# ZYGOTIC SELECTION UNDER MIXED RANDOM MATING AND SELF-FERTILIZATION: THEORY AND PROBLEMS OF ESTIMATION

P. L. WORKMAN AND S. K. JAIN

*Department of Agronomy, University of California, Davis*

R ECENT studies on experimental populations of barley and lima beans (JAIN and ALLARD 1960; ALLARD and WORKMAN 1963; HARDING, ALLARD and SMELTZER, in preparation) have provided estimates of the intensity of the selection maintaining genetic variation in these predominantly self-fertilized species. In each experiment, selective values were estimated from census data on genotypic frequencies in successive generations, by means of estimation formulas based upon the model of mixed random mating and selfing with selection described by HAYMAN (1953). It was assumed that all zygotic selection had occurred prior to the determination of the genotypic frequencies and consequently the estimated fitness values were considered to represent the so-called net Darwinian fitnesses defined in terms of the relative number of progeny left by different genotypes.
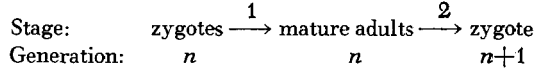
However, PROUT (1965) has pointed out that if selection has not been completed at the time of the census, then the estimated viabilities may not represent the net fitness values. In addition, the selective differences can arise in many different ways (differential viability at one or more zygotic stages, differential fecundity of one or both sexes, gametic selection, etc.), and consequently appropriate estimators must depend both on the mode of selection and the stage at which the genotypic proporitons are determined (seed, seedling, or adult). For example, in the barley populations, a preliminary analysis of fitness components provided evidence for differential seedling emergence at locus $b$ (lemma color), fertility differenccs at locus $r$ (awn texture) and a combination of gametic and zygotic selection at the aleurone color loci $bl_1$ and $bl_2$ (JAIN 1961, and unpublished data). Analysis of multistage selection may also require the use of loci which can be scored at several different stages of the life cycle.

In this paper we shall consider three different selection schemes appropriate for a population in which there is mixed random mating and self-fertilization. Their theoretical properties will be briefly examined and the formulas for the estimation of relative viabilities derived from each model will be compared by an analysis of the barley and lima bean data and by application to data on multistage selection provided by computer simulation.

## Theoretical models of zygotic selection

Models which describe the effects of zygotic selection on a population mating by mixed random mating and selfing can differ with respect to the time of action

of the selection, the stage of the life cycle at which the genotypic frequencies are determined, or both. The following schematic representation of the life cycle between generations $n$ and $n + 1$ will help to demonstrate the differences between the three models which will be discussed in this paper.

$$\text{Stage:} \quad \text{zygotes} \xrightarrow{\;1\;} \text{mature adults} \xrightarrow{\;2\;} \text{zygote}$$
$$\text{Generation:} \quad n \qquad\qquad n \qquad\qquad n+1$$

Zygotic selection could result either from differential viability at stage 1 or from differential fecundity or variation in the mating system at stage 2. Under model I, both selection and the determination of genotypic frequencies occur at stage 1 but it is assumed that all of the selection has occurred prior to scoring. For example, data based on seed characters (e.g. loci $b$, $s$, $g$ in barley) are best analyzed by this model because these characters are determined by the maternal genotype. Model II permits selection at either stage 1 or 2, but requires that scoring be done soon after mating and that no selection occurs between mating and scoring. This model would be appropriate for an analysis of the aleurone color loci ($bl_1$, $bl_2$) in barley. Under model III, all selective differences result from differential fecundity (stage 2) determined by the maternal genotype alone and the scoring is the same as in model II. Thus, each of the three models is based upon the assumptions that selection acts during a single prescribed stage and that the scoring of genotypic frequencies does not occur at a partially selected stage.

For each model we shall consider only a single locus with two alleles, say $A_1$ and $A_2$, in an indefinitely large population. Let $(P, R, Q)$ and $(P', R', Q')$ denote the genotypic proportions of $(A_1A_1, A_1A_2, A_2A_2)$ in generations $n$ and $(n+1)$. Let the relative viabilities of the genotypes be in the ratios $(x{:}1{:}y)$, (this applies to male and female parents in models I and II, female parents only in model III). Further, it is assumed that each individual has a constant probability $s$ of selfing and $t = 1 - s$ of random outcrossing.

Model I: As shown by HAYMAN (1953) the proportionalities relating genotypic frequencies in successive generations are given by

$$P' \propto x\{s(P + \tfrac{1}{4}R) + t(P + \tfrac{1}{2}R)^2\}$$
$$R' \propto \tfrac{1}{2}sR + 2t(P + \tfrac{1}{2}R)(Q + \tfrac{1}{2}R) \qquad (1)$$
$$Q' \propto y\{s(Q + \tfrac{1}{4}R) + t(Q + \tfrac{1}{2}R)^2\}$$

Now, if we define $F = 1 - R/2pq$ where $p = (P + \tfrac{1}{2}R)$ and $q = (Q + \tfrac{1}{2}R) = (1 - p)$, then the genotypic distribution at any generation, $n$, viz. $(P_n, R_n, Q_n)$ can be written as

$$P_n = p_n{}^2 + p_n q_n F_n$$
$$R_n = 2p_n q_n (1 - F_n) \qquad (2)$$
$$Q_n = q_n{}^2 + p_n q_n F_n .$$

If the only force causing genotypic frequencies to depart from Hardy-Weinberg expectations is that of inbreeding, then $F_n$, is the coefficient of inbreeding at gen-

eration, $n$, as discussed by WRIGHT (1965). If selection is also present, then $F_n$ denotes the joint effects of inbreeding and selection and in WRIGHT's terminology (1965), $F_n$ is called the Fixation Index. At equilibrium, both $\Delta P = 0$, and $\Delta F = 0$ (or $dp/dt = 0$, $dF/dt = 0$) must be true. This transform involving only allelic frequencies, $p_i$, and $F$ can be extended to multiple alleles or multiple loci and has been found to be extremely useful in describing equilibrium conditions. A general discussion of this extended use of $F$ statistics will be presented elsewhere.

At equilibrium, $(P', R', Q') = (P, R, Q)$ and, substituting for $P, R, Q$ by (2), proportionalities (1) become

$$A_1A_1 \ : \ p^2 + pqF \propto x \{p^2 + \tfrac{1}{2}spq(1 + F)\}$$

$$A_1A_2 \ : \ 2pq(1 - F) \propto \{2pq - spq(1 + F)\} \tag{3}$$

$$A_2A_2 \ : \ q^2 + pqF \propto y \{q^2 + \tfrac{1}{2}spq(1 + F)\}$$

From (3) we can show that at equilibrium the following relations must both hold.

$$p = \frac{(1-y) - \tfrac{1}{2}s(1-x)(1+F)}{(2-x-y) \, [1-\tfrac{1}{2}s(1+F]} \tag{4a}$$

$$sF^2(1-x)(1-y) + F\{2(xy-1) + s(2-x-y)\} - 2(1-x)(1-y) + s(1-xy) = 0 \tag{4b}$$

Equations which describe the equilibrium state were also given by HAYMAN (1953) but they appear in a more complicated form than (4a) and (4b) and are much less simple to handle. The equilibrium formula for $p_e$ under mixed random mating and selfing and zygotic selection has also been derived by LI (1955) and LEWONTIN (1958) as $p_e = [(1-y)-(1-x)F]/[(2-x-y)(1-F)]$. Their results correspond to the equilibrium state just after mating, given by model II for which the correct equilibrium value for $F$ is given by formula (6b) and not by $F = s/(2-s)$.

Model II: The recursions relating genotypic proportions in successive generations are given by

$$P' = \frac{s(xP + \tfrac{1}{4}R)}{xP + yQ + R} \quad + \quad \frac{t(xP + \tfrac{1}{2}R)^2}{(xP + yQ + R)^2}$$

$$R' = \frac{\tfrac{1}{2}sR}{xP + yQ + R} \quad + \quad \frac{2t(xP + \tfrac{1}{2}R)(yQ + \tfrac{1}{2}R)}{(xP + yQ + R)^2} \tag{5}$$

$$Q' = \frac{s(yQ + \tfrac{1}{4}R)}{xP + yQ + R} \quad + \quad \frac{t(yQ + \tfrac{1}{2}R)^2}{(xP + yQ + R)^2} \ .$$

Using the transform given by (2), at equilibrium it can be shown that the following relations must both hold:

$$p = \frac{(1-y) - F(1-x)}{(2-x-y)(1-F)} \tag{6a}$$

$$2F^2(1-x)(1-y) + F\{(x+y-2) - t(x+y-2xy)\} + (x+y-2xy)(1-t) = 0 \tag{6b}$$
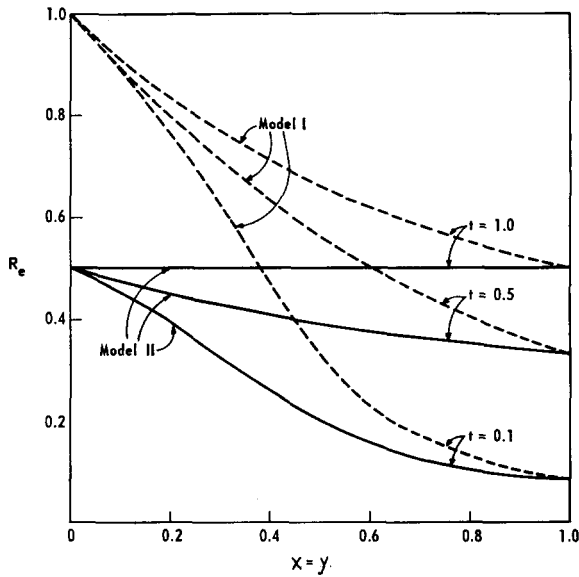
FIGURE 1.—Equilibrium proportions of heterozygotes $(R_e)$ for equal homozygote viabilities $(x = y)$ under models I and II.

A comparison of (6a, b) and (4a, b) shows that if $F_I$, $F_{II}$ are the equilibrium values of $F$ in the two models, then $F_{II} = \frac{1}{2}s(1 + F_I)$. This result is, of course, excepted since the model II differs only operationally from model I and represents genotypic frequencies derived from model I by one generation of mixed random mating and selfing.

If scoring is done after mating (and before selection), then the maximum heterozygosity which can be observed is always less than or equal to 50%, whereas scoring before mating permits the observed heterozygosity to reach 100% $(x = y = 0)$. Figure 1 compares the equilibrium proportion of observed heterozygosity for symmetric heterozygote advantage $(x = y)$ and different amounts of outcrossing $(t)$ under models I and II.

Model III: The recursive proportionalities relating the genotypic frequencies in successive generations are:

$$P' \propto s(Px+\tfrac{1}{4}R) + t(Px+\tfrac{1}{2}R)(P+\tfrac{1}{2}R)$$

$$R' \propto \tfrac{1}{2}sR + t\{(Px+\tfrac{1}{2}R)(Q+\tfrac{1}{2}R) + (Qy+\tfrac{1}{2}R)(P+\tfrac{1}{2}R)\} \qquad (7)$$

$$Q' \propto s(Qy+\tfrac{1}{4}R) + t(Qy+\tfrac{1}{2}R)(Q+\tfrac{1}{2}R)$$

Following substitution of equations (2), we find at equilibrium,

$$p = \frac{(1-y) - F(1-x)}{(2-x-y)(1-F)} \qquad (8a)$$

$$2F^2(1-y)(x-1) + Ft(x+y-2xy) + F(2-x-y) + (x+y-2xy)(t-1) = 0 \qquad (8b)$$

Equations (6a, b) and (8a, b) are identical, and therefore, under models II and III the gene and genotypic frequencies at equilibrium are identical. However, as can be shown these two models agree only at this equilibrium point.

*The approach to equilibrium:* The equilibrium values of $p$ and $F$, as noted earlier, are identical at the same stage of the life cycle (for given $x$, $y$, and $s$) for all three models. However, there are differences in the approach to equilibrium as shown in Figures 2 and 3. For $x = .50$, $y = .75$ and $(P_o, R_o, Q_o) = (.25, .50, .25)$, changes in $p$ and $F$ are plotted against time. Note that $p$ goes monotonically to its equilibrium value, but $F$ can approach its equilibrium value by complex paths. Table 1 shows changes in $\Delta F$ for several cases which demonstrate both variation in the sign of $\Delta F$ and differential rates of change for varying $s$.

*Phase diagrams.* In order to describe equilibrium populations it is useful to examine phase diagrams (Figures 4, 5) as described by HAYMAN and MATHER (1953) and HAYMAN (1953). These show the equilibrium populations for different magnitudes of the relative viabilities $(x,y)$ of the two homozygotes given some particular value of $s$ (the amount of selfing).

The areas marked A and B correspond to populations homozygous for $A_1A_1$ and $A_2A_2$ respectively; in C, heterozygotes are present but in a frequency lower than in a population mating at random without selection; in D there are more heterozygotes than in a random mating population.

For all three models, the boundary conditions describing the borders are identical. Since the equilibrium gene frequencies are identical for given $x$, $y$, $t$,
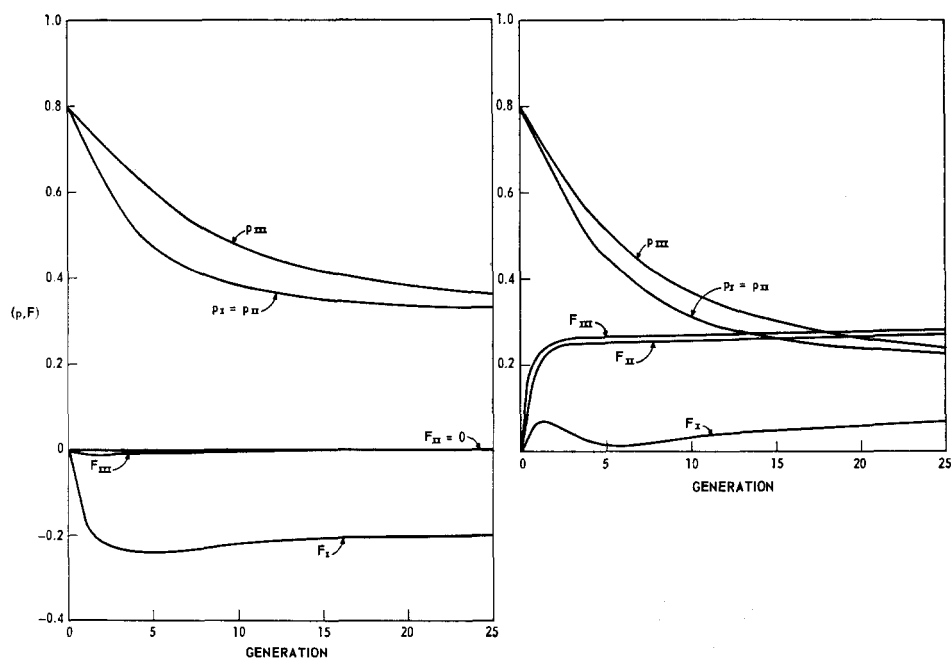


FIGURES 2 and 3.—Changes in the values of $p$ and $F$ for viabilities $(x = .50, y = .75)$ and $s = 0$ (Figure 2, left) and $s = .50$ (Figure 3, right).

TABLE 1

*Values of* Δ F *given by changes under selection* (x = .50, y = .75)

| Generation | Model I | | | | Model II | | | |
|---|---|---|---|---|---|---|---|---|
| | s = 0 | 0.2 | 0.5 | 0.9 | s = 0 | 0.2 | 0.5 | 0.9 |
| 1 | —.1754 | —.0793 | .0725 | .2902 | 0 | .0825 | .2061 | .3711 |
| 2 | —.0375 | —.0427 | —.0106 | .1340 | 0 | .0044 | .0426 | .1577 |
| 3 | —.0205 | —.0248 | —.0227 | .0616 | 0 | —.0016 | .0064 | .0758 |
| 4 | —.0067 | —.0103 | —.0160 | .0256 | 0 | —.0009 | —.0007 | .0379 |
| 5 | .0010 | —.0011 | —.0069 | .0077 | 0 | —.0001 | —.0008 | .0188 |
| 10 | .0038 | .0043 | .0049 | —.0016 | 0 | .0004 | .0012 | .0004 |
| 15 | .0013 | .0018 | .0027 | .0009 | 0 | .0002 | .0007 | .0005 |

this was to be expected. But models II and III are based on genotypic proportions determined immediately after mating and consequently they permit no region $D$. Obviously, however, at equilibrium the genotypic proportions just prior to mating would be the same for all three models.

The formulas for the boundary curves have been derived by HAYMAN and may be easily obtained from the equilibrium formulas for $p$ and $F$. The boundary between A and C, obtained by setting $p = 1$ in the equilibrium formula (4a) for $p$ and then eliminating $F$ from the two equations (4a,b) is given by

$$2sxy - 2x^2 + 2x - sx - sy = 0 \qquad (9)$$

The boundary between B and C is obtained by interchanging $x$ and $y$ in (9). The boundary between C and D (model I only) is obtained by setting $F = 0$, into the equilibrium formula for $F$, is

$$(2 + s)xy - 2x - 2y + 2 - s = 0. \qquad (10)$$

HAYMAN (1953) presented several examples of phase diagrams for various values
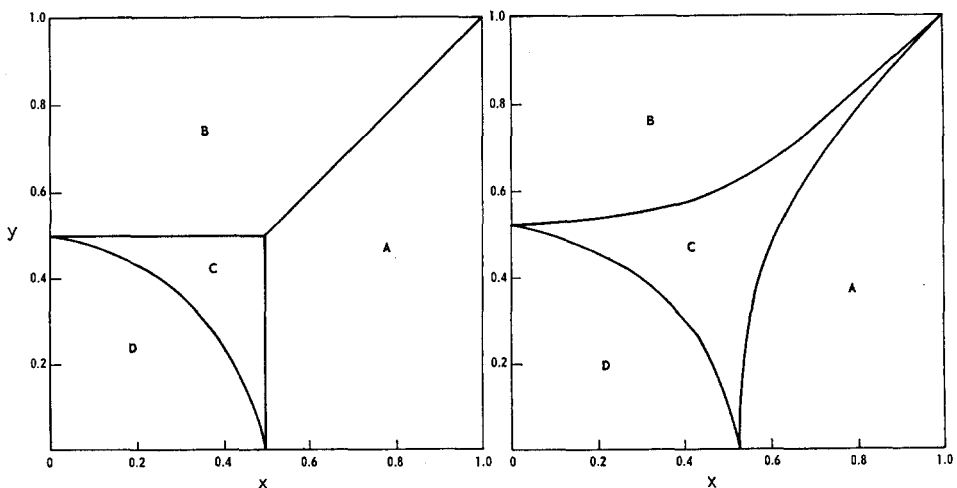


FIGURES 4 and 5.—Phase diagrams under model I for $s = 1.00$ (Figure 4, left) and $s = 0.95$ (Figure 5, right).

of s; only two cases ($s = 1.00$, $s = .95$) are given here (Figures 4, 5) which are of particular interest in relation to our results in predominantly inbreeding species. Note that any outcrossing ($t = 1 - s > 0$) in the system allows region C to extend along the entire range of x and y, although the width of region C is restricted.

*Application to experimental data:* It follows from the preceding discussion of changes in genotypic frequencies under these models that differences in the stages of counting, or in the mode of selection necessitate different estimators of the selective values ($x, y$). Given an independent estimate of the amount of outcrossing ($t, s = 1 - t$), the maximum likelihood estimators of ($x, y$) which can be derived for each of the three models are:

Model I:

$$\hat{x}_I = \frac{P'[\frac{1}{2}sR + 2t(P + \frac{1}{2}R)(Q + \frac{1}{2}R)]}{R'[s(P + \frac{1}{4}R) + t(P + \frac{1}{2}R)^2]},$$

$$\hat{y}_I = \frac{Q'[\frac{1}{2}sR + 2t(P + \frac{1}{2}R)(Q + \frac{1}{2}R)]}{R'[s(Q + \frac{1}{4}R) + t(Q + \frac{1}{2}R)^2]};$$

(11)

(ALLARD and WORKMAN 1963) or, rewriting in terms of $p = P + \frac{1}{2}R$, $q = Q + \frac{1}{2}R$, and $F$,

$$\hat{x}_I = \frac{P'[2pq - spq(1 + F)]}{R'[p^2 + \frac{1}{2}spq(1 + F)]},$$

$$\hat{y}_I = \frac{Q'[2pq - spq(1 + F)]}{R'[q^2 + \frac{1}{2}spq(1 + F)]}.$$

*Model II:*

Using the relations $(P, R, Q) = (p^2 + pqF, 2pq(1 - F), q^2 + pqF)$, and similarly, $(P', R', Q') = (p'^2 + p'q'F', 2p'q'(1 - F'), q'^2 + p'q'F')$, the estimators are:

$$\hat{x}_{II} = \frac{2q(1 - F)}{1 - q(1 - F)} \left\{ \frac{s - 2q'(s - F')}{4q'(s - F')} \right\},$$

$$\hat{y}_{II} = \frac{2p(1 - F)}{1 - p(1 - F)} \left\{ \frac{s - 2p'(s - F')}{4p'(s - F')} \right\}.$$

(12)

*Model III:*

$$\hat{x}_{III} = \frac{P'Q'(B_1B_3 - B_2B_3) + P'R'(B_3C_1 - B_1C_2) + Q'R' \; A_2B_2 - R'^2 \; A_2C_1}{P'Q'(B_2^2 - B_1^2) - R' \; B_2(Q'A_1 + P'C_1) + R'^2 \; A_1C_1},$$

$$\hat{y}_{III} = \frac{P'Q'(B_2B_3 - B_1B_3) + Q'R'(A_1B_3 - A_2B_2) + P'R' \; B_1C_2 - R'^2 \; A_1C_2}{P'Q'(B_1^2 - B_2^2) - R' \; B_1(P'C_1 + Q'A_1) + R'^2 \; A_1C_1}$$

(13)

where

$A_1 = P(s + tp)$, $A_2 = \frac{1}{2}R(\frac{1}{2}s + tp)$,
$B_1 = Ptq$, $B_2 = Qtp$, $B_3 = \frac{1}{2}R$,
$C_1 = Q(s + tq)$, $C_2 = \frac{1}{2}R(\frac{1}{2}s + tq)$.

In all cases, the variances of ($x, y$) can be derived from the matrix of expected values of $\frac{\partial^2 \ln L}{\partial x^2}$, $\frac{\partial^2 \ln L}{\partial y^2}$, $\frac{\partial^2 \ln L}{\partial x \partial y}$ as shown by ALLARD and WORKMAN (1963)

for model I. The estimators for models II and III are valid for $0 < s \leq 1$. When $s = 0$, $F = 0$ for these models and the appropriate formulas for this special case can be easily obtained.

The differences among these estimators can be illustrated by applying them to data obtained from experimental studies on lima beans (ALLARD and WORKMAN 1963) and on barley (JAIN and ALLARD 1960). The decision as to which model would be most appropriate for given data, of course, depends both on the stage of census and on the knowledge, if any, of the nature of the selective forces, presumably obtained from an independent study of the various components of fitness. For example, at locus $b$ in barley, evidence suggested differential emergence of $B–$ and $bb$ individuals. Furthermore, the general experience with bulk populations of cereals suggests that a large part of the fitness differential results from plants failing to flower, rather than from differential fecundity as determined by the seed number produced per plant. It would therefore appear appropriate to use model I on data related to flowering or seed stage characters (e.g. loci $b$, $s$, $g$, $r$, $e$ in barley). On the other hand, as pointed out earlier, model II is certainly more appropriate for aleurone color data based on a census of the newly formed zygotes.

Table 2 gives estimates of $(x,y)$, under models I and II, for a sample of data from barley populations. It is interesting to note that the bias arising from applying model II when model I would be more appropriate is usually small and does not alter inferences drawn regarding the direction and magnitude of the selective

TABLE 2

*Genotypic frequencies in barley population CCV (JAIN and ALLARD 1960) and the estimates of* (x,y)

| Locus* and genotype | Observed frequency | | Selective values with $s = .98$ | | | |
| | | | Model I | | Model II | |
| | $F_3$ | $F_5$ | $\hat{x}_1$ | $\hat{y}_1$ | $\hat{x}_{II}$ | $\hat{y}_{II}$ |
|---|---|---|---|---|---|---|
| GG | .357 | .556 | | | | |
| gg | .357 | .350 | 1.04 | 0.82 | 1.00 | 0.75 |
| Gg | .286 | .094 | | | | |
| SS | .072 | .148 | | | | |
| ss | .594 | .745 | 0.81 | 0.96 | 0.68 | 0.93 |
| Ss | .334 | .107 | | | | |
| BB | .059 | .074 | | | | |
| bb | .810 | .906 | 1.06 | 1.31 | 1.18 | 1.42 |
| Bb | .131 | .020 | | | | |
| EE | .349 | .270 | | | | |
| ee | .535 | .616 | 0.47 | 0.59 | 0.40 | 0.54 |
| Ee | .116 | .114 | | | | |
| RR | .266 | .370 | | | | |
| rr | .614 | .557 | 0.82 | 0.68 | 0.73 | 0.64 |
| Rr | .120 | .073 | | | | |

* $G/g$, dentate *vs.* nondentate lemma; $S/s$, short *vs.* long haired rachilla; $B/b$, black *vs.* white pericarp color; $E/e$, long *vs.* short glume awn; $R/r$, rough *vs.* smooth awn.

TABLE 3

*Genotypic frequencies in lima bean population 53 (*ALLARD *and* WORKMAN *1963)*
*and the estimates of* (x,y)

| Model | Genotype | Genotypic frequencies in generation | | | | | | | | | |
|-------|----------|------|------|------|------|------|------|------|------|------|------|
| | | $F_2$ | $F_3$ | $F_4$ | $F_5$ | $F_6$ | $F_7$ | $F_8$ | $F_9$ | $F_{10}$ | $F_{11}$ |
| | SS | .237 | .326 | .477 | .445 | .380 | .393 | .343 | 3.65 | .380 | .151 |
| | ss | .250 | .335 | .360 | .383 | .450 | .525 | .495 | .521 | .530 | .762 |
| | Ss | .514 | .339 | .163 | .173 | .171 | .081 | .162 | .114 | .089 | .087 |
| | | Selective values of homozygotes | | | | | | | | | |
| I | $\hat{x}_I$ | | 0.72 | 1.36 | 0.52 | 0.50 | 1.25 | 0.33 | 0.87 | 0.88 | 0.30 |
| | $\hat{y}_I$ | | 0.71 | 1.00 | 0.58 | 0.68 | 1.42 | 0.36 | 0.88 | 0.87 | 1.08 |
| II | $\hat{x}_{II}$ | | 0.54 | 1.47 | 0.41 | 0.39 | 1.42 | 0.20 | 0.82 | 0.82 | 0.19 |
| | $\hat{y}_{II}$ | | 0.54 | 1.03 | 0.45 | 0.56 | 1.62 | 0.23 | 0.84 | 0.82 | 0.82 |
| III | $\hat{x}_{III}$ | | 0.54 | 1.49 | 0.41 | 0.39 | 1.42 | 0.20 | 0.82 | 0.82 | 0.21 |
| | $\hat{y}_{III}$ | | 0.54 | 1.03 | 0.46 | 0.56 | 1.63 | 0.23 | 0.84 | 0.82 | 0.92 |

differences. In fact, model II suggests greater under- or overdominance than that obtained using model I. The estimators for each of the three models are compared in Table 3, using data at locus $s$ governing patterns of seed coat color in lima beans. Estimates from models II and III are very similar to each other and they differ from model I in the same general way. The differences are similar to those between the estimates from models I and II in the barley data (Table 2). The stage of census is the same in models II and III, so it appears that, in general, the stage of census is a more important source of possible bias than is the actual model of selection (viability *vs.* fecundity). In these examples, model I estimates are probably most appropriate for reasons already given, except in the case of locus $r$, in which independent investigations suggested that differential fertility may be an important factor.

Next, consider the problem of estimating selective values from observations taken after partial selection has occurred. Clearly, one needs to observe genotypic frequencies at more than one stage during a single generation cycle in order to estimate the individual components of fitness independently. Thus, none of the three models discussed above would give exact estimators for data taken at a partially selected stage. such as that from a census on seedling or adult plant characters for which the net fitness values depended upon both viability and fecundity differences. For practical purposes it is important to know the magnitude of the bias in the estimates of $(x,y)$ resulting from applying these models to data taken at a partially selected stage. For this purpose we have simulated several examples of two- and three-stage selection from which the selective values were estimated, using the estimates derived from models I, II and III. A sample of the results is given in Tables 4 and 5. Under two-stage selection, we have considered examples with or without differential fecundity. Note that seed stage data using model I give the same values of $(\hat{x},\hat{y})$ as obtained with model II, which simply

## TABLE 4

*Estimates of $(x_i, y_i)$ under two-stage selection*\*

| Input selective values† | | | Selling (s) | Estimates under | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Model I adult stage | | Model I seedlings | | Model II‡ | | Model III | |
| $(x_1, y_1)$ | $(x_2, y_2)$ | $(x_3, y_3)$ | | $\hat{x}$ | $\hat{y}$ | $\hat{x}$ | $\hat{y}$ | $\hat{x}$ | $\hat{y}$ | $\hat{x}$ | $\hat{y}$ |
| (1, 1) | (.55, .75) | (.80, .90) | .50 | .4400 | .6750 | .5261 | .7236 | .4400 | .6750 | .3731 | .6366 |
| | | | .95 | .4400 | .6750 | .4773 | .6939 | .4400 | .6750 | .4439 | .6807 |
| (1, 1) | (.55, .75) | (.95, .80) | .50 | .5225 | .6000 | .5728 | .6626 | .5225 | .6000 | .5159 | .5956 |
| | | | .95 | .5225 | .6000 | .5376 | .6188 | .5225 | .6000 | .5228 | .6012 |
| (.55, .75) | (.80, .90) | (1, 1) | .50 | .7002 | .8259 | .7002 | .8259 | .4400 | .6750 | .3731 | .6366 |
| | | | .95 | .5433 | .7354 | .5433 | .7354 | .4400 | .6750 | .4439 | .6807 |
| (.55, .75) | (.95, .80) | (1, 1) | .50 | .7655 | .7902 | .7665 | .7902 | .5225 | .6000 | .5159 | .5966 |
| | | | .95 | .6180 | .6739 | .6180 | .6739 | .5225 | .6000 | .5228 | .6012 |

\* Taking initial $P_o = .25 = O_o$, $R_o = .50$, these estimates were obtained after ten generations. It should be noted that during the first 20 generations simulated for estimation, the values of $x_i$, $y_i$ slightly vary from one generation to another.
† $x_1$, $y_1$ at stage 1 (fertility differences), $x_2$, $y_2$ at stage 2 (seed to seedling viability), $x_3$, $y_3$ at stage 3 (seedling to adults viability).
‡ Estimates for model I, seed stage, are equivalent to estimates derived from model II.

follows from the fact that no matter how many stages of zygotic selection are involved, census data just prior to or just after matings would yield correct estimates in all cases. In fact, these are always the products $\pi\, x_i$, $\pi\, y_i$, for $i = 1, 2, \ldots n$ different stages.

Data recorded during seedling stage in all such models and on adult plants in models involving fertility differences would give biased estimates, the amount of bias depending on the intensity and direction of selection, the level of selfing and the degree to which the genotypic frequencies differ from those expected at

## TABLE 5

*Estimates of (x, y) under three-stage selection*

| Input selective values\* | | | Selfing (s) | Estimates under | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Model I adult stage | | Model I seedlings | | Model II | | Model III | |
| $(x_1, y_1)$ | $(x_2, y_2)$ | $(x_3, y_3)$ | | $\hat{x}$ | $\hat{y}$ | $\hat{x}$ | $\hat{y}$ | $\hat{x}$ | $\hat{y}$ | $\hat{x}$ | $\hat{y}$ |
| (.8, .8) | (.8, .8) | (.9, .9) | .50 | .679 | .679 | .733 | .733 | .576 | .576 | .576 | .576 |
| | | | .95 | .615 | .615 | .635 | .635 | .576 | .576 | .576 | .576 |
| (.8, .9) | (.8, .9) | (.8, .9) | .50 | .609 | .784 | .720 | .849 | .512 | .729 | .444 | .693 |
| | | | .95 | .553 | .757 | .597 | .791 | .512 | .729 | .522 | .741 |
| (.75, .80) | (.85, .80) | (.90, .95) | .50 | .703 | .723 | .748 | .762 | .574 | .608 | .572 | .607 |
| | | | .95 | .624 | .654 | .642 | .669 | .574 | .608 | .574 | .608 |
| (.55, .75) | (.95, .80) | (.85, .70) | .50 | .626 | .570 | .736 | .712 | .444 | .420 | .444 | .420 |
| | | | .95 | .510 | .469 | .539 | .516 | .444 | .420 | .444 | .420 |
| (.55, .75) | (1.10, 1.20) | (.85, .70) | .50 | .770 | .816 | .942 | .963 | .514 | .630 | .499 | .620 |
| | | | .95 | .617 | .707 | .665 | .756 | .514 | .630 | .516 | .633 |

\* See note for Table 4.

equilibrium. However, this bias appears to become progressively smaller as the mating system approaches one of complete selfing. Model III gives estimates successively closer to those obtained for model II as the equilibrium is approached. In this connection it should be noted that these estimates of $(x,y)$, despite the use of constant selective values in the simulation program, vary among generations. Thus, from estimation of viabilities at a partially selected stage, one might mistakenly infer that selection is gene-frequency dependent.

DISCUSSION

Both the trajectories of changes in genotypic frequencies and the use of estimators of selective values for three different models (strictly speaking, models I and II are merely different experimental schemes) show that the stage of scoring frequencies is an important factor to consider in estimating selective forces. PROUT (1965) showed that estimation at a partially selected stage is likely to cause rather serious problems for the case of random mating. He also suggested that estimates at a partially selected stage would be incorrect for a system of mixed random mating and selfing. However, as the mating system approaches one of predominent selfing, our results clearly show that whichever model (I, II, or III) is used, the estimates of the selective values are not affected too seriously over a wide range of conditions. Therefore, for most experimental problems on predominantly selfed populations, it should be sufficient to use model I for adult or seed coat characters and model II for zygotes and early life stages when selfing is as high as 95 or 98%. Of course, wherever additional information is available to suggest the actual mode of selection, say gametic selection, or some sort of incompatibility factors, (e.g. ROWLANDS 1958, in Vicia; LEWIS and CROWE 1956, in Origanum; JAIN, WORKMAN, and ALLARD 1966, in barley), these models have to be modified accordingly and the appropriate estimation formulas derived to take other than simple zygotic selection into account.

An alternative approach to the problem of multi-stage selection is to gather data on genotypic frequencies at various stages during the life cycle. Two such studies have now been undertaken in wild oats and barley populations to estimate the individual components of fitness (JAIN, unpublished). Results in barley, lima bean and wild oat populations suggest marked heterozygote advantage at several of the chromosome segments investigated and it appears very likely that these estimates represent the net differences arising cumulatively at various stages. Among other reports on multiple stage selection which may be cited are the work of ROWLANDS (1958) in Vicia, of COTTER (1963) in Ephestia, and of SHEPPARD and COOK (1962) and WORKMAN (1964) in Panaxia. In fact, in the case of Panaxia, it was suggested that even a model based on partial dissortative mating, differential female fertility and differential viability might not be sufficient to explain the wing color pattern polymorphism, but the analysis illustrated certain complex relationships between the mating system and the selective forces. The knowledge of individual components of fitness would of course be very useful in studying the problems of efficiency of selection, substitutional cost and in matters

related to determining the ecological strategy in evolution (e.g. see SKELLAM 1951; LEWONTIN 1965). Despite a lack of knowledge of the individual components of the selective forces, estimation of the net fitnesses based on genotypic frequencies in successive generations appears to be, in general, a satisfactory approximation.

## SUMMARY

Three one-locus models of zygotic selection in populations mating by mixed random mating and selfing were compared. The models differed either in the mode of selection or in the stage of the life cycle at which genotypic frequencies were determined. It was shown that the equilibrium stage, and the approach to equilibrium could be described solely in terms of gene frequency and an extended use of WRIGHT's Fixation Index, $F$.—For each model, formulas for estimation of viabilities were compared by analysis of data from experimental barley and lima bean populations and from multi-stage selection data obtained by computer simulation. Estimation based on frequencies determined at a partially selected stage provides incorrect estimates of the net fitness values. However, the deviation from correct estimates decreases with an increasing amount of selfing, and appears to be of little practical importance in the analysis of predominantly self-fertilized populations.

## LITERATURE CITED

ALLARD, R. W., and P. L. WORKMAN, 1963  Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. Evolution **18:** 470–480.

COTTER, W. B., JR., 1963  Population genetic studies of alleles at the $a^+$ locus in *Ephestia Kühniella*. Evolution **17:** 233–248.

HARDING, J. A., R. W. ALLARD, and D. G. SMELTZER  Population studies in predominantly self-pollinated species. VIII. Frequency dependent selection in *Phaseolus lunatus*. Proc. Natl. Acad. Sci. U.S. (To be submitted.)

HAYMAN, B. I., 1953  Mixed selfing and random mating when homozygotes are at a disadvantage. Heredity **7:** 185–192.

HAYMAN, B. I., and K. MATHER, 1953  The progress of inbreeding when homozygotes are at a disadvantage. Heredity **7:** 165–183.

JAIN, S. K., 1961  Studies on the breeding of self-pollinating cereals. The composite cross bulk population method. Euphytica **10:** 315–324.

JAIN, S. K., and R. W. ALLARD, 1960  Population studies in predominantly self-pollinated species. I. Evidence for heterozygote advantage in a closed population of barley. Proc. Natl. Acad. Sci. U.S. **46:** 1371–1377.

JAIN, S. K., P. L. WORKMAN, and R. W. ALLARD  Evidence for gametic selection at the aleurone color loci in barley. (In preparation.)

LEWIS, D., and L. K. CROWE, 1956  The genetics and evolution of gynodioecy. Evolution **10:** 115–125.

Lewontin, R. C., 1958 A general method for investigating the equilibrium of gene frequency in a population. Genetics **43**: 419–434. —— 1964 Selection for colonizing ability. pp. 79–94. *The Genetics of Colonizing Species*, Edited by H. G. Baker and G. L. Stebbins. Academic Press, New York.

Li, C. C., 1955 The stability of an equilibrium and the average fitness of a population. Am. Naturalist **89**: 281–296.

Prout, T., 1965 The estimation of fitness from genotypic frequencies. Evolution **19**: 546–551.

Rowlands, D. G., 1958 The nature of the breeding system in the field bean (*V. faba* L.) and its relationship to breeding for yield. Heredity **12**: 113–126.

Sheppard, P. M., and L. M. Cook, 1962 The manifold effects of the Medionigra gene of the moth *Panaxia dominula* and the maintenance of a polymorphism. Heredity **17**: 415–426.

Skellam, J. G., 1951 Random dispersal in theoretical populations. Biometrika **38**: 196–218.

Workman, P. L., 1964 The maintenance of heterozygosity by partial negative assortative mating. Genetics **50**: 1369–1382.

Wright, S., 1965 The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19**: 395–420.