THE RELATION BETWEEN FITNESS COMPONENTS AND POPULATION PREDICTION IN DROSOPHILA. II: POPULATION PREDICTION

TIMOTHY PROUT

Department of Life Sciences, University of California, Riverside 92502

Received October 22, 1970

THIS is the second of two articles (PROUT 1971) describing an experimental system for estimating fitness components in Drosophila. The system is principally concerned with the fitness components operating in the adult phase of the life cycle where fertility and mating effects can generate complications in the pattern of selection. The objective is the estimation of a sufficient set of adult components for the purpose of predicting genotype frequencies in a population. In the first article the experimental system is described in detail and then applied to certain fourth chromosome mutants of *Drosophila melanogaster*. In this article the estimates of fitness components are incorporated into recurrence equations which are then used to predict the behavior of experimental populations segregating for these same mutants.

MATERIALS AND METHODS

Drosophila melanogaster fourth chromosome recessive mutants eyeless (ey^2) and shaven (sv^n) in repulsion linkage constitute the genetic system used to test the method. Because of the negligible recombination within fourth chromosomes, the two mutants are treated as segregating alleles which complement to produce a wild-type heterozygote. The three genotypes will be denoted *ee*, *e/s*, *ss*, for eyeless, heterozygote, and shaven, respectively.

Five populations were started with *e/s* flies. Each population was maintained in a half-pint culture bottle containing medium. The eggs laid were allowed to develop at 25°C for two weeks, at the end of which time the adult flies were transferred to a new bottle for 24 hr of egg laying, after which the adult flies were removed, and a sample of 100 (ignoring sex) was classified and counted. The eggs resulting from the 24 hr of egg laying constituted the beginning of the next generation.

The five populations, designated A through E, were maintained for 25 generations. As will be shown presently, all populations quickly established a polymorphic equilibrium. In order to provide data for testing the theory, artificial perturbations of the equilibrium were introduced twice during the history of the populations, at generation 9 and at generation 21.

A perturbation was accomplished in the following way: For a given population at the end of the 2-week period, instead of transferring the flies as usual, the entire population was etherized and all flies were discarded except the females of one of the homozygous phenotypes (eyeless or shaven). Then only these females, most of whom had previously mated, were used as founders of the next generation by the usual 24 hr of egg laying. The next and subsequent generations were treated routinely.

At generation 9, populations A and B were selected for shaven (against eyeless), D and E were selected in favor of eyeless (against shaven), while population C was left unperturbed. At generation 21, A and B were selected for eyeless, D and E were selected for shaven, and C was left unperturbed.

RESULTS

The fitness component estimates described in detail in the first article are recorded here again in Table 1. All components are normalized to the heterozygote *e/s*. Here and elsewhere the three genotypes *ee*, *e/s*, *ss* are indexed $i = 1, 2, 3$, respectively, for females; and $j = 1, 2, 3$ for males. Viability measures the egg-to-adult survival component. The adult components are termed fecundity for females and virility for males and encompass adult survival as well as fertility and mating effects. It can be seen that the virility components, V_{ij}, vary with the female partner to which the males are mated. On the other hand it was found that fecundity, F_i, was independent of male partner and so the females are characterized by one rather than three sets of fitness components. These are the fitness components which will be used presently for predicting the behavior of the populations.

The culture regime for the populations resulted in population sizes between 200 and 400 individuals. Genotypic composition based on samples of 100 (usually) are shown in Table 2. The two points at which the populations were perturbed, PI and PII, are indicated in the table. The obvious effect of the method of perturbing can be readily seen in the absence of that homozygous

TABLE 1

Components of fitness estimates

Females			Males		
VIABILITY					
$L_{1♀}$	1	$L_{3♀}$	$L_{1♂}$	1	$L_{3♂}$
.865 (.039)	1	.934 (.039)	.839 (.036)	1	.777 (.038)

ADULT COMPONENTS

Fecundity				Virility		
F_1	1	F_3	constant ♀ i =	V_{i1}	1	V_{i3}
1.037 (.122)	1	.458 (.068)	1	.363 (.074)	1	.039 (.033)
			2	.243 (.042)	1	.122 (.037)
			3	.135 (.036)	1	-.018 (.030)

$L♀_i$, $L♂_j \equiv$ female and male viabilities, respectively; $F_i \equiv$ adult female components or fecundity; $V_{ij} \equiv$ adult male components or virility of males of genotype j when mated to female constant parent of genotype i. The genotypes *ee*, *e/s*, *ss* are indexed $i = 1,2,3$, respectively for females, and $j = 1,2,3$ for males. Standard error in parentheses.

TABLE 2

Observed absolute frequencies in populations A through E of genotypes ee, e/s, *and* ss

	Population A					Population B				
Generation	ee	e/s	ss	Total	Pe	ee	e/s	ss	Total	Pe
0	0	100	0	100	.50	0	100	0	100	.50
1	10	35	31	76	.36	11	62	14	87	.48
2	19	53	28	100	.46	29	52	19	100	.55
3	22	54	22	98	.50	24	57	24	105	.50
4	34	44	20	98	.57	24	52	24	100	.50
5	30	57	13	100	.59	9	18	9	36	.50
6	34	46	20	100	.57	27	44	29	100	.49
7	28	56	16	100	.56	26	60	14	100	.56
8	48	44	8	100	.70	33	54	13	100	.60
9 PI	0	34	33	67	.25	PI 0	74	26	100	.37
10	14	64	22	100	.46	16	55	29	100	.44
11	20	66	14	100	.53	23	54	23	100	.50
12	21	58	21	100	.50	36	46	18	100	.59
13	31	47	22	100	.55	23	50	27	100	.48
14	27	55	18	100	.55	33	57	10	100	.62
15	40	48	12	100	.64	33	52	15	100	.59
16	43	42	15	100	.64	37	53	10	100	.64
17	37	47	16	100	.61	20	65	15	100	.53
18	34	49	14	97	.60	31	54	15	100	.58
19	29	62	9	100	.60	28	60	12	100	.58
20	36	51	13	100	.62	28	56	16	100	.56
21 PII	75	25	0	100	.88	PII 49	51	0	100	.75
22	50	44	6	100	.72	50	42	8	100	.71
23	47	45	8	100	.70	36	45	19	100	.59
24	30	57	13	100	.59	24	53	23	100	.51
25	31	53	16	100	.58	24	59	17	100	.51

class which is opposite to the homozygous mothers which founded the perturbed generations (9 and 21).

The allele frequencies of eyeless are shown in Figure 1. The mean of all five populations is shown except for perturbed points and several subsequent generations where individual populations are shown. This figure displays the following general characteristics of the system: First, it is clear that the perturbations were effective. Secondly, it is clear that there is an equilibrium which favors eyeless somewhat over shaven. And finally, the equilibrium appears to be very stable judging by the rapid return of the perturbed populations.

TABLE 2—Continued

	Population C						Population D				
Generation	ee	e/s	ss	Total	Pe		ee	e/s	ss	Total	Pe
0	0	100	0	100	.50		0	100	0	100	.50
1	12	44	17	73	.47		18	46	18	82	.50
2	16	68	16	100	.50		34	51	15	100	.56
3	38	78	21	137	.57		49	43	8	100	.71
4	26	50	24	100	.51		35	50	15	100	.60
5	32	48	20	100	.56		25	59	16	100	.55
6	42	43	15	100	.64		45	45	10	100	.68
7	31	51	18	100	.57		47	41	12	100	.68
8	60	36	4	100	.78		52	42	6	100	.73
9	30	54	16	100	.57	PI	62	38	0	100	.81
10	28	54	18	100	.55		41	50	9	100	.66
11	37	52	11	100	.63		43	42	15	100	.64
12	31	52	17	100	.57		32	56	12	100	.60
13	33	50	17	100	.58		38	48	14	100	.62
14	25	60	15	100	.55		27	57	16	100	.56
15	26	57	17	100	.55		38	51	11	100	.64
16	23	63	14	100	.55		30	53	17	100	.57
17	25	56	19	100	.53		32	57	11	100	.54
18	23	59	18	100	.53		34	56	10	100	.62
19	20	57	10	87	.56		22	57	21	100	.51
20	31	59	10	100	.61		25	54	21	100	.52
21	39	50	11	100	.64	PII	0	53	47	100	.27
22	37	48	15	100	.61		13	41	46	100	.27
23	28	61	11	100	.59		16	49	35	100	.41
24	36	49	15	100	.61		26	59	15	100	.56
25	35	47	18	100	.59		22	65	13	100	.55

THEORY

It is now necessary to incorporate the fitness component estimates of Table 1 into a model which will provide the recurrence relationship between genotype frequencies of successive generations.

Table 3 schematically presents the major steps in the derivation, and equations (1), below, are the final resulting recurrence equations.

As indicated in Table 3, it is convenient to divide the process whereby an egg of one generation produces an egg in the next generation into two phases—a survival phase and a mating phase.

TABLE 2—Continued

		Population E				
Generation		ee	e/s	ss	Total	Pe
0		0	100	0	100	.50
1		20	48	15	83	.53
2		23	60	17	100	.53
3		22	57	24	103	.49
4		38	49	13	100	.63
5		43	35	22	100	.61
6		39	50	11	100	.64
7		50	36	14	100	.68
8		28	58	14	100	.56
9	PI	60	40	0	100	.80
10		52	42	6	100	.73
11		37	52	11	100	.63
12		44	45	8	100	.67
13		38	56	6	100	.66
14		31	59	10	100	.61
15		36	51	13	100	.62
16		36	53	11	100	.63
17		31	61	8	100	.62
18		34	55	11	100	.62
19		23	60	17	100	.53
20		23	69	8	100	.59
21	PII	0	54	46	100	.27
22		9	73	18	100	.41
23		19	60	21	100	.49
24		24	58	18	100	.53
25		26	61	13	100	.57

$P_e \equiv$ relative frequency of the allele, eyeless.
PI, PII $\equiv$ the first and second perturbations, respectively.

The survival phase is sex dependent not only because actual survival or viability is sex dependent, but also because on the female side, the fecundity components are evidently independent of male partner, and so may be treated as though they were measures of survival.

These surviving female and male genotypes, q_i and p_j, then enter the mating phase to produce mated pairs which are shown in expanded form at the bottom of Table 3. It can be seen that the virility parameters, V_{ij}, are simply used in weighting and normalizing the male proportions, p_j, according to each type of

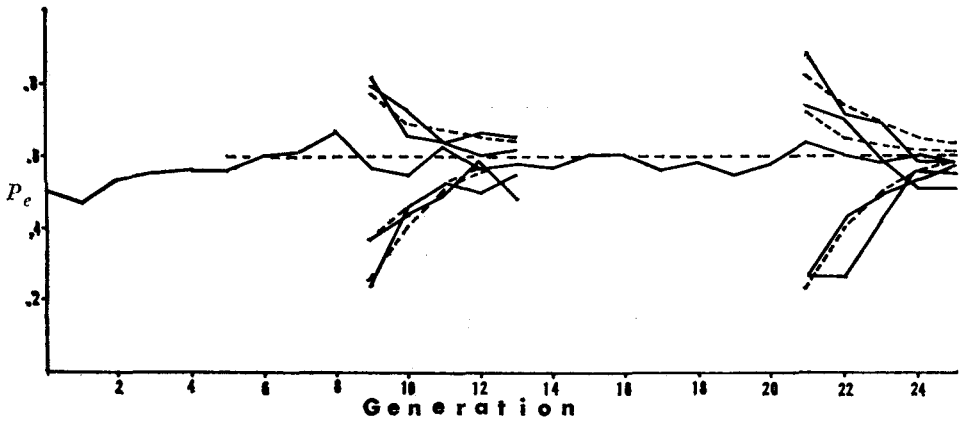

FIGURE 1.—History of the experiment as shown by eyeless allele frequency, P_e. Except for the perturbed generations 9–13 and 21–25, only the means of the five populations are shown. The dashed lines (– – –) represent the theoretical predictions.

female partner, *i*. (The assumptions involved in this step will be examined presently.)

The following recurrence equations are obtained by generating Mendelian expectations from each cell of the table of mated pairs and grouping according to genotype.

$$\tilde{Q}_1' = \bar{X}^{-1}\{\tilde{Q}_1 X_1 (\tilde{Q}_1 Y_{11} + \tfrac{1}{2}\tilde{Q}_2)\bar{Y}_1^{-1} + \tfrac{1}{2}\tilde{Q}_2(\tilde{Q}_1 Y_{21} + \tfrac{1}{2}\tilde{Q}_2)\bar{Y}_2^{-1}\} \qquad (1\text{-}1)$$

$$\tilde{Q}_2' = \bar{X}^{-1}\{\tilde{Q}_1 X_1 (\tilde{Q}_3 Y_{13} + \tfrac{1}{2}\tilde{Q}_2)\bar{Y}_1^{-1} + \tfrac{1}{2}\tilde{Q}_2 + \tilde{Q}_3 X_3 (\tilde{Q}_1 Y_{31} + \tfrac{1}{2}\tilde{Q}_2)\bar{Y}_3^{-1}\} \qquad (1\text{-}2)$$

$$\tilde{Q}_3' = \bar{X}^{-1}\{\tilde{Q}_3 X_3 (\tilde{Q}_3 Y_{33} + \tfrac{1}{2}\tilde{Q}_2)\bar{Y}_3^{-1} + \tfrac{1}{2}\tilde{Q}_2(\tilde{Q}_3 Y_{23} + \tfrac{1}{2}\tilde{Q}_2)\bar{Y}_2^{-1}\} \qquad (1\text{-}3)$$

where $\tilde{Q}_i' \equiv$ frequency of genotype i among fertilized eggs of the next generation

$$X_i = L_{\varphi i} F_i$$
$$\bar{X} = \sum_i X_i \tilde{Q}_i$$
$$Y_{ij} = L_{\sigma j} V_{ij}$$
$$\bar{Y}_i = \sum_i Y_{ij} \tilde{P}_j$$

The new parameters X_i and Y_{ij} simply represent the collapse of survival (viability) with adult parameters as indicated.

The way in which the virility parameters V_{ij} are incorporated requires some justification.

In the first article it was shown that the virility parameters could be interpreted in terms of a simple collision model of mating behavior: If a mixture q_i of virgin females is placed with a mixture of males p_j, then after time t, the proportion of mated females q_{ijt} is given by

$$q_{ijt} = \frac{q_i p_j a_{ij}}{\bar{w}_i}(1 - e^{-C\bar{w}_i t}) \qquad (2)$$

where $C \equiv$ fraction of virgin females which encounter males in unit time;

TABLE 3

Schematic presentation of notation and some relationships necessary for construction of recurrence equations (1)

SURVIVAL PHASE

	Female	Male
Fertilization genotype frequencies by sex. When $i = j$, $Q_i = P_j$	$\tilde{Q}_i$	$\tilde{P}_j$
Viability parameters	$L_{\varphi i}$	$L_{\sigma j}$
Fecundity parameters	F_i	--

MATING PHASE

Genotype frequencies at start of mating

$$q_i = \frac{L_{\varphi i} F_i \tilde{Q}_i}{\sum_i L_{\varphi i} F_i \tilde{Q}_i} \qquad p_j = \frac{L_{\sigma j} \tilde{P}_j}{\sum_j L_{\sigma j} \tilde{P}_j}$$

Mating parameters V_{ij}

$\bar{V}_i = \sum_j P_j V_{ij}$

Mated pairs

Females \ Males	P_1	P_2	P_3
q_1	$\frac{q_1 p_1 V_{11}}{\bar{V}_1}$	$\frac{q_1 p_2 V_{12}}{\bar{V}_1}$	$\frac{q_1 p_3 V_{13}}{\bar{V}_1}$
q_2	$\frac{q_2 p_1 V_{21}}{\bar{V}_2}$	$\frac{q_2 p_2 V_{22}}{\bar{V}_2}$	$\frac{q_2 p_3 V_{23}}{\bar{V}_2}$
q_3	$\frac{q_3 p_1 V_{31}}{\bar{V}_3}$	$\frac{q_3 p_2 V_{32}}{\bar{V}_3}$	$\frac{q_3 p_3 V_{33}}{\bar{V}_3}$

$a_{ij} \equiv$ conditional probability of mating, given an encounter;
$t \equiv$ time from the start;
$\bar{w}_i \equiv \sum_j a_{ij} p_j$.

And when all females have mated ($t \to \infty$)

$$q_{ij(t\to\infty)} = \frac{q_i p_j a_{ij}}{\bar{w}_i} \tag{3}$$

The mating parameter a_{ij} can be normalized to the male heterozygote ($j=2$) for each female, i

$$q_{ij(t\to\infty)} = \frac{q_i p_j A_{ij}}{\bar{W}_i}$$

where

$$\overline{W}_i = \sum_j A_{ij} p_j$$

$$A_{ij} = \frac{a_{ij}}{a_{iz}}$$

The substitution of the measured virility components V_{ij} for A_{ij}, as has been done in Table 3, assumes the following to be true:

1. All of the surviving females are mated. Also, the above collision model of mating behavior strictly assumes that each female mates just once. However, this is equivalent to allowing for multiple matings (involving the parameter a_{ij} in each one) so long as a female's fecundity is independent of the number of matings.

2. The females have the same pattern of male preference when in the presence of females of different genotypes as they do in the virility experiments where they are in the presence of other females of their own genotype only (see first article).

3. The different male genotypes have no differential effect on female egg laying. The estimation of the virility components, V_{ij}, was based on the relative contribution of different male genotypes to the progeny when mated to a given female genotype, i. If the males differentially diminish or enhance female egg laying by a factor M_j (see first article), this effect is confounded with the factor A_{ij} so that, as measured,

$$V_{ij} = A_{ij} M_j$$

However, for the construction of the recurrence equation it is necessary to assume that all fecundity differences are determined by the female genotype, measured by F_i, and that therefore $M_j = 1$ for all j. This assumption is further illustrated by the fact that the rows of the mating table in Table 3 sum to q_i, since all of the fecundity effects have already been accounted for by F_i. That this assumption must be made in order to write the recurrence equation constitutes proof that if $M_j \neq 1$, then the values of this component must be known explicitly.

The populations to which the model will be applied accumulate adults for four days and so this model of mating can be regarded as applying to each cohort of newly hatched females. An obvious defect is that the model does not take into account differential development time.

Before equations (1) can be applied to the data, two modifications are necessary. First, genotype frequencies were actually determined on partially selected stages, namely after the operation of viability within each generation. This is a common situation in selection experiments and simply requires that the recurrence equation to be fitted to the data must be advanced up to the stage of observation (PROUT 1969). This means that both the female and male viability components which are collapsed in equations (1) must, in fact, be resolved again explicitly. The second modification arises from the fact that the population data of this experiment were collected without regard to sex which means that the viability components must be averaged between the sexes, as follows:

$$L_1 = \tfrac{1}{2}\ (L_{1\,♀} + L_{1\,♂}) = .852$$
$$L_3 = \tfrac{1}{2}\ (L_{3\,♀} + L_{3\,♂}) = .856$$

This should introduce only a very small error in this case since the viability effects are very small and so also is their sex dependence. When these two corrections are applied to equations (1), and the component values of Table 1 are substituted, the following working equations result:

$$\begin{aligned} q_1' &= .852T^{-1}(1.037q_1(.363q_1+.5q_2)\overline{V}_1^{-1}+.5q_2(.243q_1+.5q_2)\overline{V}_2^{-1}) \\ q_2' &= T^{-1}(1.037q_1(.039q_3+.5q_2)\overline{V}_1^{-1}+.5q_2+.458q_3(.135q_1+.5q_2)\overline{V}_3^{-1}) \\ q_3' &= .856T^{-1}(.458q_3+.5q_2\overline{V}_3^{-1}+.5q_2(.122q_3+.5q_2)\overline{V}_2^{-1}) \end{aligned} \qquad (4)$$

where $q_i, q_i' \equiv$ partially selected (observed) frequencies in successive generations, respectively;

$T \equiv$ normalizing factor such that $\Sigma\, q_i' = 1$;

$\overline{V}_1 = .363q_1 + q_2 + .039q_3$

$\overline{V}_2 = .243q_1 + q_2 + .122q_3$

$\overline{V}_3 = .135q_1 + q_2$, assuming $V_{33} = 0$.

ANALYSIS OF DATA

The theory just developed will now be applied to the population data. This analysis will be carried out in two parts: first the equilibrium generations will be examined, and, second, the path of return to equilibrium by the perturbed generations will be examined.

1. *Equilibrium:* The theoretical equilibrium frequencies were obtained by simply iterating equations (4) until genotype frequencies stopped changing. (Because of the powerful overdominance in the virility components, it was assumed that the single equilibrium point was globally stable.) The results of such iteration give the following theoretical equilibrium values:

Genotype frequencies			Gene frequencies	
ee	*e/s*	*ss*	*e*	*s*
.351	.506	.143	.604	.396

It is interesting to note that the Hardy-Weinberg frequencies for these gene frequencies are (.365, .478, .157) for (*ee*, *e/s*, *ss*, respectively), and that these are not drastically different from those given by the theory in which, of course, powerful selection is taking place.

In order to compare this equilibrium prediction with the population data, it was arbitrarily decided that equilibrium was achieved in generation 5 after the start of the experiment, as well as the 5th generation after a perturbation. Thus the following generations were used for comparison with the theoretical equilibrium: generations 5–8 and 14–20 for populations A,B,D, and E; and generations 5–25 for the unperturbed population C.

In Table 4 the totals are given for each genotype over these generations for each population. Also shown, on the right, are the results of a chi-square test for homogeneity as well as a chi-square testing the totals against the theory. It can be seen that all but population A showed significant heterogeneity, a result not

TABLE 4

Equilibrium generation totals

Population	*ee*	*e/s*	*ss*	Total	χ^2_H	df	χ^2_{th}	df
A	386	557	154	1097	28.9	20	.06	2
B	305	573	158	1036	32.8*	20	14.7**	2
C	672	1106	309	2087	79.9**	40	7.7*	2
D	377	572	151	1100	53.1**	20	.9	2
E	374	587	139	1100	51.7**	20	4.2	2
Grand totals	2114	3395	911	6420				

	ee	*e/s*	*ss*
Proportion in Grand Total	.329	.529	.142
Theoretical	.351	.506	.143
Heterogeneity among Population Totals	$\chi^2 = 12.2$		df = 8
Comparison between grand total and Theory	$\chi^2 = 15.6$**		df = 2

$\chi^2_H \equiv$ chi-square for heterogeneity among generations; $\chi^2_{th} \equiv$ chi-square for comparison between total and theory; df $\equiv$ degrees of freedom; * $\equiv$ significant at 5% level; ** $\equiv$ significant at 1% level.

unexpected for this type of experiment. Three of the five populations showed good agreement with theory (A, D, and E), and only population B showed a strong departure. The grand totals, over all the populations, are shown. A test for homogeneity *among the totals* over the five populations gave a $\chi^2 = 12.2$ which, with 8 degrees of freedom, is not significant, but the grand totals showed a significant departure from theory: $\chi^2 = 15.3$ with 2 degrees of freedom.

Finally in Table 4, the grand totals are expressed as decimal fractions for comparison with the predicted equilibrium. There it can be seen that actually the discrepancy between the two is not very great. One means of evaluating this discrepancy is to determine how large a change in the parameters of the theoretical equation is required to produce the actual equilibrium observed. There are ten such parameters. However, it is a reasonable guess that the viability parameters, L_1 and L_3, might be different under the population conditions (higher larval densities) as opposed to the conditions under which they were estimated (lower larval densities). Whether this is true or not, it is still interesting to note that if $L_1 = .797$ and $L_3 = .800$, then the theoretical equation, with the adult parameters as before, would give the observed equilibrium frequencies noted for the grand totals in Table 4. This small downward adjustment of viabilities as compared with the former estimates ($L_1 = .852, L_3 = .856$) would account for the discrepancy between observed and predicted equilibrium.

This procedure amounts to using the population equilibrium data to estimate viabilities, and since the grand totals possess only two degrees of freedom, both were used to estimate the two viabilities and none are left for a test of goodness of fit. However, the perturbed generations provide additional degrees of freedom, and these new estimates of viability will be discussed again in connection with the analysis in the following section.

2. *The perturbed generations:* The starting point for each perturbation was produced by the progeny of nonvirgin homozygous females, as explained earlier. This progeny generation was sampled in the same manner as other generations, so that the true starting point was unknown except to the extent that one homozygous class was missing. Therefore, in each case, maximum likelihood estimates of the starting points were obtained. The likelihood function was constructed by choosing a starting point and using this as generation zero in equations (4), which were then used to produce four succeeding generations. The values for five generations (including generation 0) were then regarded as the true population values of which the observed data (including generation 0) were samples. The likelihood function obtained from this relationship was then maximized with respect to the starting point value.

Graphical presentation of the results of this fitting procedure are shown in Figures 1 and 2. In Figure 1, showing the history of the populations expressed as the eyeless allele frequencies, the dashed lines show the theoretical allele frequencies obtained from equations (4). However, since Hardy-Weinberg frequencies are not necessarily expected, the full description of each generation requires two dimensions. In Figure 2, a–f, the paths of population points are presented by expressing the shaven homozygote q_3 as a function of the eyeless homozygote q_1. Figures 2a–2d each show two of the perturbed generations with the theoretical and actual paths. In Figure 2e the equilibrium generations are shown as unconnected points together with the theoretical path to equilibrium from the beginning of the experiment. Finally, Figure 2f shows several theoretical paths together with the Hardy-Weinberg parabola,

$$q_3 = 1 - 2\sqrt{q_1} + q_1$$

In this last figure it can be seen that this system very rapidly approaches an equilibrium path which closely parallels the Hardy-Weinberg parabola as the population rapidly moves toward the equilibrium point.

In Table 5 the first three columns show for each perturbation the estimated starting points and immediately below these their observed sample values. The fourth column gives the chi-squares testing goodness of fit. The eight degrees of freedom for these chi-squares are obtained as follows: Including the starting point there are five generations. The starting generation has only two classes giving one degree of freedom, while the remaining four generations have three classes giving two degrees of freedom each. From these 9 degrees of freedom one is deducted for estimating the starting point. The first set of four perturbations indicates a good fit. The first population, A, gave a significant chi-square; but since the first four perturbations were done at the same time, the sum of four chi-squares can be considered, and this χ^2 is not statistically significant. On the other hand the second set of perturbations shows a poor fit. The fifth and sixth columns of Table 5 are the results of the same kind of analysis except that the viabilities $L_1 = .797$ and $L_3 = .800$, obtained from the equilibrium data, were used. The fifth column shows the estimated heterozygote starting frequency, and the sixth column has the resulting χ^2. The same pattern is revealed as before between the first and second perturbation experiments. The effect of using the lower viabilities was to

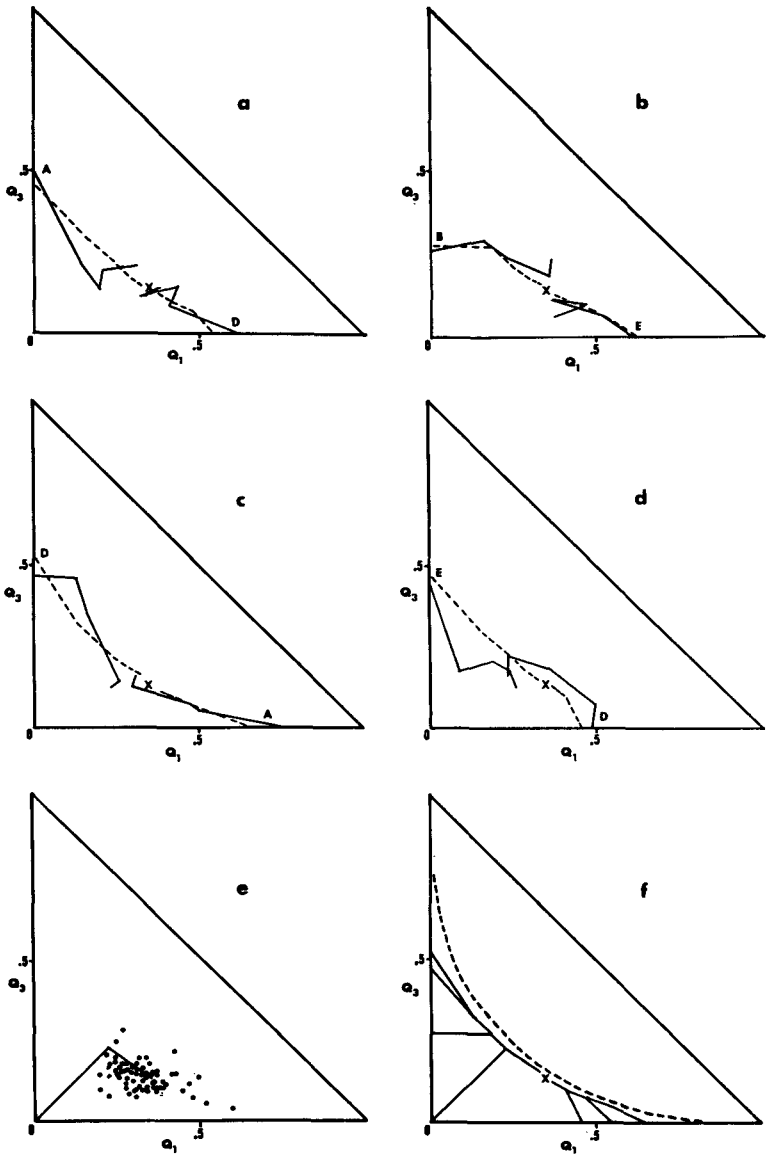

FIGURE 2.—a through d: Theoretical (----) and observed (——) paths of perturbed population points. The coordinates of a point are the two homozygote frequencies Q_3 and Q_1 for shaven and eyeless, respectively. The particular populations A, B, D, E are indicated. Graphs a and b represent the first perturbation and c and d the second. $\times \equiv$ theoretical equilibrium point. Graph e shows the observed equilibrium points (unconnected) and the theoretical path (——) from the start of the experiment. Graph f shows several theoretical paths (——) and the Hardy-Weinberg parabola (----).

reduce most of the 8 chi-squares, thus indicating agreement between the perturbation and equilibrium data on the suggestion that viabilities were lower in the populations than in the viability experiments.

TABLE 5

Chi-squares and maximum likelihood estimates resulting from fitting equations (4) to perturbed generations

Population		Starting frequencies ee	e/s	ss	(4) $L_1 = .797$ $L_3 = .800$ χ^2 (df)	(5) Start e/s	(6) χ^2 (df)	(7) Start e/s	(8) MLE Viabilities L_1	L_3	(10) χ^2 (df)
FIRST PERTURBATION											
A	MLE	0	.52	.48	17* (8)	.53	14 (8)	.53	.67	.68	11 (6)
	OBS	0	.51	.49							
B	MLE	0	.73	.27	7 (8)	.73	8 (8)	.74	.85	.99	5 (6)
	OBS	0	.74	.26							
D	MLE	.55	.45	0	9 (8)	.44	8 (8)	.42	.80	.95	7 (6)
	OBS	.62	.38	0							
E	MLE	.60	.40	0	8 (8)	.38	8 (8)	.41	.83	.64	6 (6)
	OBS	.60	.40	0							
					41 (32)		38 (32)				30 (24)
SECOND PERTURBATION											
D	MLE	0	.47	.53	27**(8)	.48	27**(8)	.51	.75	.97	20**(6)
	OBS	0	.53	.47							
E	MLE	0	.53	.47	21**(8)	.53	16* (8)	.54	.59	.56	6 (6)
	OBS	0	.54	.46							
A	MLE	.66	.34	0	14 (8)	.33	10 (8)	.27	.68	.88	4 (6)
	OBS	.75	.25	0							
B	MLE	.46	.54	0	28**(8)	.53	25**(8)	.46	.71	1.11	10 (6)
	OBS	.49	.51	0							
					91**(32)		79**(32)				39* (24)

Columns (1)–(3) give the observed (OBS) and maximum likelihood estimate (MLE) of the starting frequencies; Columns (5) and (6) give the maximum likelihood estimate of the heterozygote starting frequency and resulting chi-square, assuming viabilities $L_1 = .797$ and $L_3 = .800$. Columns (7)–(10) give the maximum likelihood estimates of heterozygote starting frequency, the two viabilities L_1 and L_3, and the resulting chi-square.

Because of the highly significant chi-squares in the second perturbation experiment, an improved fit was attempted by obtaining maximum likelihood estimates of not only the starting point but also the two viabilities for each individual population. The final four columns of Table 5 show the results of this investigation. The estimates of the starting point and individual viabilities are shown and finally the resulting chi-squares, whose degrees of freedom are reduced to 6 due to the estimation of the two viability parameters. This last procedure resulted in a considerably better fit in the second perturbation experiment. Only one of the chi-squares remains significant, and the total of the four chi-squares is just barely significant at the 5% level.

This result suggests that there may have been some differentiation of the populations with respect to viability toward the end of their history. However, no correlation could be found between the estimated viabilities for individual popula-

tions when in equilibrium (not shown) and the viabilities estimated for the same population after the final perturbation. Because of this lack of correlation, because of the small sample sizes in each generation (100), and because components other than viability could be causing the discrepancies, it was decided that a serious investigation of population differentiation was not warranted.

DISCUSSION

The principal conclusion is that this method of estimating fitness components, discussed in detail in the first article, is capable of producing reasonably good population prediction. One probable reason for this success is that it was possible to estimate the fitness components under conditions fairly similar to those of the rather simple population regime. However the conditions were not identical. In the population, during the 4 days of hatching (days 10–14), an age structure developed; while the component experiments were done with one age class. In the population, mating was carried out in a genotypic mixture of both sexes; whereas the virility components were estimated using one female genotype at a time. Finally, in the population there was greater larval crowding, which fact provided some justification for new viability estimates from the population itself.

Nevertheless, the reasonably good population fits suggest that one may have a certain degree of confidence that this method of estimating fitness components can provide information concerning a relatively intricate mode of selection occurring in a population whose ecology is somewhat more complicated than conditions of the component experiments.

It is interesting to note how little information would be gained in this particular case from a study of the viability component alone or by a "field test" for agreement with Hardy-Weinberg proportions. I have emphasized the weakness of these tests for selection elsewhere (PROUT 1969) and here the point is illustrated by a system with powerful selection in the adult components, but because of the relatively weak viability effect, the population point moves and settles down always close to the Hardy-Weinberg parabola (Figure 2f).

It is possible in this particular case that a simpler model of selection would fit the data equally well and that a proper analysis of the population data alone (ANDERSON 1969; WILSON 1968) would be sufficient, thus rendering component experiments unnecessary. A study of this question was not done partly because the small scale of the population data did not seem to warrant it, but mainly because the emphasis of these two articles is on the methodological principle of performing component experiments in such a way that the measures of component effects are designed for incorporation into recurrence equations for population prediction. It is for this purpose that the recurrence equations (1) and (4) are displayed and used for prediction with the full level of complexity that the component experiments indicated.

Also this shaven–eyeless system may be regarded as a living case which begins to approach the kind of selection which, for all practical purposes, is inaccessible through analysis of population data alone (PROUT 1969). If the virility inter-

actions were stronger, this plus the powerful sex dependence in the adult phase would produce frequency dependence of fitnesses and generally poor fits in any sort of curve-fitting procedure performed on the data of large-scale population experiments.

The model, equations (1), used for prediction in these experiments is a rather general one allowing for nonrandom mating. It may be of some interest to examine the relationship between this model and other nonrandom mating models which have appeared in the literature. There appears to be two types of such models: those in which it is assumed all surviving females are mated, and those which assume equal likelihood of mating by the two sexes.

The first type, to which the model of this study belongs, has been termed "asymmetrical" by WATTERSON (1959) or "limited selection" by SPIETH (1970), and special cases of this model have been used for theoretical studies of negative assortative mating and incompatibility systems by the above authors and others (see review by KARLIN 1969).

The second type of model has been termed "symmetrical" (WATTERSON 1959) or "mass action" (KARLIN 1969) and special cases of this model have been examined by BODMER (1965), CANNING (1969), PROUT (1969), KEMPTHORNE and POLLAK (1970) and LEWONTIN, KIRK and CROW (1968). In its most general form this latter model simply attaches a single parameter, W_{ij}, to random pairs of unselected females $\tilde{Q}_i$ and males $\tilde{P}_j$, thus

$$\tilde{Q}_i \tilde{P}_j W_{ij}$$

The next generation is produced by applying Mendelian expectations to the above expression. An appealing aspect of this symmetrical model is that all aspects of selection and mating are encompassed by the eight W_{ij} parameters (assuming normalization). On the other hand, equations (1), representing the asymmetrical model, do not permit the collapse of the two female parameters, X_i with the six male parameters, Y_{ij}. Thus the two models have the same number of parameters but are not formally equivalent.

It is interesting to note that equation (2), describing mating behavior as a repeated collision or "mass action" process, can be used to link the two models, because when $t = 1/c$, then by assuming $e^{-\bar{w}_i} \cong 1 - \bar{w}_i$, equation (2) becomes

$$q_{ijt} \cong q_i p_j a_{ij}$$

which leads to the symmetrical model; and when $t \to \infty$, equation (2) gives equation (3) which leads to the asymmetrical model.

Which of the two models is appropriate (if either) depends, of course, on the mating biology of the species in question. As WORKMAN (1964) has pointed out, the asymmetrical model might best apply to short-lived insects which store sperm, such as Drosophila, and to plants where there is excess of pollen so that seed set is determined by the seed parent only (F_i, in the present notation). Whereas the symmetrical model would better describe monogamous species, or organisms where each mating leads to the production of a brood, after which both male and female become generally available again to produce another brood etc., as perhaps in mammals and other vertebrates.

In fact, it is conceivable that a systematic study of the varieties of mating

biology would be rewarded by the generalization that for most of them, either the symmetrical or the asymmetrical model is a sufficient description for purposes of population prediction.

Finally, there is the question of robustness. It could be that although these two models are formally different, they are fairly robust with respect to either each other or simpler models, not only for the purposes of theoretical study in the sense of LEVINS (1966) but also for the purpose of fitting experimental data. Here too, further study is required.

The writer wishes to acknowledge the invaluable assistance of Dr. DAVID GIBO and Mrs. JOAN MCFARLANE. This work was supported by the National Institutes of Health grant GM-06174.

SUMMARY

This is the second of two articles on the estimation of fitness components and the use of such estimates to predict population genotype frequencies. In the first article an experimental system was described for estimating a set of fitness components sufficient for population prediction. The system was applied to certain fourth chromosome mutants in *Drosophila melanogaster* revealing that these mutants affected all components: viability, fertility, and mating behavior. In this article these components were incorporated into recurrence equations for predicting population genotype frequencies. These equations were tested on experimental populations segregating for these mutants. Several opportunities for test were provided by artificial perturbations of the equilibrium these populations tended to seek. Reasonably good fits were obtained by a chi-square criterion.— The recurrence equations used for prediction entail a repeated collision model of mating behavior such that eventually all females are mated, but not randomly so. This asymmetrical model is compared with other symmetrical models which have appeared in the literature and which take into account nonrandom mating. It is suggested that either this model or the "mass action" model might constitute sufficient descriptions of a great variety of different mating biologies.

LITERATURE CITED

ANDERSON, W. W., 1969 Selection in experimental populations. I: Lethal genes. Genetics **62**: 653–672.

BODMER, W. F., 1965 Differential fertility in population genetics models. Genetics **51**: 411–424.

CANNING, C., 1969 The study of multiallelic genetic systems by matrix methods. Genet. Res. **14**: 167–183.

KARLIN, S., 1969 Equilibrium behavior of population genetic models with nonrandom mating. J. Appl. Probability **5**: 231–313.

KEMPTHORNE, O. and E. POLLAK, 1970 Concepts of fitness in Mendelian populations. Genetics **64**: 125–145.

LEVINS, R., 1966 The strategy of model building in population biology. Am. Scientist **54**: 421–431.

LEWONTIN, R., D. KIRK and J. CROW, 1968 Selective mating, assortative mating, and inbreeding: definitions and implications. Eugenics Quart. **15**: 141–143.

PROUT, T., 1969 The estimation of fitnesses from population data. Genetics **63**: 949–967. ——, 1971 The relation between fitness components and population prediction in Drosophila. I: The estimation of fitness components. Genetics **68**: 127–149.

SPIETH, P., 1970 Ph.D. dissertation. Dept. Biol. University of Oregon, Eugene.

WATTERSON, G. A., 1959 Nonrandom mating, and its effect on the rate of approach to homozygosity. Ann. Human Genet. **23**: 204–220.

WILSON, J., 1968 Experimental determination of fitness interactions in *Drosophila melanogaster* by the method of marginal populations. Genetics **59**: 501–511.

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