but $N = 20$.

3. RESULTS

(i) *Asymptotic rates of response*

The effects of changes in chromosome length on the asymptotic rate of response are shown in Figs 2–4. Each figure relates to one population size ($N = 10$, 20 or 40) and in each figure results are given for a range of values of standard deviation of selective value ($\sigma_s = 0.05$, 0.1, 0.2 and 0.4) and of mutational variance $\sigma_M^2/\sigma_s^2 = \lambda \sigma_s^2/4 = 0.001/3$, 0.001 and 0.003 (see Equation 4). The intermediate value of mutational variance was chosen to reflect the total variance for Drosophila bristles (see above), the lower value that per Drosophila chromosome, and the higher value for illustration. In Fig. 5 some of the results are replotted as a function of population size, with additional values for $N = 80$. Results in these Figs 2–5 are
not complete for all parameter combinations, because computing time increases with \( N \), reduces with \( \sigma_s \) (more genes segregating) and increasing with number of crossovers per chromosome.

It was noted earlier that with no linkage, and a symmetric distribution of mutant effects, the expected rate of response is \( 2N\sigma_M^2/\sigma_P \). The actual rates (Figs. 2–5) are somewhat less, particularly for the higher values of \( \sigma_s \). This could be anticipated from fixation probabilities \( u \) computed using the same simulation program but with fixed values of \( s \) and low mutation rates (Table 1), where it can be seen that \( u < s = \alpha k/\sigma_P \), particularly for large values of \( s \). This is due partly to the non-linear relation between \( u \) and \( s \) in the diffusion approximation and partly to inadequacies in the diffusion approximation. There is also a less than linear increase in response rates with increase in population size as \( N \) becomes very large even for unlinked loci (Figs. 4 and 5), presumably because the increased variance in fitness maintained in the population leads to a reduced effective population size (Nei & Murata, 1966).

For given population size, tight linkage has more effect the greater the
mutational variance and, for given mutational variance, linkage effects are greater with larger populations (Fig. 5). The effects of linkage and of population size depend rather little on the magnitude of $\sigma_r$, despite the 8-fold range shown in $\sigma_r$ or 64-fold range in $\sigma^2_r$ and thus number of mutants ($\lambda$). Effects of linkage are small for chromosomes of more than one crossover in length, and most of the reduction in $\sigma_r = 0.2$

![Graph showing response against population size](image)

Fig. 5. Asymptotic rates of response plotted for different values of chromosome length and mutational variance against population size, with $\sigma_r = 0.2$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>10</th>
<th>20</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.075 ± 0.008</td>
<td>0.066 ± 0.007</td>
<td>0.045 ± 0.006</td>
</tr>
<tr>
<td>0.1</td>
<td>0.094 ± 0.008</td>
<td>0.086 ± 0.008</td>
<td>0.087 ± 0.008</td>
</tr>
<tr>
<td>0.2</td>
<td>0.140 ± 0.011</td>
<td>0.148 ± 0.011</td>
<td>0.159 ± 0.011</td>
</tr>
<tr>
<td>0.4</td>
<td>0.264 ± 0.014</td>
<td>0.269 ± 0.015</td>
<td>0.272 ± 0.015</td>
</tr>
<tr>
<td>0.8</td>
<td>0.370 ± 0.017</td>
<td>0.383 ± 0.017</td>
<td>0.380 ± 0.017</td>
</tr>
</tbody>
</table>

Table 1. Fixation probabilities of genes of specified selective value in the simulation procedure when appearing at low mutation rates ($2N\lambda = 0.2$)
Fig. 6. Patterns of response and variance for one simulated population with $N = 20$, $\sigma_M^2 / \sigma_s^2 = 3.3 \times 10^{-4}$, one crossover per chromosome and $\sigma_s = 0.2$. The upper graph shows the population mean, the lower graph the variance components within the population: — genic, - - - gametic. The bursts of response in most cases coincide with a large increase in both genic and gametic variance, indicative of effects of a single gene.
Fig. 7. As Fig. 6, but $\sigma_z = 0.02$. The small increases in response rate are sometimes associated with an increase in gametic but not genic variance, indicative of breakdown of negative linkage disequilibrium.
response occurs with very short chromosomes. The larger the value of population size, the shorter the chromosome length at which linkage effects are important (cf. Figs. 2 and 4).

(ii) Individual response patterns

Responses for 100 generations of selection using parameters appropriate for experiments with Drosophila bristle number are given in Figs 6 and 7, using $\sigma^2_M / \sigma^2_p = 3.3 \times 10^{-4}$ (one chromosome only) and one crossover per chromosome (allowing for no recombination in males). In Fig. 6 a high standard deviation of gene effect is assumed ($\sigma_s = 0.2$) and in Fig. 7 a low value ($\sigma_s = 0.02$). Also shown in the figures are the gametic and genic variances. Most bursts of response for $\sigma_s = 0.2$ coincide with the occurrence and fixation of genes of large effect, the genic and genotypic variances changing together. For the lower $\sigma_s$ value the response pattern is much more regular and in some cases an increase in gametic variance occurs without an increase in the genic variance, indicating a release of 'hidden' variation due to crossing-over.

4. DISCUSSION

(i) Simulation model

In previous simulations of directional selection with linked loci many parameters in the model had to be specified, for example the number of loci, the effects and initial frequency of each gene and, in most cases, the recombination fraction between adjacent loci. In our model all that have to be specified, in addition to the basic assumption of additive gene action, are: (i) the distribution and variance of individual mutant effects or selective values, for which we have only scant information from experimental data, though there is some indication for bristle number in Drosophila that the shape of the distribution which we assumed was a reasonable one (Shrimpton, 1982); (ii) the mutational variance, for which data are available on Drosophila bristles; and (iii) the known parameters of map length, number of chromosomes, population size and selection intensity. Thus the model simulated is, we believe, of some interest per se. If studies of populations initially segregating are to be made, then these can be done by allowing mutation to occur either without selection or with, for example, stabilizing selection for sufficient generations so that the variance reaches steady state, and then commencing directional selection.

(ii) Distribution of mutant effects

One of the most surprising results to us was that the influence of linkage on the asymptotic rate of response depended very little on the variance of mutant effects, $\sigma^2_s$, for a specified mutational variance (Figs 2–4). Thus there is a 64-fold range in number of mutants appearing per generation, and we had anticipated that linkage effects would have been greater with more mutants. It appears, however, that the greater influence of one mutant on another when they both have large effects (i.e. higher $\sigma_s$) (Hill & Robertson, 1966) compensates for the greater probability of having another gene segregating nearby (i.e. higher $\lambda$ and lower $\sigma_s$). Some compensation was to be anticipated, but no mathematical argument has been
developed to show why it should be so nearly exact. It can probably be expressed in terms of inherited variability in fitness, tighter linkage causing greater correlations over generations. Theoretical analyses are being undertaken on this point.

In view of the lack of dependence of the rate of $s$ it is not surprising that alterations in the shape of the distribution of mutant effects or selective value from the double-gamma to the much less leptokurtic normal distribution with the same

<table>
<thead>
<tr>
<th>$\sigma_s$</th>
<th>Model</th>
<th>0</th>
<th>0.1</th>
<th>1</th>
<th>5</th>
<th>$\rightarrow\infty$</th>
<th>Approx. s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>$AN$</td>
<td>3.78</td>
<td>4.43</td>
<td>5.92</td>
<td>7.36</td>
<td>7.51</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>$AG$</td>
<td>3.91</td>
<td>4.33</td>
<td>6.24</td>
<td>6.88</td>
<td>6.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$MG$</td>
<td>4.09</td>
<td>4.47</td>
<td>5.97</td>
<td>6.34</td>
<td>7.12</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>$AN$</td>
<td>3.84</td>
<td>4.61</td>
<td>5.89</td>
<td>6.73</td>
<td>6.98</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$AG$</td>
<td>3.87</td>
<td>4.39</td>
<td>5.30</td>
<td>6.38</td>
<td>6.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$MG$</td>
<td>4.27</td>
<td>4.91</td>
<td>5.47</td>
<td>6.26</td>
<td>6.55</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>$AN$</td>
<td>4.07</td>
<td>4.72</td>
<td>5.24</td>
<td>5.25</td>
<td>5.95</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>$AG$</td>
<td>3.30</td>
<td>4.25</td>
<td>4.54</td>
<td>5.00</td>
<td>5.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$MG$</td>
<td>4.03</td>
<td>4.53</td>
<td>5.29</td>
<td>4.99</td>
<td>4.75</td>
<td></td>
</tr>
</tbody>
</table>

values of $\sigma_s$ made little difference to the asymptotic response. Some comparisons of simulations made with the double-gamma distribution (cf. Fig. 2) and normal distribution are given in Table 2. Noticeable differences between them occur only for $\sigma_s = 0.4$, when very large selective values can be sampled in the double-gamma case and the linear relation between fixation probability and $s$ breaks down (Table 1).

The analysis has been entirely for the case of a symmetric distribution of mutant effects. Some limited analyses have shown that the influence of linkage increases as a higher proportion of mutants take negative values: essentially the population has more difficulty in ridding itself of deleterious mutants, as in a Muller’s ratchet effect (Muller, 1964; Felsenstein, 1974).

Population size. In previous analyses of limits it was found that linkage had more influence, the smaller the population size (Hill & Robertson, 1966; Robertson, 1970). Mathematical arguments can be invoked to explain this, but the result can be viewed as due to the fact that in large populations recombinants have more time to appear before genes are lost by chance. The results here are quite different, however: linkage has more effect in larger populations. Why is this so?

The figures show clearly that, for given population size, the higher the mutational variance, $\sigma_M^2$, the greater the effects of linkage. This could be expected because more genes are segregating and the mean distance between adjacent genes is less. Previous analyses have shown that linkage always reduces the fixation probability of additive genes initially in linkage equilibrium (Hill & Robertson,
1966, Robertson, 1970) and in the present model mutants in single copy are necessarily in disequilibrium, but this is equally likely to be positive or negative. For a constant value of mutational variance, $\sigma^2_M$, per haploid set, the total number of mutants produced per generation increases in proportion to the population size and, furthermore, they can remain segregating longer. The presence of many more segregating genes in larger populations clearly increases the effects of the linkage.

It is, perhaps, pertinent to remark here that the effects of linkage do not act solely through the reduction in variance contributed by artificial selection or fertility selection with additive genes (Felsenstein, 1965; Bulmer, 1971). Simulation using the same model but with multiplicative effects over loci such that fitness is proportional to $\exp(s_j)$, which does not lead immediately to disequilibrium (Felsenstein, 1965), gave very similar effects of linkage (see examples in Table 2), confirming our previous studies (Hill & Robertson, 1966; Robertson, 1970). It seems that the critical point about recombination in finite populations is that it acts by increasing the probability of the favourable coupling gamete being formed from repulsion types before either positive gene is lost.

**Number of chromosomes: practical consequences.** The results shown in the figures refer to variation on single chromosomes, but those for many chromosomes are readily inferred and illustrated in these figures. Thus, for Drosophila bristles, $\sigma^2_M/\sigma^2_p = 0.001$ in all, or $0.00033$ per major chromosome approximately. The expected responses are therefore about three times those shown for the lowest value of $\sigma^2_M$ and it is clear that the linkage effects are very much less than when all genes are on the same chromosome. In mammals, therefore, with 20 or so chromosomes with map lengths around 1 Morgan (see Avery & Hill, 1978), very substantial values of $\sigma^2_M$ or large populations would be required before the effects of linkage on rates of response would be appreciable. Calculations of rates based on unlinked loci would therefore seem quite adequate, at least where it can be assumed that the distribution of mutant effects is not too asymmetric.

**Short-term responses.** It should be emphasized that though lines were simulated for up to 3000 generations this was merely a computing convenience and not what we expect in the real world. The contribution of mutation to response in early generations was investigated in a previous paper (Hill, 1982b) and several points emerged: if there are expected to be many mutants of small effect, the contribution increases with the square of generation number in a fairly steady way, but is likely to be trivial for at least twenty generations; if $\sigma_M$ is large, however, the variation between replicates and generations is high and sudden bursts of response might come at any time. The simulations reported here with linkage included in the model generally indicate that such bursts of response are much more likely to be due to the occurrence of a single mutant of large effect than to a recombinational event. For the latter, it is necessary for two genes of roughly equal and quite large effect to be segregating in repulsion. What the model and simulation suggest is that this is likely to be a very rare event indeed, and long-lasting negative disequilibrium is more often associated with many genes of small effect.

**Generalizations.** There are many simplifying assumptions in the present model, some likely to be more important than others. The mutational events are taken to be independent over time; perhaps there are waves of mutations associated with
transpositions following, for example, infection with P-elements associated with hybrid dysgenesis. Such a situation would imply that the quantity $\sigma_N^2$ has no fixed value, that estimates from one population have no relevance to another, and that bursts of response from mutation are possible. Recombination or, equivalently, density of mutations has been taken to be uniform over the chromosome. Our results would indicate that substantial non-uniformity would be necessary before the linkage effects became much more important, especially for organisms with many chromosomes. All genes have been assumed to be additive: since the fate of mutants is largely determined by their effect in the heterozygote, the linkage effects would seem likely to apply qualitatively to dominant or partially dominant mutants; and because favourable recessives are rarely fixed, we are perhaps entitled to ignore their contribution. The joint effects of mutants on the metric trait under selection and on fitness, whether the effects are pleiotropic or through linkage, require a more substantial analysis.

We are grateful to Joe Felsenstein for helpful comments and discussion and Jon Rasbash for computational assistance.

REFERENCES


