The Alternative Fitness Sets Which Preserve Allele Trajectories: A General Treatment

Michael Lachmann-Tarkhanov* and Sahotra Sarkar†

**Department of Biological Sciences, Stanford University, Stanford, California 94305 and †Dibner Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

Manuscript received August 13, 1993
Accepted for publication August 17, 1994

ABSTRACT

A general solution is presented of the problem of specifying all alternative, generally frequency-dependent, (absolute) fitness sets which give rise to the same allele frequency changes and population dynamics as a given fitness set. The one- and two-locus cases are analyzed in detail and the method is then extended to the n-locus case. It is shown that if biological constraints can be used to specify the mean fitness of the population and the relative fitnesses of the heterozygotes, then the allele frequency trajectories determine a unique fitness set.

SEWALL WRIGHT (WRIGHT and DOBZHANSKY 1946) was probably the first to notice that two different fitness sets could give rise to essentially the same allele frequency trajectories in a one-locus-two-allele model. Such alternative fitness sets are thus *equivalent* in this sense. They are important because, when they exist, allele frequency trajectories do not suffice to specify a particular (genotypic) fitness set uniquely. DENNISTON and CROW (1990) have since provided several procedures for constructing such alternative, but equivalent, fitness sets. For one-locus models they have shown that there exist multiplicative and additive transformations, linear combinations of these, and miscellaneous other transformations between alternative fitness sets which conserve allele frequency trajectories and, sometimes, also the total population dynamics. However, their treatment was episodic, rather than systematic: they presented no general procedure that could generate all equivalent alternatives to a given fitness set. For two-locus models, they exhibited examples of similar procedures. Although they did not assume linkage equilibrium in finding the transformations, all their examples did.

This paper reports more general and systematic solutions of the problem of constructing alternative fitness sets from a given one. *First*, the one-locus case is fully solved for an arbitrary number of alleles. *Second*, the two-locus case is also fully solved without assuming linkage equilibrium. *Finally*, the general method is illustrated for the n-locus case with arbitrary numbers of alleles at the different loci. This constitutes a full solution to the problem. The most interesting result is that the mean fitness and the relative fitnesses of the heterozygotes can be arbitrarily specified in the new fitness sets, but once the functional forms of these are specified, the new equivalent fitness sets are exactly fixed.

ONE LOCUS

Basic notation: In all our models, like DENNISTON and CROW (1990), we assume random mating, discrete generations, very large populations (no stochastic effects), and viability selection (or, equivalently, multiplicative fertilities). For one-locus models with n alleles A_i, $i = 1, \ldots, n$, let p_i $(1 \leq i \leq n)$ be the frequency of allele A_i, and p be the vector of allele frequencies. Let the absolute fitness of the genotype A_iA_j be $w_{ij}(p)$ where $i, j = 1, \ldots, n$. Then the marginal fitness of the A_i allele is defined by $w_i(p) = \sum_{j=1}^{n} p_j w_{ij}(p)$. Let $\bar{w}(p)$ be the mean absolute fitness of the population. Then $\bar{w}(p) = \sum_{i,j} p_i p_j w_{ij}(p)$. The set of absolute genotypic fitnesses $\{w_{ij}(p) : i, j = 1, \ldots, n\}$ is the *fitness set* of interest. Let $\{w^*_{ij}(p) : i, j = 1, \ldots, n\}$ be an alternative fitness set. Let $w^*_i(p) = \sum_j p_j w^*_{ij}(p)$. Then $\bar{w}^*(p) = \sum_{i,j} p_i p_j w^*_{ij}(p) = \sum_j p_j w^*_j(p)$. Throughout, we will assume that $w_{ij}(p) = w_{ji}(p)$ and $w^*_{ij}(p) = w^*_{ji}(p)$. All the $w_{ij}(p)$, $w^*_{ij}(p)$, $w_i(p)$ and $w^*_i(p)$ are functions of the allele frequencies p_i. Two alternative fitness sets will be called *equivalent* if they give rise to the same allele frequency trajectories. They will be called *strongly equivalent* if they *also* give rise to the same dynamics for the entire population, that is, the changes in population size or allele numbers are also conserved.

The problems investigated here are those of constructing the $\{w^*_{ij}\}$ from $\{w_{ij}\}$ and of determining the conditions under which they are equivalent or strongly equivalent. For expository convenience we will fully analyze a one-locus-two-allele model first. It will serve to illustrate our general methods.

Two alleles: Here $n = 2$. Let p be the frequency of the allele A_1, in some generation. Then $1 - p$ is the frequency of A_2 in that generation. If p' is the frequency of A_1 in the next generation,

$$p' = \frac{p^2 w_{11}(p) + p(1-p) w_{12}(p)}{\bar{w}(p)} \qquad (1)$$

where $\bar{w}(p) = p^2w_{11}(p) + 2p(1-p)w_{12}(p) + (1-p)^2w_{22}(p)$. Then $1-p'$ will be the frequency of A_2 in that generation. Let $f(p) \equiv p'$. If $v^*_{11}(p)$, $v^*_{12}(p)$, and $v^*_{22}(p)$ are *relative* fitnesses of A_1A_1, A_1A_2, and A_2A_2 respectively, which correspond to a new equivalent fitness set $\{w^*_{ij}(p) : i,j = 1,2\}$, then the v^*_{ij} have to satisfy the following equations:

$$f(p) = p^2v^*_{11}(p) + p(1-p)v^*_{12}(p); \tag{2a}$$

$$1 - f(p) = p(1-p)v^*_{12}(p) + (1-p)^2v^*_{22}(p). \tag{2b}$$

Since there are three unknowns–$v^*_{11}(p)$, $v^*_{12}(p)$, and $v^*_{22}(p)$–and only two equations, one of the three functions $v^*_{11}(p)$, $v^*_{12}(p)$, and $v^*_{22}(p)$ can be chosen arbitrarily. It is convenient, as will be apparent from the examples below, to let this be $v^*_{12}(p)$. Then we can express the other two functions as:

$$v^*_{11}(p) = \frac{f(p)}{p^2} - \frac{1-p}{p}v^*_{12}(p); \tag{3a}$$

$$v^*_{22}(p) = \frac{1-f(p)}{(1-p)^2} - \frac{p}{1-p}v^*_{12}(p). \tag{3b}$$

As should be expected for relative fitnesses, $p^2v^*_{11}(p) + 2p(1-p)v^*_{12}(p) + (1-p)^2v^*_{22}(p) = 1$.

In general, $\bar{w}^*(p) = p^2w^*_{11}(p) + 2p(1-p)w^*_{12}(p) + (1-p)^2w^*_{22}(p)$ can be arbitrary. This is incorporated into the equations for the transformed absolute fitnesses using $w^*_{ij}(p) = \bar{w}^*(p)v^*_{ij}(p) : i,j = 1,2$ (since the $v^*_{ij}(p)$ are only relative fitnesses). The resulting equations are:

$$w^*_{11}(p) = \bar{w}^*(p)\left(\frac{f(p)}{p^2} - \frac{1-p}{p}v^*_{12}(p)\right); \tag{4a}$$

$$w^*_{12}(p) = \bar{w}^*(p)v^*_{12}(p); \tag{4b}$$

$$w^*_{22}(p) = \bar{w}^*(p)\left(\frac{1-f(p)}{(1-p)^2} - \frac{p}{1-p}v^*_{12}(p)\right). \tag{4c}$$

Two functions, v^*_{12} and $\bar{w}^*(p)$ can be chosen arbitrarily. Now, since we were solving for three functions, $w^*_{11}(p)$, $w^*_{12}(p)$, and $w^*_{22}(p)$, and were subject to exactly one constraint, viz. the functional form of $f(p)$, there are at most two functions that could be arbitrarily specified. Therefore Equations 4a–c capture all the alternative fitness sets that are equivalent to $\{w_{11}(p), w_{12}(p), w_{22}(p)\}$. Of course, the form of these equations would have been different if we had chosen functions other than $v^*_{12}(p)$ and $\bar{w}^*(p)$ as the arbitrarily specified ones. If strong equivalence is desired, then $\bar{w}^*(p)$ must be equal to $\bar{w}(p)$. Then only one function, $v^*_{12}(p)$, can be arbitrarily chosen. We thus see that the mean fitness $\bar{w}^*(p)$ and the relative fitness of the heterozygote $v^*_{12}(p)$ are exactly the two functions which are required to fix the new fitness set. This pattern will recur throughout the models considered in this paper.

However, the fact that $\{w^*_{ij}\}$ has to be biologically interpreted as a set of absolute fitnesses imposes the following constraints: (i) $\bar{w}^*(p) \geq 0$; and (ii) $w^*_{ij}(p) \geq 0 : i,j = 1,2$. These are not independent constraints since (ii) implies (i). From (ii), the following constraints result from Equations 4a–c:

$$\frac{f(p)}{p^2} - \frac{1-p}{p}v^*_{12}(p) \geq 0; \tag{5a}$$

$$v^*_{12}(p) \geq 0; \tag{5b}$$

$$\frac{1-f(p)}{(1-p)^2} - \frac{p}{1-p}v^*_{12}(p) \geq 0; \tag{5c}$$

These restrict the $v^*_{12}(p)$ as follows:

$$0 \leq v^*_{12}(p) \leq \frac{\frac{1}{2} - |f(p) - \frac{1}{2}|}{p(1-p)} \tag{6}$$

if $f(p)$ is given. If the $w_{ij}(p)$ are given this can be written as:

$$0 \leq v^*_{12}(p) \leq \frac{w_{12}(p)}{\bar{w}(p)} + \min\left\{\frac{w_{11}(p)}{\bar{w}(p)}\frac{p}{1-p}, \frac{w_{22}(p)}{\bar{w}(p)}\frac{1-p}{p}\right\}. \tag{7}$$

A sufficient condition for this is clearly

$$v^*_{12}(p) \leq \frac{w_{12}(p)}{\bar{w}(p)}. \tag{8}$$

However, note that this is not a necessary condition.

Ordinarily, biological assumptions about $\{w^*_{ij} : i,j = 1,2\}$ will determine the forms of $\bar{w}^*(p)$ and $v^*_{12}(p)$. The importance of this method is that such biological assumptions can often simply be translated into constraints on these functions. We can also turn the constraints in equations such as (5a–c) around and ask: what are the constraints on the dynamics of the system that are imposed on a model with certain biological assumptions? The following examples illustrate this process:

1. Suppose that the heterozygote is to be lethal according to the new fitness set. Then $w^*_{12}(p) = 0$. Set $v^*_{12}(p) = 0$. This gives:

$$w^*_{11}(p) = \bar{w}^*(p)\frac{f(p)}{p^2}; \tag{9a}$$

$$w^*_{12}(p) = 0; \tag{9b}$$

$$w^*_{22}(p) = \bar{w}^*(p)\frac{1-f(p)}{(1-p)^2}; \tag{9c}$$

and $\bar{w}^*(p)$ can still be chosen arbitrarily. Since $v^*_{12}(p) = 0 \leq w_{12}(p)/\bar{w}(p)$, Equation 8 is automatically satisfied. This also means that any dynamics can be explained by a frequency-dependent model in which the heterozygote is lethal.

2. Suppose we assume dominance of A_1 over A_2 (with respect to fitness). Then, $w^*_{11}(p) = w^*_{12}(p)$. Then $v^*_{12}(p) = f(p)/p$ and

$$w^*_{11}(p) = w^*_{12}(p) = \bar{w}^*(p)\frac{f(p)}{p}; \tag{10a}$$

$$w^*_{22}(p) = \bar{w}^*(p)\frac{1-2f(p)+pf(p)}{(1-p)^2}. \tag{10b}$$

TABLE 1

Function choices for alternative fitness sets presented by DENNISTON and CROW (1990)

	Model A			Transformations			Model B		
Set	w_{11}	w_{12}	w_{22}	$\bar{w}$	v_{12}^*	$\bar{w}^*$	w_{11}^*	w_{12}^*	w_{22}^*
a	1	1	$1 - s$	$1 - s(1-p)^2$	$\frac{p^2}{\bar{w}^*}$	$p(1 - s(1-p)^2)$	$1 - p(1-p)$	p^2	$p^2 + p(1-s)$
b[a]	$1 - p$	1	p	$3p(1-p)$	$\frac{1 + p(1-p)}{\bar{w}^*}$	$3p(1-p)$	$p(1-p)$	$1 + p(1-p)$	$p(1-p)$
c	$\frac{a-(1-p)}{p}$	1	$\frac{b-p}{1-p}$	$ap + b(1-p)$	$\frac{ab}{\bar{w}^*}$	$[ap + b(1-p)]^2$	a^2	ab	b^2
d[a]	$\frac{2a-(a-b+1)(1-p)}{p}$	1	$\frac{2b+(a-b-1)p}{1-p}$	$2[ap + b(1-p)]$	$\frac{a+b}{\bar{w}^*}$	$2[ap + b(1-p)]$	$2a$	$a + b$	$2b$
e	$\frac{a+bp}{a-b(1-p)}$	1	$\frac{a-b(1-p)}{a+bp}$	$\frac{a^2}{(a+bp)(a-b(1-p))}$	$\frac{a}{\bar{w}^*}$	$(a-b) + 2bp$	$a + b$	a	$a - b$

The fitness models A and B in each row give rise to the same allele frequency dynamics, that is, they are equivalent. The transformations in the middle column are from model A to model B. (Reverse transformations would not have the same functional form.) The biological interpretation of these models is discussed in DENNISTON and CROW (1990). In model A, fitness sets (**a**) and (**b**) always satisfy the biological constraint that the fitness functions be non-negative. For (**c**) the condition is a, $b \geq 1$; for (**d**) a, $b \geq 0$ and $a + b \geq 1$; for (**e**) $a \leq -|b|$ or $a \geq |b|$, a and b are not both 0. In model B fitness set (**b**) always satisfies the non-negativity constraint. Set (**a**) has to satisfy $s \leq 1$; (**c**) has to satisfy $ab \geq 0$; (**d**) a, $b \geq 0$; and for (**e**) $a \geq |b|$.

[a] These equivalent models are also strongly equivalent. This is seen from the equality of $\bar{w}$ and $\bar{w}^*$.

Obviously a similar set can be obtained if $w_{12}^*(p)$ is assumed to be equal to $w_{22}^*(p)$, that is A_2 is dominant over A_1. The constraint of Equation 6 is equivalent to $f(p) \leq 1/(2 - p)$. If this constraint is not satisfied then no model which exhibits dominance of A_1 over A_2 can explain the dynamics. Using Equation 1, this constraint can be written in the form:

$$p^2(w_{11}(p) - w_{12}(p)) \leq (1-p)w_{22}(p) \qquad (10c)$$

which can be used to determine the possibility of the existence of alternative equivalent fitness sets with the dominance of A_1 over A_2.

3. Suppose that A_2 is lethal in the homozygous state. Set $w_{22}^*(p) = 0$. This gives $v_{12}^*(p) = (1 - f(p))/p(1-p)$. Thus:

$$w_{11}^*(p) = \bar{w}^*(p)\frac{2f(p) - 1}{p^2}; \qquad (11a)$$

$$w_{12}^*(p) = \bar{w}^*(p)\frac{1 - f(p)}{p(1-p)}; \qquad (11b)$$

$$w_{22}^* = 0. \qquad (11c)$$

Once again, $\bar{w}^*(p)$ can be chosen arbitrarily. Equation 6 is now equivalent to the requirement that $f(p) \geq \frac{1}{2}$ Using Equation 1, this reduces to the rather simple requirement that:

$$p^2 w_{11}(p) \geq (1-p)^2 w_{22}(p). \qquad (11d)$$

The asymmetry between the alleles seen here reflects the asymmetry of the requirement that the A_2 allele is lethal in (and only in) the homozygous state.

4. Suppose the A_1 allele increases fitness (viability) additively. Then $w_{12}^*(p) - w_{22}^*(p) = w_{11}^*(p) - w_{12}^*(p)$, and $v_{12}^*(p) = f(p)(1-p)/p + (1 - f(p))p/(1-p)$, giving:

$$w_{11}^*(p) = \bar{w}^*(p)\left(\frac{2f(p)}{p} - 1\right); \qquad (12a)$$

$$w_{12}^*(p) = \bar{w}^*(p)\left(f(p)\frac{1-p}{p} + (1 - f(p))\frac{p}{1-p}\right); \qquad (12b)$$

$$w_{22}^*(p) = \bar{w}^*(p)\left(\frac{2(1 - f(p))}{1-p} - 1\right). \qquad (12c)$$

Equation 6 now becomes equivalent to $(p/2) \leq f(p) \leq (1 + p)/2$. In terms of the initial fitness set, this can be written as $p/2 \leq p^2 w_{11}(p)/\bar{w}(p) + p(1-p)w_{12}(p)/\bar{w}(p) \leq (p + 1)/2$.

In all these examples $\bar{w}^*(p)$ could be arbitrarily chosen. It could, therefore, be set equal to $\bar{w}(p)$. The new fitness sets that would then result would preserve the total population dynamics besides the allelic trajectories. These alternative fitness sets should thus be strongly equivalent. However, as DENNISTON and CROW (1990) have pointed out, there are examples of alternative fitness sets that are equivalent but not strongly equivalent. The choices of $v_{12}^*(p)$ and $\bar{w}^*(p)$ for their examples are presented in Table 1. Only the transformations (**b**) and (**d**) there maintain strong equivalence. (The effects of the non-negativity constraint (Equation 6) are indicated in the caption to the table.)

It is clear from Equations 4a–c that, given any fitness set, an infinite number of equivalent or strongly equivalent frequency-dependent alternative fitness sets can be constructed. However, it is also clear that, in general, there need not always be a frequency-independent (or constant) fitness set equivalent, let alone strongly equivalent, to an arbitrarily given fitness set. Equa-

tions 4a–c require the estimation of *three* quantities: $w^*_{11}(p)$, $w^*_{12}(p)$, $w^*_{22}(p)$, and at most *only two* functions can be arbitrarily chosen. In general, exactly two of the new fitnesses can be taken to be constant by setting any two of the Equations 4a, 4b and 4c equal to (different) constants. This is consistent with the fact that for some types of allele frequency trajectories, such as cycles, there can be no constant genotypic fitness set that can give rise to them: for one locus models, if fitnesses are constant, the mean fitness of a population ($\bar{w}$) is non-decreasing [see, *e.g.*, EWENS (1979), pp. 40–45]. However, if the allele frequencies oscillate, so must the mean fitness. Thus there can be no constant (*i.e.*, frequency-independent) fitness set that would give rise to such dynamics. Another consequence of having exactly two functions that can be arbitrarily chosen is that DENNISTON and CROW's (1990) linear combination transformation, which allows two independent parameters, A and B, to be arbitrary functions of p, constitutes a complete specification of allowed transformations for the two-allele case, though not in general.

***n* Alleles:** The extension of these arguments to models with n alleles at one locus is straightforward and amounts to little more than a change of notation. Assuming, as usual, that $w_{ij}(p) = w_{ji}(p)$, there are $n(n+1)/2$ different function $w^*_{ij}(p)$ to be determined. The equations for allele frequency change impose $(n-1)$ independent constraints on the system. There are thus $n(n+1)/2 - (n-1) = n(n-1)/2 + 1$ functions that can be chosen arbitrarily. Let p'_i be the frequency of the i^{th} allele in the next generation. Then, for each i,

$$p'_i = \frac{\sum_{j=1}^{n} p_i p_j w_{ij}(p)}{\sum_{k=1}^{n}\sum_{j=1}^{n} p_k p_j w_{kj}(p)} = \frac{w_i(p)p_i}{\bar{w}(p)}. \qquad (13)$$

In analogy to the two-allele case, define $f_i(p) \equiv p'_i$. Introducing new relative fitnesses $v^*_{ij}(p)$ as before:

$$f_i(p) = \sum_{j=1}^{n} p_i p_j v^*_{ij}(p) = \sum_{\substack{j=1 \\ j\neq i}}^{n} p_i p_j v^*_{ij}(p) + p_i^2 v^*_{ii}(p). \qquad (14)$$

It is trivially true that $\sum_{i,j} p_i p_j v^*_{ij}(p) = \sum_i f_i(p) = 1$. Suppose that the $v^*_{ij}(p) : i, j = 1, \ldots, n;\ i \neq j$, are arbitrarily chosen. There are $n(n-1)/2$ such functions. Finally, with $w^*_{ij}(p) = \bar{w}^*(p)v^*_{ij}(p)$, we get $\bar{w}^*(p) = \sum_{i,j} p_i p_j w^*_{ij}(p)$. $\bar{w}^*(p)$ can also be chosen as an arbitrary function of the frequencies (which provides a measure of the population growth). There are thus exactly $n(n-1)/2 + 1$ arbitrary functions, as required.

From Equation 13, the remaining functions, that is, the $v^*_{ii} : i = 1, \ldots, n$, are calculated:

$$v^*_{ii}(p) = \frac{f_i(p)}{p_i^2} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{p_j}{p_i} v^*_{ij}(p). \qquad (15)$$

The final equations for generating the equivalent fitness set $\{w^*_{ij} : i, j = 1, \ldots, n\}$ now become:

$$w^*_{ij}(p) = \bar{w}^*(p)v^*_{ij}(p) : i, j = 1, \ldots, n, i \neq j; \qquad (16a)$$

$$w^*_{ii}(p) = \bar{w}^*(p)\left(\frac{f_i(p)}{p_i^2} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{p_j}{p_i} v^*_{ij}(p)\right) : i = 1, \ldots, n. \qquad (16b)$$

As before, fixing the mean fitness and the relative fitnesses of the heterozygotes specifies this set fully.

Once again, the relevant biological constraints are: (i) $\bar{w}^*(p) \leq 0$; and (ii) $w^*_{ij}(p) \geq 0$: $i, j = 1, \ldots, n$. The second constraint implies the first and requires:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} p_j v^*_{ij}(p) \leq \frac{f_i(p)}{p_i}. \qquad (17)$$

The argument that led to Equation 8 now shows that $v^*_{ij}(p) \leq w_{ij}(p)/\bar{w}(p)$ is always a sufficient (but not a necessary) criterion.

As before, additional biological information can help to fix the $v^*_{ij}(p)$. In particular, it is easy to see that if, for every pair of alleles, one is dominant over the other (with respect to fitness), the $v^*_{ij}(p)$ are completely specified because such an assumption adds $n(n-1)/2$ independent constraints of the form $v^*_{ij}(p) = v^*_{ii}(p)$ (with $v^*_{ij}(p) = v^*_{ji}(p)$ for $i \neq j$). This, together with an arbitrary $\bar{w}^*(p)$, fixes the $w^*_{ij}(p)$. If $\bar{w}^*(p)$ is chosen equal to $\bar{w}(p)$, then the original $w_{ij}(p)$ are recovered.

In this generalized scheme, the multiplicative and additive transformations of DENNISTON and CROW (1990) are obtained from:

Additive:

$$v^*_{ij}(p) = \frac{w_i(p)}{\bar{w}(p)} + \frac{w_j(p)}{\bar{w}(p)} - 1 : i, j = 1, \ldots, n;\ i \neq j; \qquad (18a)$$

Multiplicative:

$$v^*_{ij}(p) = \frac{w_i(p)w_j(p)}{\bar{w}(p)} : i, j = 1, \ldots, n;\ i \neq j. \qquad (18b)$$

In either case $\bar{w}^*(p)$ can be chosen arbitrarily. These transformations lead to strongly equivalent alternative fitness sets, if $\bar{w}^*(p)$ is chosen to be equal to $\bar{w}(p)$. To satisfy the non-negativity constraint, note that Equation 17 is equivalent, for the multiplicative case, to $\bar{w}(p) - w_i(p)p_i \leq 1$ $(i = 1, \ldots, n)$. This implies that $\bar{w}(p) \leq n/(n-1)$ for all p. For the additive case, $\bar{w}(p) \leq 2w_i(p)$ $(i = 1, \ldots, n)$ for all p is a necessary and sufficient condition for the transformation to be a valid one. This, in turn, is equivalent to the requirement that $p_i/2 \leq f_i(p)$ $(i = 1, \ldots, n)$ for all p.

TWO LOCI

Two Alleles: In multiple locus models, allele frequencies do not suffice to characterize the population dynamics. The basic variables are the gametic or haplotype

frequencies. Once again we will first solve the two-allele case fully in order to illustrate our methods. Let A_1 and A_2 be the two alleles at locus 1, B_1 and B_2 be those at locus 2. Let h_i: $i = 1, \ldots, 4$ be the haplotypes A_1B_1, A_1B_2, A_2B_1, and A_2B_2, respectively. Genotypes can be written in the form $h_i \mid h_j$. (According to the usual notation, the ij-th genotype would consist of both the $h_i \mid h_j$ and the $h_j \mid h_i$ genotypes in our notation. We will follow the same convention for n loci.) Let H be the set of haplotypes, that is, $H = \{A_1B_1, A_1B_2, A_2B_1, A_2B_2\}$. Let $p(h_i)$ be the frequency of the haplotype h_i, and $p(h_i \mid h_j)$ be the frequency of the genotype $h_i \mid h_j$ with $i, j = 1, \ldots, 4$. During gamete formation, recombination of the alleles at the two loci either takes place or does not. Let the probability of recombination taking place be c. Let $p'(h_i)$ be the frequency of the i-th haplotype in the next generation. Then, for example,

$$\begin{aligned} p'(A_1B_1) = (1/\bar{w})[&c(p(A_1B_1 \mid A_1B_1)w(A_1B_1 \mid A_1B_1) \\ &+ p(A_1B_2 \mid A_1B_1)w(A_1B_2 \mid A_1B_1) \\ &+ p(A_1B_1 \mid A_2B_1)w(A_1B_1 \mid A_2B_1) \\ &+ p(A_1B_2 \mid A_2B_1)w(A_1B_2 \mid A_2B_1)) \\ &+ (1-c)(p(A_1B_1 \mid A_1B_1)w(A_1B_1 \mid A_1B_1) \\ &+ p(A_1B_1 \mid A_2B_1)w(A_1B_1 \mid A_2B_1) \\ &+ p(A_1B_1 \mid A_1B_2)w(A_1B_1 \mid A_1B_2) \\ &+ p(A_1B_1 \mid A_2B_2)w(A_1B_1 \mid A_2B_2))] \end{aligned} \tag{19}$$

where $w(h_i \mid h_j)$ is the fitness of $h_i \mid h_j$ and $\bar{w}$ is again the mean fitness. In order to facilitate our discussion of the n loci case, we will now introduce some new notation in this more familiar context. Let $r_1(h_i \mid h_j) = h_k \mid h_l$: $i, j, k, l = 1, \ldots, 4$ be the genotype that results when recombination takes place between h_i and h_j, and $r_2(h_i \mid h_j) = h_i \mid h_j$ be the resultant genotype when no recombination takes place. This *defines* the recombination processes r_1 and r_2. Thus, for example:

$$r_1(A_1B_1 \mid A_2B_2) = A_1B_2 \mid A_2B_1; \tag{20a}$$

$$r_2(A_1B_1 \mid A_2B_2) = A_1B_1 \mid A_2B_2. \tag{20b}$$

Since only two loci are involved, r_1 and r_2 are the only possible recombination processes. The probability of these recombination processes will be indicated by $c(r_1) = c$, $c(r_2) = 1 - c$. Using this notation we can rewrite Equation 19 as:

$$\begin{aligned} p'(h_1) = (1/\bar{w})[&c(r_1)(p(r_1(h_1 \mid h_1))w(r_1(h_1 \mid h_1)) \\ &+ p(r_1(h_1 \mid h_2))w(r_1(h_1 \mid h_2)) \\ &+ p(r_1(h_1 \mid h_3))w(r_1(h_1 \mid h_3)) \\ &+ p(r_1(h_1 \mid h_4))w(r_1(h_1 \mid h_4))) \\ &+ c(r_2)(p(r_2(h_1 \mid h_1))w(r_2(h_1 \mid h_1)) \\ &+ p(r_2(h_1 \mid h_2))\ w(r_2(h_1 \mid h_2)) \\ &+ p(r_2(h_1 \mid h_3))w(r_2(h_1 \mid h_3)) \\ &+ p(r_2(h_1 \mid h_4))w(r_2(h_1 \mid h_4)))]. \end{aligned} \tag{21}$$

In general, the $p(h_i)$ are given by:

$$p'(h_i) = \frac{1}{\bar{w}} \sum_{k=1}^{2} (c(r_k) \sum_{j=1}^{4} p(r_k(h_i \mid h_j))w(r_k(h_i \mid h_j))). \tag{22}$$

Note that Equation 22 makes no assumptions about linkage equilibrium. As before, define $f_i(p) \equiv p'(h_i)$, where p is now a vector of haplotype frequencies. (Our notation underscores the fact that the $w(h_i \mid h_j)$ are frequency-dependent.) Let $v^*(h_i \mid h_j)$ be the relative fitness of $h_i \mid h_j$ corresponding to a new fitness set $\{w^*(h_i \mid h_j)\}$. We wish to solve for all possible fitness sets $\{w^*(h_i \mid h_j) : i, j = 1, \ldots, 4\}$ equivalent or strongly equivalent to $\{w(h_i \mid h_j)\}$. Then

$$\begin{aligned} f_i(p) &= \sum_{k=1}^{2} \left(c(r_k) \sum_{j=1}^{4} p(r_k(h_i \mid h_j))v^*(r_k(h_i \mid h_j)) \right) \\ &= \sum_{k=1}^{2} \left(c(r_k) \sum_{\substack{j=1 \\ j \neq i}}^{4} p(r_k(h_i \mid h_j))v^*(r_k(h_i \mid h_j)) \right) \\ &\quad + \sum_{k=1}^{2} c(r_k)p(r_k(h_i \mid h_i))v^*(r_k(h_i \mid h_i)) \\ &= \sum_{k=1}^{2} \left(c(r_k) \sum_{\substack{j=1 \\ j \neq i}}^{4} p(r_k(h_i \mid h_j))v^*(r_k(h_i \mid h_j)) \right) \\ &\quad + p(h_i \mid h_i)v^*(h_i \mid h_i). \end{aligned} \tag{23}$$

This gives:

$$v^*(h_i \mid h_i) = \frac{f_i(p) - \sum_{k=1}^{2} (c(r_k) \sum_{\substack{j=1 \\ j \neq i}}^{4} p(r_k(h_i \mid h_j))v^*(r_k(h_i \mid h_j)))}{p(h_i \mid h_i)}. \tag{24}$$

As in the one-locus case, if $w^*(h_i \mid h_j) = \bar{w}^*(h_i \mid h_j)$, assuming $w^*(h_i \mid h_j) = w^*(h_j \mid h_i)$, there are $4^2 - (4 \times 3)/2 = 10$ such functions to be calculated. Since the equations for the $f_i(p)$ put three constraints, seven of these functions can be arbitrarily specified. Now suppose the $v^*(h_i \mid h_j)$, : $i \neq j$ are specified arbitrarily. This gives six functions. If $\bar{w}^*$ is also chosen arbitrarily, finally, the transformations are fully specified. The final equations are:

$$w^*(h_i \mid h_i) = \bar{w}^* \frac{f_i(p) - \sum_{k=1}^{2} (c(r_k) \sum_{\substack{j=1 \\ j \neq i}}^{4} p(r_k(h_i \mid h_j))v^*(r_k(h_i \mid h_j)))}{p(h_i \mid h_i)}, \tag{25a}$$

$$w^*(h_i \mid h_j) = \bar{w}^* v^*(h_i \mid h_j)\text{: } i \neq j. \tag{25b}$$

Thus, once again, fixing the mean fitness and the relative fitnesses of the heterozygotes specifies the new fitness set uniquely and strong equivalence is maintained if the new mean fitness equals the old one. DENNISTON and CROW (1990) suggest the same procedure for this case (see their Table 2). As before,

the necessary and sufficient biological criterion is that the fitnesses $\bar{w}^*$ and $\bar{w}^*(h_i|h_j)$ be non-negative. The latter constraint implies the former and can be written as:

$$\sum_{\substack{j=1\\ j\neq i}}^{4} p(h_i|h_j)v^*(h_i|h_j) \leq \frac{f_i(p)}{p(h_i|h_i)}. \tag{25c}$$

Consider a model with complete dominance with respect to fitness. In this case, the 10 different genotypes split into four sets, in each of which the fitness of each of the members is the same. These sets are:

(i) $\{A_iB_i|A_1B_1,\ A_1B_1|A_1B_2,\ A_1B_1|A_2B_1,$
$A_1B_1|A_2B_2,\ A_1B_2|A_2B_1\}$

(ii) $\{A_1B_2|A_2B_2,\ A_1B_2|A_1B_2\}$

(iii) $\{A_2B_1|A_2B_2,\ A_2B_1|A_2B_1\}$

(iv) $\{A_2B_2|A_2B_2\}$,

where A_1 and B_1 are dominant over A_2 and B_2, respectively. A set of k genotypes, which all have the same fitnesses, imposes $k-1$ constraints on the transformed fitnesses. This gives $4+1+1=6$ constraints. Assuming we want to have $\bar{w}=\bar{w}^*$, we had freedom to choose six arbitrary functions. The 6 new constraints exactly specify the six $v^*(h_i|h_j): i\neq j$ which were up for arbitrary choice. For instance, from the equality of all the fitnesses of the members of set (i), we obtain with α being any member of set (i):

$$v^*(\alpha) = v^*(A_1B_1|A_1B_1) \tag{26}$$
$$= \frac{f_{A_1B_1}(p) - \sum_{r\in R}\sum_{\substack{h\in H\\ h\neq A_1B_1}} c(r)p(r(A_1B_1|h))v^*(r(A_1B_1|h))}{p(A_1B_1|A_1B_1)}$$
$$= \frac{f_{A_1B_1}(p)}{p(A_1B_1|A_1B_1)} - \frac{\mathcal{A}}{p(A_1B_1|A_1B_1)} - \frac{\mathcal{B}}{p(A_1B_1|A_1B_1)}$$

where

$$\mathcal{A} = p(A_1B_1|A_1B_2)v^*(A_1B_1|A_1B_2) + p(A_1B_1|A_2B_1)v^*(A_1B_1|A_2B_1)$$
$$\mathcal{B} = cp(A_1B_2|A_2B_1)v^*(A_1B_2|A_2B_1) - (1-c)p(A_1B_1|A_2B_2)v^*(A_1B_1|A_2B_2)$$

This gives, after simplification:

$$v^*(\alpha) = \frac{f_{A_1B_1}(p)}{p(A_1B_1) + c(p(A_1B_2|A_2B_1) - p(A_1B_1|A_2B_2))}. \tag{27a}$$

Similarly, we get, for β a member of our set (ii) above, and γ a member of set (iii):

$$v^*(\beta) = \frac{f_{A_1B_2}(p)}{p(A_1B_2) - c(p(A_1B_2|A_2B_1) - p(A_1B_1|A_2B_2))}; \tag{27b}$$

$$v^*(\gamma) = \frac{f_{A_2B_1}(p)}{p(A_2B_1) - c(p(A_1B_2|A_2B_1) - p(A_1B_1|A_2B_2))}. \tag{27c}$$

Denniston and Crow (1990) present partial solutions of this dominance-preserving model. However, they assume linkage equilibrium which we do not.

Arbitrary numbers of alleles: Let a be the number of alleles at the A locus and b be the number of alleles at the B locus. Then the number of haplotypes is $m=ab$. Let H be the set of haplotypes (which can be written as $H=\{h_i: i=1,\ldots,m\}$), R be $\{r_1, r_2\}$, where r_1 and r_2 are the two recombination processes defined above. Then, following the same reasoning as above,

$$p'(h_i) = \frac{1}{\bar{w}}\sum_{k=1}^{2}\sum_{j=1}^{m} c(r_k)p(r_k(h_i|h_j))w(r_k(h_i|h_j)), \tag{28}$$

where $w(h_i|h_j)$ is the fitness of the genotype $h_i|h_j$, and $\bar{w}=\sum_{i,j}p(h_i|h_j)w(h_i|h_j)$. As before, let $f_i(p)\equiv p'(h_i): i=1,\ldots,m$. Then:

$$f_i(p) = \frac{1}{\bar{w}}\sum_{r\in R}\sum_{h\in H} c(r)p(r(h_i|h))w(r(h_i|h)). \tag{29}$$

Now there are $m(m+1)/2$ different functions $w^*(h_i|h_j)$ to be solved for, and $(m-1)$ constraints from the $(m-1)$ independent equations for the $f_i(p)$. This leaves $m(m-1)/2+1$ functions that can be arbitrarily chosen. Thus we can again solve for $w^*(h_i|h_i)$, with $v^*(h_i|h_j): i\neq j$ and $\bar{w}^*$ as arbitrarily chosen functions. The equations for $w^*(h_i|h_j)$ then are:

$$w^*(h_i|h_i) = \bar{w}^*\frac{f_i(p) - \sum_{k=1}^{2} c(r_k)\left(\sum_{\substack{j=1\\ j\neq 1}}^{m} p(r_k(h_i|h_j))v^*(r_k(h_i|h_j))\right)}{p(h_i|h_i)}, \tag{30a}$$

$$w^*(h_i|h_j) = \bar{w}^*v^*(h_i|_j): i\neq j. \tag{30b}$$

Thus once again the mean fitness and the relative fitnesses of the heterozygotes fix the new alternative set. Since $\bar{w}^*$ and $w^*(h_i|h_j)$ are fitnesses, these functions are constrained to be non-negative. The form of the constraint is the same as Equation 25c. That complete dominance fully specifies the $v^*(h_i|h_j)$ can be shown by an easy extension of the argument given above for two alleles.

n LOCI

Following the notation of the two-locus, multiple allele case, let H be the set of haplotypes, and R the set of

all recombination processes. If the i-th locus has a_i alleles, then there are $m = \prod_{i=1}^{n} a_i$ haplotypes. There are 2^{n-1} elements in R. [R has the mathematical structure of a group where the identity element, I, is the process $I(h_i|h_j) = (h_i|h_j)$, that is the case when during meiosis no recombination takes place, and each element its own inverse: $r(r(h_i|h_j)) = (h_i|h_j)$, because imposing the same recombination transformation twice gives the starting genome back (that is, $r^2 = I$). Actually R is a representation of S_2^{n-1}.] Let $p(h_i|h_j)$ denote the frequency of genotype $(h_i|h_j)$ in the population. Let $p(h_i)$ denote the frequency of haplotype h_i. Let $c(r)$, for $r \in R$, denote the probability of the recombination event r to occur. Thus $\sum_{r\in R} c(r) = 1$. Let $w(h_i|h_j)$ be the fitness of an organism with the genome $(h_i|h_j)$. For notational convenience, as in the two-locus case, we assume that fitnesses are generally frequency-dependent without indicating that fact explicitly in our notation. The equation for the haplotype frequency change is:

$$p'(h_i) = \frac{\sum_{r\in R}\sum_{h_j\in H} c(r)p(r(h_i|h_j))w(r(h_i|h_j))}{\bar{w}} \tag{31}$$

where

$$\begin{aligned}\bar{w} &= \sum_{h_i\in H}\sum_{r\in R}\sum_{h_j\in H} c(r)p(r(h_i|h_j))w(r(h_i|h_j))\\ &= \sum_{r\in R} c(r)\sum_{h_i\in H}\sum_{h_j\in H} p(r(h_i|h_j))w(r(h_i|h_j))\\ &= \sum_{r\in R} c(r)\sum_{h_i\in H}\sum_{h_j\in H} p(h_i|h_j)w(h_i|h_j)\\ &= \sum_{h_i\in H}\sum_{h_j\in H} p(h_i|h_j)w(h_i|h_j).\end{aligned} \tag{32}$$

As before, let $f_i(p) \equiv p'(h_i)$: $i = 1, \ldots, m$. Once again, for a given set of fitnesses $w(h_i|h_j)$ we want to find an alternative set of fitnesses $w^*(h_i|h_j)$ which give the same allele frequency trajectories and population dynamics. Since m is the total number of possible haplotypes, we have to solve for $m(m+1)/2$ functions $w^*(h_i|h_j)$, with the equations for the allele trajectories imposing $m-1$ independent constraints, leaving $m(m-1)/2 + 1$ functions to be arbitrarily chosen. These can again be the functions $v^*(h_i|h_j)$ with $i \neq j$, and $\bar{w}^*$, the population fitness. As before, $v^*(h_i|h_i)$ will be calculated from:

$$f_i(p) = \sum_{r\in R}\sum_{h_j\in H} c(r)p(r(h_i|h_j))v^*(r(h_i|h_j)). \tag{33}$$

Since, for all $h_i \in H$, $r \in R$ we have $r(h_i|h_i) = (h_i|h_i)$ and for all $h_l, h_k \in H$, $l \neq k$, we have $r(h_l|h_k) \neq (h_i|h_i)$ we can rewrite Equation 24 as:

$$\begin{aligned}f_i(p) &= \sum_{r\in R}\sum_{h_j\neq h_i} c(r)p(r(h_i|h_j))v^*(r(h_i|h_j))\\ &\quad + \sum_{r\in R} c(r)p(r(h_i|h_i))v^*(r(h_i|h_i))\\ &= \sum_{r\in R}\sum_{h_j\neq h_i} c(r)p(r(h_i|h_j))v^*(r(h_i|h_j))\\ &\quad + \sum_{r\in R} c(r)p(h_i|h_i)v^*(h_i|h_i)\\ &= \sum_{r\in R}\sum_{h_j\neq h_i} c(r)p(r(h_i|h_j))v^*(r(h_i|h_j))\\ &\quad + p(h_i|h_i)v^*(h_i|h_i).\end{aligned} \tag{34}$$

Thus

$$v^*(h_i|h_i) = \frac{f_i(p) - \sum_{r\in R}\sum_{h_j\neq h_i} c(r)p(r(h_i|h_j))v^*(r(h_i|h_j))}{p(h_i|h_i)} \tag{35}$$

where the right hand side includes only the $v^*(h_i|h_j)$ where $i \neq j$. Choosing a function $\bar{w}^*$, we get the general equations for $w^*(h_i|h_j)$:

$$w^*(h_i|h_j) = \bar{w}v^*(h_i|h_j)\text{: } (i \neq j); \tag{36a}$$

$$w^*(h_i|h_i) = \bar{w}\frac{f_i(p) - \sum_{r\in R}\sum_{h_j\neq h_i} c(r)p(r(h_i|h_j))v^*(r(h_i|h_j))}{p(h_i|h_i)}. \tag{36b}$$

Note, as before, the mean fitness and the relative fitnesses of the heterozygotes fix the new fitness set. As before, the biological interpretation of $\bar{w}^*$ and the $w^*(h_i|h_j)$ as fitnesses requires that they all be non-negative. In general this will put significant limits on the alternative equivalent fitness sets.

DISCUSSION

Our analysis underscores the point made by DENNISTON and CROW (1990) that allele (or haplotype) frequency trajectories alone do not suffice to specify the absolute or relative fitnesses of genotypes uniquely, provided that the conditions of discrete generations, random mating, infinite populations, and viability selection (or multiplicative fertilities) are met. This does not mean that these absolute fitnesses can never be estimated. However, more information than only the allele (or haplotype) frequency dynamics is necessary. Biological assumptions about the system can help this process. What our analysis shows is that one way in which biological assumptions could be used to obtain a unique fitness set from allele (or haplotype) trajectories is to fix the mean fitness of the population and the fitness of the heterozygotes. In principle, the former can be directly measured. The latter can sometimes be specified from what is known about the (developmental) genetics of an organism. For instance, if it is known that the fitness depends on some phenotypic trait that exhibits domi-

nance, the equality of some of the fitness functions $w^*(h_i|h_j)$ can potentially be used to compute the $v^*(h_i|h_j)$.

Note, however, that we were *not required* to choose the $\bar{w}^*$ and $v^*(h_i|h_j)$, $i \neq j$, as the arbitrary functions in our analysis. We chose them for convenience in the one-locus case and because these turned out to constitute the exact number of functions that could be arbitrarily specified. In certain situations it could turn out to be more convenient to work with the $v^*(h_i|h_i)$. If, for example, $v^*(h_i|h_i)$ can be directly measured, then our equations could be rewritten with $v^*(h_i|h_i)$ as the "independent" variables. However, the general procedure will remain the same. The constraint that all fitnesses (mean as well as individual) have to be non-negative puts a severe constraint on the possible transformations that give rise to alternative equivalent fitness sets. It is rare that a transformation will automatically satisfy this constraint for all parts of an allele (or haplotype) trajectory.

Finally, it is far from clear what the relaxation of our conditions would do to this analysis. In general overlapping generations would introduce entirely different types of dynamical equations. So would fertility selection with non-multiplicative fertilities. Assortative mating is usually modelled using genotypic frequencies. There is no *prima facie* reason to suppose that our manipulations, involving allele and haplotype frequencies, will be generally conserved when assortative mating is introduced. Finally, DENNISTON and CROW (1990) have observed that some of these transformations do not preserve the allele frequency dynamics in some stochastic one-locus models such as the Ethier-Nagylaki model (NAGYLAKI 1990). Whether any fitness set transformation preserves allele frequency dynamics in general stochastic models is an open question.

This work was partly supported by a grant to S.S. from the Sidney M. Edelstein Centre, Hebrew University, Jerusalem. We would like to thank JAMES F. CROW, WARREN J. EWENS and THOMAS NAGYLAKI for helpful comments on earlier versions of this analysis.

LITERATURE CITED

DENNISTON, C., and J. F. CROW, 1990 Alternative fitness models with the same allele frequency dynamics. Genetics **125:** 201–205.

EWENS W. J., 1979 *Mathematical Population Genetics.* Springer-Verlag, Berlin.

NAGYLAKI T., 1990 Models and approximations for random genetic drift. Theor. Popul. Biol. **37:** 192–212.

WRIGHT, S., and TH. DOBZHANSKY, 1946 Genetics of natural populations. XII. Experimental reproduction of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura.* Genetics **31:** 125–156.

Communicating editor: W. J. EWENS