PROPERTIES OF EQUILIBRIA IN MULTI-LOCUS GENETIC SYSTEMS

WARREN J. EWENS

Department of Biology, University of Pennsylvania

AND

Department of Mathematics, Monash University, Victoria, Australia

AND

GLENYS THOMSON*

Genetics Laboratory, University of Oxford, England

Manuscript received April 13, 1977
Revised copy received August 4, 1977

ABSTRACT

The classical mathematical theory of population genetics considered, for simplicity, almost exclusively one-locus systems. In the last two decades much work has been done on two-locus and, less frequently, multi-locus systems. This research has usually involved investigating properties of systems with given, and usually rather special, fitness parameters. Real genetic fitness systems are undoubtedly multi-locus and seldom will possess simplifying characteristics. One aim of this paper is to study generalized systems where no special assumptions are made about fitness structure, the number of alleles at each locus, the number of loci involved or the recombination structure between loci. A second aim is to consider marginal properties (often one-locus properties) of complex systems: the fact that many observations involve data from only one locus makes this second aim relevant.

THE classical mathematical theory of population genetics considers largely one-locus systems where fixed fitnesses are assigned to various one-locus genotypes. This is at best an approximation: the fitness of a genotype at one locus will almost always depend on the genetic constitution of the remainder of the genome (as well as, no doubt, on the particular environment that the individual is placed in). Leaving aside the environmental question, real fitness schemes are multi-locus, and furthermore no practical assessment can usually be made of the number of loci involved.

An aim here is to open up a line of research that assumes as its starting point that fitnesses depend on an unknown (and possibly large) number of loci with an unknown number of alleles at each locus, unknown recombination structure between loci and unknown genotypic fitness values. We assume only fixed fitnesses (although for many of our conclusions fitnesses can be frequency-dependent), random mating and that the population size is so large that all

---

* Present address: Genetics Department, 345 Mulford Hall, University of California, Berkeley, California 94720.

Genetics **87**: 807–819 December, 1977.

processes can be treated as deterministic. In this way we hope to approach more realistically the effects of fitness differentials in actual populations.

A second aim also has a practical motivation. Many experiments or observations involve data from a very small number of loci (often only one), and we thus consider the marginal properties of a k-locus system when only l loci $(l < k)$ are considered. This will involve the definition of marginal or induced l-locus fitness values as appropriately weighted averages of fitnesses in the full k-locus system. These induced marginal fitnesses change from generation to generation until equilibrium is reached, but are nevertheless, in any generation, the values which will be calculated by the experimenter from his observation of the l-locus subsystem. One of our results is to show that for any general k-locus system all "lower order" marginal systems (*i.e.*, involving 1,2,3, . . . , k-2 or k-1 loci) satisfy the appropriate lower-order set of recurrence relations with these marginal fitnesses used in the standard equations. One consequence of this is that if the k-locus system is in equilibrium, then all marginal l-locus systems must exhibit equilibrium behavior, and in particular, all gene frequencies must be equal to the values predicted from marginal gene fitnesses.

We also discuss the question of stability of equilibria. It is easy to construct examples where a k-locus system is at a stable equilibrium, and each l-locus marginal subsystem is also at a stable equilibrium for its induced marginal fitnesses. The same is true for unstable k- and l-locus systems. We give an example of a k-locus system at an unstable equilibrium point for which all $l(l < k)$ subsystems satisfy the stability criteria for their subsystems. We do not know if a k-locus system at a stable equilibrium can have l-locus subsystems at apparently unstable equilibria. For $k = 2$, the many simulations of KARLIN and CARMELLI (1975) suggest that this is unlikely, and KARLIN (1975), in a paper that is the main stimulus for this research, claims that it is impossible for a sufficiently small recombination fraction between the loci. For larger values of k, the question is still quite open. This question is of practical interest since at least two reported cases (LEWONTIN and WHITE 1960; CHRISTIANSEN *et al.* 1974) show systems at apparently unstable equilibrium points of (respectively) two- and one-locus systems. It is still an open question as to whether these could be the result of a system at a stable equilibrium for a three (or more) locus fitness system.

Finally, we pay some attention to a quantity of fundamental importance in theoretical population genetics, namely the additive genetic variance in fitness. (Note that since fitness is the only characteristic we are interested in, the words "in fitness" will be dropped below). We show that the additive genetic variance is always equal to a second important quantity, namely the additive gametic variance. The classical one-locus result that a system has zero additive genetic variance if and only if it is at equilibrium is shown to be no longer true for many loci, although a restricted version of this conclusion will be proved. Relations between the k-locus additive genetic variance and the induced l-locus additive genetic variances will also be considered.

ONE-LOCUS SYSTEMS

We begin by discussing in more quantitative fashion some of the problems raised in the Introduction.

The simplest classical theory of population genetics considers a locus "A" at which occur genotypes A_1A_1, A_1A_2 and A_2A_2, with fitnesses w_{11}, w_{12} and w_{22}. We defer for the moment the meaning of the word fitness in this context. Then, under certain assumptions the frequencies of these genotypes at the time of zygote formation are in Hardy-Weinberg form $x^2, 2x(1-x), (1-x)^2$ (where x is the frequency of A_1) and in the simplest discrete time model a recurrence relation of the form

$$x' = (w_{11}x^2 + w_{12}x(1-x))/(w_{11}x^2 + 2w_{12}x(1-x) + w_{22}(1-x)^2) \qquad (1)$$

is found. It is a simple consequence of this relation that the frequency of A_1 is at an equilibrium if

$$x = x^* = (w_{12} - w_{22})/(2w_{12} - w_{11} - w_{22}) \ . \qquad (2)$$

The point (2) will be allowable (*i.e.*, $0 \leqslant x^* \leqslant 1$) if and only if

$$w_{12} > w_{11}\,, w_{22} \qquad (3)$$

or

$$w_{12} < w_{11}\,, w_{22} \ . \qquad (4)$$

The equilibrium (2) is stable if and only if (3) obtains.

There is one potential problem with these arguments. Equations (1) through (4) are arrived at under the assumption that the fitnesses w_{ij} are fixed constants, while on the other hand it is accepted that such single-locus fitnesses cannot be fixed constants and will depend at the very least on the genetic constitution of each individual at the remaining loci. But if this is so, the recurrence relations connecting genome genotypic frequencies between consecutive generations will be very complex, and it is not immediately clear that they will imply the truth of equations (1) through (4) when the w_{ij} are defined by some suitable averaging procedure. We shall indeed see later, for example, that genome genotypic frequencies can be at an unstable equilibrium point even though (3) obtains for all marginal fitness values. To discuss such problems in more detail we must turn to an examination of the complex genotypic recurrence relations arising when fitness depends on an arbitrarily large number of loci.

THE k-LOCUS SYSTEM

We consider a k-locus genetic system with an arbitrary number of alleles possible at each locus, arbitrary recombination structure and arbitrary (fixed) fitnesses for the various possible genotypes. Random mating is assumed. It is a standard result of population genetics theory that the evolution of such a k-locus system must be described by considering the frequencies of the various k-locus gametes possible in the system. We suppose that these gametes are arranged in some dictionary order and denote their frequencies $x_1, x_2, \ldots$. Let the fitness of the genotype defined by gametes i and j be w_{ij}. (We reserve the symbols i, j and h to denote arbitrary k-locus gametes). The marginal fitness w_i of gamete i is defined, according to standard practice, as

$$w_i = \sum_j x_j w_{ij} , \tag{5}$$

and the mean fitness $\bar{w}$ is defined by

$$\bar{w} = \Sigma \Sigma x_i x_j w_{ij} . \tag{6}$$

Then trivially

$$\Sigma x_i (w_i - \bar{w}) = 0 , \tag{7}$$

and EWENS (1976) has shown that, at any equilibrium point of the system,

$$\Sigma^* x_i (w_i - \bar{w}) = 0 , \tag{8}$$

where Σ^* refers to summation restricted to those gametes containing any arbitrary but specified allele at any arbitrary but specified locus.

One of our interests in this paper concerns various components of the total variance in fitness

$$\sum_i \sum_j x_i x_j (w_{ij} - \bar{w})^2 = \sum_i \sum_j x_i x_j w^2_{ij} - \bar{w}^2 . \tag{9}$$

Because of the role played by the gamete frequencies in describing the evolution of the k-locus genetic system, it is natural to consider the component of this variance that measures the extent to which the marginal fitnesses of the various gametes differ from each other. This is the "total" gametic variance, defined as $2\Sigma x_i (w_i - \bar{w})^2$. This variance does not, however, have particular evolutionary importance. To find a variance which does, we must partition the "total" gametic variance into two components, the so-called "additive" and "non-additive" components. We do this as follows. Suppose all the alleles at all the k loci are written down in some dictionary order $A_1, A_2, \ldots, A_T$ and let y_t be the frequency of the t^{th} of these alleles. (Note that $\sum_t y_t = k$). Attach additive parameters $\alpha_1, \alpha_2, \ldots, \alpha_T$ to the alleles, where the α's are constrained by the requirement

$$\Sigma y_t \alpha_t = 0 , \tag{10}$$

the summation being true over all alleles at any one of the k loci (so that (10) implies k separate equations). Subject to this constraint, we attempt to minimize (with respect to $\alpha_1, \alpha_2, \ldots, \alpha_T$) the expression

$$2 \sum_i x_i (w_i - \bar{w} - \Sigma c_{it} \alpha_t)^2 , \tag{11}$$

where $c_{it} = 1$ if the i^{th} gamete contains the allele A_t, but is zero otherwise. It is not difficult to see that the minimizing values satisfy the equations

$$\Sigma^{(t)} x_i (w_i - \bar{w}) = y_t \hat{\alpha}_t + \sum_{u \neq t} y_{tu} \hat{\alpha}_u , \tag{12}$$

where $\Sigma^{(t)}$ implies summation over all those gametes containing the allele A_t and, for any two alleles A_t and A_u at different loci, y_{tu} is the total frequency of those gametes containing both A_t and A_u. Furthermore, least squares theory shows that the additive gametic variance in fitness (AGV), namely the component of the total gametic variance removed by fitting the parameters $\alpha_1, \ldots, \alpha_T$, is

$$2 \sum_{t=1}^{T} \{\Sigma^{(t)} x_i(w_i - \bar{w})\} \hat{\alpha}_t \ . \tag{13}$$

Of more classical interest is the so-called "additive genetic variance", defined in the following way. Subject to the constraints (10) we attempt to minimize the expression

$$\sum_i \sum_j x_i x_j \{w_{ij} - \bar{w} - \Sigma_t \alpha_t n_{t(ij)}\}^2 \ , \tag{14}$$

where $n_{t(ij)}$ is the number of times (0,1 or 2) that allele A_t appears in the genotype defined by gametes i and j. Appropriate differentiation in (14) leads to the equations

$$\begin{aligned} \Sigma^{*(t)} n_{t(ij)} x_i x_j (w_{ij} - \bar{w}) = \hat{\alpha}_t \Sigma^{*(t)} x_i x_j (n_{t(ij)})^2 \\ + \sum_u \Sigma^{*(t,u)} x_i x_j n_{t(ij)} n_{u(ij)} \hat{\alpha}_u \ , \\ t = 1,2, \ldots T \ , \end{aligned} \tag{15}$$

where $\Sigma^{*(t)}$ implies summation over all those genotypes (ij) containing the allele A_t and $\Sigma^{*(t,u)}$ implies summation over all those genotypes (ij) containing both alleles A_t and A_u. It is easy enough to see that the left-hand side in (15) reduces to twice the left-hand side in (12), while the right-hand side is just

$$\begin{aligned} &\alpha_t [4y_t^2 + 2y_t(1-y_t)] + 2y_t\Sigma^{(1)} \hat{\alpha}_u y_u \\ &+ 2 \Sigma^{(2)} \hat{\alpha}_u (y_{tu} + y_t t_u) \\ &= 2 y_t \hat{\alpha}_t + 2 \Sigma^{(2)} \hat{\alpha}_u y_{tu}) \ , \end{aligned}$$

where $\Sigma^{(1)}$ implies summation over all alleles $A_u (u \neq t)$ at the same locus as A_t, and $\Sigma^{(2)}$ implies summation over all alleles A_u at a different locus than A_t. It follows that equations (12) and (15) are identical and from this that the sum of squares removed in (14) is identical to that removed in (11). In other words we have reached:

RESULT 1. *The additive genetic variance in fitness for a general* k-*locus system is equal to the additive gametic variance in fitness.*

This conclusion was found by KIMURA (1965) for two loci and two alleles and by NAGYLAKI (1976) for two loci with an arbitrary number of alleles. Clearly the result is true for any character, not only fitness.

We consider below various properties of the AGV and the conclusion just reached shows that identical properties will hold for the additive genetic variance. Note also that the conclusion of EWENS (1976) that the AGV is zero at any equilibrium point of the k-locus system immediately implies a parallel property for the additive genetic variance.

We next ask a converse question, namely, what equilibrium properties obtain if the equation

$$\mathrm{AGV} = 0 \tag{16}$$

is true? The truth of (16) implies the truth of (8), but this does *not* necessarily imply that the system is at an equilibrium. A genetic system can have curves or even surfaces in gamete frequency space on which (16) is true, with only one point on the curve (or surface) being an equilibrium point. An example is given by the trajectory computed by MORAN (1964): at all points on this trajectory equation (16) holds, and yet only one point on it is an equilibrium point. Although this example refers to un unlikely biological circumstance, it is sufficient to illustrate:

RESULT 2. *While equation (16) does not necessarily imply that the* k*-locus genetic system is at an equilibrium, it does nevertheless imply that, at least for one generation, there will be no change in allelic frequencies.*

The proof of this statement follows immediately from the fact that (16) implies (8), which implies the statement made.

The above conclusions are intended mainly as background for the main aim of this paper, to which we now turn, namely to discuss the properties of any l-locus $(l < k)$ marginal subsystem of the system under consideration and to ask, in particular, what properties of this subsystem can be deduced from those of the full k-locus system, what properties of the k-locus system can be deduced from those of l-locus subsystems, and what properties of an l-locus subsystem can be found from those of the subsystem itself.

l-LOCUS SUBSYSTEMS

Consider any l-locus subset of the k-locus system, $(l < k)$. These l loci will define l-locus gametes whose frequencies we denote $z_1, z_2, \ldots, z_R$. If S_P is the collection of k-locus gametes having the same alleles as the p^{th} of these gametes at the l loci under consideration, then

$$z_p = \sum_{i \varepsilon S_p} x_i \, . \tag{17}$$

We use suffixes p, q and r throughout to denote l-locus gametes and gametic frequencies.

Now consider the l-locus genotype formed by the p^{th} and q^{th} l-locus gametes. Following KARLIN (1975) we would define the induced marginal fitness $\bar{w}_{pq}$ of this genotype (we use the bar notation throughout to denote induced l-locus fitnesses) as the fitness obtained by averaging over all genotypic combinations at the remaining k-l loci, appropriately weighted by their frequencies and fitnesses. It is thus the average fitness of all k-locus genotypes exhibiting the l-locus genotype of interest, and so

$$\begin{aligned} \bar{w}_{pq} &= \sum_{i \varepsilon S_p} \sum_{j \varepsilon S_q} x_i x_j w_{ij} \Big/ \sum_{i \varepsilon S_p} \sum_{j \varepsilon S_q} x_i x_j \\ &= \sum_{i \varepsilon S_p} \sum_{j \varepsilon S_q} x_i x_j w_{ij} / z_p z_q \, . \end{aligned} \tag{18}$$

KARLIN (1975, equation 2.4) gives examples of such induced fitnesses for $k = 2$, $l = 1$. The induced marginal fitness $\bar{w}_p$ of the p^{th} l-locus gamete is defined in a manner identical to that of equation (5) by

$$\bar{w}_p = \sum_q \bar{w}_{pq} z_q \ , \tag{19}$$

and from (18) this becomes

$$\begin{aligned} \bar{w}_p &= \sum_q \sum_{j \varepsilon S_q} \sum_{i \varepsilon S_p} x_i x_j w_{ij} / z_p \\ &= \sum_j \sum_{i \varepsilon S_p} x_i x_j w_{ij} / z_p \ , \end{aligned} \tag{20}$$

since for any function ϕ_j,

$$\sum_q \sum_{j \varepsilon S_q} \phi_j = \sum_j \phi_j \ . \tag{21}$$

Equations (20) and (21) now show that

$$z_p \bar{w}_p = \sum_{i \varepsilon S_p} x_i w_i \tag{22}$$

$$\sum z_p \bar{w}_p = \sum_j \sum_i x_i x_j w_{ij} = \bar{w} \tag{23}$$

so that the induced gametic fitnesses $\bar{w}_1, \ldots, \bar{w}_R$ obey an equation parallel to (2). We note also that since

$$\begin{aligned} \sum_p \sum_q z_p z_q \bar{w}_{pq} &= \sum_p \sum_q \sum_{i \varepsilon S_p} \sum_{j \varepsilon S_q} x_i x_j w_{ij} \\ &= \sum_i \sum_j x_i x_j w_{ij} \\ &= \bar{w} \ , \end{aligned} \tag{24}$$

the mean fitness, as computed from marginal l-locus gamete frequencies and fitnesses, is identical to the true k-locus mean fitness. It is well known that in a k-locus system ($k \geq 2$), decreases in mean fitness can occur, whereas the fundamental theorem of natural selection shows that under the assumptions we make this is impossible for a genuine one-locus system. On the other hand, equation (24) shows that the apparent one-locus mean fitness in a k-locus system, being identical to the k-locus mean fitness, can decrease. This can happen even when gene frequencies at this locus are constant. There is of course no anomaly between these two facts, since in the k-locus case the induced one-locus fitnesses are not constant, but normally change with changing k-locus gametic frequencies.

We now wish to establish what l-locus analogues exist for certain k-locus results noted above. We first prove:

RESULT 3. *The marginal* l-*locus AGV is zero at any equilibrium point of the* k-*locus system.*

From the discussion centering around equations (8) and (16) it is sufficient to prove that, at an equilibrium of the k-locus system,

$$\sum_p{}^{*} z_p (\bar{w}_p - \bar{w}) = 0 \ , \tag{25}$$

where the summation Σ^* is over all l-locus gametes containing any arbitrary specified allele at an arbitrary one of the l loci defining these gametes. To do this, it is sufficient, from (8), to prove that

$$\sum_p^* z_p \bar{w}_p = \sum_j^* z_j w_j \ , \tag{26}$$

where the summation on the right-hand side is over all k-locus gametes containing the allele in question.

Use of (5) and (22) shows that the left hand side in (26) is

$$\sum_p^* \sum_j \sum_{i \varepsilon S_p} x_i x_j w_{ij}$$

(using an equation identical to (21) with Σ^* replacing Σ)

$$= \sum_i^* x_i w_i \ , \tag{27}$$

which is the right-hand side in (26).

Thus, equation (26) is true and, from the discussion following equation (7), the induced l-locus AGV is zero at any equilibrium point of the k-locus system.

We now wish to prove a stronger result. The fact that the apparent AGV is zero does not necessarily imply an induced l-locus equilibrium point at any k-locus equilibrium, except in the case $l = 1$. We now show, however, that if the k-locus system is at an equilibrium point, so will also be any induced l-locus system. (It is perhaps important to stress the distinction between "real" and "induced" equilibria. If a k-locus system is at equilibrium, then necessarily any l-locus system is at a "real" equilibrium. We are here interested, however, in something different (and more relevant for practical purposes), namely whether the recurrence relations signifying an equilibrium of the induced system are in fact satisfied at an equilibrium of the k-locus system).

The typical recurrence relation for gamete frequencies in the k-locus system is

$$\bar{w} x_i' = w_i x_i - \Sigma^{(1)} w_{ij} x_i x_j f(i,j \rightarrow h) + \Sigma^{(2)} w_{hj} x_h x_j f(h,j \rightarrow i) \tag{28}$$

where $\bar{w}$ is the mean fitness of the population, x_i is the frequency of gamete i in the current generation, x_i' the frequency in the next generation, w_i is defined by (1), $\Sigma^{(1)}$ implies summation over all gametes j and h with $i, j \neq h$, $\Sigma^{(2)}$ implies a parallel summation with h and i interchanged in role; while $f(i,j \rightarrow h)$ is one-half the probability of a recombinational event in an (i,j) individual such that one of the two gametes produced after this recombination is gamete h.

At equilibrium, $x_i' = x_i$. Now sum equation (28), with x_i' replaced by x_i, over all $i \varepsilon S_p$. We find

$$\begin{aligned} \bar{w} z_p = \bar{w}_p z_p &- \sum_{i \varepsilon S_p} \Sigma^{(1)} w_{ij} x_i x_j f(i, j \rightarrow h) \\ &+ \sum_{i \varepsilon S_p} \Sigma^{(2)} w_{hj} x_h x_j f(h, j \rightarrow i) \ , \end{aligned} \tag{29}$$

the first term on the right hand side arising from equation (22). It is sufficient to consider $\Sigma^{(1)}$ for those gametes i and j differing at two or more loci. If these differences occur only at the k-l loci outside the l-locus system considered, then an exact compensation and cancellation will occur between a term in $\Sigma^{(1)}$ and a term in $\Sigma^{(2)}$ when summation over $i \varepsilon S_p$ takes place. It is not difficult to see that a corresponding statement applies when exactly one of the loci at which gametes i and j differ is included in the l-locus system. Consider now cases where gametes i and j differ at two or more loci in the l-locus system. Any recombinational event not involving these loci will again lead to a cancellation of terms in $\Sigma^{(1)}$ and $\Sigma^{(2)}$. Consider then, finally, any recombinational event involving two or more loci in the l-locus system. We assume that the frequency of such recombination is independent of the genetic constitution of the loci involved. The gamete x_j will define a class S_q of l-locus gametes different from the class S_p defined by gamete i, and the sum of the second right-hand side terms in (28) becomes

$$f(i, j \rightarrow h) \sum_{i \varepsilon S_p} \sum_{j \varepsilon S_q} x_i x_j w_{ij}$$

$$= f(p,q \rightarrow r)\, \bar{w}_{pq} z_p z_q$$

from (18), where the symbol $f(p,q \rightarrow r)$ is one-half the frequency of a recombinational even in the l-locus system such that one of the l-locus gametes produced by a (p,q) individual is gamete r, where gamete r differs from gametes p and q at a minimum of two loci. Parallel arguments apply to the summation $\Sigma^{(2)}$ and the summation of (16) over $i \varepsilon S_p$ then yields

$$\bar{w} z_p = \bar{w}_p z_p - \Sigma^{(1)} \bar{w}_{pq} z_p z_q f(p, q \rightarrow r) + \Sigma^{(2)} \bar{w}_{qr} z_q z_r f(q, r \rightarrow p) \ . \qquad (30)$$

We have proved:

RESULT 4. *If the* k-*locus system is in equilibrium, then so also is any marginal* l-*locus subsystem (in the sense of apparent, as contrasted to real, equilibrium as discussed above).*

A further result follows almost immediately. The derivation of (30) remains true if we replace x_i and z_p throughout on the left-hand side by x_i' and z_p' where these are respectively the frequencies of the i^{th} k-locus gamete and the p^{th} l-locus gamete in the following generation in an evolving system. From this observation, and using (30) with z_p on the left-hand side replaced by z_p', we find:

RESULT 5. *Changes in frequencies of* l-*locus gametes can be calculated for one generation in advance knowing only the current* l-*locus gametic frequencies and marginal fitnesses. (In particular, this is true of allele frequencies). However, to predict two or more generations in advance, the full* k-*locus system must, in general, be studied.*

It follows in particular that equation (1) is correct, with appropriate definitions of the w_{ij}, but is useless for predicting more than one generation in advance.

k-LOCUS PROPERTIES FROM l-LOCUS PROPERTIES

In the above we have considered what properties obtain in marginal l-locus systems in a k-locus system whose behavior is assumed. It is perhaps of more practical relevance to ask what properties of a k-locus system must hold, given observed properties of induced l-locus subsystems. In this direction, we first show:

RESULT 6. *If the marginal AGV is zero for each single locus in a* k*-locus system, then the* k*-locus AGV must be zero.*

Indeed, a stronger result is true, namely that if the AGV is zero for every l-locus subsystem (for any fixed value of l) then the k-locus AGV is zero. The proof is straightforward. Suppose, for any given number l, each l-locus subsystem has induced AGV of zero. Then equation (25), with Σ^* as defined below the equation, holds. But

$$\begin{aligned}\sum_i{}^* x_i w_i &= \sum_p{}^* z_p \bar{w}_p \qquad \text{[from (26)]}\\ &= \bar{w} \sum_p{}^* z_p \\ &= \bar{w} \sum_i{}^* x_i\end{aligned}$$

so that

$$\sum_i{}^* x_i (w_i - \bar{w}) = 0 \ .$$

This is equation (8), and hence equation (16) holds.

The case $l=1$ is of particular interest. Since "AGV $= 0$" is equivalent to "no changes in allele frequencies," we can say: if, by considering each locus separately, we expect (at all loci) no changes in allele frequencies, then indeed there will be no changes in allele frequencies. In view of the results given below, this does not seem to be a trivial conclusion.

If equation (16) implied that the k-locus system is at an equilibrium point, we could conclude from the above that if all one-locus systems were at an equilibrium, so also would be the k-locus equilibrium. Unfortunately, (16) does not imply equilibrium and we now exhibit an example where a k-locus system is *not* at an equilibrium, even though all $l(l < k)$ induced subsystems *are* at equilibrium points.

To demonstrate this, consider the $(k = 2)$ two-locus system of MORAN (1964). At every point on the trajectory computed by him we have

$$\begin{aligned}\text{frequency } (A_1) &= \text{frequency } (A_2) = 0.5 \ ,\\ \text{frequency } (B_1) &= \text{frequency } (B_2) = 0.5 \ .\end{aligned}$$

Furthermore, at each point on the trajectory,

$$\begin{aligned}\text{induced fitness } (A_1A_1) &= \text{induced fitness } (A_2A_2) \ ,\\ \text{induced fitness } (B_1B_1) &= \text{induced fitness } (B_2B_2) \ .\end{aligned}$$

It follows that throughout the evolution of the system, in which gametic frequencies are constantly changing, both induced one-locus systems are at equilibrium points. The entire system, on the other hand, is obviously not at equilibrium.

We turn now to a consideration of the additive genetic variance and ask to what extent its (k-locus) value can be found from a knowledge of the k constituent single locus additive genetic variances calculated using marginal fitness values. The original derivations of FISHER (1918) at least suggest that the k-locus value can be found by simple summation of the k single locus values and this also appears to be more or less explicitly stated by WRIGHT (1969, page 439).

Now standard theory shows that if we compute the additive genetic variance at a single locus by using the marginal fitness values, the estimates of the additive parameter α_t is

$$\hat{\alpha}_t = \bar{w}_t - \bar{w} \ , \tag{31}$$

where $\bar{w}_t$ is the weighted marginal fitness of allele A_t. [This conclusion can also be reached from equation (12)]. The additive genetic variance at the locus at which A_t appears is then computed, from (13), as

$$2 \Sigma y_t (\bar{w}_t - \bar{w})^2 \ , \tag{32}$$

the summation being over all alleles at the locus in question. Adding over all loci gives the expression (32) if we now interpret the summation as being over all alleles at all k loci. This is not necessarily equal to the true k-locus additive genetic variance given by (13). Comparison of (13) and (31) and use of (10) shows that a sufficient condition for equality of the two expressions is that

$$y_{tu} = y_t y_u \tag{33}$$

for all alleles A_t and A_u at different loci. Condition (29) may be stated as:

RESULT 7. *A sufficient condition that the true additive genetic variance can be found by summing single locus marginal genetic variances is that all pairwise linkage disequilibria be zero.*

Note that in practice this situation is most likely to arise approximately for loci determining characters not associated with fitness, and is less likely to arise for those loci which do affect fitness.

We turn finally to k-locus systems which are at an equilibrium point and consider stability properties of the equilibrium. It is trivially easy to find k-locus stable equilibria with marginal fitnesses satisfying the stability conditions at all $l (l < k)$ subsystems, and a parallel remark applies for unstable equilibria ("additive" fitnesses will serve in both cases). It is also possible for a k-locus system to be at an unstable equilibrium and yet the stability conditions be satisfied at all l-locus $(l < k)$ marginals. A far more difficult question is whether a k-locus stable equilibrium can lead to apparent l-locus unstable equilibria. KARLIN (1975) mentions an unpublished proof that this is impossible for small recombination values and KARLIN and CARMELLI (1975), based on a very large number of simulations, conjecture it is impossible for any recombination structure, at least for $k = 2$, $l = 1$ and two alleles per locus. (KARLIN (1975) mentions an unpublished proof that for some special fitness configurations marginal underdominance is impossible for any recombination value when $k = 2$.) Although we believe the conjecture of KARLIN and CARMELLI may well be correct for $k = 2$, $l = 1$, it is interesting to note that two published accounts of natural populations (LEWONTIN and WHITE 1960; CHRISTIANSEN *et al.* 1974) apparently show behavior contrary to the general spirit of this conjecture, that it is not true if fitnesses are frequency-dependent and that for larger values of k and l schemes mimicking frequency-dependent-fitnesses might mean that the conjecture cannot be generalized.

The results we have achieved do at least allow us to draw some conclusions, and these are most easily discussed when two alleles only occur at each locus. First, if the k-locus system is at an equilibrium of any kind (stable or unstable), the induced systems at each single locus must exhibit equilibrium behavior and hence reveal either underdominance or overdominance. Put in contrapositive form, if there exists any one locus for which induced underdominance or overdominance does not appear, then the k-locus system cannot be at an equilibrium. This provides a reasonably powerful test for k-locus equilibria, which we regard as particularly useful in view of our aim to state facts about a k-locus system from observed single-locus behavior. More strongly still, for the k-locus system to be in equilibrium, not only must underdominance or overdominance hold; the allele frequencies must take the required equilibrium values as defined by the induced one-locus fitnesses. The further requirement of marginal stability would require all the above to hold with now only overdominance allowed. This leads to further intuitive support of the KARLIN-CARMELLI conjecture for $k = 2$, although for larger values of k the picture is quite complex and intuitive arguments are less valuable.

An illuminating example of some of these points is provided by the symmetric viability matrix of EWENS (1969) (pages 104–105) with

$$s = 0.21, \ t = 0.4, \ u = 0.1.$$

The theory in EWENS (1969, page 105) shows that if R is the recombination fraction between the two loci, then for this fitness scheme there are two equilibria with

$$x_1 = x_4 = 0.25 \pm 0.25 \ (1 - R/0.0725)^{1/2} \ ,$$
$$x_1 = x_3 = 0.25 \mp 0.25 \ (1 - R/0.0725)^{1/2} \ .$$

Both these equilibria exist only when $R < 0.07250$, and are stable whenever $R < 0.05756$. The marginal fitnesses at the B locus always exhibit overdominance at the equilibrium point, but those at the A locus do so only if $R < 0.05971$. Thus we conclude, in this case, that if the two-locus system is at a stable equilibrium, both marginal loci must exhibit overdominance, whereas it is possible (when $0.05756 < R < 0.05971$) that both loci exhibit equilibrium and overdominance and yet the two-locus system is not at a stable equilibrium.

GENERAL REMARKS

We have assumed above that the k-locus fitnesses are fixed constants. This assumption can be relaxed, at least for many of our conclusions. Thus results 1,2,3,4,6 and 7 are true if fitnesses are frequency or density dependent. Result 5 is not true in such cases since the gene frequencies themselves are not in general sufficient to determine the w_{ij}. Note that further interesting conclusions are possible for frequency-dependent selection schemes, for example that they admit equilibria showing marginal underdominance (LEWONTIN 1958), whereas we have mentioned our belief that this is impossible for fixed fitness schemes. The equilibria of LEWONTIN and WHITE (1960) and CHRISTIANSEN *et al.* (1974) could well be due to frequency-dependent selection schemes.

We have also assumed fixed recombination rates throughout. Possible inversion systems or the evolution of recombination rates will necessitate later generalizations of our conclusions.

Note also that our analysis is deterministic and assumes effectively infinite population sizes. This implies not only that we ignore the possible effects of stochastic fluctuation but also the fact that, in finite populations, only a very small fraction of the possible k locus genotypes will, for large k, be represented in the population. These and other complications will be taken up in a subsequent paper.

We wish to acknowledge the assistance of O. MAYO, who has contributed several helpful suggestions. W.J.E. was supported by Public Health Service grant PHS1 RO1 GM21335 at the University of Pennsylvania. G.T. was supported by an IBM Research Fellowship at Oxford University.

LITERATURE CITED

CHRISTIANSEN, F. B., O. FRYDENBERG, A. O. GLYDENHOLM and V. SIMONSEN, 1974 Genetics of *Zoarces* populations. VI. Further evidence, based on age group samples, of a heterozygote deficit in the *EST* III polymorphism. Hereditas **77**: 225–236.

EWENS, W. J., 1969 *Population Genetics.* Methuen, London. ——, 1976 Remarks on the evolutionary effect of natural selection. Genetics **83**: 601–607.

FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edin. **52**: 399–433.

KARLIN, S., 1975 General two-locus selection models: some objectives, results and interpretations. Theoret. Pop. Biol. **7**: 364–398.

KARLIN, S. and D. CARMELLI, 1975 Numerical studies on two-loci selection models with general viabilities. Theoret. Pop. Biol. **7**: 399–421.

KIMURA, M., 1965 Attainment of quasi-linkage equilibrium when gene frequencies are changing by natural selection. Genetics **52**: 875–890.

LEWONTIN, R. C., 1958 A general method for investigating the equilibrium of gene frequency in a population. Genetics **43**: 419–433.

LEWONTIN, R. C. and M. J. D., WHITE, 1960 Interaction between inversion polymorphisms of two chromosome pairs in the grasshopper, *Moraba scurra.* Evolution **14**: 116–129.

MORAN, P. A. P., 1964 On the nonexistence of adaptive topographies. Ann. Hum. Genet. **27**: 383–393.

NAGYLAKI, T. 1976 The evolution of one- and two-locus systems. Genetics **83**: 583–600.

WRIGHT, S., 1969 *Evolution and the Genetics of Populations. Vol. 2. The Theory of Gene Frequencies.* Chicago University Press.

Corresponding editor: R. C. LEWONTIN