Genet. Res. Camb. (1996), **67**, *pp.* 77–87 *With 6 text-figures*

Interactions of selection, linkage and drift in the dynamics of polygenic characters

FRÉDÉRIC HOSPITAL[1]* AND CLAUDE CHEVALET[2]
[1] *Station de Génétique Végétale, INRA/UPS/INA-PG, Ferme du Moulon, 91190 Gif sur Yvette, France*
[2] *Laboratoire de Génétique Cellulaire, INRA, BP27, 31326 Castanet Tolosan Cedex France*

(*Received 11 November 1994 and in revised form 15 July 1995*)

Summary

We study the dynamics under directional truncation selection of the genetic variability of a quantitative character controlled by a finite number of possibly linked loci with additive effects. After the first generation of selection, the build-up of linkage disequilibria (Bulmer effect) is analytically demonstrated from a genetical point of view in an infinite population. In the following generations, the dynamics of the system in a finite population are predicted using analytic recurrences under a multi-normal approximation, and computer simulations. The effects of recombination on the dynamics of linkage disequilibria induced by selection and drift, and the consequences for the additive genetic variance are then analysed and discussed from the simulation results. Compared to the rapid exploitation of genetic variability promoted by high recombination rates, low recombination rates promote an early storage of genetic variability in repulsion associations of alleles and a possible late release of genetic variance in the population, so that the variability of the character may be maintained over a longer period of time. In some cases, favourable recombination events in tightly linked systems induce an increase of the additive variance of the character, which may explain some results observed in long-term selection experiments. Our results emphasize that the joint effects of selection, linkage and drift must not be neglected in theoretical quantitative genetics, and require further investigation.

1. Introduction

From a formal point of view, the genetics of quantitative characters, i.e. polygenic characters, should use the theoretical tools provided by population genetics for the study of mono- or oligogenic characters. This would give reliable and precise predictions for the behaviour of allele frequencies at all the loci involved. Unfortunately, even in the simplest case of a character determined by genes of purely additive effects, the complete analytic description of the system is not possible when several loci are taken into account. This was clearly shown for example by Turelli & Barton (1990). This impossibility stems from the explosive increase in the number of parameters that must be taken into account to describe all moments of any orders between gene effects at several loci. Hence, approximations are often made, to reduce the number of parameters, assuming Gaussian distributions of breeding values (Pearson, 1896; Bulmer, 1980), or of every genetic contribution (Lande, 1976). Gaussian distributions of breeding values are obtained by assuming that the genes involved are unlinked and contribute small effects to the phenotype (infinitesimal model, Fisher, 1918). Even though selection deviates the distribution away from normality, it seems that in the case of unlinked genes, deviations from normality might be neglected even under directional truncation selection (Turelli & Barton, 1994) and the Gaussian approximation may hold.

Although any linkage values may *a priori* be expected between the genes controlling a given character, most of the theory in quantitative genetics has been obtained for unlinked loci, assuming that the effects of linkage may be neglected. This is not always true and we have shown in a previous paper that taking both linkage and drift into account may lead to important discrepancies with some classical results obtained under the assumptions of the infinitesimal model (Hospital & Chevalet, 1993). Here, we aim to study the effects causing these discrepancies, by

* Corresponding author: Frédéric Hospital, Station de Génétique Végétale, INRA/UPS/INA-PG, Ferme du Moulon, 91190 Gif sur Yvette, France. Phone: (33) (1) 69 33 23 36; Fax: (33) (1) 69 33 23 40; E-mail: fred@moulon.inra.fr.

investigating the joint effects of selection, linkage and drift on the variability of quantitative characters. Under selection for a polygenic trait, linkage disequilibria are generated between pairs of quantitative trait loci. In quantitative genetics, this effect was not clearly understood before Bulmer (1971) made it popular, even if it can be argued that Pearson (1903) discovered the effect at the beginning of the century, or that Lush (1945) wrote it down explicitly. In population genetics, the build-up of linked repulsion complexes of genes as a result of optimizing selection was first predicted by Fisher (1930) and Mather (1941), then further investigated by Lewontin (1964*a*, *b*, 1974) at equilibrium in an infinite population. In any case the main body of the literature on quantitative genetics did not take into account the role of linkage disequilibrium in the understanding of selection. Our approach is aimed at going from quantitative genetics to population genetics by studying the transient dynamics of additive genetic variance and linkage disequilibria under directional truncation selection in a finite population. The question is addressed using approximate analytic models and stochastic simulations.

2. Models and methods

(i) *Analytic models*

(*a*) *General assumptions*

We consider a single character controlled by a finite number of possibly linked loci evenly spread on a chromosome segment (cluster). The genotypic value (or breeding value) is the sum of $2n$ random variables $g_i^{(m)}$ and $g_i^{(f)}$ representing the contributions of genes carried by the chromosomes of paternal and maternal origins at locus i ($i = 1, \ldots, n$). The phenotype P is the sum of the genotypic value and a normally distributed environmental effect, E.

$$P = \sum_i (g_i^{(m)} + g_i^{(f)}) + E.$$

Matings are assumed to be at random at each generation, so that the variance covariance matrix of gene effects in a zygote takes the form

$$\begin{pmatrix} G & 0 \\ 0 & G \end{pmatrix}, \tag{1}$$

where G is the variance covariance matrix between gene effects of a gamete drawn from the reproducing individuals in the previous generation. The elements of matrix G are $G_{ij} = Cov(g_i^{(m)}, g_j^{(m)}) = Cov(g_i^{(f)}, g_j^{(f)})$. The first two moments of trait P are:

$$\bar{P} = 2\sum_i \bar{g}_i \tag{2}$$

$$V_P = V_A + V_E \tag{3}$$

$$= 2\sum_i \sum_j G_{ij} + V_E, \tag{4}$$

where $\bar{P}$ is the phenotypic mean, $\bar{g}_i$ is the mean of gametic values $g_i^{(m)}$ and $g_i^{(f)}$, V_P is the phenotypic variance, V_A is the additive genetic variance, and V_E the variance of environmental effects.

The initial population is at Hardy–Weinberg equilibrium at each locus, and all loci are in linkage equilibrium. Thus, contributions $g_i^{(m)}$ and $g_i^{(f)}$ of locus i to the genotypic value of an individual are independent random variables with variance $\frac{1}{2}\sigma_i^2$. The initial matrix $G^{(0)}$ is diagonal with $G_{ii}^{(0)} = \frac{1}{2}\sigma_i^2$ and the initial genetic variance is $V_A^{(0)} = \Sigma_i \sigma_i^2$. If the number n of loci is large, and the contribution σ_i^2 of any locus i to the total phenotypic variance $V_P^{(0)} = \sigma^2$ is small ($\sigma_i^2/\sigma^2 \sim 1/n$), the distribution of genotypic values is approximately Gaussian.

Truncation selection is performed, with intensity $\imath$, corresponding to truncation at the point

$$P_0 = \xi\sqrt{V_P},$$

assuming that the mean of the population is 0, and that ξ is the truncation point in the reduced scale. Assuming that P is normally distributed, the relevant parameters used in analytical derivations are

(i) The selection intensity $\imath$, relating the change in mean phenotypic value to the standard phenotypic deviation:

$$\bar{P}_{(s)} - \bar{P} = \imath\sqrt{V_P}.$$

(ii) The relative change $\imath(\imath - \xi)$ in the variance, such that:

$$V_{P(s)} = (1 - \imath(\imath - \xi))\, V_P,$$

where subscript (s) refers to values among selected individuals.

(*b*) *One generation of selection*

The effect of selection on the genetic variance available for selection has been established by Bulmer (1971) in a statistical setting (at the level of the breeding value). We give in the Appendix a genetical derivation of this effect (at the level of genes) that we obtained under simple hypotheses: infinite population size, initial linkage equilibrium and Hardy–Weinberg frequencies of alleles at several loci, additive effects of alleles on the breeding value. These hypotheses are less restrictive than Bulmer's (see Discussion).

(*c*) *Several generations of selection*

Going on to the next generation cannot follow the simple rationale of the Appendix. The special joint distribution corresponding to linkage equilibrium (i.e. independence between allelic states at different loci) can no longer be accepted, so that fitnesses of 1- or 2-locus genotypes cannot be calculated. More precisely, the step from equation (A 1) to equation (A 2) is no longer allowed. Describing the dynamics of polygenic systems by means of the moments of distributions has been considered by Turelli & Barton (1990). A first

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

level of approximation uses only the first and second moments of distributions. This may be obtained by assuming that distributions are Gaussian, as done by Lande (1976) and Chevalet (1988, 1994). Although Gaussian distributions are not conserved under the processes of recombination and random mating (Felsenstein, 1977), they allow the genetic modifications induced by selection on phenotype and on genotype to be calculated from standard regression equations.

The predicted changes in the means, variances and covariances of gene effects after one cycle of selection are, from one generation (t) to the next ($t+1$) (Lande, 1976; Chevalet, 1988):

$$E(\bar{g}_i^{(t+1)}) = E(\bar{g}_i^{(t)}) + \frac{\imath}{\sqrt{V_P^{(t)}}} G_{i\cdot}^{(t)} \tag{5}$$

$$G_{ij}^{(t+1)} = \left(1 - \frac{1}{2N} - r_{ij}\right) G_{ij}^{(t)} - \left(1 - \frac{1}{N}\right) \frac{\imath(\imath - \xi)\, G_{i\cdot}^{(t)}\, G_{\cdot j}^{(t)}}{V_P^{(t)}}, \tag{6}$$

where r_{ij} is the recombination fraction between loci i and j, and $G_{i\cdot}$ and $G_{\cdot j}$ are such that:

$$G_{i\cdot}^{(t)} = G_{\cdot i}^{(t)} = \sum_{j=1}^{j=n} G_{ij}^{(t)}$$

(ii) *Simulation model*

The computer program simulates selection in a finite population. A diploid individual is represented by n pairs of loci. Each locus has two alleles with additive effects 0 and 1, so that the genotypic value of an individual is the sum of all 1's at his $2n$ loci. The phenotypic value is the sum of the genotypic value and a random normal variable with 0 mean and appropriate variance V_E to achieve the specified heritability h^2 in the first generation. The population is split into two groups (sexes) of equal size. In each group, at each generation (t), a fixed proportion of individuals is selected. For each zygote, one parent is drawn at random from each selected group, and produces a gamete after crossing-overs have been generated. The procedure is repeated until the required number of zygotes is obtained (half males and half females). Generations are non-overlapping, and the process starts with an initial population obtained by drawing at random two alleles at each locus for all individuals, with a specified frequency (0·5 for the results shown here) of favourable alleles (alleles with effects equal to 1). Crossing-overs are assumed to be Poisson distributed with no interference, and the Haldane mapping function is used. Linkage may be represented either by the recombination fraction r between adjacent loci, or by the map length L (in Morgans) on which the loci are evenly distributed, with one locus standing at each end of the chromosome segment so that:

$$L = -\tfrac{1}{2}(n-1)\ln(1-2r).$$

For each set of parameters, the program simulates selection from a random initial population ($t = 0$) until the requested number of generations or complete fixation at all loci is achieved, then the whole run is replicated. The simulation data shown in Figs. 1 to 4 are the means of the studied parameters over 400 replications. The data shown in Figs. 5 and 6 correspond to one single replicate.

3. Results

(i) *The reduction of additive genetic variance*

(*a*) *One generation of selection*

The corresponding calculations are detailed in the Appendix. Assuming that the total genetic variance V_A is the sum of the variances σ_i^2 contributed by loci i, it turns out that after one cycle of truncation selection the contributions g_i and g_j of genes at two loci i and j become correlated, so that:

$$Cov^*(g_i, g_j) = -\imath(\imath - \xi)\frac{\sigma_i^2\, \sigma_j^2}{4V_P}.$$

Although they are all small, these covariances sum up to a macroscopic effect, so that in the next generation one gets:

$$V_A^* = V_A(1 - \tfrac{1}{2}\imath(\imath - \xi)h^2), \tag{7}$$

where h^2 is the initial heritability of the selected trait.

The detailed calculation of the Appendix shows how the first generation of selection develops systematic negative covariances between the effects of contributing loci. From the viewpoint of population genetics these covariances are linkage disequilibria, indicating a tendency of gametes to carry alleles of opposite effects on the selected trait.

The expression (7) is derived here directly under standard hypotheses of normal distribution of the phenotypic values and small effects of individual loci. It could also be obtained from the general approach of Turelli & Barton (1990), when assuming normal distributions of phenotypes.

(*b*) *Several generations of selection*

Considering a finite number of loci contributing additively to the variability of quantitative trait, the Gaussian approximation recalled in the Models and methods section (eqns (5) and (6)), allows one to derive the values of macroscopic parameters (eqns (2) and (4)) at each generation, the expected mean value of phenotypes $\bar{P}^{(t)} = 2\Sigma_i \bar{g}^{(t)}$ and the additive genetic variance

$$V_A^{(t)} = 2\sum_i \sum_j G_{ij}^{(t)} = V_g^{(t)} + C^{(t)}, \tag{8}$$

where $V_g^{(t)} = 2\Sigma_i G_{ii}^{(t)}$ is the genic variance, i.e. the additive genetic variance if all loci were in linkage

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

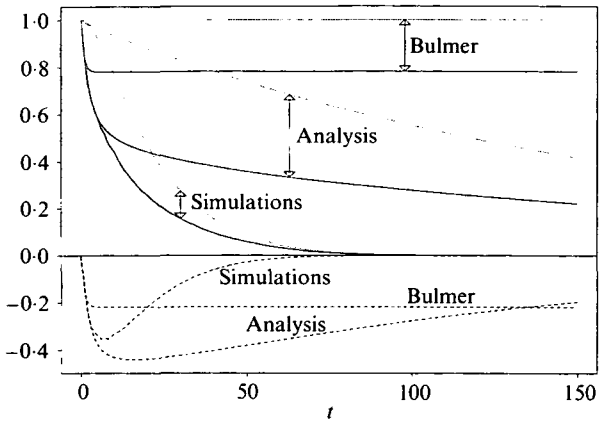

Fig. 1. Predictions for genetic variances from analytic and simulation models as compared to the predictions of Bulmer (1980, eqn 9.33). Abscissa: time in generations. Ordinate: ratio of variances at t over value in initial generation. Positive values: for each of the three models: genic variance V_g (dotted line, up arrow) and additive genetic variance V_A (solid line, down arrow). Negative values: difference $C = V_A - V_g$ for each of the three models (dashed lines). $n = 51$ loci spread on 50 cM, $r = 0{\cdot}01$, 100 individuals selected among 200 at each generation, $h^2 = 0{\cdot}5$. The simulation results are averaged over 400 replicates.

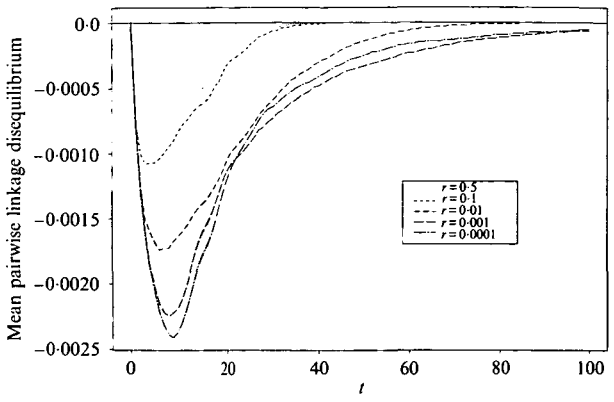

Fig. 2. Mean linkage disequilibria between all pairs of loci. Simulation results averaged over 400 replicates. Variation with time of $\{2/[n(n-1)]\}\Sigma_{i>j} D_{ij}$ for various recombination rates r between adjacent loci. $n = 51$ loci, 100 individuals selected among 200 at each generation, $h^2 = 0{\cdot}5$. To be compared to the C values plotted in Fig. 1, the scale of the present ordinate must be multiplied by 200.

equilibrium, and $C^{(t)}$ is a reduction term. These equations were solved numerically.

Results from the analytical approximation and from simulations are presented in Fig. 1 where predictions for $V_A^{(t)}$ and $V_g^{(t)}$ from both models are plotted as well as their differences $C^{(t)}$ and can be compared with the predictions of Bulmer (1980, eqn 9.33).

As stated in the Appendix, the additive genetic variance after one generation of selection does not depend on the recombination fraction between loci, provided the initial population is at linkage equilibrium, so that the value predicted by the analytic model is the same as Bulmer's and the value predicted by the simulations is very close. This is no longer the case in the following generations. In Bulmer's model, V_g remains constant and V_A reaches a positive equilibrium value, while in our models both V_g and V_A tend towards zero, the decrease being much faster in the simulations. Note that reduction C of variance predicted by the analytic model and the simulations is larger than Bulmer's prediction in the first generations, then becomes smaller after a number of generations while tending towards 0.

In the additive model considered here, with allelic effects 0 and 1 at each locus, covariances G_{ij} between gene effects are equal to linkage disequilibria D_{ij} in the sense of Lewontin & Kojima (1960). So, the genetic variance of the character is completely determined by the variation of gene frequencies (controlling the genic variance V_g), and by linkage disequilibria between genes at different loci (reducing V_g to its actual value V_A). We will in the following focus on the dynamics of linkage disequilibria and their consequences for the genetic variance.

Since simulations make it easier to study these values without any approximation, and since they provide additional features unpredicted by the analytic approach, the following will refer to the results of our simulation model.

(ii) *Dynamics of pairwise linkage disequilibria*

We have chosen to note positively the disequilibria corresponding to loci in coupling, and negatively the disequilibria corresponding to loci in repulsion. The linkage disequilibria addressed in this paper being always negative (on an average), strong disequilibria will mean low and negative values.

Figure 2 gives the values of $\{2/[n(n-1)]\}\Sigma_{i<j} D_{ij}$ measured from the simulations for $n = 51$ loci and five different recombination rates. As predicted by the analytic model, it is found that directional selection generates negative linkage disequilibria between the loci controlling the character. The stronger the linkage, the stronger are linkage disequilibria and the longer they persist in the population. But whereas linkage disequilibria increase at first, they later undergo a reduction. The values plotted in Fig. 2 correspond to the sum of linkage disequilibria between all pairs of different loci, including disequilibria equal to zero between pairs for which one or both loci is fixed. The simulation results show that the time when linkage disequilibria start decreasing corresponds to the time when some loci reach fixation (data not shown). Fixations, and the dependence of any measure of linkage disequilibria on the variation of gene frequencies (Lewontin, 1988) explain that all values plotted in Fig. 2 eventually tend towards zero though selection still goes on.

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

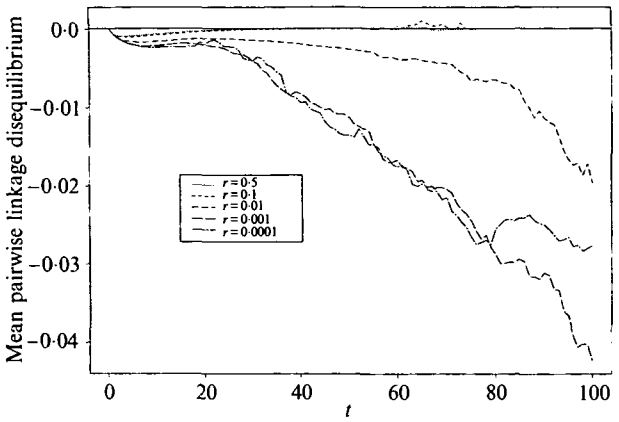

Fig. 3. Mean linkage disequilibria between pairs of polymorphic loci. Same as in Fig. 2 restricted to the pairs for which neither of the loci is fixed.

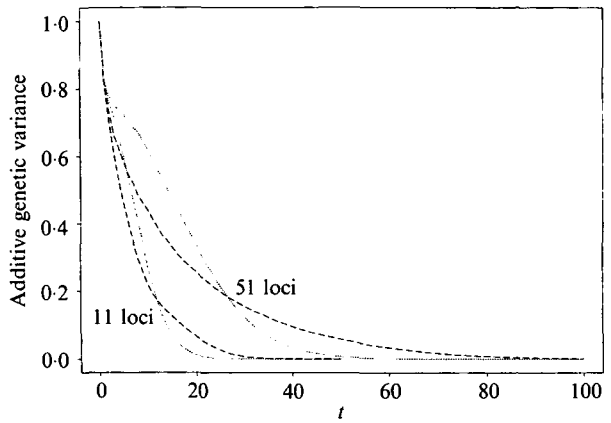

Fig. 4. Dynamics of additive genetic variance for unlinked versus linked genetic systems. Simulation results averaged over 400 replicates. Variation with time t of ratio $V_A^{(t)}/V_A^{(0)}$ for two numbers of loci (11 and 51) and two recombination rates ($r = 0{\cdot}5$, dotted lines; $r = 0{\cdot}01$, dashed lines). 100 individuals selected among 200 at each generation, $h^2 = 0{\cdot}5$.

The effect of fixations can be seen by comparing Fig. 2 to Fig. 3, where only the pairs of polymorphic loci are taken into account (pairs for which neither of the loci is fixed). This shows that, although the difference $V_A - V_g$ tends towards zero, linkage disequilibria between polymorphic loci still increase as a consequence of directional selection. The same qualitative conclusion may be drawn by studying the D' measure of linkage disequilibrium proposed by Lewontin (1964) (results not shown). In the mid and long term (Fig. 3), linkage disequilibrium values corresponding to the lowest recombination rates increase even above the value for no recombination (data not shown), indicating that linkage disequilibria generated by selection are then stronger than the maximum possible value generated by the random sampling of gametes in the initial generation.

(iii) Consequences on genetic variance

The effects of linkage between quantitative loci on the genetic variance can be seen by comparing the

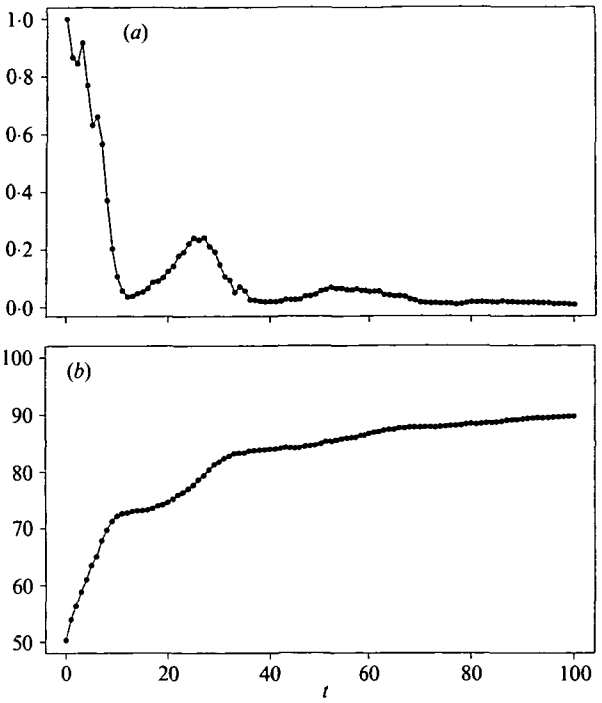

Fig. 5. Dynamics of additive genetic variance and (*b*) mean of a tightly linked genetic system in a single replicate. Simulation results. (*a*): variation with time t of ratio $V_A^{(t)}/V_A^{(0)}$. (*b*): variation with time t of the mean of the character in the population. One given replicate for $n = 51$ loci, $r = 0{\cdot}001$, 100 individuals selected among 200 at each generation, $h^2 = 0{\cdot}5$.

dynamics of V_A for linked versus unlinked genetic systems. This is done in Fig. 4 for two numbers of loci. In the first generations, the decrease of V_A is faster for the linked systems than for the unlinked ones. On the contrary, in the last generations, whereas the genetic variability of the unlinked systems is almost exhausted, the additive genetic variance of the linked systems is sustained at a low but higher level until the end of the process. The same qualitative behaviour is observed in the analytical Gaussian approximation, although the process is slower, so that crosses occur around generations 200 and 450 (data not shown), instead of 12 and 28 in Fig. 4.

The data shown in Fig. 4 are the average of $V_A^{(t)}$ over 400 replicates. In a single replicate, the genetic variance for tightly linked loci may be not only maintained, but actually increase in the population as is illustrated in Fig. 5*a*, with each increase of variance corresponding to an acceleration of response following a plateau on the mean (Fig. 5*b*). Note that selection intensity is kept constant. This phenomenon cannot be predicted by the Gaussian approximation.

4. Discussion

(i) *The Bulmer effect*

Bulmer's result was formalized in a purely statistical

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

way: the expression for the reduction of V_A due to selection (Bulmer, 1980, eqn 9.32) is obtained by linear homoscedastic regression under the hypotheses of the infinitesimal model (infinite number of unlinked loci), and of joint Gaussian distribution between phenotypic and breeding values. We show in the Appendix that in the first generation of selection the same result may be obtained rigorously under less restrictive hypotheses assuming a large number of loci and the normality of phenotypic values only. It is important to note that the hypotheses of the Appendix are different from the ones of the analytic model of Fig. 1. Although the hypotheses of the Appendix are more general, they only allow results to be obtained for the first generation of selection, assuming that the initial population is at linkage equilibrium. If linkage disequilibrium is present in the initial population, additional approximations are needed to go further, all the more so since linkage and drift are taken into account.

Our results in Fig. 1 emphasize the importance of the Bulmer effect in the dynamics of the additive genetic variance. However, they show two main differences with Bulmer's predictions.

Both the genic and genetic variances tend towards 0 in our models, whereas V_g is constant in Bulmer's analysis, leading to a positive equilibrium value for V_A (Bulmer, 1980). This property holds only if both the population size and the number of loci are infinite. In our models, both numbers are finite.

In the first generations, the reduction C of variance predicted by our models is more important than predicted by Bulmer. This is an effect of linkage. It can be seen from Fig. 2 that in the case of unlinked loci the reduction predicted by our simulation model remains below or equal to Bulmer's prediction. This is also true for the Gaussian model (Chevalet, 1994).

The faster decrease of variances shown by the simulations as compared to the analytic results (Fig. 1) may be due to the latter relying on the multinormal approximation assuming at each locus a distribution of allelic effects that is unskewed and of infinite range, whereas only two alleles per locus are considered in the completely finite simulation model and it is known that selection promotes skewed distributions of allelic effects at each locus (Barton & Turelli, 1987). Also, fixation events are taken into account in the Gaussian equations only implicitly with the parameter N, the actual size of the population (eqn (6)), whereas true fixation events occur in the simulations under both drift and selection.

(ii) *Dynamics of polygenic systems under selection*

The consequences of selection, linkage and drift on the genetic variance of the character can be summarized as follows. The aim of directional selection is to accumulate all the favourable alleles in the same genotype as fast as possible. In a finite population, the probability that this goal is achieved in the first few generations is low (unless a small number of unlinked loci is considered), and in practice it will seldom be the case. Hence, the gametes selected for are not those with the best possible values, but they are taken from best genotypes present in the population at the time. These gametes may bear favourable alleles at a large number of loci, and possibly unfavourable alleles at other loci, so that linkage disequilibrium corresponding to gametes in repulsion will increase in the selected parent population. These disequilibria will be reduced in the next offspring generation as a function of the recombination rate.

If the recombination rate is low, i.e. if linkage is tight, linkage disequilibria will be slightly reduced, and repulsion associations of alleles will remain in the offspring population. Hence, selection for the favourable alleles will increase the frequencies of both favourable and unfavourable alleles at different loci (hitch-hiking effect). A large amount of genetic variability will then be hidden in negative linkage disequilibria, and the additive genetic variance V_A will be strongly reduced. In the following generations, unfavourable alleles will be fixed at some loci, due to hitch-hiking effects. This will definitively reduce the level of response to selection that might be achieved at the end of the process. At other loci where the frequency of the unfavourable allele is high, favourable alleles might nevertheless appear in coupling on rare gametes of high genetic value, due to recombination. These gametes are then likely to be selected for, so that the frequency of the favourable alleles will increase at such loci, leading to a release of genetic variability in the population.

On the contrary, if linkage is loose, linkage disequilibria will be low, corresponding to weak hitch-hiking effects, only a small amount of genetic variability will be hidden, and frequent recombination events will lead to both a rapid exhaustion of genetic variability and a high level of response to selection.

This intuitive reasoning gives an interpretation of the behaviour of the simulation model, as illustrated in Figs. 2 to 4, and of the results observed by Hospital & Chevalet (1993): whereas unlinked systems may withstand strong selection intensities and still achieve high response to selection, linked systems require low selection intensities to reduce the rate of fixation of unfavourable alleles due to hitch-hiking effects, and to allow enough recombination events to take place releasing as much genetic variability as possible. This analysis strengthens the relevance of recurrent selection schemes to avoid the fixation of unfavourable alleles, and to increase the ultimate response to selection.

The value of the previous analysis is confirmed by the behaviour of genetic structures in single replicates.

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

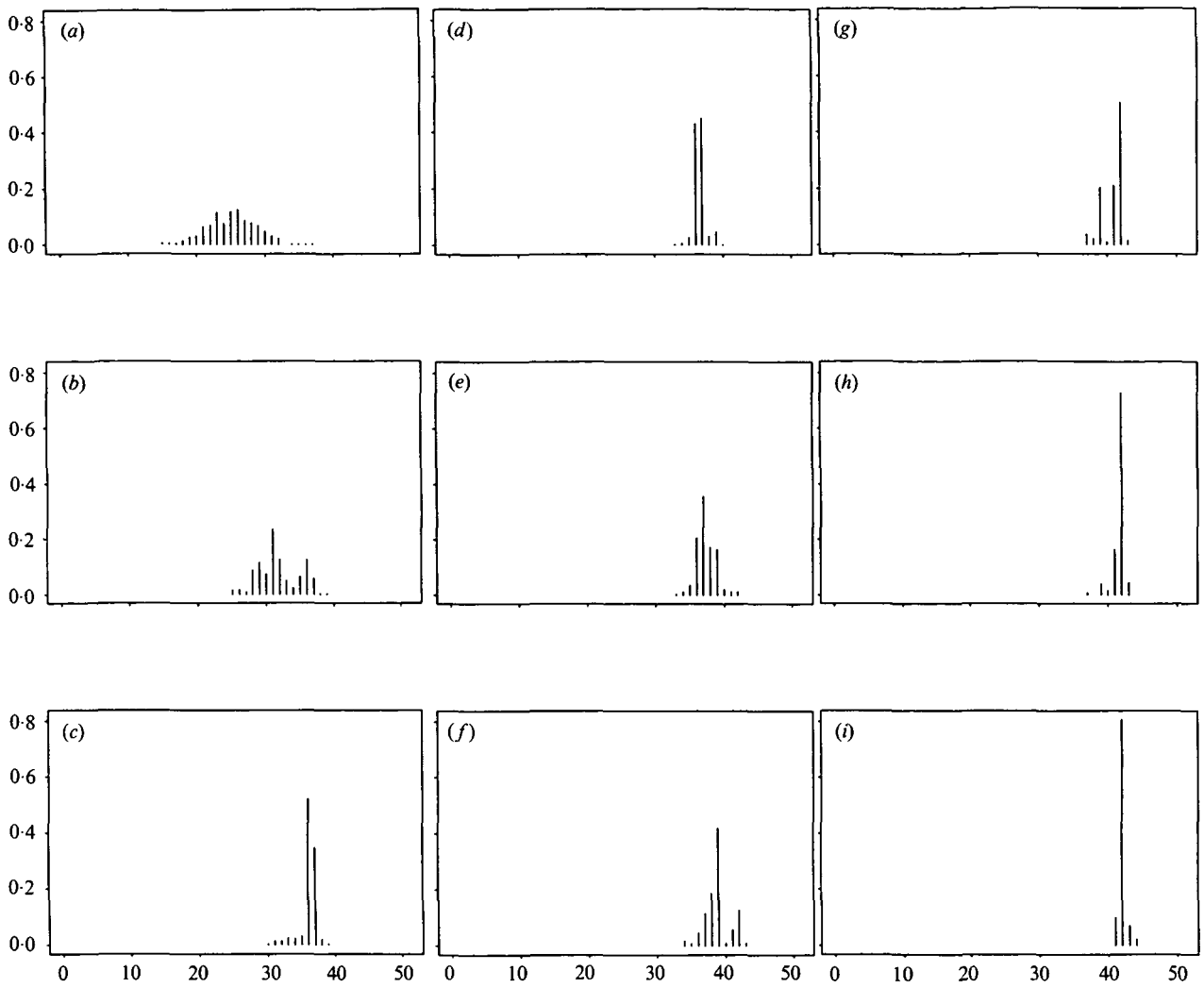

Fig. 6. Frequency distribution of gametic types. Simulation results at various generations in the same replicate as in Fig. 5. Abscissa: gametes values (number of '1' alleles). Ordinate: frequency of the gametic type in the population. Only non-zero frequencies are plotted. (*a*), Generation 0; (*b*), generation 5; (*c*), generation 10; (*d*), generation 15; (*e*), generation 20; (*f*), generation 25; (*g*), generation 30; (*h*), generation 35; (*i*), generation 40.

As illustrated in Fig. 5, the release of genetic variability due to recombination may lead to an increase of V_A. The distribution of gamete frequencies corresponding to the same replicate is shown in Fig. 6. It is seen that the increase of variance in Fig. 5 corresponds to the invasion of an advantageous gamete in the population, after it appeared at a low frequency through recombination. From generation 0 to 10, the initial, approximately normal distribution of gametic types becomes skewed, two gametic types (of values 36 and 37) becoming predominant in the population. Then, during the first peak of variance between generations 10 and 40, these gametic types are replaced by a gamete of higher value (42) not present in the initial population. Note that if one gametic type is predominant in the population, the zygote bearing two copies of it is also likely to be predominant so that the same results are obtained for zygotic types (data not shown).

Although the example shown in Fig. 5 and 6 was chosen to be particularly characteristic, it is important to note that such an increase of variance was often observed in a given replicate in our simulations, under various conditions provided that the recombination rate is not too high, the population size and the number of loci are not too small, the heritability and selection intensity are not too high. The magnitude of the increase of variance may be different among replicates with the same initial conditions. More importantly the increase may take place at various times so that the phenomenon is masked under a slow continuous decrease of V_A when the results of several replicates are averaged.

The relevance of this increase of variance is twofold. Firstly, it is not taken into account in any present analytic model. Secondly, without calling upon mutations, the phenomenon provides an interpretation of the stop and start response observed in some long-term selection experiments (selection for wing vein length in *Drosophila*, Scharloo, 1987; selection

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

for oil and protein in maize, Dudley & Lambert, 1992).

Although we emphasized the role of linkage, as compared to the hypotheses of the infinitesimal model, the above observations are clearly consequences of the joint effects of linkage, drift and selection. If the population size is too small, or if selection is too strong, hitch-hiking effects will be of great importance and hardly balanced by favourable recombination events, so that few or no genetic variability may be released in the population, and the gametic type fixed at the limit may be close to the best gametic type present in the initial population.

(iii) *Other effects*

In the present paper, we focused on the control of genetic variance by linkage disequilibria. It is important to keep in mind that other effects influencing the dynamics of the genetic variance were not studied in detail although they were implicitly present in the simulation model.

(*a*) *Dynamics of the genic variance*

Whereas linkage disequilibria control the difference between the additive genetic variance V_A and the genic variance V_g, the variations of V_g itself are of importance. Most models in quantitative genetics either assume that V_g is constant (Bulmer, 1980) or consider only the variations of V_g due to drift (Robertson, 1970). The genic variance is determined by the variation of gene frequencies at all the loci, which are affected not only by drift, but also by the various effects of selection investigated in this paper. Further theory is hence needed to predict the joint dynamics of gene frequencies in polygenic systems undergoing selection.

(*b*) *Departure from random mating*

Since random mating was assumed in our models, covariances between the effects of genes located on different gametes are equal to zero, leading to the relatively simple expressions (1) and (8). It can be checked from the simulations that the difference between the genetic and the genic variance (Fig. 1) equals the sum $2\Sigma_{i\neq j} D_{ij}$ of linkage disequilibria between all pairs of different loci (Fig. 2). Covariances between gene effects at loci on different gametes may be neglected, as far as the means of these parameters over several replicates are considered. It is important to note that this is not true in any given replicate, especially for the first generations. This is due to stochastic departures from random mating between males and females. Whereas these effects do not affect our results, they should be taken into consideration when single replicates are considered, as is the case in practical experiments, or in the case of non random mating.

(iv) *The models*

In the present paper, we referred to two different models: an analytic model based on the multi-normal approximation of the distribution of gene effects, and a stochastic simulation model with two alleles per locus. Both models give some insight into the joint effects of selection, linkage and drift on the variability of quantitative characters over several generations. Yet, both models involve hypotheses that keep them away from what might be the real conditions. The stochastic process of drift may be considered as better dealt with in the simulations where a true finite number of events is drawn at random amongst possibilities, whereas the Gaussian approximation introduces a virtual infinite range of states and deals only with expected values. A consequence is that rare genetic combinations of high value are always taken into account, even at low frequency, in the analytic approximation, whereas their low probability will almost always prevent them from being realized in the simulation process. Also, a deterministic modelling cannot be used to characterize the properties of unique trajectories, as it is possible in simulations (Figs. 5 and 6).

On the contrary, the simulation process deals with the limited amount of variability present in the initial population, so that the studied effects may only be evaluated in the way they affect the ability of selection to produce the best possible genetic combinations before complete loss of variability due to drift. In this, the simulation model lacks generality. New variability should be added in the system, for example by mutation as considered by Keightley & Hill (1987), or many more loci should be considered (Robertson, 1977; Robertson & Hill, 1983).

5. Conclusion

Most of the mechanisms investigated in the present paper, such as the hitch-hiking effect generated by drift and linkage disequilibrium, are the same as the ones intensively studied by theoretical population genetics in the one and two locus cases. However, analytic extensions of such available approaches to polygenic characters reach the limits of mathematical tractability, and do not deal with the stochastic features of the system. Yet, it could be argued that natural, or at least artificial selection is remarkably efficient regardless of the mathematical complexity of the problem. Hence, instead of making the problem more complex, one may wish to simplify it by seeking some general laws that would both describe the dynamics of the macroscopic parameters of the system

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

without referencing to individual genes, and be consistent with the results of the present paper. In this, we follow the analogy of Lewontin (1974) with thermodynamics. In the present paper, our approach started from an analytic model, which predictions were checked and analysed by simulations. It appears that some features of the simulation results were not predicted by the analytic model. In a forthcoming paper, we plan to proceed from these simulations to seek simple macroscopic parameters to be used to refine the analytic predictions.

We wish to thank I. Goldringer, Y. Michalakis, M. Slatkin and N. Barton for helpful comments on earlier versions of this manuscript. Thanks are also due to J. Deatrick for trying to improve our English.

Appendix

Linkage disequilibrium under selection for a polygenic trait: a genetic approach to Bulmer's effect

We give here a genetic demonstration of Bulmer's effect for the first generation of selection assuming a large population size and a large number n of loci. In this population, contributions $g_i^{(m)}$ and $g_i^{(f)}$ are independent random variables taking values a_{il} with probability q_{il} where q_{il} is the frequency of the l-th allele A_{il} at locus i. The variance of $g_i^{(m)}$ is equal to $\frac{1}{2}\sigma_i^2$. For the sake of simplicity we assume in the following that $g_i^{(m)}$, $g_i^{(f)}$ and P have zero mean, and the phenotypic variance V_P is denoted σ^2.

Calculations are performed in three steps: (i) evaluate the changes in allele frequencies and in haplotype frequencies due to selection, (ii) calculate the new allele and haplotype frequencies in the next generation, and (iii) derive the new additive genetic variance in the next generation.

Previous derivations (Griffing, 1960) have given the first-order approximations for the changes in allele frequencies (step (i)). However, second-order approximations (in $1/\sqrt{n}$) are needed to exhibit the joint effects of many small second-order contributions of selection to the covariances between gene effects at different loci, the sum of which is finite and yield the total contribution of gamete disequilibria to the additive genetic variance. The important point in step (ii) is linkage equilibrium in the initial population, that makes the result in first generation independent of recombination fractions between loci.

(i) *Change of allele frequencies under selection*

An individual with phenotype P has a fitness function equal to $W(P) = Y(P-\xi\sigma)$ where $Y(.)$ is the step function. The relative fitness of a zygote carrying genotype $A_{il}A_{im}$ is:

$$w_{i,lm} = \frac{Pr(P > \xi\sigma \mid A_{il}A_{im})}{Pr(P > \zeta\sigma)}.$$

We can write

$$\begin{aligned} &Pr(P > \xi\sigma \mid A_{il}A_{im})\, Pr(A_{il}A_{im}) \\ &\quad = Pr(\{P > \xi\sigma\} \wedge \{A_{il}A_{im}\}) \\ &\quad = Pr(\{P > \xi\sigma\} \wedge \{g_i^{(m)} + g_i^{(f)} = a_{il} + a_{im}\} \\ &\qquad \wedge \{A_{il}A_{im}\}) \\ &\quad = Pr(\{P - (g_i^{(m)} + g_i^{(f)}) > \xi\sigma - (a_{il} + a_{im})\} \\ &\qquad \wedge \{A_{il}A_{im}\}). \end{aligned} \tag{A 1}$$

Since the distribution of $P-(g_i^{(m)}+g_i^{(f)})$ is independent of the state at locus i, we have:

$$\begin{aligned} &Pr(P > \xi\sigma \mid A_{il}A_{im}) \\ &\quad = Pr(P - (g_i^{(m)} + g_i^{(f)}) > \xi\sigma - (a_{il} + a_{im})) \end{aligned} \tag{A 2}$$

$P-(g_i^{(m)}+g_i^{(f)})$ is a Gaussian variable with mean 0 and variance $\tau^2 = \sigma^2 - \sigma_i^2$, so that:

$$\begin{aligned} w_{i,lm} &= \left(\frac{1}{\sqrt{2\pi}\tau}\int_{\xi\sigma - a_{il} - a_{im}}^{\infty} e^{-\frac{1}{2}x^2/\tau^2} dx\right) \\ &\quad \div \left(\frac{1}{\sqrt{2\pi}\sigma}\int_{\xi\sigma}^{\infty} e^{-\frac{1}{2}x^2/\sigma^2} dx\right) \\ &= \left(\int_{\xi^*}^{\infty} e^{-\frac{1}{2}x^2} dx\right) \Big/ \left(\int_{\xi}^{\infty} e^{-\frac{1}{2}x^2} dx\right), \end{aligned} \tag{A 3}$$

where

$$\xi^* = \frac{\xi\sigma - a_{il} - a_{im}}{\tau}.$$

Noting that $\sigma/\tau - 1$ is of order $1/n$, and that a_{il}/σ is of order $1/\sqrt{n}$, a development to the second order in $1/\sqrt{n}$ gives:

$$\xi^* = \xi - \frac{a_{il} + a_{im}}{\sigma} + \frac{1}{2}\xi\frac{\sigma_i^2}{\sigma^2} + \mathcal{O}\left(\frac{1}{n^{3/2}}\right).$$

Equation (A 3) becomes:

$$w_{i,lm} \simeq 1 + \imath\frac{a_{il} + a_{im}}{\sigma} + \frac{1}{2}\imath\xi\frac{(a_{il} + a_{im})^2 - \sigma_i^2}{\sigma^2} + \ldots \tag{A 4}$$

Similarly, the fitness of a single allele A_{il} is

$$w_{i,l} \simeq 1 + \imath\frac{a_{il}}{\sigma} + \frac{1}{2}\imath\xi\frac{a_{il}^2 - \frac{1}{2}\sigma_i^2}{\sigma^2} + \ldots$$

and that of the haplotype $A_{il}A_{jk}$ for two loci i and j is:

$$w_{i,l;j,k} \simeq 1 + \imath\frac{a_{il} + a_{jk}}{\sigma} + \frac{1}{2}\imath\xi\frac{(a_{il} + a_{jk})^2 - \frac{1}{2}\sigma_i^2 - \frac{1}{2}\sigma_j^2}{\sigma^2} + \ldots$$

(note that fitnesses of two-loci genotypes such as $A_{il}A_{jk}/A_{im}A_{jh}$ can be computed in the same way). From the previous expressions of fitnesses are derived the following starred frequencies of alleles and of haplotypes among the reproducing individuals:

$$q_{il}^* = q_{il}\, w_{il}$$

$$Q_{i,l;j,k}^* = q_{il}\, q_{jk}\, w_{i,l;j,k}.$$

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

(ii) *Allele and haplotype frequencies in the next generation*

The hypothesis that linkage equilibrium holds in the initial generation allows us to write down directly that the frequencies of gametic types produced by the selected individuals are equal to those in the selected population of parents, so that the previous expressions q^* and Q^* are also the frequencies in the gametes uniting to form the next generation.

This can be checked as follows. Consider two loci i and j, with recombination fraction r. Let $P_{i,lm;j,kh}$ be the frequency of selected individuals of genotype $A_{il}A_{jk}/A_{im}A_{jh}$ carrying a paternal chromosome of type $A_{il}A_{jk}$ and a maternal chromosome of type $A_{im}A_{jh}$. Such a genotype produces gametes of type $A_{il}A_{jk}$ with probability equal to $\frac{1}{2}(1-r)$, while the symmetric genotype $A_{il}A_{jk}/A_{im}A_{jk}$ produces gametes of type $A_{il}A_{jk}$ with probability equal to $\frac{1}{2}r$. Hence, the frequency $P'_{i,l;j,k}$ of gametes of type $A_{il}A_{jk}$ is:

$$\begin{aligned}P'_{i,l;j,k} = {} & \tfrac{1}{2}(1-r)\sum_{m\neq l}\sum_{j\neq k} P_{i,lm;j,kh} \\ & +\tfrac{1}{2}r\sum_{m\neq l}\sum_{h\neq k} P_{i,lm;j,hk} \\ & +\tfrac{1}{2}\sum_{h\neq k} P_{i,ll;j,kh}+\tfrac{1}{2}\sum_{h\neq k} P_{i,ll;j,hk} \\ & +\tfrac{1}{2}\sum_{m\neq l} P_{i,lm;j,kk}+\tfrac{1}{2}\sum_{m\neq l} P_{i,ml;j,kk} \\ & + P_{i,ll;j,kk}.\end{aligned}$$

One can note that the frequency of chromosome types $A_{il}A_{jk}$ among parents is given by the same expression, provided that r is taken equal to 0. The difference between this chromosome frequency and $P'_{i,l;j,k}$ is:

$$\tfrac{1}{2}r\sum_{m\neq l}\sum_{h\neq k}(P_{i,lm;j,kh}-P_{i,lm;j,hk}),$$

and is zero in the case of linkage equilibrium in the parent population, for any value of the recombination rate. Hence the frequencies of gametic types are equal to the previous Q^* expressions, since fitnesses of symmetrical genotypes $A_{il}A_{jk}/A_{im}A_{jh}$ and $A_{il}A_{jh}/A_{im}A_{jk}$ are equal.

(iii) *Genetic variance in the next generation*

We can now calculate the mean and variance of genotypic values G in the next generation, in which frequencies of alleles and haplotypes have changed. The first two moments of G are, E^* standing for expectations in the new generation

$$E^*(G) = 2\sum_i E^*(g_i)$$

$$Var^*(G) = 2\sum_i Var^*(g_i)+2\sum_i\sum_{j\neq i} Cov^*(g_i,g_j).$$

Taking account of the new frequencies of alleles we have

$$\begin{aligned}E^*(g_i) &= \sum_l a_{il}q^*_{il} \\ &= 0+\imath\frac{1}{2}\frac{\sigma_i^2}{\sigma}+\frac{1}{2}\imath\xi\frac{\sum_l a_{il}^3 q_{il}}{\sigma^2}+\dots\end{aligned}$$

Note that this expansion introduces the third moment of the distribution of gene effects, which is nearly zero under the assumptions that the distribution of G is nearly Gaussian. The variance of g_i is derived in the same way

$$E^*(g_i^2) = \tfrac{1}{2}\sigma_i^2+\frac{\imath}{\sigma}\sum_l a_{il}^3 q_{il}+\tfrac{1}{2}\imath\xi\frac{\sum_l a_{il}^4 q_{il}-\frac{1}{4}\sigma_i^4}{\sigma^2}+\dots$$

which introduces the fourth moment. Under the hypothesis of small effects of loci, the genetic distribution is nearly Gaussian, so that the reduced fourth moment is small. Grouping terms according to their order of magnitude, we have

$$\begin{aligned}Var^*(g_i) = {} & \sigma_i^2+\frac{\imath}{\sigma}\sum_l a_{il}^3 q_{il} \\ & -\imath(\imath-\xi)\frac{\sigma_i^4}{4\sigma^2}+\tfrac{1}{2}\imath\xi\frac{\sum_l a_{il}^4 q_{il}-\frac{3}{4}\sigma_i^4}{\sigma^2}+\dots\end{aligned}$$

Note that the second term is of order $\sigma^2/n^{3/2}$ and the last two terms are of order σ^2/n^2, so that summing up n variance terms will not contribute a significant change to the genic variance – the sum of variances contributed by all loci. Concerning covariances we have

$$\begin{aligned}E^*(g_ig_j) = {} & \sum_l\sum_k q_{il}q_{jk}a_{il}a_{jk} \\ & +\frac{\imath}{\sigma}\left(\sum_l\sum_k q_{il}q_{jk}a_{il}a_{jk}(a_{il}+a_{jk})\right) \\ & +\frac{1}{2}\frac{\imath\xi}{\sigma^2}\left(\sum_l\sum_k q_{il}q_{jk}a_{il}a_{jk}\right. \\ & \left.\times[(a_{il}+a_{jk})^2-\tfrac{1}{2}\sigma_i^2-\tfrac{1}{2}\sigma_j^2]\right)+\dots \\ = {} & 0+0+\frac{1}{2}\frac{\imath\xi}{\sigma^2}\sum_l\sum_k 2q_{il}q_{jk}a_{il}^2a_{jk}^2+\dots \\ = {} & \frac{1}{2}\frac{\imath\xi}{\sigma^2}(\tfrac{1}{2}\sigma_i^2\sigma_j^2)+\dots\end{aligned}$$

so that

$$Cov^*(g_i,g_j) \simeq \frac{\imath\xi}{4\sigma^2}\sigma_i^2\sigma_j^2-E^*(g_i)E^*(g_j),$$

which yields if third and higher moments are neglected

$$Cov^*(g_i,g_j) \simeq -\imath(\imath-\xi)\frac{\sigma_i^2\sigma_j^2}{4\sigma^2}.$$

For each variance and covariance, the change developed under selection is thus very small, of order σ^2/n^2. The change on the variance at each locus is negligible, as told above. The same effect on the covariances between gene effects at different loci has

https://doi.org/10.1017/S0016672300033498 Published online by Cambridge University Press

on the contrary a macroscopic effect obtained by summing up $n(n-1) \simeq n^2$ contributions. We get:

$$\begin{aligned} Var^*(G) &= 2\sum_i Var^*(g_i) + 2\sum_i \sum_{j \neq i} Cov^*(g_i, g_j) \\ &= \sum_i \sigma_i^2 - 2t(t-\xi)\sum_i \sum_j \frac{\sigma_i^2 \sigma_j^2}{4\sigma^2} + \dots \\ &\simeq h^2\sigma^2 - \tfrac{1}{2}t(t-\xi)h^4\sigma^2 \\ &= h^2\sigma^2(1 - \tfrac{1}{2}t(t-\xi)h^2), \end{aligned}$$

which is exactly Bulmer's expression.

References

Barton, N. H. & Turelli, M. (1987). Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genetical Research* **23**, 337–370.

Bulmer, M. (1971). The effect of selection on genetic variability. *American Naturalist* **105**, 201–211.

Bulmer, M. (1980). *The Mathematical Theory of Quantitative Genetics*. Oxford: Clarendon.

Chevalet, C. (1988). Control of genetic drift in selected populations. In *Proceedings 2nd International Conference on Quantitative Genetics* (ed. B. Weir, E. Eisen, M. Goodman and G. Namkoong), pp. 379–394.

Chevalet, C. (1994). An approximate theory of selection assuming a finite number of quantitative trait loci. *Genetics Selection Evolution* **26**, 379–400.

Dudley, W. J. & Lambert, R. J. (1992). Ninety generations of selection for oil and protein in maize. *Maydica* **37**, 81–87.

Felsenstein, J. (1977). Multivariate normal genetic models with a finite number of loci. In *Proceedings of the International Conference on Quantitative Genetics* (ed. E. Pollak, O. Kempthorne and T. B. Barley Jr), pp. 227–246. Ames, IA: Iowa State University Press.

Fisher, R. A. (1918). The correlation between relatives under the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*, 1st Ed. Oxford: Clarendon Press.

Griffing, B. (1960). Theoretical consequences of truncation selection based on the individual phenotype. *Australian Journal of Biological Science* **15**, 307–343.

Hospital, F. & Chevalet, C. (1993). Effect of population size and linkage on optimal selection intensity. *Theoretical and Applied Genetics* **86**, 775–780.

Keightley, P. D. & Hill, W. G. (1987). Directional selection and variation in finite populations. *Genetics* **117**, 573–582.

Lande, R. (1976). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical Research* **28**, 221–235.

Lewontin, R. C. (1964*a*). The interaction of selection and linkage. I. General considerations. Heterotic models. *Genetics* **49**, 49–67.

Lewontin, R. C. (1964*b*). The interaction of selection and linkage. II. Optimum models. *Genetics* **50**, 757–782.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Lewontin, R. C. (1988). On measures of gametic disequilibrium. *Genetics* **120**, 849–852.

Lewontin, R. C. & Kojima, K. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.

Lush, J. L. (1945). *Animal Breeding Plans*, 2nd Ed. Ames, IA: Iowa State University Press.

Mather, K. (1941). Variation and selection of polygenic characters. *Journal of Genetics* **41**, 159–193.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philosophical Transaction of the Royal Society of London Series A* **187**, 253–318.

Pearson, K. (1903). Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlations of organs. *Philosophical Transaction of the Royal Society of London Series A* **200**, 1–66.

Robertson, A. (1970). Some optimal problems in individual selection. *Theoretical Population Biology* **1**, 120–127.

Robertson, A. (1977). Artificial selection with a large number of linked loci. (ed. E. Pollak, O. Kempthorne and T. B. Barley Jr), In *Proceedings of the International Conference on Quantitative Genetics* pp. 307–322. Ames, IA: Iowa State University Press.

Robertson, A. & Hill, W. G. (1983). Population and quantitative genetics of many linked loci in finite populations. *Proceedings of the Royal Society of London Series B* **219**, 253–264.

Scharloo, W. (1987). Constraints in selection response. In *Genetic Constraints on Adaptative Evolution* (ed. V. Loeschcke), pp. 125–140. Berlin: Springer-Verlag.

Turelli, M. & Barton, N. (1990). Dynamics of polygenic characters under selection. *Theoretical Population Biology* **38**, 1–57.

Turelli, M. & Barton, N. H. (1994). Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? *Genetics* **138**, 913–941.