On the Determination of Relative Fitness from Frequencies of Genotypes in Subsequent Generations[*]

Part I. Algebra of Reproduction and Selection in Populations with Discrete Generations

R. J. LORENZ

Institut für Biologie, Lehrstuhl für Genetik, der Universität Tübingen
und Bundesforschungsanstalt für Viruskrankheiten der Tiere, Tübingen (GFR)

Summary. An algebraic treatment of reproduction and selection processes in populations of diploids with discrete generation cycles is presented. The main objective was a determination of the conditions under which the selection coefficients for the various genotypes can be determined solely on the basis of known frequencies of genotypes in two subsequent generations. In part two, to be published later, the statistical properties of a special fitness estimator, proposed by Hayman, are considered. Confidence intervals and tests of significance concerning hypotheses about fitness are established and checked by computer simulation studies.

Introduction

This paper was stimulated by extensive studies in predominantly selfpollinated species by Allard and his coworkers. These studies were concerned with generation-by-generation estimates of two kinds of parameters whose values have implications concerning population structure and the evolutionary potential of the populations under question. The first is the proportion of selfing versus outcrossing, the second are the relative viabilities of the two homozygotes and the heterozygote at a diallelic locus.

The experimental basis for this task was obtained by annual censuses of frequencies of the three genotypes together with a separate experiment for the estimation of the proportion of selfing, which is, however, not independent in the statistical sense, from the censuses.

The present author took special interest in two types of problems. The first is concerned with the inherent logic of inference of experiments such as Allard's. The total process leading from the genotype composition in one generation to that in the next one is a sequence of different reproductive and selective componental processes. A given pair of arrays of genotype frequencies in the parental and offspring generation can — in general — be explained by quite different hypotheses about what has happened between the censuses, if further conditions cannot be assumed to be satisfied. Unfortunately the censuses alone give no indication with respect to this. In other words: it doesn't seem to be possible to prove the conditions for the estimation procedure and, at the same time, to estimate the parameters from the body of census data alone.

A further point of interest in the same context is, that the estimation of outcrossing presupposes, that fitnesses, which just are under question, satisfy certain conditions. On the other hand, the estimation of fitnesses presupposes that, for example, the amount of outcrossing is the same for each genotype. This last condition is certainly not satisfied in general (Harding and Tucker, 1964).

In part one of this paper an algebraic representation of reproductive and selective processes is given by means of operators, which make the transformation process for the arrays of genotype frequencies more transparent. The problem of whether unique inference can be drawn on the type of selection force at work then reduces to considering the algebraic properties of the resulting systems of linear or nonlinear equations.

In part two the statistical properties of the fitness estimator, which is due to Hayman (1953), is considered in more detail. This part will be published in a subsequent paper. Especially the power functions of some tests of significance for the null hypotheses, that the fitnesses of the genotypes do not differ, are evaluated by computer simulations. Furthermore, confidence intervals are established for fitnesses.

1. Genotypic Vectors and Operators of Reproduction

Consider two consecutive discrete generations n and $n + 1$ of an infinite population of diploids. The frequencies of genotypes (with respect to any number of loci) are represented by the genotype vectors

$$F_n = [f_n^{(1)}, f_n^{(2)}, \ldots, f_n^{(k)}] \tag{1.1}$$

and

$$F_{n+1} = [f_{n+1}^{(1)}, f_{n+1}^{(2)}, \ldots, f_{n+1}^{(k)}] \tag{1.2}$$

where

a) $\sum_{i=1}^{k} f_n^{(i)} = 1$ and   b) $\sum_{i=1}^{k} f_{n+1}^{(i)} = 1$.   (1.3)

We have $k = 3$ for one locus and $k = 10$ or 9 for two loci according to whether or not attraction and repulsion heterozygotes are considered different classes. Sometimes the frequencies of genotypes will be labeled by letters instead of integers, such as $f_n^{(x)}, f_n^{(y)}, f_n^{(z)}$.

The vector F_n corresponds to a point in a space of k dimensions, the position of which is characterized by the condition $f_n^{(i)} \geq 0$ $(i = 1, \ldots, k)$ together with (1.3.a). Thus this point is situated in the $(k-1)$-dimensional hyperplane with equation (1.3.a), in particular in the closed region $f_n^{(i)} \geq 0$ of that plane. In the simplest case (one locus, two alleles: $k = 3$) this is the set of points inside of an equilateral triangle (including its boundary), the edges of which coincide with the points $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ of a three-dimensional cartesian coordinate system. F_{n+1} corresponds to a point in the same set.

The transition from generation n to generation $n + 1$ may be considered to be composed of two steps. In the first a vector of gamete genotypes

$$G_n = [g_n^{(1)}, g_n^{(2)}, \ldots, g_n^{(r)}]$$ (1.4)

can be derived from F_n. We have $r = 2$ in the one locus case, $r = 4$ for two loci, and $r = 2^t$ for t loci, each with two alleles. Every $g_n^{(j)}$ is a function of some of the $f_n^{(i)}$ and possibly of some further parameters ϑ_i (for instance the frequency of recombination in case of two or more loci). In general $g_n^{(j)}$ is therefore of the form

$$g_n^{(j)} = g_n^{(j)} (f_n^{(1)}, \ldots, f_n^{(k)}; \vartheta_1, \vartheta_2, \ldots) .$$ (1.5)
$$(j = 1, 2, \ldots, r)$$

It is

a) $\sum_{j=1}^{r} g_n^{(j)} - 1 = 0$ and   b) $g_n^{(j)} \geq 0$.   (1.6)

This system of equations (1.5) may be regarded as an operator Γ, which, when applied to F_n, produces G_n:

$$G_n = \Gamma(F_n) .$$ (1.7)

G_n corresponds to a point in the $(r-1)$-dimensional hyperplane (1.6.a) which in turn is part of a space of r dimensions. The effect of Γ is therefore to map a $(k-1)$-dimensional set of points into a $(r-1)$-dimensional set.

The second step is the transition from G_n to F_{n+1}. Every $f_{n+1}^{(i)}$ is a function of some $g_n^{(j)}$ and possibly of some parameters ϑ_i':

$$f_{n+1}^{(i)} = f_{n+1}^{(i)}(g_n^{(1)}, \ldots, g_n^{(r)}; \vartheta_1', \vartheta_2', \ldots)$$ (1.8)
$$(i = 1, 2, \ldots, k) .$$

This system of equations may be regarded as another operator Z, which, when applied to G_n, produces F_{n+1}:

$$F_{n+1} = Z(G_n) .$$ (1.9)

The effect of Z is to map the $(r-1)$-dimensional set of points, in which G_n is contained, into the $(k-1)$-dimensional set with F_{n+1}. Altogether we have

$$F_{n+1} = Z * \Gamma(F_n) = \Phi(F_n) ,$$ (1.10)

where the symbolic "multiplication" means consecutive execution of two operators (from right to left).

The effect of $\Phi = Z * \Gamma$ is thus to map the original set of points into itself. In detail the system of equations $F_{n+1} = \Phi(F_n)$ may be written as

$$f_{n+1}^{(i)} = \varphi_i(f_n^{(1)}, \ldots, f_n^{(k)}; \vartheta_1, \vartheta_2, \ldots)$$ (1.11)
$$(i = 1, \ldots, k) .$$

If selection is present, some of the ϑ_i may be declared as fitnesses. To distinguish these from other parameters, the i-th fitness is symbolized by $w^{(i)}$. Thus

$$f_{n+1}^{(i)} = \varphi_i(f_n^{(1)}, \ldots, f_n^{(k)}; w^{(1)}, \ldots, w^{(k)}; \vartheta_1, \vartheta_2, \ldots) .$$
(1.12)

It will be useful to associate different operators with different components of selection or fitness. Selection may act upon zygotes or gametes or both. In any case it should be noticed which traits are affected by selection. With zygotic selection, for example, the viability of the zygote until maturity and/or fertility of the mature zygote may be considered.

2. Selection Operators

When selection acts upon *fertility*, the fitnesses of the different genotypes (mature) are labeled with the index f. It is assumed that the individuals of the various genotypes contribute different mean numbers of gametes and that these mean numbers are in the proportions $w_f^{(1)} : w_f^{(2)} : \cdots : w_f^{(k)}$. This is equivalent to replacing the $f_n^{(i)}$ by

$$f_n'^{(i)} = w_f^{(i)} f_n^{(i)} / \overline{w}_f , \quad \text{where} \quad \overline{w}_f = \sum w_f^{(i)} f_n^{(i)} . \quad (1.13)$$

The $f_n'^{(i)}$ may be regarded as "effective" genotype frequencies, with the individuals of any genotype now contributing the same numbers of gametes. The substitutions $f_n^{(i)} \to f_n'^{(i)}$ form an operator S_f with

$$F_n' = S_f(F_n)$$ (1.14)

which also has a simple matrix representation. Application of Φ to F_n' gives the operator equation for F-selection:

$$F_{n+1} = \Phi(F_n') = \Phi(S_f(F_n)) = \Phi * S_f(F_n) .$$ (1.15)

Viability selection (V-selection) means that the probabilities of survival of fresh zygotes until maturity depend on their genotypes and are assumed to be in the proportions $w_v^{(1)} : w_v^{(2)} : \cdots : w_v^{(k)}$. Starting with adult genotypes in the parental generation (F_n) after V-selection has terminated in the n-th generation, we have $F_n \to F_{n+1}'$ according to (1.10). Now S_v is to be applied to F_{n+1}'. The result is the operator equation for V-selection

$$F_{n+1} = S_v(F_{n+1}') = S_v * \Phi(F_n) .$$ (1.16)

Apart from the labels f and v of the selection operators, (1.15) and (1.16) differ only in the order of S and Φ. This difference, however, is essential, since in general the exchange of two operators produces quite different results. If Φ and S are matrices, this follows simply from the noncommutativity of matrix multiplication.

The meaning of *gametic selection* is that parental individuals of the various genotypes contribute the same numbers of gametes, but that the efficiency of a single gamete depends on its genotype. Let the chance of a gamete of the j-th type to be successful be proportional to $w_g^{(j)}$. This is equivalent to replacing the gametic frequencies $g_n^{(j)}$ by

$$g_n'^{(j)} = w_g^{(j)} g_n^{(j)} / \bar{w}_g, \quad \text{where} \quad \bar{w}_g = \sum_{j=1}^{r} w_g^{(j)} g_n^{(j)} \quad (1.17)$$

before applying Z according to (1.9). An operator S_g is established by these equations and

$$G_n' = S_g(G_n) , \quad (1.18)$$

from which

$$F_{n+1} = Z(G_n') = Z(S_g(G_n)) = Z(S_g(\Gamma(F_n))) =$$
$$= Z * S_g * \Gamma(F_n) . \quad (1.19)$$

It is seen that the order of succession of the various operators is important. With regard to the description of selection processes it is therefore essential, which phase of the generation cycle is observed. This applies even more if different components of selection are expected to operate. Strictly speaking, F_n and F_{n+1} should be supplied with an index for the point of time of inspection. Let $F_n^{(0)}$ and $F_{n+1}^{(0)}$ be the frequency vectors just after birth of the zygotes in generation n and $n+1$ respectively, and $F_n^{(1)}$ and $F_{n+1}^{(1)}$ the same vectors at maturity (just before mating). If the selection operators are the same in both generations, the following equations hold for *V-selection*:

$$F_n^{(1)} = S_v(F_n^{(0)}) , \quad (1.20.a)$$

$$F_{n+1}^{(1)} = S_v(F_{n+1}^{(0)}) , \quad (1.20.b)$$

$$F_{n+1}^{(0)} = \Phi(F_n^{(1)}) . \quad (1.20.c)$$

For *F-selection* we have only

$$F_n^{(1)} = F_n^{(0)} , \quad F_{n+1}^{(1)} = F_{n+1}^{(0)} . \quad (1.21)$$

With these relations in mind the following operator equations are easily seen to be true

$$F_{n+1}^{(0)} = \Phi * S_v(F_n^{(0)}) , \quad (1.22.a)$$

$$F_{n+1}^{(1)} = S_v * \Phi(F_n^{(1)}) , \quad (1.22.b)$$

$$F_{n+1}^{(0)} = F_{n+1}^{(1)} = \Phi * S_f(F_n^{(1)}) = \Phi * S_f(F_n^{(0)}) , \quad (1.23)$$

$$F_{n+1}^{(0)} = \Phi * S_f * S_v(F_n^{(0)}) , \quad (1.24.a)$$

$$F_{n+1}^{(1)} = S_v * \Phi * S_f(F_n^{(1)}) . \quad (1.24.b)$$

These equations remain correct if, in addition, gametic selection is operating; then $\Phi = Z * S_g * \Gamma$. Furthermore, the case

$$F_{n+1} = Z * S_g * \Gamma(F_n) \quad (1.25)$$

should be considered separately. With this all selection models which make sense when use is made of F_n, F_{n+1}, G_n; Φ, Z, Γ; S_v, S_g, and S_f are enumerated. Among these the model (1.24.a) is of special interest, since both fitness components involved may be concentrated to a total fitness by simple matrix multiplication. $F_{n+1}^{(0)} = \Phi * S_f * S_v(F_n^{(0)})$ then reduces to the type (1.22.a): $F_{n+1} = \Phi * S_{fv}(F_n)$.

The operators Φ, Z, and Γ are the essence of all mating and recombination modalities which may be described by algebraic equations. In contrast, the operators S_v, S_f, and S_g are parameters by nature, while F_n and F_{n+1} are observable entities.

3. Determination of Fitnesses

We now face the problem of making statements about S_v, S_f, and S_g from merely the observation of F_n and F_{n+1} (at defined periods of the generation cycle) and from certain assumptions about Γ and Z. This is equivalent to solving one of the systems of equations (1.22.a) to (1.24.b) for the unknown elements of one (or more than one) S-matrix. This problem has in general no solution. Φ consists of k equations together with (1.3.b). But one S_v- or S_f-matrix already contains k unknowns which are linearly related. For example, the system (1.24.b) has $2k - 2$ unknowns and therefore no solution of practical value exists. This is more than ever true if G-selection is present along with V- and/or F-selection.

A unique solution will be available, if one can be sure that at most one component of selection is present (unless case (1.25) is under question, where F- and V-selection are condensed to a combined fitness). Even if this is given, it must be known which component is involved, since $F_n \rightarrow F_{n+1}$ may be accomplished in many different ways. F_n and F_{n+1} alone give no indication which sort (or sorts) of selection is involved. In other words, it must be clear which of the operator equations given above is to be chosen as the appropriate model. Applying an inadequate model would possibly lead to a "solution" which is of course numerically correct but at the same time would misinterprete the observed transition $F_n \rightarrow F_{n+1}$.

In this situation there remain three types of model:

$$F_{n+1} = \Phi * S(F_n) , \quad (1.26)$$

where $S = S_f$ (see (1.15)) or $S = S_v$ (see (1.22.a)) or $S = S_{fv}$.

$$F_{n+1} = S * \Phi(F_n) , \quad (1.27)$$

where $S = S_v$ (see (1.16)), and

$$F_{n+1} = Z * S_g * \Gamma(F_n) . \quad (1.28)$$

The properties of the solutions of these systems of equations depend on the nature of Φ, Γ, and Z respectively and will be different in the various cases. Some of the possible situations are examined in more detail in the subsequent paragraphs.

4. Selection of Type $F_{n+1} = \Phi * S(F_n)$

F-selection according to (1.15) and V-selection according to (1.22.a), i.e. V-selection under the condition that genotype frequencies are observed between reproduction and beginning of selection, are special cases of this type.

a) *Panmixia, one locus.* For simplification the frequencies of

genotypes AA Aa aa

are now represented by $(x_n, \quad y_n, \quad z_n) = F_n$,

$\qquad\qquad\qquad\qquad\quad (x_{n+1}, \quad y_{n+1}, \quad z_{n+1}) = F_{n+1}$.

Their fitnesses be w_x $w_y = 1$ w_z .

After selection we have

$F_n' = S(F_n)$:

$$x_n' = \frac{w_x}{\bar{w}} x_n , \qquad y_n' = \frac{1}{\bar{w}} y_n , \qquad z_n' = \frac{w_z}{\bar{w}} z_n , \quad \Big|$$

$$\text{where} \quad \bar{w} = x_n w_x + y_n + z_n w_z . \qquad\qquad (1.29)$$

F' produces the vector of gametic frequencies

$G_n = \Gamma(F_n')$:

$$p_n(A) = x_n' + \frac{1}{2} y_n' = \frac{1}{\bar{w}}\Big(x_n w_x + \frac{1}{2} y_n\Big),$$

$$q_n(a) = z_n' + \frac{1}{2} y_n' = \frac{1}{\bar{w}}\Big(z_n w_z + \frac{1}{2} y_n\Big)$$

and from this we obtain

$F_{n+1} = Z(G_n)$:

$$x_{n+1} = p_n^2 \qquad = \frac{1}{\bar{w}^2}\Big(x_n w_x + \frac{1}{2} y_n\Big)^2,$$

$$y_{n+1} = 2 p_n q_n = \frac{1}{\bar{w}^2}\Big(x_n w_x + \frac{1}{2} y_n\Big)\Big(z_n w_z + \frac{1}{2} y_n\Big),$$

$$z_{n+1} = q_n^2 \qquad = \frac{1}{\bar{w}^2}\Big(z_n w_z + \frac{1}{2} y_n\Big)^2.$$

With regard to (1.29) this appears to be a system of quadratic equations in the unknowns w_x and w_z with the following pattern of coefficients

Since these equations sum up to zero, only two of them are independent. Consider (g_1) and (g_3), which correspond to conic sections in the w_x, w_z-plane; but each of them is degenerated to a pair of straight lines (with a real point of intersection). This is easily realized from the vanishing determinants of coefficients of (g_1) and (g_3). Furthermore in both cases the 2×2-minors of the coefficients of the quadratic terms are negative. (g_1) as well as (g_3) is therefore the product of two linear factors, which are found to be

$$L_{11} \cdot L_{12} \equiv$$

$$\Big[x_n \big(1 + \sqrt{x_{n+1}}\big) w_x + z_n \sqrt{x_{n+1}} w_z + y_n\Big(\frac{1}{2} + \sqrt{x_{n+1}}\Big)\Big] \times$$

$$\Big[x_n\big(1 - \sqrt{x_{n+1}}\big) w_x - z_n \sqrt{x_{n+1}} w_z + y_n\Big(\frac{1}{2} - \sqrt{x_{n+1}}\Big)\Big] = 0 ,$$

$$(1.31.a)$$

$$L_{21} \cdot L_{22} \equiv$$

$$\Big[x_n \sqrt{z_{n+1}} w_x + z_n \big(1 + \sqrt{z_{n+1}}\big) w_z + y_n\Big(\frac{1}{2} + \sqrt{z_{n+1}}\Big)\Big] \times$$

$$\Big[-x_n\sqrt{z_{n+1}} w_x + z_n\big(1 - \sqrt{z_{n+1}}\big) w_z + y_n\Big(\frac{1}{2} - \sqrt{z_{n+1}}\Big)\Big] = 0 .$$

$$(1.31.b)$$

Both pairs of straight lines intersect in one and the same point with coordinates

$$w_x^* = -\frac{y_n}{2 x_n} , \qquad w_z^* = -\frac{y_n}{2 z_n} .$$

No point with coordinates $w_x > 0$, $w_z > 0$ will satisfy the equations L_{11} or L_{21}; the corresponding straight lines therefore do not pass through the $(++)$-quadrat of the w_x, w_z-plane. L_{12} and L_{22}, on the other hand, enter this quadrant from (w_x^*, w_z^*). Moreover, they coincide with one another and have the same equation

$$x_n \sqrt{z_{n+1}} w_x - z_n \sqrt{x_{n+1}} w_z + y_n\Big(\frac{1}{2} - \sqrt{x_{n+1}}\Big) = 0 .$$

$$(1.32)$$

All pairs of w_x, w_z, which are positive and satisfy (1.32) solve the problem. Unless $\sqrt{x_{n+1}} = \frac{1}{2}$ or $y_n = 0$, there is an infinite set of solutions which are essentially different, since L_{12} (and L_{22}) does not cross the origin. If $\sqrt{x_{n+1}} = \frac{1}{2}$ or $y_n = 0$ then the ratio $w_x : w_z$ is determined from (1.32).

w_x^2	w_z^2	$w_x w_z$	w_x	w_z	right hand	
$x_n^2 (1 - x_{n+1})$	$- z_n^2 x_{n+1}$	$- 2 x_n z_n x_{n+1}$	$x_n y_n (1 - 2 x_{n+1})$	$- 2 z_n y_n x_{n+1}$	$y_n^2\Big(x_{n+1} - \frac{1}{4}\Big)$	(g_1) ,
$- x_n^2 y_{n+1}$	$- z_n^2 y_{n+1}$	$2 x_n z_n (1 - y_{n+1})$	$x_n y_n (1 - 2 y_{n+1})$	$z_n y_n (1 - 2 y_{n+1})$	$y_n^2\Big(y_{n+1} - \frac{1}{2}\Big)$	(g_2) ,
$- x_n^2 z_{n+1}$	$z_n^2 (1 - z_{n+1})$	$- 2 x_n z_n z_{n+1}$	$- 2 x_n y_n z_{n+1}$	$z_n y_n (1 - 2 z_{n+1})$	$y_n^2\Big(z_{n+1} - \frac{1}{4}\Big)$	(g_3) .

$$(1.30)$$

It follows that a unique determination of the relative fitnesses is not available from the model given in the headline of this chapter. There is an infinite number of different fitnesses with which one is able to describe the transition $F_n \to F_{n+1}$ equally well!

Example: $F_n = (.3, .3, .4)$, $F_{n+1} = (.16, 48, 36)$. With these values (1.32) is given to be $18\,w_x -$ $- 16\,w_z + 3 = 0$. Among the totality of solutions (with $w_y = 1$) there are, for example

1) $w_x = 1$, $w_y = 1$, $w_z = 1.3125$,
2) $w_x = 2$, $w_y = 1$, $w_z = 2.4375$,

which naturally account for quite different selection processes, but at the same time produce the same transition $F_n \to F_{n+1}$.

b) *Selfing, one Locus.* In contrast to the foregoing result, (1.26) has a unique solution if the population consists of true selfing individuals. Here $F_n' = S(F_n)$ such as in (1.29). But

$F_{n+1} = Z(F_n')$:

$$\left.\begin{array}{l} x_{n+1} = x_n' + \dfrac{1}{4}\,y_n' = \dfrac{x_n\,w_x + 1/4\,y_n}{\bar{w}}\,, \\[2mm] y_{n+1} = \dfrac{1}{2}\dfrac{1}{\bar{w}}\,y_n\,, \\[2mm] z_{n+1} = z_n' + \dfrac{1}{4}\,y_n' = \dfrac{z_n\,w_z + 1/4\,y_n}{\bar{w}}\,. \end{array}\right\} \quad (1.33)$$

These are linear dependent equations for w_x and w_z

$$\left.\begin{array}{l} (x_{n+1} - 1)\,x_n w_x + x_{n+1} z_n w_z + y_n\left(x_{n+1} - \dfrac{1}{4}\right) = 0 \quad (a), \\[2mm] y_{n+1}\,x_n\,w_x + y_{n+1}\,z_n\,w_z + y_n\left(y_{n+1} - \dfrac{1}{2}\right) = 0 \quad (b), \\[2mm] z_{n+1}\,x_n\,w_x + (z_{n+1} - 1)\,z_n w_z + y_n\left(z_{n+1} - \dfrac{1}{4}\right) = 0 \quad (c). \end{array}\right\} \quad (1.34)$$

From two of them we get the unique solution (with $w_y = 1$)

$$w_x = \dfrac{y_n\,x_{n+1} - 1/2\,y_{n+1}}{x_n\quad 2\,y_{n+1}}\,, \qquad w_z = \dfrac{y_n\,z_{n+1} - 1/2\,y_{n+1}}{z_n\quad 2\,y_{n+1}}\,. \quad (1.35)$$

5. Selection of Type $F_{n+1} = Z * S * \Gamma(F_n)$

a) *Panmixia.* $F_n = (x_n, y_n, z_n)$,

$G_n = \Gamma(F_n)$: $p_n(A) = x_n + \dfrac{1}{2}\,y_n$, $q_n(a) = z_n + \dfrac{1}{2}\,y_n$.

Let the fitnesses of gamete types A and a be in the proportion $1 : w$. With $\bar{w} = p + w_n\,q_n$ we get

$G_n' = S_g(G_n)$: $p_n' = \dfrac{1}{\bar{w}}\,p_n$, $q_n' = \dfrac{w}{\bar{w}}\,q_n$,

$F_{n+1} = Z(G_n')$: $x_{n+1} = p_n'^2$, $y_{n+1} = 2\,p_n'\,q_n'$,
$\qquad\qquad\qquad z_{n+1} = q_n'^2$.

If the last set of equations is written down in detail, we have three quadratic equations for the one

unknown w:

$$x_{n+1}\,q_n^2\,w^2 + 2\,x_{n+1}\,p_n\,q_n\,w + p_n^2\,(x_{n+1} - 1) = 0\,, \quad (1.36.a)$$

$$y_{n+1}\,q_n^2\,w^2 + 2\,(y_{n+1} - 1)\,p_n\,q_n\,w + y_{n+1}\,p_n^2 = 0\,, \quad (1.36.b)$$

$$q_n^2\,(z_{n+1} - 1)\,w^2 + 2\,z_{n+1}\,p_n\,q_n\,w + z_{n+1}\,p_n^2 = 0\,. \quad (1.36.c)$$

Each of them is solved by the common root

$$w = \dfrac{p_n\,q_{n+1}}{q_n\,p_{n+1}} \quad (1.37)$$

since $p_{n+1} = p_n'$. This w gives the unique solution of the problem.

b) *Selfing.* We assume that only pollen is subjected to selection. Fitnesses of types A and a again are as $1 : w$, meaning that fertilization by A- and a-pollen occurs in proportion $1 : w$. The composition of the offspring of Aa-parents is as follows

$$\begin{array}{ccc} \female & \times & \male \end{array}$$

$$\left[\dfrac{1}{2}\,(A) + \dfrac{1}{2}\,(a)\right]\left[\dfrac{1}{2}\,y_n(A) + \dfrac{w}{2}\,y_n(a)\right] =$$

$$= \dfrac{1}{4}\,y_n(AA) + \dfrac{1 + w}{4}\,y_n(Aa) + \dfrac{w}{4}\,y_n(aa)\,.$$

Therefore

$$F_{n+1}: \quad \left.\begin{array}{l} x_{n+1} = \dfrac{1}{\bar{w}}\left(x_n + \dfrac{1}{4}\,y_n\right)\,, \\[2mm] y_{n+1} = \dfrac{1 + w}{4\,\bar{w}}\,y_n\,, \\[2mm] z_{n+1} = \dfrac{w}{\bar{w}}\left(z_n + \dfrac{1}{4}\,y_n\right)\,, \end{array}\right\} \quad (1.38)$$

where

$$\bar{w} = x_n + \dfrac{1 + w}{2}\,y_n + w\,z_n = p_n + w\,q_n\,.$$

If F_n and F_{n+1} are given, we have three linear equations for one unknown w

$$w\,q_n\,x_{n+1} + p_n\,x_{n+1} + \dfrac{1}{4}\,y_n - p_n = 0\,, \quad (1.39.a)$$

$$w\left(q_n\,y_{n+1} - \dfrac{1}{4}\,y_n\right) + p_n\,y_{n+1} - \dfrac{1}{4}\,y_n = 0\,, \quad (1.39.b)$$

$$w\left(q_n\,z_{n+1} + \dfrac{1}{4}\,y_n - q_n\right) + p_n\,z_{n+1} = 0\,, \quad (1.39.c)$$

which have a sum of zero. Thus from two of them, (a) and (c) for example, w may be determined if and only if

$$p_n\,q_n\,y_{n+1} - \dfrac{1}{4}\,y_n\left(1 - p_n\,x_{n+1} - q_n\,z_{n+1} - \dfrac{1}{4}\,y_n\right) = 0\,. \quad (1.40)$$

If F_{n+1} arises from F_n by selfing and if selection is acting upon pollen, (1.40) is always satisfied; and F_{n+1} can result from F_n under selfing with pollen selection only if this condition holds. (1.40) may therefore be used as a test criterion for this condition. If (1.40) holds, the determination of w from one of the equations (1.39) is unique.

6. Selection of Type $F_{n+1} = S * \boldsymbol{\Phi}(F_n)$

Zygotic viability selection ($S = S_v$) is of this type, if genotype frequencies are observed just before reproduction and after selection has been brought to an end. The following exposition is restricted to this model and some of its specializations. Returning to the notation of (1.12), these equations now have the simple form

$$f_{n+1}^{(i)} = \frac{w_i}{\bar{w}} \varphi(f_n^{(1)}, \ldots, f_n^{(k)}; \vartheta_1, \ldots)$$

$$(i = 1, \ldots, k), \quad (1.41)$$

where

$$\bar{w} = \sum w_i \varphi_i. \quad (1.41.a)$$

Here w_i is the fitness of the i-th genotype. If F_n and F_{n+1} are known, as supposed, (1.41) together with (1.41.a) is a system of k linear homogeneous equations in the w_i

$$(f_{n+1}^{(1)} - 1)\, \varphi_1 w_1 + f_{n+1}^{(1)}\, \varphi_2 w_2 \quad + \cdots + f_{n+1}^{(1)}\, \varphi_k w_k \quad = 0,$$

$$f_{n+1}^{(2)}\, \varphi_1 w_1 \quad + (f_{n+1}^{(2)} - 1)\, \varphi_2 w_2 + \cdots + f_{n+1}^{(2)}\, \varphi_k w_k \quad = 0,$$

$$\cdots \cdots \cdots \cdots \cdots$$

$$f_{n+1}^{(k)}\, \varphi_1 w_1 \quad + f_{n+1}^{(k)}\, \varphi_2 w_2 \quad + \cdots + (f_{n+1}^{(k)} - 1)\, \varphi_k w_k = 0. \quad (1.42)$$

Let $\mathfrak{A}$ be the matrix of coefficients. If $\varphi_i \neq 0$ for all i, then the rank of $\mathfrak{A}$ is $k-1$. If $\varphi_i = 0$ for some i, (1.42) has no solution except $w_i = 0$. $\varphi_i = 0$ means that the expected frequency of the i-th genotype in generation $n + 1$ equals zero, provided there is no selection. This can happen under some very special conditions, which may be regarded as degenerate. This will therefore be excluded here. Then, if $\varphi_i \neq 0$, there exists a non-vanishing vector $(w_1, \ldots, w_k)$ of fitnesses satisfying (1.42). But if c is an arbitrary constant, then $(c w_1, \ldots, c w_k)$ is also a solution of (1.42). Therefore, the fitnesses are determined up to a constant factor only.

It follows that, from knowledge of genotype frequencies of two subsequent generations, only the ratios of fitnesses can be determined, i.e. the *relative* fitnesses. This is evident because, if the probabilities of the various genotypes to survive are multiplied by the same factor, the proportions of surviving genotypes remain the same.

Because of (1.41) the general solution of (1.42) has the form

$$w_i = c\, \frac{f_{n+1}^{(i)}}{\varphi_i}. \qquad (i = 1, \ldots, k) \qquad (1.43)$$

The relative fitness of a genotype is, therefore, proportional to the ratio of the observed frequency in generation $n + 1$ to that frequency which is expected when selection is absent. c may be chosen at will. Sometimes it is useful to take the relative fitness of a certain genotype (heterozygotes or double heterozygotes) as 1. Let this genotype be the h-th one. Then $w_h = 1$, and $c = \varphi_h/f_{n+1}^{(h)}$, from which

$$w_i = \frac{f_{n+1}^{(i)}}{\varphi_i}\, \frac{\varphi_h}{f_{n+1}^{(h)}}. \qquad (1.44)$$

More simply, $c = 1$, and

$$w_i = \frac{f_{n+1}^{(i)}}{\varphi_i}. \qquad (1.44.a)$$

The following considerations are based on this definition of relative fitness. The variable $\bar{w} = \sum w_i \varphi_i$ is called the mean fitness. It is $\bar{w} = c$ under (1.43) and $\bar{w} = \varphi_h/f_{n+1}^{(h)}$ under (1.44).

In case of *mixed selfing and random mating* with one locus (two alleles) the expressions for φ_1, φ_2, and φ_3 are as follows

$$\varphi_1 = t\left(f_n^{(1)} + \frac{1}{2}f_n^{(2)}\right)^2 + (1 - t)\left(f_n^{(1)} + \frac{1}{4}f_n^{(2)}\right),$$

$$\varphi_2 = 2\,t\left(f_n^{(1)} + \frac{1}{2}f_n^{(2)}\right)\left(f_n^{(3)} + \frac{1}{2}f_n^{(2)}\right) + \frac{1-t}{2}f_n^{(2)},$$

$$\varphi_3 = t\left(f_n^{(3)} + \frac{1}{2}f_n^{(2)}\right)^2 + (1 - t)\left(f_n^{(3)} + \frac{1}{4}f_n^{(2)}\right).$$

$$(1.45)$$

where t is the rate of outcrossing.

Zusammenfassung

Für Reproduktions- und Selektionsvorgänge in Populationen diploider Organismen mit diskreter Generationenfolge wird eine algebraische Darstellung gegeben. Das Interesse konzentriert sich dabei auf die Frage, unter welchen Bedingungen die Selektionskoeffizienten der verschiedenen Genotypen aus den Genotypenfrequenzen zweier aufeinander folgender Generationen allein bestimmt werden können. Im zweiten Teil der Arbeit, der später veröffentlicht wird, werden die statistischen Eigenschaften einer speziellen, von HAYMAN aufgestellten Schätzfunktion untersucht. Hierbei werden Konfidenzintervalle und Signifikanzteste für Fitnesswerte betrachtet und mit Hilfe von Simulationsstudien geprüft.

Literature

1. Allard, R. W., Workman, P. L.: Population studies in predominantly selfpollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. Evolution 17, 470—480 (1963). — 2. Aroian, L. A.: The probability function of the product of two normally distributed variables. Ann. math. Statist. 18, 265—271 (1947). — 3. Cramér, H.: Mathematical methods of statistics. Princeton: Princeton Univ. Press 1946. — 4. Curtiss, J. H.: On the distribution of the quotient of two chance variables. Ann. math. Statist. 12, 409—421 (1941). — 5. Fieller, E. C.: The distribution of the index in a normal bivariate population. Biometrika 24, 428—440 (1932). — 6. Fieller, E. C.: A fundamental formula in the statistics of biological assay, and some applications. J. Pharm. Pharmac. 17, 117—123 (1944). — 7. Harding, J., Tucker, C. L.: Quantitative studies on mating systems. I. Evidence for the non-randomness of outcrossing in *Phaseolus Lunatus*. Heredity 19, 369—381 (1964). — 8. Hayman, B. I.: Mixed selfing and random mating when homozygotes are at a disadvantage. Heredity 7, 185—192 (1953). — 9. Jain, S. K., Allard, R. W.:

Population studies in predominantly self-pollinated species. I. Evidence for heterozygote advantage in a closed population of barley. Proc. nat. Acad. Sci. U.S.A. **46**, 1371—1377 (1960). — 10. Jain, S. K., Marshall, D. R.: Genetic changes in a barley population analyzed in terms of some life cycle components of selection. Genetics **38**, 356—374 (1967). — 11. Merrill, A. S.: Frequency distribution of an index when both the components follow the normal law. Biometrika **20A**, 53—63 (1928). — 12. Prout, T.: The estimation of fitnesses from genotypic frequencies. Evolution **19**, 546—551 (1965). — 13. Tucker, C. L., Harding, J.: Quantitative studies on mating systems.

II. Estimation of fitness parameters in a population of *Phaseolus Lunatus*. Heredity **20**, 393—402 (1965). — 14. Wöhrmann, K.: Fremdbefruchtungsrate und genotypische "fitness". Züchter/Genet. Breed. Res. **37**, 142 to 145 (1967). — 15. Workman, P. L., Allard, R. W.: Population studies in predominantly selfpollinated species. III. A matrix model for mixed selfing and random outcrossing. Proc. nat. Acad. Sci. U.S.A. **48**, 1318—1325 (1962). — 16. Workman, P. L., Jain, S. K.: Zygotic selection under mixed random mating and self-fertilization: Theory and problems of estimation. Genetics **54**, 159—171 (1966).

Dipl.-Math. R. J. Lorenz
Bundesforschungsanstalt für
Viruskrankheiten der Tiere
Postfach 1149
7400 Tübingen (BRD)