Dynamics of a Multilocus Trait

D. WAXMAN

*Center for the Study of Evolution, School of Biological Sciences,
University of Sussex, Brighton, Sussex, UK*

A detailed theoretical study of the dynamical behaviour of a quantitative trait under stabilizing selection is given for an effectively infinite sexual population. The trait is controlled by a finite number of loci and the exact equation obeyed by the distribution of allelic effects in gametes is derived and investigated. Results are derived for the effects of selection over one generation when the population is initially in linkage equilibrium, allelic effects are normally distributed, but the strength of selection is arbitrary. When weak stabilizing selection is operative, a more general analysis is presented. This includes explicitly identifying the linkage disequilibrium generating aspect of selection. In addition, the way quantities may be obtained from summing over effective one-locus haploid loci, is derived from the explicit dynamics. Numerical tests of dynamical results over 10^4 generations and equilibrium results are presented.

Keywords: Quantitative trait, dynamics, stabilising selection, source of linkage disequilibrium

1. Introduction and overview

The dynamical behaviour of a multilocus trait is the outcome of a number of different processes and is complex. In this work we consider a single multilocus trait of a sexual population of diploid individuals. The aim of this work is to provide theoretical insight into the interplay between the various processes that affect the dynamics. To achieve this we establish the exact equation for the dynamics and show how various features follow directly from the dynamics.

The model we adopt for our analysis is that introduced by Lande (1976), in which a quantitative trait is controlled by a finite number of linked loci (with non-zero, but otherwise arbitrary, recombination fractions between loci). At each locus, there are assumed to be an effectively infinite number of alleles and this is modelled by treating possible alleles as having a continuous set of effects (Crow and Kimura, 1964). The alleles contribute additively to the trait, which, in the present work, is taken to be

Corresponding author: Dr. D. Waxman, Center for the Study of Evolution, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, Sussex, UK, E-mail: D.Waxman@sussex.ac.uk

subject to stabilizing selection. The population is taken to be sufficiently large that stochastic effects may be neglected.

Lande (1976) analyzed his model using a Gaussian distribution of allelic effects. This was based on an approximation of Kimura (1965), which was subsequently shown to not apply for sexual populations under normal circumstances (Turelli, 1984). Fleming (1979) also analyzed Lande's model, under what is essentially the Gaussian approximation of Kimura (1965).

Bulmer (1980) analyzed a related model, where discrete effect alleles at an effectively infinite number of loci, each of infinitesimal effect (the infinitesimal model), control a quantitative trait. The central limit theorem allowed him to invoke normality of the distribution of genotypic values and to infer an approximate form for the contribution of linkage disequilibrium to the genetic variance. Generally, however, there are not an extremely large number of small effect loci controlling a trait and the distribution of genotypic values deviates from a Gaussian (Turelli and Barton, 1994). There are additional artifacts of the infinitesimal model. In particular, selection does not cause changes in gene frequencies, but only generates pairwise correlations (linkage disequilibria) between genes in gametes of common

parental origin. In practise gene frequencies do change after a modest number of generations (Turelli and Barton, 1994).

The exact equation obeyed by the distribution of allelic effects in gametes in the Lande model incorporates mutation, selection and recombination. We partially analyse this in one particular case when selection is arbitrary and provide a systematic analysis when selection is weak. In the weak selection case, various terms in the equation are separated according to their effects. We explicitly identify the terms in the selection function that generate linkage disequilibrium as well as terms which simply sum contributions from individual alleles. The framework we present is distinct from that in other recent work on multilocus systems and is complementary to these investigations (Bürger, 1993; Bürger and Hofbauer, 1994; Barton and Turelli, 1991, 1994; Nagylaki, 1993; Christiansen, 1999). We also believe that the mathematical approach to the dynamics presented here may be generalizable to more complex situations.

We adopt the convention that summations with unspecified limits cover the full range of the summation index and that all integrals range from $-\infty$ to ∞. Generally, but not always, we use the convention that quantities represented by a capital letter denote a random variable, and those in lower case denote a particular realization of that variable. Additionally, the only correlations we consider in this work are those between alleles at two different loci, i.e. pairwise linkage disequilibria, and for brevity we shall omit the adjective "pairwise" throughout the paper.

2. Model

Consider an effectively infinite population of diploid, randomly mating dioecious sexual organisms. Their lifecycle takes place in discrete generations. One generation consists of: (i) the random union of gametes to form zygotes, (ii) viability selection on a single polygenic trait, (iii) the production, by surviving individuals, of gametes and the death of these individuals shortly afterwards.

The trait under consideration is controlled by L linked loci where $2L$ alleles reside. These alleles combine, both across and between loci, in a purely additive manner i.e. without dominance or epistasis.

The phenotypic value of the trait, Z, is the sum of the genotypic value, G, and an environmental effect $\mathcal{E}$. Thus $Z = G + \mathcal{E}$ where

$$G = \sum_{i=1}^{L}(X_i + Y_i) \qquad (1)$$

and X_i (Y_i) is the contribution of the paternally (maternally) inherited allele at locus i. Each X_i and Y_i is continuous (Crow and Kimura, 1964) and can take any value from $-\infty$ to ∞. The environmental effect, $\mathcal{E}$, is a random Gaussian variable that is independent of G. It has a mean of zero, and a variance of V_e.

Each allele in an individual is a copy of an allele in one or other of its parents. Any difference between a parental allele and the allele passed on to an offspring is due to mutations. Alleles are taken to mutate independently of all other alleles and all loci are equally mutable. The probability of a mutation per replication of an allele is denoted by μ and the effect of a mutated allele is taken to be normally distributed around the parental allele of which it is a copy (Kimura, 1965). Given the occurrence of a mutation that e.g. alters the allele of paternal origin at locus i, the probability that the mutated offspring will inherit an allele, with effect X_i in the infinitesimal interval $u + du > X_i > u$, is $f(u - x^*)du$, where x^* is the paternal value of X_i,

$$f(x) = (2\pi m^2)^{-1/2} \exp(-x^2 / (2m^2)) \qquad (2)$$

and m^2 is the variance of mutated effects around the parental value.

Fertility is taken to be independent of phenotype and selection acts only on viability, which is determined by an individual's phenotypic value. The viability of an individual, with phenotypic value z, is taken to be proportional to

$$\exp[-(z - \theta)^2 / (2V)] \qquad (3)$$

(Haldane, 1954) where $V > 0$. This corresponds to normalizing selection where the optimal phenotypic value is θ (a constant). The quantity V^{-1} is a measure of the strength of selection on phenotypic values; small V^{-1} corresponds to weak selection and vice versa. Averaging (3) over environmental effects yields the viability of individuals in terms of their genotypic value. An individual whose genotypic value is $\sum_{i=1}^{L}(x_i + y_i)$ has a viability proportional to

$$w(\mathbf{x}, \mathbf{y}) \overset{\text{def}}{\equiv} \exp(-[\sum_{i=1}^{L} (x_i + y_i) - \theta]^2 / (2V_s)) \quad (4)$$

where

$$V_s \overset{\text{def}}{\equiv} V + V_e. \quad (5)$$

In (4), we have introduced a convenient vector notation, so generally, a function depending on $\mathbf{x}$ will depend on the L variables $x_1, x_2, ..., x_L$. We shall refer to $w(\mathbf{x}, \mathbf{y})$ as the fitness of individuals with genotypic value $\sum_{i=1}^{L} (x_i + y_i)$.

3. Compact representation of the dynamics

For the model outlined, the population, in generation $\tau = 0, 1, 2, ...$, is characterized by the distribution (probability density) of allelic effects of paternal gametes

$$\Phi_\tau (\mathbf{x}) \equiv \Phi_\tau (x_1, x_2, ..., x_L). \quad (6)$$

Assuming no sexual dimorphism, the distribution of maternal gametes is identical to that of paternal gametes, namely $\Phi_\tau(\mathbf{x})$. A property of $\Phi_\tau(\mathbf{x})$, since it is a probability density, is

$$\int \Phi_\tau (\mathbf{x}) d^L x = 1 \quad (7)$$

where $d^L x = dx_1 dx_2...dx_L$.

In generation τ, the distribution of allelic effects of zygotes, under random mating, is $\Phi_\tau(\mathbf{x})\Phi_\tau(\mathbf{y})$. Thus prior to selection, the expectation of an arbitrary quantity $H(\mathbf{X},\mathbf{Y})$ that depends on the effects of paternally and maternally inherited alleles, $\mathbf{X} = (X_1, X_2, ..., X_L)$ and $\mathbf{Y} = (Y_1, Y_2, ..., Y_L)$ is denoted by $E_\tau [H (\mathbf{X}, \mathbf{Y})]$ and is given by $E_\tau [H (\mathbf{X}, \mathbf{Y})] = \int H(\mathbf{x}, \mathbf{y})\Phi_\tau(\mathbf{x})\Phi_\tau(\mathbf{y})d^L x d^L y$. (The expectation *after* selection, in generation τ, is $E_\tau[w(\mathbf{X}, \mathbf{Y}) H(\mathbf{X}, \mathbf{Y})]/ E_\tau[w(\mathbf{X}, \mathbf{Y})]$.)

The dynamics determines how $\Phi_\tau(\mathbf{x})$ changes from one generation to the next. The equation for this is

$$\Phi_{\tau+1}(\mathbf{x}) = \frac{\int K(\mathbf{x}; \mathbf{u}, \mathbf{v}) w(\mathbf{u}, \mathbf{v}) \Phi_\tau(\mathbf{u})\Phi_\tau(\mathbf{v}) d^L u d^L v}{\int w(\mathbf{u}, \mathbf{v}) \Phi_\tau(\mathbf{u})\Phi_\tau(\mathbf{v}) d^L u d^L v} \quad (8)$$

where $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ incorporates mutation and recombination. This equation is derived Appendix B and, in equations (B1) to (B5) of this Appendix, we give a specification of $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$. In Appendix C some properties of $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ are derived.

4. Results following from the exact dynamics

Equation (8) contains an exact statement of the dynamics. It is possible to extract a small number of exact results from this and we shall present these before considering an approximate treatment.

Consider the effects of a single round of selection on a population that, in generation 0, is in linkage equilibrium. We take the distribution of allelic effects in gametes to be given

$$\Phi_0 (x) = \prod_{j=1}^{L} \chi_j (x_j) \quad (9)$$

$$\chi_j (x_j) = (2\pi\alpha_j^2)^{-1/2} \exp[-(x_j - c_j)^2 / (2\alpha_j^2)] \quad (10)$$

corresponding to the effects of alleles at locus j being normally distributed around $x_j = c_j$ and having a variance of α_j^2. Let us define

$$\overline{G}_0 \overset{\text{def}}{\equiv} 2\sum_{j=1}^{L} c_j, \quad V_{g,0} \overset{\text{def}}{\equiv} 2\sum_{j=1}^{L} \alpha_j^2 \quad (11)$$

then in Appendix D we show that the following hold, with no approximation:

$$\int x_j \Phi_1 (\mathbf{x}) d^L x = c_j - (\overline{G}_0 - \theta)\alpha_j^2 / (V_s + V_{g,0}) \quad (12)$$

$$\int x_j^2 \Phi_1 (\mathbf{x}) d^L x - (\int x_j \Phi_1 (\mathbf{x}) d^L x)^2$$
$$= \mu m^2 + \alpha_j^2 - \alpha_j^4 / (V_s + V_{g,0}) \quad (13)$$

$$\int x_j x_k \Phi_1 (\mathbf{x}) d^L x - \int x_j \Phi_1 (\mathbf{x}) d^L x \int x_k \Phi_1 (\mathbf{x}) d^L x$$
$$= -\alpha_j^2\alpha_k^2 / (V_s + V_{g,0}), \quad j \neq k. \quad (14)$$

These three results hold for arbitrary L and V_s^{-1}, the latter being a measure of the strength of selection.

Let Var_τ denote the variance of a quantity in generation τ and E_τ the corresponding expectation. Then since $\alpha_j^2 = \text{Var}_0(X_j)$ is the variance of X_j in generation 0, we can write (14) as

$$E_1 (X_j X_k) - E_1 (X_j)E_1 (X_k) =$$
$$- \text{Var}_0 (X_j) \text{Var}_0 (X_k) / (V_s + V_{g,0}). \quad (15)$$

This quantity is a measure of pairwise linkage disequilibrium. We shall compare this with an approximate result in a later section.

5. Strength of selection

Let us estimate the strength of selection in the above model when loci are unlinked. Without loss of generality, quantities may be scaled so $V_e = 1$. With this scaling, typical values of the parameters m and V_s as estimated by Turelli (1984), are in the vicinity of $V_s = 20$ and $m = 0.2$. We furthermore take $\mu = 10^{-5}$ as a typical allelic mutation rate (Griffiths et al., 1996). The typical values of taken for V_s, m and μ indicate that $\mu V_s/m^2 \ll 1$ and under this condition a "House of Cards" approximation applies (Turelli, 1984). This approximation is related to a mutation scheme of Kingman (1978). The essence of the approximation is that the distribution of allelic effects is much narrower than the typical change in allelic effect caused by a mutation. Let E denote the expectation operator for equilibrium. The mean equilibrium genotypic value coincides with the fitness optimum: $E[G] = \theta$. Furthermore, the mean fitness of the population is $E[\exp(-(G - \theta)^2/(2V_s))]$ and expanding the exponent yields

$$E[\exp(-(G-\theta)^2 / (2V_s))]$$

$$= 1 - V_G / (2V_s) + E[(G - E[G])^4] / (8V_s^2) + \dots \quad (16)$$

where $V_G \overset{\text{def}}{\equiv} E[(G - E[G])^2]$ is the equilibrium genetic variance. Results from Appendix A, namely (A2) and (A3), yield $V_G/(2V_s) \approx 2L\mu$ and $E[(G - E[G])^4] / (8V_s^2) \approx (2L\mu)m^2 / (4V_s) + (3/2)(2L\mu)^2 (1 - 1/(2L))$. When the ratio of quartic to quadratic moments in (16) is small, namely when $m^2 / (4V_s) \ll 1$ and $2L\mu \ll 1$, selection is weak and the fitness of individuals with genotypic value $\sum_{i=1}^{L} (x_i + y_i)$ is

$$w(\mathbf{x}, \mathbf{y}) \approx 1 - [\sum_{i=1}^{L} (x_i + y_i) - \theta]^2 / (2V_s). \quad (17)$$

6. Approximation of the dynamical equations when selection is weak

Assuming selection is weak in typical dynamical situations, as well as in equilibrium, allows us to write $w(\mathbf{x}, \mathbf{y}) \approx 1 - s(\mathbf{x}, \mathbf{y})$ where

$$s(\mathbf{x}, \mathbf{y}) = [\sum_{i=1}^{L} (x_i + y_i) - \theta]^2 / (2V_s). \quad (18)$$

For unlinked loci under the "House of Cards" approximation, $s(\mathbf{x}, \mathbf{y})$ has an approximate mean equilibrium value of $2L\mu$.

An approximate analysis of (8) when $s(\mathbf{x}, \mathbf{y})$, (or its mean value) and μ are treated as small quantities of comparable magnitude can be obtained by neglecting terms of order $s(\mathbf{x}, \mathbf{y}) \times s(\mathbf{x}, \mathbf{y})$, $s(\mathbf{x}, \mathbf{y}) \times \mu$ and $\mu \times \mu$ along with higher order terms.

To begin the approximate analysis, expand $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$, of (8), to linear order in μ:

$$K(\mathbf{x}; \mathbf{u}, \mathbf{v}) \approx K(\mathbf{x}; \mathbf{u}, \mathbf{v})|_{\mu=0} + \mu \frac{\partial K(\mathbf{x}; \mathbf{u}, \mathbf{v})}{\partial \mu}\bigg|_{\mu=0}$$

$$\overset{\text{def}}{\equiv} K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) + \mu K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}). \quad (19)$$

Appendix C is devoted to the functions $K_0(\mathbf{x}; \mathbf{u}, \mathbf{v})$ and $K_1(\mathbf{x}; \mathbf{u}, \mathbf{v})$ and some important properties of these are given in (C2) and (C3).

Equation (8), on expansion to the stated order, takes the form

$$\Phi_{\tau+1}(\mathbf{x}) \approx \int K_0(\mathbf{x}; \mathbf{u}, \mathbf{v})\{1 - [s(\mathbf{u}, \mathbf{v}) - \bar{s}_\tau]\}$$

$$\times \Phi_\tau(\mathbf{u})\Phi_\tau(\mathbf{v}) d^L u\, d^L v$$

$$+ \mu \int K_1(\mathbf{x}; \mathbf{u}, \mathbf{v})\Phi_\tau(\mathbf{u})\Phi_\tau(\mathbf{v}) d^L u\, d^L v \quad (20)$$

where $\bar{s}_\tau = \int s(\mathbf{x}, \mathbf{y})\Phi_\tau(\mathbf{x})\Phi_\tau(\mathbf{y}) d^L x\, d^L y$.

Equation (20) can be written in a form where the detailed significance of different terms can be readily appreciated. We achieve this by separating the selective difference, $s(\mathbf{u}, \mathbf{v}) - \bar{s}_\tau$ into various components related to their influence.

7. Decomposition of the selection coefficient

Selection in the diploid stage of the lifecycle has a number of different effects that arise because selective deaths, due to genotypic differences, result in changes in the genetic constitution of a population. The most obvious of these results in changes in allele frequencies of non-equilibrium populations. In addition there are changes in correlations with the following origins.

(i) At the start of a generation, immediately after random union of gametes, alleles of different parental origin are uncorrelated and at this point the population is in Hardy Weinberg equilibrium. During the

course of a generation, selective deaths cause correlations between the alleles of different parental origin that exist in surviving individuals. Via recombination, some of these alleles may end up in the same gamete and as a result, correlations of alleles in gametes are generated (this is one source of linkage disequilibrium). These correlations may then be passed on to the next generation.

(ii) At the start of a generation, newly formed individuals, containing alleles of the same parental origin, may already contain correlations. Selective deaths may generate additional correlations between these alleles. Some of these alleles may also end up in the same gamete. Selection may thus generate additional correlations between alleles in gametes and these will then affect the next generation (this is another source of linkage disequilibrium).

We shall shortly see which terms in the selection coefficient are responsible for changes in frequency or production of linkage disequilibrium.

Part of the selection coefficient is comprised of a sum of contributions from the separate alleles present in an individual and is written as $s_g(\mathbf{u}, \mathbf{v})$. Generally, this causes changes in gene frequencies. In the language used by Bulmer (1989) when discussing genetic variance, $s_g(\mathbf{u}, \mathbf{v})$ is a "genic" contribution, obtained by summing contributions from individual genes.

The part of the selection coefficient that produces correlations between alleles in gametes originating in a single parent is written as $s_L(\mathbf{u}, \mathbf{v})$. This term is responsible for the generation of linkage disequilibrium, and is termed a "linkage" contribution.

Using the notation

$$\overline{G}_\tau \overset{\text{def}}{\equiv} E_\tau[G], \qquad \overline{X}_{i,\tau} \overset{\text{def}}{\equiv} E_\tau[X_i]$$

$$\delta_\tau u_i \overset{\text{def}}{\equiv} u_i - \overline{X}_{i,\tau}, \qquad \delta_\tau v_i \overset{\text{def}}{\equiv} v_i - \overline{X}_{i,\tau} \qquad (21)$$

we find we can write

$$s(\mathbf{u}, \mathbf{v}) - \overline{s}_\tau = [s_g(\mathbf{u}, \mathbf{v}) - \overline{s}_{g,\tau}]$$

$$+ [s_L(\mathbf{u}, \mathbf{v}) - \overline{s}_{L,\tau}] \qquad (22)$$

where

$$s_g(\mathbf{u}, \mathbf{v}) = \frac{1}{2V_s} \sum_i (u_i + \overline{G}_\tau - \overline{X}_{i,\tau} - \theta)^2 +$$

$$+ \frac{1}{2V_s} \sum_i (v_i + \overline{G}_\tau - \overline{X}_{i,\tau} - \theta)^2 \qquad (23)$$

$$s_L(\mathbf{u}, \mathbf{v}) = \frac{1}{2V_s} \sum_{i,j(i \neq j)} (\delta_\tau u_i \delta_\tau u_j + \delta_\tau v_i \delta_\tau v_j) +$$

$$+ \frac{1}{2V_s} \sum_{i,j} 2\delta_\tau u_i \delta_\tau v_j \qquad (24)$$

$$\overline{s}_{g,\tau} = \int s_g(\mathbf{u}, \mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v,$$

$$\overline{s}_{L,\tau} = \int s_L(\mathbf{u}, \mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L x d^L y. \qquad (25)$$

The form of the genic contribution, $s_g(\mathbf{u}, \mathbf{v})$, may be appreciated if it is noted that $\overline{G}_\tau - \overline{X}_{i,\tau}$ is the averaged effect of all alleles controlling the trait, with the exclusion of the effect of the paternal allele at locus i. Thus $[u_i + \overline{G}_\tau - \overline{X}_{i,\tau} - \theta]^2 / (2V_s)$ corresponds to the selection coefficient against individuals carrying a paternal allele at locus i with effect u_i, when all other alleles have their mean effects and a quadratic approximation to the selection coefficient applies.

It is not immediately obvious that the linkage contribution, $s_L(\mathbf{u}, \mathbf{v})$, is a term that generates linkage disequilibrium. We shall shortly show, however, that it does indeed have this property.

8. Weak selection dynamics

The approximate dynamical equation (20), can now be written as

$$\Phi_{\tau+1}(\mathbf{x}) \approx \int K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v$$

$$+ \int \{\mu K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) - K_0(\mathbf{x}; \mathbf{u}, \mathbf{v})[s_g(\mathbf{u}, \mathbf{v}) - \overline{s}_{g,\tau}]\}$$

$$\times \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v$$

$$\int K_0(\mathbf{x}; \mathbf{u}, \mathbf{v})[s_L(\mathbf{u}, \mathbf{v}) - \overline{s}_{L,\tau}] \Phi_\tau(\mathbf{u})$$

$$\Phi_\tau(\mathbf{v}) d^L u d^L v. \qquad (26)$$

We shall sometimes use a shorthand notation where terms in this equation are written as

$$\Phi_{\tau+1} = \int K_0 \Phi_\tau \Phi_\tau + \int (\mu K_1 - K_0[s_g - \overline{s}_g]) \Phi_\tau \Phi_\tau$$

$$- \int K_0[s_L - \overline{s}_L] \Phi_\tau \Phi_\tau.$$

8.1. Interpretation of terms

(i) The term $\int K_0 \Phi_\tau \Phi_\tau$ contains only effects of recombination [see (C4) for K_0]. Thus if all terms on

the right-hand side of (26), with the exception of $\int K_0 \Phi_\tau \Phi_\tau$, were absent, the linkage disequilibrium between alleles at locus i and locus j would decrease by a factor of $(1 - r_{ij})$ each generation, where r_{ij} is the recombination fraction between these loci. Thus for $i \neq j$, the term $\int K_0 \Phi_\tau \Phi_\tau$ yields, with $C_{ij}(\tau) = = E_\tau(X_i X_j) - E_\tau(X_i)E_\tau(X_j)$,

$$C_{ij}(\tau + 1) = (1 - r_{ij})C_{ij}(\tau)$$

(result of recombination term).   (27)

This result may be derived directly from (26) by multiplying (26) by either x_i or $x_i x_j$, integrating over $\mathbf{x}$, and using the properties of $K_0(\mathbf{x}; \mathbf{u}, \mathrm{v})$ derived in Appendix C.

(ii) The term $\int(\mu K_1 - K_0[s_g - \bar{s}_g])\Phi_\tau \Phi_\tau$ contains two parts, one of which carries a factor of the mutation rate, μ. It follows that $\int \mu K_1 \Phi_\tau \Phi_\tau$ is responsible for the mutation of individual alleles.

The term $\int(\mu K_1 - K_0[s_g - \bar{s}_g])\Phi_\tau \Phi_\tau$ also contains the part $\int K_0 s_g \Phi_\tau \Phi_\tau$ which is the genic selective force that consists of a sum of $2L$ selection coefficients against the $2L$ alleles controlling the trait [see the form of s_g in (23)].

It follows that $\int(\mu K_1 - K_0[s_g - \bar{s}_g])\Phi_\tau \Phi_\tau$ contains opposing evolutionary forces that are each a sum of terms from individual alleles.

(iii) The term $\int K_0 s_L \Phi_\tau \Phi_\tau$ is unique in that it is the only term that generates correlations between alleles that end up in a single gamete. In (24), s_L is split into two sums, the first of which is responsible for generating correlations between alleles originating in just one of the parents of the individual producing the gamete. The second sum in (24) generates correlations between alleles originating in different parents of the individual producing the gamete, but which, due to recombination, find themselves in the same gamete. Thus $\int K_0 s_L \Phi_\tau \Phi_\tau$ is responsible for the production of linkage disequilibrium. To see this explicitly, consider a situation of complete linkage equilibrium in generation τ : $C_{ij}(\tau) = 0$. In Appendix E it is shown that the next generation, when $i \neq j$,

$$C_{ij}(\tau + 1) \approx -\mathrm{Var}_\tau(X_i)\mathrm{Var}_\tau(X_j)/V_s$$

(result of the linkage term).   (28)

This result can be compared with the exact, but somewhat specialized result of (15) which was de-

rived when, in generation 0, the population was in linkage equilibrium with normally distributed allelic effects. By comparison, it follows that (28) is a valid approximation when $2\sum_{i=1}^{L} \mathrm{Var}_\tau(X_i) \ll V_s$.

8.2. Dynamics of correlations

We shall assume that all recombination fractions between loci, r_{ij}, are large compared with the allelic mutation rate μ. Thus if $r_{\min}$ is the smallest recombination fraction between loci controlling the trait: $r_{\min} = \min_{i,j}(r_{ij})$, $(i \neq j)$, we assume

$$r_{\min} \gg \mu .   (29)$$

A simple picture of the dynamics of correlations then emerges. This comes about since there are then two distinct timescales in the problem. There is the short timescale $\sim r_{\min}^{-1}$, associated with recombination processes. The other timescale is longer, very roughly of order μ^{-1}, and is associated with changes caused by mutation and selection processes.

To obtain evidence for this two-timescale picture, we combine the results for correlations, (27) and (28), namely

$$C_{ij}(\tau + 1) = E_{\tau+1}[X_i X_j] - E_{\tau+1}[X_i]E_{\tau+1}[X_j]$$

$$\approx (1 - r_{ij})C_{ij}(\tau) - \mathrm{Var}_\tau(X_i)\mathrm{Var}_\tau(X_j)/V_s, \ i \neq j.   (30)$$

Combining results is justified if linkage disequilibrium remains small and this is automatically satisfied if $2\sum_{i=1}^{L}\mathrm{Var}_\tau(X_i) \ll V_s$, in which case $\int K_0 s_L \Phi_\tau \Phi_\tau$ only affects, to leading order, the assumed very large linkage equilibrium aspect of $\Phi_\tau(\mathbf{x})$. The solution of (30) is

$$C_{ij}(\tau) \approx (1 - r_{ij})^\tau C_{ij}(0) - \sum_{n=0}^{\tau-1}(1 - r_{ij})^n$$

$$\times \mathrm{Var}_{\tau-n-1}(X_i)\mathrm{Var}_{\tau-n-1}(X_j)/V_s.   (31)$$

The correlations associated with linkage disequilibrium therefore change because of recombination [which causes the various powers of $(1 - r_{ij})$ in (31)] and also because $\mathrm{Var}_\tau(X_i)$ depends on τ, so there are dynamical changes in the variance of allelic effects.

8.3. Dynamics of allelic effects

The result given in (31) requires the variance of effects of alleles (of e.g. paternal origin), $\text{Var}_\tau(X_i)$, at different loci and for different τ. The assumption we have previously made, that linkage disequilibrium is small, allows the calculation of the allelic variances under the approximation of linkage equilibrium. At locus i the distribution of effects of alleles of paternal origin, is, in generation τ, defined by

$$\psi_{i,\tau}(x_i) = \int \delta(x_i - y_i)\Phi_\tau(\mathbf{y})d^L y$$
$$= \int \Phi_\tau(\mathbf{x})dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_L. \quad (32)$$

The equation obeyed by $\psi_{i,\tau}(x_i)$ follows by multiplying (26), for $\Phi_{\tau+1}(\mathbf{y})$, by $\delta(x_i - y_i)$ and integrating over all y. After some work we obtain

$$\psi_{i,\tau+1}(x_i) =$$

$$= \psi_{i,\tau}(x_i) + \mu[\int f(x_i - u)\psi_{i,\tau}(u)du - \psi_{i,\tau}(x_i)]$$

$$-[S_i(x_i) - \overline{S}_{i,\tau}]\psi_{i,\tau}(x_i)$$

$$+ \text{linkage disequilibrium contributions} \quad (33)$$

where

$$S_i(x_i) \overset{\text{def}}{=} (x_i + \overline{G}_\tau - \overline{X}_{i,\tau} - \theta)^2 / (2V_s) \quad (34)$$

$$\overline{S}_{i,\tau} \overset{\text{def}}{=} \int S_i(x_i)\psi_{i,\tau}(x_i)dx_i. \quad (35)$$

The linkage disequilibrium contributions in equation (33) may, in situations close to equilibrium, be estimated to be of order $2L\mu$ smaller than other terms present and is thus neglectable when $2L\mu \ll 1$.

Equation (33) is a dynamical equation that determines the change of $\psi_{i,\tau}(x_i)$ with τ. With the neglect the linkage disequilibrium contributions, (33) has the form associated with a one locus haploid problem where $S_i(x_i)$ is the selection coefficient against alleles of effect x_i. We can solve the resulting equation for $\psi_{i,\tau}(x_i)$ and hence can determine $\text{Var}_\tau(X_i) = \int(x_i - \overline{X}_{i,\tau})^2\psi_{i,\tau}(x_i)dx_i$.

8.4. Considerations of general dynamics

Equation (26) contains the general dynamics of the distribution function $\Phi_\tau(\mathbf{x})$ under conditions of weak selection. We shall not consider the general dynamics in any appreciable detail but shall simply point out that equations may be obtained from (26) that relate quantities in generation $\tau + 1$ to quantities in generation τ. For example, the mean genotypic value in generation $\tau + 1$, $E_{\tau+1}[G] = E_{\tau+1}[\sum_i(X_i + Y_i)] \equiv 2E_{\tau+1}[\sum_i X_i]$, may be obtained by multiplying (26) by $\sum_i x_i$ and integrating over all x_i. Details of the calculation are given in Appendix F. One virtue of the analysis is that the dynamical origin of the different terms, whether from genic or linkage parts of selection, is immediately known.

9. Approach to equilibrium

From (31) and (33) we can infer the following picture of the approach to equilibrium.

(i) On short times, of order $r_{\min}^{-1}$, all correlations associated with linkage disequilibrium settle down to a quasi-equilibrium value, given by the balance between recombination and linkage disequilibrium generating aspects of selection. During changes over time-intervals small compared with μ^{-1} we may imagine that genic changes (associated with changes in individual allele frequencies) vary so little that they may be treated as a constant. Thus when $\mu^{-1} \gg \tau \gg r_{\min}^{-1}$ we can replace the sum in (31) by

$$\sum_{n=0}^{\tau-1}(1 - r_{ij})^n \text{Var}_{\tau-n-1}(X_i)\text{Var}_{\tau-n-1}(X_j)/V_s \approx$$

$\text{Var}_\tau(X_i)\text{Var}_\tau(X_j)/(r_{ij}V_s)$. Thus when $\mu^{-1} \gg \tau \gg r_{\min}^{-1}$

$$E_\tau(X_iX_j) - E_\tau(X_i)E_\tau(X_j)$$

$$\approx -\text{Var}_\tau(X_i)\text{Var}_\tau(X_j)/(r_{ij}V_s). \quad (36)$$

The rapidity with which recombination causes the quasi-equilibrium result of (36) to be achieved (namely over $\sim r_{\min}^{-1}$ generations) indicates recombination is generally a very strong genetic force. The equilibrium value of the right-hand side of (36), which, under the "House of cards approximation," is

of order $\mu^2 V_s / r_{\min}$, indicates that recombination produces a high degree of linkage equilibrium.

(ii) After the relatively rapid "recombination" generated approach to quasi-equilibrium, genic effects undergo their approach to equilibrium on the longer timescale of $\sim \mu^{-1}$. Over this longer timescale, (36) indicates that correlations associated with linkage disequilibrium are controlled by (or are slaves of) the changes in genic quantities.

10. Equilibrium genetic variance

In equilibrium, quantities do not have any τ dependence and the genetic variance is given by $V_G \stackrel{\text{def}}{=} \text{Var}[G] = \text{Var}[\sum_{i=1}^{L} (X_i + Y_i)]$. This can be split into the following parts (Bulmer, 1989), $V_G = V_g + C_{HW} + C_L$ where

$$V_g = 2\sum_{i=1}^{L} \text{Var}(X_i) \tag{37}$$

$$C_{HW} = 2\sum_{i=1}^{L}\sum_{j=1}^{L} (E[X_i Y_j] - E[X_i]E[Y_j]) \tag{38}$$

$$C_L = 2\sum_{\substack{i=1 \\ (i \neq j)}}^{L}\sum_{j=1}^{L} (E[X_i X_j] - E[X_i]E[X_j]) \tag{39}$$

V_g arises from contributions of individual genes and is called the genic variance (Bulmer, 1989). In the absence of linkage disequilibrium, the genetic variance coincides with the genic variance.

C_{HW} is a Hardy Weinberg disequilibrium term that arises from correlations between pairs of alleles originating in different parents. Since C_{HW} is calculated immediately after random union of gametes, alleles originating in different parents are statistically independent and $C_{HW} = 0$.

C_L arises from linkage disequilibrium, i.e. from statistical correlations between pairs of alleles at different loci, the alleles having the same parental origin.

To calculate these quantities, we proceed by first calculating equilibrium quantities associated with alleles at a single location within a locus (each locus, being diploid, has two locations where alleles can reside) and then feed these into the results already obtained for linkage disequilibrium correlations. In this way we are able to determine the genetic variance.

10.1. Genic variance

We first calculate the variance allelic of effects. The selection function, $S_i(x_i)$, of (34) may be simplified since equilibrium entails $\bar{G} = \theta$. Thus (33) may be written as

$$\left[\frac{(x_i - \bar{X}_i)^2}{2V_s} - \frac{1}{2V_s}\int (u - \bar{X}_i)^2 \psi_i(u)du + \mu\right]\psi_i(x_i)$$
$$-\mu\int f(x_i - u)\psi_i(u)du \approx 0. \tag{40}$$

As Lande (1976) pointed out, the $\bar{X}_i$'s in equilibrium need not be zero. Apart from the constraint $2\sum_{i=1}^{L} \bar{X}_i = \theta$, they have arbitrary values which are ultimately determined by initial conditions. Let $\Psi(x_i)$ be the equilibrium solution of (40) when all $\bar{X}_i = 0$ (so $\Psi(x_i)$ is real, positive, symmetric and $\int \Psi(x_i)dx_i = 1$). Then for all i, $\psi_i(x_i) = \Psi(x_i - \bar{X}_i)$. We can thus write the linkage equilibrium approximation to $\Phi(\mathbf{x})$ as $\prod_{i=1}^{L} \Psi(x_i - \bar{X}_i)$ where $2\sum_{i=1}^{L} \bar{X}_i = \theta$, and $\Psi(x)$ obeys $([x^2 - \int u^2 \Psi(u)du]/(2V_s) + \mu) \times \Psi(x) - \mu\int f(x-u)\Psi(u)du \approx 0$. Under the assumptions about parameters that we have made, $\Psi(x)$ may quite accurately be determined in the "House of Cards Approximation" (Turelli, 1984), which approximates $\int f(x-u)\Psi(u)du$ by $f(x)$ because $\Psi(x)$ is appreciable only over a small range compared with m (the standard deviation of mutant effects). The net result is $\Psi(x) \approx 2\mu V_s f(x)/(x^2 + \lambda^2)$ where $\lambda \approx \sqrt{2\pi}\mu V_s / m$. This distribution has a mean of zero and variance $\int x^2 \Psi(x)dx \approx 2\mu V_s$. We thus have, for all i,

$$\text{Var}(X_i) \approx 2\mu V_s \tag{41}$$

which is the result found by Turelli (1984).

With (36) and (41) we have all of the ingredients needed to calculate the equilibrium genetic variance.

In particular, using (41), the genic contribution to the genetic variance is

$$V_g = 2\sum_{i=1}^{L} \text{Var}(X_i) \approx 4L\mu V_s.$$

10.2. Linkage contribution

Using (36) and (39) we obtain an equilibrium linkage contribution of

$$C_L = 2 \sum_{\substack{i=1 \\ (i \neq j)}}^{L} \sum_{i=1}^{L} C_{ij}$$

$$\approx -2 \sum_{\substack{i=1 \\ (i \neq j)}}^{L} \sum_{i=1}^{L} \text{Var}(X_i)\text{Var}(X_j)/(r_{ij}V_s) \quad (42)$$

Let us define

$$\left\langle \frac{1}{r_{ij}} \right\rangle = \frac{\sum_{\substack{i=1 \\ (i \neq j)}}^{L} \sum_{j=1}^{L} \frac{1}{r_{ij}}}{\sum_{\substack{i=1 \\ (i \neq j)}}^{L} \sum_{j=1}^{L} 1}$$

$$= \frac{1}{L(L-1)} \sum_{i=1}^{L} \sum_{\substack{j=1 \\ (i \neq j)}}^{L} \frac{1}{r_{ij}}. \quad (43)$$

Then using $\text{Var}(X_i) = V_g/(2L)$, we can write (42) as

$$C_L \approx -\frac{1}{2} \frac{V_g^2}{V_s} \left\langle \frac{1}{r_{ij}} \right\rangle \left(1 - \frac{1}{L}\right) \quad (44)$$

[cf. the result found by Bulmer (1974) for the infinitesimal model].

10.3. Genetic variance

From (44) we obtain $V_G \approx V_g + C_L$ i.e.

$$V_G \approx V_g \left[1 - \frac{1}{2} \frac{V_g}{V_s} \left\langle \frac{1}{r_{ij}} \right\rangle \left(1 - \frac{1}{L}\right) \right]. \quad (45)$$

This equation closely corresponds with Bulmer's equilibrium result for an infinite population [e.g. equation (8) of Bulmer (1989)] when linkage is present and an infinite number of equivalent loci with discrete alleles exist. Note that at no point have we based any approximations on the value of the number of loci, L, thus (45) applies for $L = 2, 3, ...$, and indeed any number of loci provided the population lies in a regime of weak selection.

11. Numerical results and discussion

The results we have presented indicate that much of the dynamics as well as equilibrium properties can be explained from knowledge of the distributions of allelic effects at the $2L$ different "haploid" locations

in the genome. In an L locus model, one such distribution of allelic effects is determined by assuming that alleles at one particular location exist in a genetic background consisting of the averaged effects of alleles at the remaining $2L - 1$ locations [see Equations (33) and (34)]. Once the distributions are known, they may be used to calculate the allelic variance, $\text{Var}_\tau(X_i)$, and from this determine the genic variance in a non-equilibrium situation:

$$V_g(\tau) = 2 \sum_{i=1}^{L} \text{Var}_\tau(X_i). \quad (46)$$

The allelic variance may also be used in (31) to determine the pairwise correlations between alleles and from this the non-equilibrium linkage contribution to the genetic variance may be obtained.

$$C_L(\tau) = 2 \sum_{\substack{i=1 \\ (i \neq j)}}^{L} \sum_{j=1}^{L} C_{ij}(\tau)$$

$$\approx 2 \sum_{\substack{i=1 \\ (i \neq j)}}^{L} \sum_{j=1}^{L} [(1 - r_{ij})^\tau C_{ij}(0)$$

$$- \sum_{n=0}^{\tau-1} (1 - r_{ij})^n \text{Var}_{\tau-n-1}(X_i)\text{Var}_{\tau-n-1}(X_j)/V_s] \quad (47)$$

and the genetic variance is simply the sum of $V_g(\tau)$ and $C_L(\tau)$.

An inspection of the approximations made indicates that the key assumptions are that selection is weak and that the smallest relevant recombination fraction, $r_{\min}$, is much larger than the allelic mutation rate μ. It thus follows that, with the necessary changes made, the results derived should also apply when allelic effects are discrete and not continuous. It may also be observed that the number of loci is not a key feature of the validity of the results (this is in sharp contrast to the infinitesimal model, which relies on the number of loci approaching infinity). This allows us to perform a numerical test on an L locus, biallelic model. We have taken the allelic effects in this model to be 0 and 1 and this amounts to having a variance in mutant effects of $m^2 \sim 1$, which is larger than that of the typical value ($m^2 \sim 0.04$). Accordingly we have inflated V_s to the value $V_s = 400$, so the value of m^2/V_s is comparable with the ratio when m^2 and V_s have typical values. Furthermore the results presented apply for $r_{\min} \gg \mu$ (29) and we have taken $\mu = 10^{-5}$. For the simulations, we have assumed all L loci run contiguously and have taken the recombination fraction between adjacent loci, r, to be ≥ 0.05. Other recombination fractions have been calculated in the absence of interference.

We have numerically solved the exact difference equations describing the gamete frequencies for this model for various numbers of loci, assuming the special case of linkage equilibrium at $\tau = 0$ [i.e. $C_L(0) = 0$]. The dynamical behaviour arising from the ex-

act difference equation is contrasted with the dynamical predictions of this work.

As a particular example, in Figures 1A and 1B, we plot the genic contribution to the genetic variance, V_g, for a recombination fraction between adja-
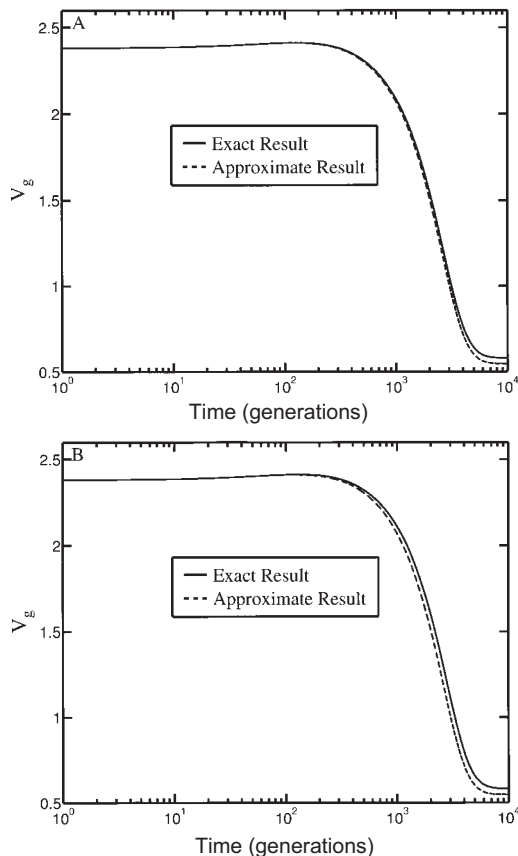

FIG. 1. Genic contribution to the genetic variance

FIG. 2. Linkage contribution to the genetic variance

In Figure 1A, the genic contribution to the genetic variance, V_g, is plotted against the number of generations, τ. The exact result (continuous line) is obtained from iteration of the exact equations determining the dynamics of the population. The approximate result (dashed line) is obtained by: (i) numerically solving the equation for the "haploid" allele distribution at the $2L$ different locations, (33) and (ii) determining, from the distributions, the various allelic variances and using them in (46). For the purposes of the simulation, the trait was determined from the additive effects of 6 biallelic loci. Possible allelic effects at each locus are 0 and 1. A normalizing fitness function of the form (4) was adopted, where $\theta = 5$ and $V_s = 400$. The allelic mutation rate of $\mu = 10^{-5}$ was used. Initial frequencies of alleles with effect "1" at the 6 loci were (0.1, 0.2, 0.3, 0.4, 0.5, 0.6) and all alleles were in linkage equilibrium in the initial generation. All 6 loci were assumed contiguous and the recombination fraction between adjacent loci was $r = 0.5$. Other recombination fractions have been calculated in the absence of interference.

In Figure 1B, V_g is plotted when the recombination fraction between adjacent loci is $r = 0.05$ and all other quantities are identical to those used in Figure 1A.

In Figure 2A, the linkage contribution to the genetic variance, C_L, is plotted against the number of generations, τ. The exact result (continuous line) is obtained from iteration of the exact equations determining the dynamics of the population. The approximate result (dashed line) is obtained by: (i) numerically solving the equation for the "haploid" allele distribution at the $2L$ different locations, (33) and (ii) determining, from the distributions, the various allelic variances and using them in (47). For the purposes of the simulation, the trait was determined from the additive effects of 6 biallelic loci. Possible allelic effects at each locus are 0 and 1. A normalizing fitness function of the form (4) was adopted, where $\theta = 5$ and $V_s = 400$. The allelic mutation rate of $\mu = 10^{-5}$ was used. Initial frequencies of alleles with effect "1" at the 6 loci were (0.1, 0.2, 0.3, 0.4, 0.5, 0.6) and all alleles were in linkage equilibrium in the initial generation. All 6 loci were assumed contiguous and the recombination fraction between adjacent loci was $r = 0.5$. Other recombination fractions have been calculated in the absence of interference.

In Figure 2B, C_L is plotted when the recombination fraction between adjacent loci is $r = 0.05$ and all other quantities are identical to those used in Figure 2A.
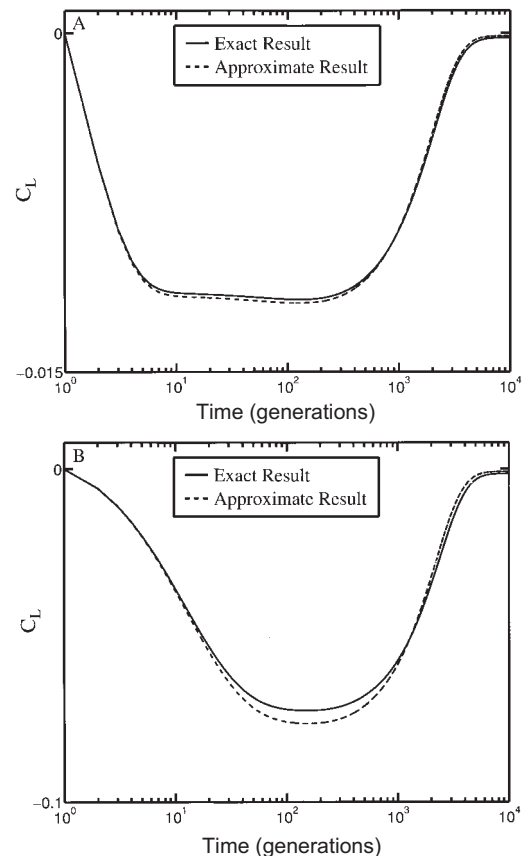
cent loci of $r = 0.5$ and $r = 0.05$, respectively. The approximate result is calculated from the "haploid" approximation used in this work and it is compared with the exact result obtained by iteration of the exact equations governing the population.

In Figures 2A and 2B we plot, for $r = 0.5$ and $r = 0.05$, respectively, the linkage contribution to the genetic variance, C_L, as calculated from the approximation of this work (47), and compare it with the exact results.

It is evident from the Figures that the genic variance, V_g, is little affected by linkage and indeed stays almost constant for of order 10^2 generations. By contrast the linkage contribution to the genetic variance, C_L, changes appreciable on a significantly smaller timescale and is substantially affected by the degree of linkage.

We note that the approximate dynamical results in Figures are derived from knowledge of variances of allelic effects at different locations and these were calculated from the dynamics of a single haploid locus in a background comprising the averaged effects of alleles at the remaining $2L - 1$ locations. Thus the approximate calculations require no knowledge of how the alleles are associated together in haploid or diploid stages of the life cycle.

C_L following from (44) is not accurate if it is calculated from the variance of allelic effects of a haploid locus in an averaged genetic background. The lack of accuracy of (44) stems not from (44) itself, but from numerical errors in those approximate allele frequencies that are close to 0 or 1. We have thus calculated the equilibrium linkage contribution to the genetic variance, C_L, from (44), when the equilibrium allele frequencies produced by the exact difference equation for the model are used. Below we give equilibrium values of C_L, as calculated from (44), along with results arising from the exact dynamics. In all cases, the optimal genotypic value $G_{opt} = 5$, $V_s = 400$ and $\mu = 10^{-5}$, and iterations were terminated when gamete frequencies changed by less than 0.5×10^{-5} in one generation.

Some environments do may not remain stationary for 10^4 generations (the number of generations the numerical studies were made for in this work). Changing environments have, however been considered elsewhere (see Waxman and Peck, 1999; Burger, 1999 and references therein).

The results presented in this work deal with the most basic situation of a population's approach to equilibrium, or the equilibrium itself. We believe it should be possible to extend the results to a variety

Recombination fraction between adjacent loci, $r = 0.5$

L	2	3	4	5	6
C_L (exact)	-1.21×10^{-5}	-9.23×10^{-4}	-1.62×10^{-3}	-2.55×10^{-3}	-3.17×10^{-3}
C_L (approx.)	-1.22×10^{-5}	-9.32×10^{-4}	-1.70×10^{-3}	-2.69×10^{-3}	-3.25×10^{-3}

Recombination fraction between adjacent loci, $r = 0.05$

L	2	3	4	5	6
C_L (exact)	-1.03×10^{-4}	-7.67×10^{-3}	-1.23×10^{-2}	-1.74×10^{-2}	-1.78×10^{-2}
C_L (approx.)	-1.23×10^{-4}	-7.95×10^{-3}	-1.32×10^{-2}	-1.85×10^{-2}	-1.97×10^{-2}

While the dynamical results produced in this work appear to be reasonably accurate over a number of generations, we have not been able to put any bounds on the error of the value of dynamical quantities.

In addition to the dynamical results, we have also investigated the equilibrium result for the linkage contribution to the genetic variance, C_L, which is given in (44). For the mutation rates we have considered, equilibrium is only achieved after relatively long times ($\sim 10^4$) generations. We have found that

of other situations from further analysis of the exact equation governing the population's dynamics (8).

D. WAXMAN

References

BARTON, N. H. and TURELLI, M. (1991): Natural and sexual selection on many loci. *Genetics* **127**:229–255.

BULMER, M. G. (1974): Linkage disequilibrium and genetic variability. *Genet. Res. Camb.* **23**:281–289.

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

BULMER, M. G. (1989): Maintenance of genetic variability by mutation-selection balance: A childs guide through the jungle. *Genome* **31**:761–767.

BÜRGER, R. (1993): Predictions of the dynamics of a polygenic character under directional selection. *J. Theor. Biol.* **162**:487–513.

BÜRGER, R. (1999): Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* **153**:1055–1069.

BÜRGER, R. and HOFBAUER, J. (1994): Mutation load and mutation selection-balance in quantitative genetic traits. *J. Math. Biol.* **32**:193–218.

CHRISTIANSEN, F. B. (1999): *Population Genetics of Multiple Loci*. Wiley, Chichester.

CROW, J. F. and KIMURA, M. (1964): The theory of genetic loads. *Proceedings of the XIth International Congress of Genetics* **2**:495–505.

FISHER, R. A. (1918): The correlation between relatives under the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**:399–433.

FLEMING, W. (1979): Equilibrium distributions of continuous polygenic traits. *SIAM J. Appl. Math.* **36**:148–168.

GRIFFITHS, A. J. F., MILLER, J. H., SUZUKI, D.T., LEWONTIN, R. C. and GELBART, W. M. (1996): *An Introduction to Genetic Analysis*. W. H. Freeman, New York.

HALDANE, J. B. S. (1954): The measurement of natural selection. In *Proceedings of the 9th International Congress of Genetics.* Bellagio, Italy, August 1953. Edited by G. Montalenti and A. Chiarugi. Part 1, 480–487.

KARLIN, S. and LIEBERMAN, U. (1978): Classifications and comparisons of multilocus recombination distributions. *Proc. Natl. Acad. Sci. USA* **75**:6332–6336.

KEARSEY, M. J. and FARQUHAR, A. G. L. (1998): QTL analysis in plants; where are we now? *Heredity* **80**:137–142

KIMURA, M. (1965): A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proc. Natl. Acad. Sci. USA* **54**:731–736.

KINGMAN, J. F. C. (1978): A simple model for the balance between selection and mutation. *J. Appl. Prob.* **15**:1–12.

LANDE, R. (1976): The maintenance of genetic variation by mutation in a polygenic character with linked loci. *Genet. Res.* **26**:221–235.

NAGYLAKI, T. (1993): The evolution of multilocus systems under weak selection. *Genetics* **134**:627–647.

TURELLI, M. (1984): Heritable genetic variation via mutation selection balance: Lerche's zeta meets the abdominal bristle. *Theor. Popul. Biol.* **25**:138–193.

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

WAXMAN, D. and PECK, J. R. (1999): Sex and adaptation in a changing environment. *Genetics* **153**:1041–1053.

APPENDIX A

In this Appendix, we provide details of the calculation for moments of the distribution of genotypic values. We deal with the case of L unlinked loci, when linkage disequilibrium is neglected and the "House of Cards" approximation (Turelli, 1984) applies.

The genotypic value is $G = \sum_{j=1}^{L} (X_j + Y_j)$ where X_j (Y_j) is the effect, at locus j, of a paternally (maternally) inherited allele.

We deal here only with variances of X_j or Y_j. The variance of e.g. X_j is unaffected by adding a constant to X_j. We thus define X_j and Y_j (and hence G) so that their mean value vanishes $E[X_j] = E[(Y_j] = E[G] = 0$ where E denotes the expectation operator appropriate to equilibrium. Then the genetic variance is $V_G = E[G^2] = \sum_{j=1}^{L} \sum_{k=1}^{L} E[(X_j + Y_j)(X_k + Y_k)]$. The neglect of linkage disequilibrium means X_j and X_k are statistically independent if $j \neq k$. Furthermore, Hardy–Weinberg equilibrium entails X_j and Y_k being statistically independent for all j and k.

It follows that since all X_j and Y_j have the same probability distribution,

$$E[e^{i\lambda G}] = [\Omega(\lambda)]^{2L}, \qquad \Omega(\lambda) \overset{\text{def}}{\equiv} E[e^{i\lambda X_j}]. \quad (A1)$$

We can use (A1) to simply determine the variance and fourth moment of G about the origin: $E[G^2]$ and $E[G^4]$.

We write $\Omega'(0) = \left. \dfrac{d}{d\lambda} \Omega(\lambda) \right|_{\lambda=0}$, $\Omega''(0) = \left(\dfrac{d}{d\lambda} \right)^2 \left. \Omega(\lambda) \right|_{\lambda=0}$ etc. and use the result that the distribution of X_i is symmetric about $X_i = 0$ so all odd derivatives of $\Omega(\lambda)$ vanish at $\lambda = 0$. Then $V_G = E[G^2] = -2L \, \Omega''(0) = 2LE(X_j^2)$ and $E[G^4] = 2L \, \Omega''(0) + 6L(2L - 1) \times [\Omega''(0)]^2 = 2LE[X_j^4] + 6L(2L - 1)(E[X_j^2])^2$. The results of Turelli (1984): $E[X_j^2] \approx 2\mu V_s$, $E[X_j^4] \approx 2\mu V_s m^2$ then yield

$$V_G \approx 4L\mu V_s \qquad (A2)$$

$$E[G^4] \approx 4L\mu V_s m^2 + 12L^2 (1 - \frac{1}{2L})(2\mu V_s)^2$$
$$\approx V_G m^2 + 3\left(1 - \frac{1}{2L}\right)V_G^2. \qquad (A3)$$

APPENDIX B

In this Appendix we derive the equation (8) for the distribution of allelic effects $\Phi_\tau(\mathbf{x})$ in a continuum of alleles model. The key quantity in this equation is the function $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ that incorporates mutation and recombination. Before deriving the above equation we give a summary of the properties of $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$.

Summary of equations defining $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$

The equations that determine the key properties of $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ are

$$K(\mathbf{x};\mathbf{u},\mathbf{v}) = \sum_{\sigma} R(\sigma) M(\mathbf{x} - \xi) \tag{B1}$$

$$\sum_{\sigma} = \sum_{\sigma_1 = \pm 1} \sum_{\sigma_2 = \pm 1} \cdots \sum_{\sigma_L = \pm 1} \tag{B2}$$

$$\sum_{\sigma} R(\sigma) = 1, \quad \sum_{\sigma} \sigma_i R(\sigma) = 0,$$

$$\sum_{\sigma} \sigma_i \sigma_j R(\sigma) = 1 - 2r_{ij} \tag{B3}$$

$$M(\mathbf{x}) = \prod_{i=1}^{L} [(1-\mu)\delta(x_i) + \mu f(x_i)] \tag{B4}$$

$$\xi_i = (1 + \sigma_i) u_i / 2 + (1 - \sigma_i) v_i / 2,$$

$$(i = 1, 2, 3, \ldots, L) \tag{B5}$$

where $\delta(x)$ denotes a Dirac delta function of argument x.

We show, in Appendix C that despite the apparent complexity of $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$, it is straightforward to calculate properties of this function.

Two locus model

We motivate the result for the dynamical equations by starting with a two locus model. The lifecycle is schematically,

Gametes $\xrightarrow{\text{random union}}$ Zygotes $\xrightarrow{\text{viability selection}}$

Adults $\xrightarrow[\text{death of adults}]{\text{meiosis with recombination}}$ Gametes.

We denote alleles at locus 1 and 2 by A_i and B_j, respectively, where i and j are discrete labels. $n_{ij}(\tau)$ represents the frequency of an $A_i B_j$ gamete in generation τ (= 0, 1, 2, ...). Ordered genotypes are written as $A_i B_j / A_k B_l$ where the gamete $A_i B_j$ ($A_k B_l$) has a paternal (maternal) origin and the frequency of zygotes of this ordered genotype is $n_{ij}(\tau) n_{kl}(\tau)$. Let $w(ij, kl) \equiv w(kl, ij)$ be the fitness of these zygotes (proportional to viability since fertility is independent of phenotype). After selection, the frequency of $A_i B_j / A_k B_l$ adults is $\rho_{ijkl}(\tau)$ where

$$\rho_{ijkl}(\tau) \stackrel{\text{def}}{=} w(ok, kl) n_{ij}(\tau) n_{kl}(\tau) / \overline{w}_\tau,$$

$$\overline{w}_\tau = \sum_{ijkl} w(ij, kl) n_{ij}(\tau) n_{kl}(\tau). \tag{B6}$$

Loci are linked, with recombination fraction r_{12} and the frequencies of an $A_{p_1} B_{p_2}$ gamete at this point in the lifecycle, written as $n^*_{p_1 p_2}(\tau)$, is

$$n^*_{p_1 p_2}(\tau) = \frac{1}{2}(1 - r_{12}) \sum_{kl} \rho_{p_1 p_2 kl}(\tau)$$

$$+ \frac{1}{2}(1 - r_{12}) \sum_{ij} \rho_{ij p_1 p_2}(\tau) + \frac{r_{12}}{2} \sum_{jk} \rho_{p_1 jk p_2}(\tau)$$

$$+ \frac{r_{12}}{2} \sum_{il} \rho_{i p_2 p_1 l}(\tau). \tag{B7}$$

Mutations result in the frequency of an $A_{a_1} B_{a_2}$ gamete in generation $\tau + 1$ being given by

$$n_{a_1 a_2}(\tau + 1) = \sum_{p_1 p_2} M_{a_1 a_2, p_1 p_2} n^*_{p_1 p_2}(\tau) \tag{B8}$$

where $M_{a_1 a_2, p_1 p_2}$ contains information on the mutation scheme. With equally mutable loci and each allele mutating independently of all others, $M_{a_1 a_2, p_1 p_2}$ is given by $M_{a_1 a_2, p_1 p_2} = \prod_{i=1}^{2} [1 - \sum_s \mu_{spi}] \delta_{a_i p_i} + \mu_{a_i p_i}]$ where μ_{ji} is the probability that an allele i undergoes a mutation to allele j.

(B8) can be written in a compact form that readily allows generalization to more than two loci, and provides a practical tool for calculation of quantities for multiple loci. Introduce two independent variables σ_1 and σ_2 that can only take the values ± 1 and define

$$R(\sigma_1, \sigma_2) = \frac{1}{2} r_{12}^{|\sigma_2 - \sigma_1|/2} (1 - r_{12})^{1 - |\sigma_2 - \sigma_1|/2} \tag{B9}$$

where r_{12} is the recombination fraction between the two loci. $R(\sigma_1, \sigma_2)$ gives the probability that a gamete produced by an individual contains a copy of an allele of particular parental origin at locus 1 and 2. For example, $R(1, -1)$ is the probability that in a gamete, the allele at locus 1 is of paternal origin, and that at locus 2 is of maternal origin. Apart from nor-

malization $[\sum_{\sigma_1=\pm1}\sum_{\sigma_2=\pm1}R(\sigma_1,\sigma_2)=1]$, $R(\sigma_1,\sigma_2)$ may be simply verified to have the properties

$$\sum_{\sigma_1=\pm1}\sum_{\sigma_2=\pm1}\sigma_1 R(\sigma_1,\sigma_2)=0,\quad i=1,2 \qquad (B10)$$

$$\sum_{\sigma_1=\pm1}\sum_{\sigma_2=\pm1}\sigma_1\sigma_2 R(\sigma_1,\sigma_2)=1-2r_{12}. \qquad (B11)$$

Furthermore, let us define, for $i=1$ and $i=2$,

$$\alpha_i\equiv\alpha_i(\sigma_i,c_i,d_i)=(1+\sigma_i)c_i/2+(1-\sigma_i)d_i/2. \quad (B12)$$

We can then write (B8) as

$$n_{a_1 a_2}(\tau+1)=$$

$$=\frac{1}{\overline{w}_\tau}\sum_{\sigma_1=\pm1}\sum_{\sigma_2=\pm1}\sum_{c_1 c_2 d_1 d_2}R(\sigma_1,\sigma_2)M_{a_1 a_2,\alpha_1\alpha_2}$$

$$\times w(c_1 c_2,d_1 d_2)n_{c_1 c_2}(\tau)n_{d_1 d_2}(\tau). \quad (B13)$$

We write (B13) in a natural vector notation as

$$n_{\mathbf{a}}(\tau+1)=\sum_\sigma\sum_{\mathbf{c},\mathbf{d}}R(\sigma)M_{\mathbf{a},\alpha}w(\mathbf{c},\mathbf{d})$$

$$\times n_{\mathbf{c}}(\tau)n_{\mathbf{d}}(\tau)/\overline{w}_\tau. \qquad (B14)$$

where $\sum_\sigma=\sum_{\sigma_1=\pm1}\sum_{\sigma_2=\pm1}$ and the remaining sums cover the full range of possible allele labels.

Generalization to L linked loci

In the form given in (B14) it is straightforward to generalize the results to L linked loci which are labelled 1, 2, ..., L.

We define

$$\mathbf{a}=(a_1,a_2,...,a_L),\qquad \alpha=(\alpha_1,\alpha_2,...,\alpha_L),$$

$$\mathbf{c}=(c_1,c_2,...,c_L),\qquad \mathbf{d}=(d_1,d_2,...,d_L), \qquad (15)$$

with α_i defined by (B12) for i now ranging over $i=1$, 2, ..., L.

Additionally, let $\sigma=(\sigma_1,\sigma_2,...,\sigma_L)$, where each element of σ takes the values ±1. Then σ can achieve 2^L different values. The variable σ is the device we use to count all 2^L different gamete recombinants that are possible from one parent (the 2^L recombinants need not be distinct). A related device has been used in previous work (Karlin and Lieberman, 1978). The probability of a gametic output associated with a given σ is given by the generalization of

(B9) to L loci: a function we denote by $R(\sigma)$, with an analogous interpretation to the 2 locus case. An explicit form for $R(\sigma)$ can be determined in case of recombination without interference between loci. For the purposes of this work, however, we shall not need the detailed form of this function, only a knowledge of a few of its moments, analogous to normalization (B10) and (B11), namely $\sum_\sigma R(\sigma)=1$, $\sum_\sigma\sigma_i R(\sigma)=0$, $\sum_\sigma\sigma_i\sigma_j R(\sigma)=(1-2r_{ij})$ where r_{ij} is the recombination fraction between locus i and locus j. These moments are calculated in Appendix C from elementary, but general principles.

Assuming again that each allele mutates independently of all others, we have $M_{\mathbf{a},\mathbf{b}}=\prod_{i=1}^{L}[(1-\sum_s\mu_{sb_i})\delta_{a_i b_i}+\mu_{a_i b_i}]$.

Let $\mathbf{c}$ denote a paternally inherited gamete that contains alleles with labels $(c_1, c_2, ..., c_L)$, and $\mathbf{d}$ a maternally inherited gamete. The fitness of an individual resulting from the union of these gametes is written $w(\mathbf{c},\ \mathbf{d})\equiv w(\mathbf{d},\ \mathbf{c})$, and $\overline{w}_\tau=\sum_{\mathbf{c},\mathbf{d}}w(\mathbf{c},\mathbf{d})\times n_{\mathbf{c}}(\tau)n_{\mathbf{d}}(\tau)$. The generalization of (B14) to L loci is

$$n_{\mathbf{a}}(\tau+1)=\sum_\sigma\sum_{\mathbf{c},\mathbf{d}}R(\sigma)M_{\mathbf{a},\alpha}$$

$$\times w(\mathbf{c},\mathbf{d})n_{\mathbf{c}}(\tau)n_{\mathbf{d}}(\tau)/\overline{w}_\tau \qquad (B16)$$

$$\alpha_i=(1+\sigma_i)c_i/2+(1-\sigma_i)d_i/2 \qquad (B17)$$

where now $\sum_\sigma\overset{\text{def}}{\equiv}\sum_{\sigma_1=\pm1}\sum_{\sigma_2=\pm1}\cdots\sum_{\sigma_L=\pm1}$.

Continuum of alleles model

The L locus model with continuous alleles may be inferred from (B16) by allowing discrete labels such as $\mathbf{a}$ to be identified with a continuous allelic effect x (where $\infty>x-\infty$) and summations replaced by integrations. It is convenient to define here $\mathbf{x}=(x_1,x_2,...,x_L)$. We identify $n_{\mathbf{a}}(\tau)$ with $\Phi_\tau(\mathbf{x})d^Lx$ where $\Phi_\tau(\mathbf{x})\equiv\Phi_\tau(x_1,x_2,...,x_L)$ is the distribution of allelic effects of gametes in generation τ and $d^Lx=dx_1 dx_2...dx_L$. The quantity $\mu_{\alpha_1\alpha_2}$ is identified with $\mu\times f(x_1-x_2)$ where μ is the probability of a mutation per allele and $f(x_1-x_2)$ is the distribution of mutation effects. This form of distribution function corresponds to mutants having allelic effects that are distributed around

their parental values (Kimura, 1965). Throughout this work we take $f(x)$ to be the normal distribution given in (2).

Lastly we define

$$M(\mathbf{x}) \overset{\text{def}}{\equiv} \prod_{i=1}^{L} [(1-\mu)\delta(x_i) + \mu f(x_i)]. \quad \text{(B18)}$$

Then with $\int d^L u \equiv \int du_1 \int du_2 \ldots \int du_L$, the continuum of alleles analogue of (B16) is $\Phi_{\tau+1}(\mathbf{x}) = \sum_\sigma \int R(\sigma) M(\mathbf{x}-\xi) w(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v / \overline{w}_\tau$ where $\overline{w}_\tau = \int w(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v$ is the mean fitness in generation τ, and ξ is an L component vector with elements

$$\xi_i = \xi_i(\sigma_i, u_i, v_i) = (1+\sigma_i) u_i / 2 + (1-\sigma_i) v_i / 2,$$

$$i = 1, 2, 3, \ldots, L. \quad \text{(B19)}$$

Setting

$$K(\mathbf{x}; \mathbf{u}, \mathbf{v}) \overset{\text{def}}{\equiv} \sum_\sigma R(\sigma) M(\mathbf{x}-\xi) \quad \text{(B20)}$$

allows (B16) to be written compactly as

$$\Phi_{\tau+1}(\mathbf{x}) = \frac{\int K(\mathbf{x}; \mathbf{u}, \mathbf{v}) w(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v}{\int w(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v}.$$

APPENDIX C

In this Appendix we derive some properties we use of the function $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ which is summarized at the beginning of Appendix B.

$K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ encapsulates the effects of mutation and recombination and although it is a complicated function all moments of it may be found with sufficient effort. Here we shall derive results for this function when it is expanded to linear order in μ:

$$K(\mathbf{x}; \mathbf{u}, \mathbf{v}) \approx K(\mathbf{x}; \mathbf{u}, \mathbf{v})|_{\mu=0} + \mu \frac{\partial K(\mathbf{x};\mathbf{u},\mathbf{v})}{\partial \mu}\Big|_{\mu=0}$$

$$\overset{\text{def}}{\equiv} K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) + \mu K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}). \quad \text{(C1)}$$

Summary

We provide a summary of results before deriving them.

$$\int K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = 1$$

$$\int x_i K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = (u_i + v_i)/2$$

$$\int x_i x_j K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = (1-r_{ij})(u_i u_j + v_i v_j)/2 +$$
$$+ r_{ij}(u_i v_j + v_i u_j)/2 \quad \text{(C2)}$$

$$\int K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = 0,$$

$$\int x_i K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = 0$$

$$\int x_i x_j K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = \mu m^2 \delta_{ij} \quad \text{(C3)}$$

where δ_{ij} is a Kronecker delta and has the value 1 for $i = j$ and is zero otherwise and r_{ij} is the recombination fraction between locus i and locus j.

Derivation

$K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ is defined in (B1)–(B5) from which,

$$K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) = K(\mathbf{x}; \mathbf{u}, \mathbf{v})|_{\mu=0}$$

$$= \sum_\sigma R(\sigma) \prod_{j=1}^{L} \delta(x_j - \xi_j)$$

$$K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) = \frac{\partial}{\partial\mu} K(\mathbf{x};\mathbf{u},\mathbf{v})\Big|_{\mu=0}$$

$$= \sum_\sigma R(\sigma) \sum_{i=1}^{L} [f(x_i - \xi_i) - \delta(x_i - \xi_i)]$$

$$\times \prod_{\substack{j=1 \\ (j \neq i)}}^{L} \delta(x_j - \xi_j) \quad \text{(C4)}$$

[$\delta(x)$ denotes a Dirac delta function of argument x].

It immediately follows, using properties of $\delta(x_j - \xi_j)$ and $f(x_j - \xi_j)$, that

$$\int K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = \sum_\sigma R(\sigma) \quad \text{(C5)}$$

$$\int x_i K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = \sum_\sigma R(\sigma)\xi_i \quad \text{(C6)}$$

$$\int x_i x_j K_0(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = \sum_\sigma R(\sigma)\xi_i\xi_j \quad \text{(C7)}$$

$$\int K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = 0 \quad \text{(C8)}$$

$$\int x_i K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = 0 \quad \text{(C9)}$$

$$\int x_i x_j K_1(\mathbf{x}; \mathbf{u}, \mathbf{v}) d^L x = \mu m^2 \delta_{ij} \sum_\sigma R(\sigma). \quad \text{(C10)}$$

Thus we need to establish the values of $\sum_\sigma R(\sigma)$, $\sum_\sigma R(\sigma)\xi_i$ and $\sum_\sigma R(\sigma)\xi_i\xi_j$.

Since ξ_i is, from (B5), given by $\xi_i \equiv \xi_i(\sigma_i, u_i, v_i) = (1 + \sigma_i)u_i/2 + (1 - \sigma_i)v_i/2$, it follows that we need to equivalently know the values of $\sum_\sigma R(\sigma)$, $\sum_\sigma R(\sigma)\sigma_i$ and $\sum_\sigma R(\sigma)\sigma_i\sigma_j$.

The first sum required, $\sum_\sigma R(\sigma)$, is unity, as follows from the probabilistic interpretation of $R(\sigma)$. Thus $\sum_\sigma R(\sigma) = 1$.

Next, we note that $\sum_\sigma R(\sigma)\sigma_i$ must vanish identically, since this sum equals $(+1) \times$ (probability the allele in a gamete at locus i is of paternal origin) + $(-1) \times$ (probability the allele in a gamete at locus i is of maternal origin) and the two probabilities are equal. Thus $\sum_\sigma R(\sigma)\sigma_i = 0$. More generally, the equivalence of paternal and maternal contributions to the gamete of an individual, means we can change σ to $-\sigma$ without changing any results. This has the consequence that all sums involving the product of an odd number of σ's vanishes identically.

Lastly consider $\sum_\sigma R(\sigma)\sigma_i\sigma_j$. We note that if $i = j$, $\sigma_i\sigma_j = 1$, and the sum is unity, by virtue of $\sum_\sigma R(\sigma) = 1$. When $i \neq j$, we have that $\sum_\sigma R(\sigma)\sigma_i\sigma_j$ is given by $(+1) \times$ (probability of no crossover between locus i and locus j) + $(-1) \times$ (probability of crossover between locus i and locus j) = $(+1) \times [1 - r_{ij}]) + (-1 \times r_{ij}) = 1 - 2r_{ij}$ where r_{ij} is the recombination fraction between locus i and locus j. The general rule for all i and j is thus $\sum_\sigma R(\sigma)\sigma_i\sigma_j = 1 - 2r_{ij}$ provided we agree to set $r_{ii} = 0$.

We can now combine the above results to straight-forwardly obtain the results of (C2) and (C3).

APPENDIX D

In this Appendix, we give a derivation of exact results following from the exact dynamical equation (8). The distribution of allelic effects in gametes, in generation 1, is $\Phi_1(\mathbf{x}) = \overline{w}_0^{-1} \int K(\mathbf{x}; \mathbf{u}, \mathbf{v})w(\mathbf{u}, \mathbf{v}) \Phi_0(\mathbf{u})\Phi_0(\mathbf{v})d^L u\, d^L v$ where $\overline{w}_0 = \int w(\mathbf{u}, \mathbf{v})\Phi_0(\mathbf{u}) \Phi_0(\mathbf{v})d^L u\, d^L v$. The form of fitness, $w(\mathbf{u}, \mathbf{v})$, is given by (4) and we write this as $w(\mathbf{u}, \mathbf{v}) = \sqrt{V_s/(2\pi)} \int \exp[i\lambda(\sum_j(u_j + v_j) - \theta) - V_s\lambda^2/2]d\lambda$. In generation 0 we consider a population in linkage equilibrium, and take $\Phi_0(\mathbf{x}) = \prod_{j=1}^L \chi_j(x_j)$ where $\chi_j(x_j) =$

$= (2\pi\alpha_j^2)^{-1/2} \exp[-(x_j - c_j)^2/(2\alpha_j^2)]$. $\chi_j(x_j)$ is the distribution of allelic effects in gametes at locus j. In terms of $\overline{G}_0 \stackrel{\text{def}}{=} 2\sum_j c_j$, $\overline{V}_{g,0} \stackrel{\text{def}}{=} 2\sum_j \alpha_j^2$, a short calculation yields

$$\overline{w}_0 = \sqrt{\frac{V_s}{V_s + V_{g,0}}} \exp\left[-\frac{(\overline{G}_0 - \theta)^2}{2(V_s + V_{g,0})}\right]. \quad \text{(D1)}$$

Using the general definition for $K(\mathbf{x}; \mathbf{u}, \mathbf{v})$ given in (B1–B5) we have

$$\Phi_1(\mathbf{x}) = \overline{w}_0^{-1}\sqrt{V_s/(2\pi)}\sum_\sigma R(\sigma)\int d\lambda \exp[-V_s\lambda^2/2 - i\lambda\theta]$$
$$\times \prod_{j=1}^L \int du\, dv \exp[i\lambda(\mathbf{u} + \mathbf{v})]$$
$$\times [(1 - \mu)\delta(x_j - \xi_j)$$
$$+ \mu f(x_j - \xi_j)]\chi_j(u)\chi_j(v). \quad \text{(D2)}$$

All moments of $\Phi_1(\mathbf{x})$ are contained in the characteristic function $\int e^{i\sum_j \eta_j x_j}\Phi_1(\mathbf{x})d^L x$. We find

$$\int e^{i\sum_j \eta_j x_j}\Phi_1(\mathbf{x})d^L x =$$
$$\overline{w}_0^{-1}\sqrt{V_s/(2\pi)}\sum_\sigma R(\sigma)$$
$$\times \int d\lambda \exp[-V_s\lambda^2/2 - i\lambda\theta]$$
$$\times \prod_{j=1}^L [1 - \mu + \mu \exp(-m^2\eta_j^2/2)]$$
$$\times \int du\, dv\, e^{i\lambda(u+v)}e^{i\eta_j \xi_j}\chi_j(u)\chi_j(v). \quad \text{(D3)}$$

Using the explicit form for ξ_j (B5), it follows that $\int du\, dv\, e^{i\lambda(u+v)}e^{i\eta_j \xi_j}\chi_j(u)\chi_j(v)$ is independent of the sign of σ_j, and we can write

$$\int du\, dv\, e^{i\lambda(u+v)}e^{i\eta_j \xi_j}\chi_j(u)\chi_j(v)$$
$$= \int du\, e^{i(\lambda + \eta_j)u}\chi_j(u)\int dv\, e^{i\lambda v}\chi_j(v)$$
$$= e^{i(2\lambda + \eta_j)c_j - (\lambda^2 + \lambda\eta_j + \eta_j^2/2)\alpha_j^2}.$$

Using this expression, and the form for $\overline{w}_0$, (D1), yields, after some calculation,

$$\int e^{i\sum_j \eta_j x_j}\Phi_1(\mathbf{x})d^L x =$$
$$\exp\left[-\frac{i(\overline{G}_0 - \theta)\sum_j \eta_j\alpha_j^2}{V_s + V_{g,0}} + \frac{(\sum_j \eta_j\alpha_j^2)^2}{2(V_s + V_{g,0})}\right]$$
$$\times \prod_{j=1}^L (1 - \mu + \mu e^{-m^2\eta_j^2/2})e^{i\eta_j c_j - \eta_j^2\alpha_j^2/2}. \quad \text{(D4)}$$

It is now straightforward to calculate moments of $\Phi_1(\mathbf{x})$ from this and in particular, we find the results given in (12), (13) and (14).

APPENDIX E

In this Appendix, we indicate how the selection term

$$-\int K_0 s_L \Phi_\tau \Phi_\tau$$

$$\equiv -\int K_0(\mathbf{x};\mathbf{u},\mathbf{v}) s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v \quad (E1)$$

which contributes in (26) to $\Phi_{\tau+1}$, is responsible for the production of linkage disequilibrium. In particular we show that if, in generation τ, a population is in linkage equilibrium, the term $-\int K_0 s_L \Phi_\tau \Phi_\tau$ generates a contribution of $-\mathrm{Var}(X_k)\mathrm{Var}(X_l) / V_s$ to $C_{kl}(\tau+1)$, the covariance in generation $\tau+1$, where $C_{kl}(\tau+1) = E_{\tau+1}[X_kX_l] - E_{\tau+1}[X_k] E_{\tau+1}[X_l]$. A population that is very close to being in linkage equilibrium would also be expected to obtain, from $-\int K_0 s_L \Phi_\tau \Phi_\tau$, an increment in its covariance in one generation that is very close to $-\mathrm{Var}(X_k)\mathrm{Var}(X_l) / V_s$.

To proceed with the derivation, we note that to determine $C_{kl}(\tau+1)$ we need to first calculate $E_{\tau+1}[X_k] = \int x_k \Phi_{\tau+1}(\mathbf{x}) d^L x$ and $E_{\tau+1}[X_kX_l] = \int x_k x_l \Phi_{\tau+1}(\mathbf{x}) d^L x$. Using (26), these quantities can be related to quantities in generation τ. Since (26) takes the form $\Phi_{\tau+1} = ... - \int K_0 s_L \Phi_\tau \Phi_\tau$, the contribution of the term on the right-hand side to $E_{\tau+1}[X_k]$ is $-\int x_k K_0(\mathbf{x};\mathbf{u},\mathbf{v}) s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v d^L x$ and its contribution to $E_{\tau+1}[X_kX_l]$ is $-\int x_k x_l K_0(\mathbf{x};\mathbf{u},\mathbf{v}) s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v d^L x$. The properties of K_0 given in (C2) yield

$$\int x_k K_0(\mathbf{x};\mathbf{u},\mathbf{v}) s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v d^L x$$

$$= \int u_k s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v \quad (E2)$$

$$\int x_k x_l K_0(\mathbf{x};\mathbf{u},\mathbf{v}) s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v d^L x$$

$$= (1-r_{kl}) \int u_k u_l s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v$$

$$+ r_{kl} \int u_k v_l s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v. \quad (E3)$$

We then use the explicit form of s_L, given in (24), make the assumption of linkage equilibrium, and write $u_k = E_\tau[X_k] + \delta_\tau u_k$. After some work we find that

$$\int u_k s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v = 0 \quad (E4)$$

$$\int u_k u_l s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v$$

$$= \mathrm{Var}_\tau(X_k)\mathrm{Var}_\tau(X_l) / V_s \quad (E5)$$

$$\int u_k v_l s_L(\mathbf{u},\mathbf{v}) \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v$$

$$= \mathrm{Var}_\tau(X_k)\mathrm{Var}_\tau(X_l) / V_s. \quad (E6)$$

We can thus combine the above results and find that $-\int K_0 s_L \Phi_\tau \Phi_\tau$ makes a contribution to $C_{kl}(\tau+1)$ of $-(1-r_{kl}) \mathrm{Var}_\tau(X_k) \mathrm{Var}_\tau(X_l) / V_s - r_{kl} \mathrm{Var}_\tau(X_k) \mathrm{Var}_\tau(X_l) / V_s$ i.e. $-\mathrm{Var}_\tau(X_k) \mathrm{Var}_\tau(X_l) / V_s$.

APPENDIX F

In this Appendix we give an example of the methods needed to establish relations from the equation of dynamics of $\Phi_\tau(\mathbf{x})$, (26). In particular, it is possible to determine the relation between quantities in generation $\tau+1$ and quantities in generation τ.

For example, the mean genotypic value in generation $\tau+1$ is $E_{\tau+1}[G]$ and

$$E_{\tau+1}[G] = E_{\tau+1}\left[\sum_i (X_i + Y_i)\right] = 2\sum_i \int x_i \Phi_{\tau+1}(\mathbf{x}) d^L x. \quad (F1)$$

Using (26), along with (C2) and (C3), yields

$$E_{\tau+1}[G] \approx 2\sum_i \int u_i \Phi_\tau(\mathbf{u}) d^L u$$

$$-2\sum_i \int u_i [s_g(\mathbf{u},\mathbf{v}) - \bar{s}_{g,\tau}] \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v$$

$$-2\sum_i \int u_i [s_L(\mathbf{u},\mathbf{v}) - \bar{s}_{L,\tau}] \Phi_\tau(\mathbf{u}) \Phi_\tau(\mathbf{v}) d^L u d^L v. \quad (F2)$$

Using the properties of s_g and s_L (23), (24) and using $\delta_\tau u_i = u_i - E_\tau[X_i]$ yields, after some simplification,

$$E_{\tau+1}[G] \approx 2\sum_i \int u_i \Phi_\tau(\mathbf{u}) d^L u$$

$$-\frac{2(\bar{G}_\tau - \theta)}{2V_s} 2\sum_{i,j} \int \delta_\tau u_i \delta_\tau u_j \Phi_\tau(\mathbf{u}) d^L u$$

$$-\frac{1}{2V_s} 2\sum_{i,j,k} \int \delta_\tau u_i \delta_\tau u_j \delta_\tau u_k \Phi_\tau(\mathbf{u}) d^L u \quad (F3)$$

and this last result can be written as

$$E_{\tau+1}[G] \approx E_\tau[G] - (\bar{G}_\tau - \theta)/V_G(\tau)/V_s$$

$$-E_\tau[(\delta_\tau G)^3] / (2V_s). \quad (F4)$$