

EXPERIMENTAL STUDIES OF THE DISTRIBUTION OF GENE
FREQUENCIES IN VERY SMALL POPULATIONS OF
DROSOPHILA MELANOGASTER. II. BAR¹

SEWALL WRIGHT AND WARWICK E. KERR

The University of Chicago and Universidade de São Paulo

Received February 23, 1954

History by generations

The results from the experiments with Bar, to be described here, present a more complicated situation than those from similar experiments with forked, described in part I of this series (Kerr and Wright, 1954) because of strong selective pressures of various sorts. A more complete analysis can be made, however, because of the possibility of distinguishing accurately all genotypes.

108 lines, each starting from 4 B/+ females and 2 B/0, 2 +/0 males, were carried either to fixation or to generation 10 by random selection of 4 females and 4 males as parents in each generation. The methods were as described in part I. In 3 cases, flies selected as parents included ones that should not have appeared under simple sex linked heredity. These could have been due to nondisjunction but also possibly to nonvirginity of the females. These 3 vials are omitted in the analysis of selection coefficients but as no important changes in gene frequency were involved, the later generations from them have been included.

Table 1 shows the distribution of frequencies of the Bar gene by generation without omission of any vials. It may be seen that wide variability in F₁ was

followed by fixation of type in 15 vials and of Bar in one vial in F₂. In generation 10, type had become fixed in 95 lines and Bar in three, while only 10 lines remained unfixed. It is evident that selection strongly favors type but that Bar can occasionally drift into fixation in spite of this.

A χ^2 test of the distributions of unfixed and newly fixed classes by generation from generation 4 to 10 (grouping generations 7 and 8, 9 and 10, classes 5 and 6, 7 to 12) yields a value of 22.2, 24 degrees of freedom and a probability between .50 and .70. It is evident that equilibrium of the form of the distribution was substantially reached by generation 4 and that an average of 22.0% of the unfixed classes became fixed per generation with respect to type and 0.7% with respect to Bar.

The selection coefficients

Table 2 shows the results from vials according to the arrays of parents. The experiment was not designed for analysis of the mode of selection but something can be deduced. The small numbers of flies counted per vial makes it necessary to consider many vials to obtain significant results but has the advantage that the results do not depend much on the peculiarities of selection in particular cases. F₁ (mothers 040, fathers 22) is here excluded as it is desirable to separate it from the others because of possible systematic differences.

Tables 3 and 4 show the numbers of Bar and type sons produced by different groups of females (no B/B in (3), B/B

¹ Experimental data by W. E. Kerr under a fellowship of the Rockefeller Foundation; mathematical analysis by S. Wright. Acknowledgment is made to Dr. Th. Dobzhansky for material and for his hospitality during the conduct of these experiments. Acknowledgment is also made to Dr. J. Crow for material and suggestions. Analysis was aided by a grant from the Wallace C. and Clara A. Abbott Memorial Fund of the University of Chicago.

TABLE 1. *The distribution of numbers of Bar genes represented in each generation among 108 lines consisting of 4 males, 4 females each. Newly fixed genes are distinguished from those fixed in previous generations.*

No. Bar genes	Generation											4-10	%	
	0	1	2	3	4	5	6	7	8	9	10			
Old	0			15	30	43	55	65	78	88	91			
New	0		15	15	13	12	10	13	10	3	4	65	22.0	
	1	8	12	10	17	10	11	7	5	3	4	57	19.3	
	2	7	13	16	14	14	6	3	1	2	2	42	14.2	
	3	24	22	10	5	9	8	6	2	4	1	35	11.9	
	4	23	20	16	10	6	4	3	2	1	1	27	9.2	
	5	21	11	13	4	2	3	4	4	2	1	20	6.8	
	6	108	17	8	5	7	4	5	2	1	—	19	6.4	
	7		3	3	3	3	3	2	2	1	1	13	4.4	
	8		3	2	3	2	—	2	1	—	—	5	1.7	
	9		2	—	1	2	3			1	1	7	2.4	
	10			1			1			—	—	1	0.3	
	11			—					1	1	—	2	0.7	
New	12			1				1		—	—	1	0.7	
Old	12				1	1	1	1	2	2	2	2		
Total		108	108	108	108	108	108	108	108	108	108	108	295	100.0

TABLE 2. *The offspring of 4 males, 4 females picked at random, from vials of each type. The column 004 refers to vials in which there were 0 B/B, 0 B/+ and 4 +/+ females and similarly with the others. The row 04 refers to vials in which there were 0 B/0 males, 4 +/0 males, and similarly with the others.*

Fathers	Offspring	Mothers													
		004	013	022	031	040	103	112	121	130	211	220	301	310	400
04	B/B		0	0	0	0	0	0	0	0		0			
	B/+		31	26	13	28	0	1	7	4		1			
	+/+		181	78	23	20	4	11	5	4		3			
	B/0		17	28	11	19	0	4	4	3		2			
	+/0		195	76	25	29	4	8	8	5		2			
13	B/B		0	1	3	2	1	0	0	0	2	4	0	3	
	B/+		11	20	35	34	14	6	7	7	8	11	6	1	
	+/+		113	111	90	44	13	10	17	13	6	5	6	0	
	B/0		0	27	39	27	15	2	7	10	8	11	7	4	
	+/0		124	105	89	53	13	14	17	10	8	9	5	0	
22	B/B		0	6	5	7	6	0	3	5	1	0	5	0	2
	B/+		26	35	58	34	31	0	7	7	12	2	12	8	9
	+/+		54	111	69	59	31	8	6	8	3	2	3	0	1
	B/0		0	12	43	40	31	0	4	7	7	0	13	3	6
	+/0		80	140	89	60	37	8	12	13	9	4	7	5	6
31	B/B		0	0	7	7	6		3	7	1		6	4	1
	B/+		4	15	32	29	9		4	17	5		7	4	3
	+/+		8	17	25	20	5		1	4	2		3	0	0
	B/0		0	0	18	27	13		4	15	5		12	8	4
	+/0		12	32	46	29	7		4	13	3		4	0	0
40	B/B			1	0		4			1	4			4	
	B/+			15	4		4			3	8			0	
	+/+			0	0		0			0	0			0	
	B/0			3	3		4			2	6			4	
	+/0			13	1		4			2	6			0	

TABLE 3. Sons from vials with no B/B mothers. Viability (V_{B0}) of B/0, relative to +/0, estimated on the hypothesis of no differential productivity. Productivity (U_{B+}) of B/+, relative to +/+, estimated on the hypothesis of no differential viability. Expectation on the hypotheses $V_{B0} = 1$, $U_{B+} = 1$.

Mothers			Sons			Sons (exp.)		SE		SE	
B	B	+	B	+	Total	V_{B0}	($V_{B0}=1$)	U_{B+}	($U_{B+}=1$)	B	+
B	+	+	0	0						0	0
0	1	3	59	485	544	.85	.11	.83	.15	68.0	476.0
0	2	2	131	301	432	1.31	.13	1.54	.17	108.0	324.0
0	3	1	105	167	272	1.05	.15	1.13	.31	102.0	170.0
0	4	0	82	90	172	.91	.13	—	—	86.0	86.0
Total (excl. 0 4 0)			295	953	1248			1.094	.105	278.0	970.0
Total (incl. 0 4 0)			377	1043	1420		1.053	.064		364.0	1056.0

TABLE 4. Sons from vials containing B/B mothers. Productivity (U_{BB}) of latter estimated on hypothesis that $U_{B+} = 1$, $V_{B0} = 1$. Expectation on these hypotheses and the hypothesis that $U_{BB} = .37$.

Mothers			Sons			Sons (exp.)		SE		SE	
Classes	B	B	+	B	+	Total	U_{BB}	SE	($U_{BB}=.37$)	B	+
	B	+	+	0	0					0	0
4	1	—	—	88	144	232	.41	.15	.14	86.0	146.0
2	2	—	—	45	31	76	.41	.19	.18	43.9	32.1
2	3	—	—	25	11	36	.41	.18	.17	24.6	11.4
Total 8				158	186	344	.408	.098	.092	154.5	189.5

present in (4)). Tables 5 to 9 show the numbers of B/B, B/+ and +/+ daughters from various groups of parents (no B/B mothers or B/0 fathers in (5), no B/B mothers or +/0 fathers in (6), at least one B/B mother present, fathers all +/0 or all B/0 in (7), no B/B mothers, mixed fathers in (8) and at least one B/B mother, mixed fathers in (9)). Table 2 must be referred to for a detailed account of the types of mating in these cases.

These are analyzed in terms of the following parameters:

In cases in which all three types of daughters are expected the analysis has been made in terms of genes: $p_B = p_{BB} + (1/2)p_{B+}$, $p_+ = p_{++} + (1/2)p_{B+}$, $n_B = 2n_{BB} + n_{B+}$ and $n_+ = 2n_{++} + n_{B+}$.

The method used has been to find the value of the parameter under consideration which maximizes the probability of the observed numbers. It has been shown that this method does not always give the best estimates in such an absolute sense as supposed by Edgeworth (1908) and Fisher (1921) (method of

	B/B	B/+	+/+	B/0	+/0
number of parents in vial	f_{BB}	f_{B+}	f_{++}	f_{B0}	f_{+0}
relative productivity of parents	U_{BB}	U_{B+}	1	U_{B0}	1
relative viability of offspring	V_{BB}	V_{B+}	1	V_{B0}	1
probability of offspring genotype	p_{BB}	p_{B+}	p_{++}	p_{B0}	p_{+0}
observed number of offspring	n_{BB}	n_{B+}	n_{++}	n_{B0}	n_{+0}

n_T is the sum of the n's in any case.

"maximum likelihood") (cf. Neyman, 1951) but in such a case as the present there seems to be no better method, with the qualification that we have not attempted any simultaneous determinations.

As the results from any one population type are here reduced to two alternatives, the quantity to be maximized is of the form $p_1^{n_1} p_2^{n_2}$. The equation to be solved for a productivity coefficient U (and similarly for a viability coefficient V) is thus

$$n_1 \frac{d}{dU} \log p_1 + n_2 \frac{d}{dU} \log p_2 = 0$$

The best estimate from a series of population types is obtained by iteration on equating the sum of such expressions to 0.

The variance from a single population type is given by the formula

$$\frac{1}{\sigma^2} = - (n_1 + n_2) \left[p_1 \frac{d^2}{dU^2} \log p_1 + p_2 \frac{d^2}{dU^2} \log p_2 \right]$$

on substituting 1 for U (or V), if it is concluded that the difference from 1 is negligible, or by substituting the value of U (or V) that has been obtained, if this seems to differ significantly from 1. The variance for the estimate from a series of population types is obtained from the sum of such terms.

As the individual fly is taken as the

$$\left[\frac{f_{B+} U_{B+}}{2(f_{B+} U_{B+} + f_{++})} \right]^{n_{B+}} \left[\frac{f_{B+} U_{B+} + 2f_{++}}{2(f_{B+} U_{B+} + f_{++})} \right]^{n_{+0}}$$

is to be maximized.

$$\Sigma \left[\left(\frac{f_{++}}{f_{B+} U_{B+} + f_{++}} \right) \left(\frac{n_{B0}}{U_{B+}} - \frac{n_{+0} f_{B+}}{(f_{B+} U_{B+} + 2f_{++})} \right) \right] = 0$$

$$\frac{1}{\sigma^2} = \Sigma \left[\frac{n_T f_{B+} f_{++}^2}{(f_{B+} U_{B+} + f_{++})^2 (f_{B+} U_{B+} + 2f_{++}) U_{B+}} \right]$$

The second entry in table 3 seems to indicate significantly superior viability of B/0 males or superior productivity of

unit, no cognizance is taken of possible heterogeneity among vials. The standard errors would be larger if such heterogeneity exists. The factor cannot be large in view of the uniform small number (4) of flies of each sex counted in each vial.

From the fourth entry in table 3, it may be seen that vials in which all 4 females were B/+ produced 82 B/0; 90 +/0 sons indicating a relative viability of .91 for B/0 which does not differ significantly from 1. In F_1 , however, there were 192 B/0: 240 +/0 sons from B/+ females, which differs from equality by 2.3 times its standard error. It indicates a relative viability of .80 for B/0. The results from the other entries in this table can be interpreted either on the basis of differential viability of the B/+ and +/0 sons or of differential productivity of the B/+ and +/+ mothers or a combination.

In estimating as differential viability it is convenient to use $r = f_{B+}/(f_{B+} + 2f_{++})$.

$\left(\frac{rV_{B0}}{1 + rV_{B0}} \right)^{n_{B0}} \left(\frac{1}{1 + rV_{B0}} \right)^{n_{+0}}$ is to be maximized.

$$\Sigma \left(\frac{1}{1 + rV_{B0}} \right) \left(\frac{n_{B0}}{V_{B0}} - n_{+0} r \right) = 0$$

$$\frac{1}{\sigma^2} = \Sigma [n_T r / V_{B0} (1 + rV_{B0})^2]$$

In estimating as differential productivity

B/+ mothers but the total for all entries does not differ significantly from 1 on either basis. Other experimenters have

found inferior viability of Bar (e.g. Petit, 1951, $V_{B0} = .68 \pm .12$, Teissier $V_{B0} = .67$ as quoted by Petit). It is possible that there may be inferior viability of B/0 but superior productivity of B/+ in our data but the adoption of even a slightly lower viability B/0 leads to highly inconsistent and on the average very high estimates of productivity of B/+ though somewhat subnormal viability of the same (e.g. if $V_{B0} = .80$, U_{B+} comes out 1.52 and V_{B+} comes out .90). It is probable that there is some sort of selection advantage of the heterozygotes as indicated by the persistence of Bar for very long periods at very low frequencies in the experiments of L'Heritier and Teissier

(1937). Petit (1951) and Teissier obtained selective values of $1.062 \pm .12$ and 1.24 respectively. We will however adopt the simplest hypothesis that is compatible with the present data, viz. that there is no differential viability of B/0 and +/0 males and no differential productivity of B/+ and +/+ females. The coefficients may of course be expected to have different values under different conditions.

The data of table 4 are interpreted on the above basis and thus as involving only possible differential productivity of B/B females relative to B/+ and +/+ collectively. The expression to be maximized in each case is

$$\left[\frac{2f_{BB}U_{BB} + f_{B+}}{2(f_{BB}U_{BB} + f_{B+} + f_{++})} \right]^{n_{B0}} \left[\frac{f_{B+} + 2f_{++}}{2(f_{BB}U_{BB} + f_{B+} + f_{++})} \right]^{n_{+0}}$$

$$\Sigma \left(\frac{f_{BB}}{f_{BB}U_{BB} + f_{B+} + f_{++}} \right) \left(\frac{n_{B0}(f_{B+} + 2f_{++})}{(2f_{BB}U_{BB} + f_{B+})} - n_{+0} \right) = 0$$

$$\frac{1}{\sigma^2} = \Sigma_{NT} \left[\frac{(f_{B+} + 2f_{++})f_{BB}^2}{(f_{BB}U_{BB} + f_{B+} + f_{++})^2(2f_{BB}U_{BB} + f_{++})} \right]$$

There was significantly low productivity of B/B ($U_{BB} = .41 \pm .10$). The number of Bar females seems to make no difference in the estimate.

We turn now to the deductions from the daughters. In table 5 with no B/B mothers, the results can be interpreted either as due to differential viability of B/+ and +/+ daughters or (again) as differential productivity of B/+ and +/+ mothers or a combination. The formulae are the same as those for the sons with obvious substitutions.

The results are much like those for the sons in table 3. The present data do not warrant any but the simplest hypothesis, viz. that there is no significant difference between B/+ and +/+ either in viability or productivity.

Table 6 shows the results for 28 females from vials in which all 4 fathers were B/0 and no B/B mothers were present. The results suggest low viability of B/B females ($.61 \pm .32$) in harmony with

Petit and Teissier ($.75 \pm .15$, $.84$ respectively) but are obviously highly inadequate. It will be assumed for simplicity that there is no differential viability in view of the apparent absence of significant viability differences among the males in our own data.

Table 7 summarizes the few cases in which at least one B/B mother was present and the males were of a single genotype. These indicate very low productivity of B/B females. On combining with the earlier estimate, we arrive at $.369 \pm .087$ as the final estimate for U_{BB} .

Table 8 summarizes the data on daughters from vials in which B/B females were absent but both B/0 and +/0 males were present. It is assumed that there are no selective differentials except in productivity of B/0 and +/0 males. The frequency of the Bar gene among the daughters is theoretically the average of that in the gametes produced by the

TABLE 5. *Daughters from vials without B/B mothers and only +/0 fathers.* Interpreted alternatively on the hypotheses of no differential productivity of B/+ and +/+, or no differential viability of B/+ and +/+. Expectation on hypotheses: $V_{B+} = 1$, $U_{B+} = 1$.

Classes	Mothers			Fathers		Daughters				V_{B+}	SE ($V_{B+}=1$)	U_{B+}	SE ($U_{B+}=1$)	Daughters (exp.)	
	B	B	+	B	+	B	B	+	Total					B	+
	B	+	+	0	0	B	+	+						+	+
	0	1	3	0	4	0	31	181	212	1.20	.24	1.24	.24	26.5	185.5
	0	2	2	0	4	0	26	78	104	1.00	.23	1.00	.34	26.0	78.0
	0	3	1	0	4	0	13	23	36	.94	.34	.87	—	13.5	22.5
	0	4	0	0	4	0	28	20	48	1.40	.29			24.0	24.0
Total (excl. 0 4 0)						0	70	282	352			1.14	.19	66.0	286.0
Total (incl. 0 4 0)						0	98	302	400	1.131	.126			90.0	310.0

TABLE 6. *Daughters from vials without B/B mothers and only B/0 fathers.* Interpreted on hypothesis of no differential productivity of B/+ and +/+. Expectation on hypothesis: $V_{BB} = 1$, $U_{B+} = 1$.

Classes	Mothers			Fathers		Daughters				V_{BB}	SE	SE $V_{BB}-1$	Daughters (exp.)			
	B	B	+	B	+	B	B	+	Total				$V_{BB}-1$	7.0	21.0	0.0
	B	+	+	0	0	B	+	+								
3	0	-	-	4	0	5	23	0	28	.61 ± .32	.47					

parents.

$$p_B = \frac{1}{2} \left[\left(\frac{f_{B0}U_{B0}}{f_{B0}U_{B0} + f_{+0}} \right) + \left(\frac{f_{B+}}{2(f_{B+} + f_{++})} \right) \right]$$

The expression to be maximized is $p_B^{nB}(1 - p_B)^{n+}$.

$$\text{Let } A = f_{B0}f_{+0}/[2(f_{B0}U_{B0} + f_{+0})^2].$$

$$\Sigma A[2n_{BB} + n_{B+} - 2n_{TPB}]/p_B(1 - p_B) = 0$$

$$\frac{1}{\sigma^2} = \Sigma [n_{TA}^2/p_B(1 - p_B)]$$

The data are classified both according to the relative numbers of B/+ and +/+ mothers and to the relative number of B/0 and +/0 fathers. In neither case is there a consistent trend. The grand average $U_{B0} = .264 \pm .040$ indicates very low productivity of B/0 males under these conditions. F_1 from 4 B/+ females; 2 B/0, 2 +/0 males in each vial, gave 28 B/B; 210 B/+; 194 +/+ daughters, indicating even lower relative productivity of B/0 males (.125) than in the later generations.

The remaining data (table 9) are those

TABLE 7. *Daughters from vials including B/B mothers and either +/0 or B/0 fathers exclusively.* Interpretation on hypothesis of no differential viability. Expectation on this hypothesis and hypothesis that $U_{BB} = .37$.

Classes	Mothers			Fathers		Daughters				U_{BB}	SE	SE ($U_{BB}=.37$)	Daughters (exp.)		
	B	B	+	B	+	B	B	+	Total				B	B	+
	B	+	+	0	0	B	+	+					+	+	+
5	1, 2	-	-	0	4	0	13	27	40				0	15.4	24.6
3	1, 3	-	-	4	0	9	11	0	20				11.3	8.7	0
8	1, 3								60	.126	.168	.276			
8	1, 3					(sons, table 4)			344	.408	.098	.092			
16	1-3								404	.369	.087	.087			

TABLE 8. *Daughters from vials without B/B mothers, classified according to numbers of B/+ and +/+ mothers or according to number of B/0 and +/0 fathers. Interpretation on the hypothesis of no differential viability. Expectation on this basis and hypothesis that $U_{B0} = .26$.*

Classes	Mothers			Fathers		Daughters				U_{B0}	SE	Daughters (exp.)		
	B	B	+	B	+	B	B	+	Total			B	B	+
	B	+	+	0	0	B	+	+				B	+	+
3	0	0	4	+	+	0	41	175	216	.36	.08	0	31.7	184.3
3	0	1	3	+	+	7	70	239	316	.18	.08	7.0	81.4	227.6
3	0	2	2	+	+	15	125	184	324	.31	.09	16.4	113.7	193.9
3	0	3	1	+	+	16	97	123	236	.15	.10	19.3	101.4	115.3
3	0	4	0	+	+	13	54	49	116	.23	.14	12.5	58.0	45.5
5	0	-	-	1	3	7	114	371	492	.25	.09	7.4	117.0	367.6
5	0	-	-	2	2	24	184	324	532	.28	.05	25.5	182.3	324.2
5	0	-	-	3	1	20	89	75	184	.24	.08	22.3	86.9	74.7
15	0	-	-	+	+	51	387	770	1208	.264	.040	55.2	386.2	766.6

in which B/B females and both type of males are present. The value $U_{BB} = .37$ is adopted from table 7 and the data used for an estimate of U_{B0} in the presence of at least one B/B female. The formulae are the same as above except that

$$P_B = \frac{1}{2} \left[\left(\frac{f_{B0}U_{B0}}{f_{B0}U_{B0} + f_{+0}} \right) + \left(\frac{2f_{BB}U_{BB} + f_{B+}}{2(f_{BB}U_{BB} + f_{B+} + f_{++})} \right) \right]$$

The results indicate significantly improved productivity of B/0 males when B/B females are present ($U_{B0} = .59 \pm .16$). Subdivision, according to the rela-

tive numbers of B/0 and +/0 males, suggest decline of productivity per B/0 male as the numbers of such males increase and subdivision according to the number of B/B females suggests increase of productivity per B/0 male as the number of B/B females increases, both in the direction of assortative mating, but not significantly.

Summing up, this analysis indicates that the present data can be interpreted adequately without assuming any differential viability, or any difference in productivity of B/+ and +/+ females but productivity of B/B females only 37% of that of B/+ or +/+, and productivity of B/0 males only 26% of that of

TABLE 9. *Daughters from vials including B/B mothers classified according to numbers of B/0 and +/0 fathers or numbers of B/B mothers. Interpretation on hypothesis of no differential viability. Expectation on this basis and hypothesis that U_{B0} in presence of B/B females is .59.*

Classes	Mothers			Fathers		Daughters				U_{B0}	SE	Daughters (exp.)		
	B	B	+	B	+	B	B	+	Total			B	B	+
	B	+	+	0	0	B	+	+				B	+	+
7	1-3	-	-	1	3	9	46	57	112	.66	.35	7.2	47.8	57.0
8	1-3	-	-	2	2	16	57	31	104	.58	.24	18.5	51.5	34.0
6	1-4	-	-	3	1	22	40	10	72	.55	.29	24.4	35.4	12.2
11	1	-	-	1-3	3-1	22	80	78	180	.55	.21	23.7	80.3	76.0
5	2	-	-	1-3	3-1	15	38	19	72	.67	.35	14.6	36.5	20.9
5	3, 4	-	-	1-3	3-1	10	25	1	36	.69	.41	11.8	17.8	6.4
Total	21					47	143	98	288	.589	.163	50.1	134.6	103.3

+0 males in the absence of B/B females, rising to 59% in the presence of at least one of the latter.

A comprehensive test can be obtained by calculating the expected numbers of each kind of offspring from each population type in table 2 on this basis, omitting F₁. These calculations are summarized in the last 2 or 3 columns of tables 3 to 9. χ^2 was calculated in each case. A condensed summary is given below.

	Classes	DF	χ^2	Probability
Tables 3, 5, 6 (all V's = 1, U _{B+} = 1)	9	9	11.41	.20-.30
Tables 4, 7 (all V's = 1, U _{BB} = .37)	5	4	1.85	.70-.80
Table 8 (all V's = 1, U _{B0} = .26)	3	5	.53	ca .99
Table 9 (all V's = 1, U _{BB} = .37, U _{B0} = .59)	3	5	2.95	.70-.80
	20	23	16.74	.80-.90

It is obvious that no more complicated theory is necessary to account for the present data.

*Rate of change of gene frequency
per generation*

The theoretical relation between gene frequencies in successive generations can

be expressed in simple formulae with sufficient accuracy for most purposes in cases in which the selective differentials are all small. The present data are not among these cases. The distribution of gene frequencies in the offspring can, however, be tabulated for the parental groups with each possible gene frequency from 1/12 to 11/12. The mean change in gene frequency (Δq) and the variance of changes ($\sigma_{\delta q}^2$) can then be expressed

as empirical functions of parental gene frequency.

Table 10 and fig. 1 give these and related statistics (excluding F₁). The observed values of Δq may be compared with those expected (c_1) on the hypothesis of no differential viability, U_{BB} = .37, U_{B0} = .26 in the absence of B/B females

TABLE 10. *Frequency of parental populations (generations 1 to 9) with each number of Bar genes, change of gene frequency (Δq) to next generation ratio, $\Delta q/q(1-q)$, variance ($\sigma_{\delta q}^2$) of frequencies in offspring generation and ratio $q(1-q)/\sigma_{\delta q}^2$. The relative constancy of the ratio $\Delta q/q(1-q)$ indicates that selection pressure is empirically of the form $\Delta q = -.35q(1-q)$. The average of the ratio $q(1-q)/\sigma_{\delta q}^2$ indicates a value of $2N_e$ of 9.74. The calculated values c_1 are those expected from the vials, classified according to the number of Bar genes among the parents on the hypotheses of no differential viability, U_{BB} = .37, U_{B0} = .26 in absence of B/B females but .59 in their presence. The calculated c_2 are from the empirical formula $\Delta q = -.35q(1-q)$.*

No. Bar genes	No. parents (gen.) (1-9)	q	Δq			$\Delta q/q(1-q)$	$\sigma_{\delta q}^2$	$q(1-q)/\sigma_{\delta q}^2$
			0	c_1	c_2			
1	84	.0833	-.0248	-.0209	-.0267	-.325	.0063	12.1
2	77	.1667	-.0314	-.0491	-.0486	-.226	.0141	9.9
3	88	.2500	-.0814	-.0738	-.0656	-.434	.0200	9.4
4	84	.3333	-.0724	-.0749	-.0778	-.326	.0226	9.9
5	63	.4167	-.0939	-.0879	-.0851	-.386	.0259	9.4
6	49	.5000	-.0884	-.1014	-.0875	-.354	.0348	7.2
7	21	.5833	-.0793	-.1060	-.0851	-.326	.0399	6.1
8	12	.6667	-.0903	-.1009	-.0778	-.406	.0119	18.6
9	10	.7500	-.1500	-.0947	-.0656	-.800	.0321	5.8
10	2	.8333	0	-.1116	-.0486	-.0	.0556	2.5
11	2	.9167	-.0417	-.0559	-.0267	-.546	.0313	2.4
	492					-.3511		9.74

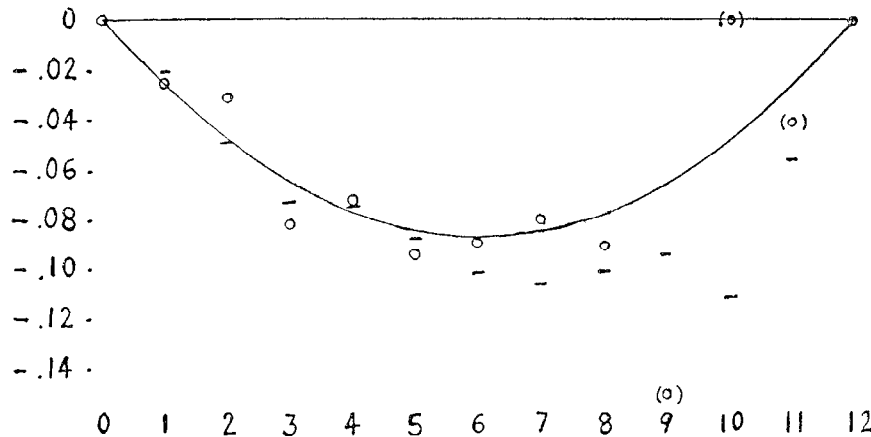


FIG. 1. The observed average rate of change in gene frequency per generation is shown in relation to the number of Bar genes per population by small circles. The rates of change, calculated on the hypothesis that $U_{BB} = .37$, $U_{BV} = .26$ in the absence of B/B females, .59 in their presence, are indicated by short lines. The rates of change calculated from the empirical formula $\Delta q = -.35q(1-q)$ are indicated by the smooth curve. Parentheses indicate observations based on highly inadequate numbers of cases.

but .59 in the presence of at least one B/B female. Calculations were made for every type of population in table 2. It may be seen that these theoretical values agree fairly well with those observed. A comparison of observed and calculated numbers of each genotype at each parental gene frequency (grouping 10/12 and 11/12) yields $\chi^2 = 17.5$ for females (about 16 degrees of freedom, probability .20-.30), $\chi^2 = 14.2$ for males (about 9 degrees of freedom, probability .10-.20), with a total χ^2 of 31.8, 25 degrees of freedom, probability .10-.20.

The observed values of Δq divided by $q(1-q)$ yield a series of values with weighted average $\bar{s} = -.351$ and only a slight trend ($s = -.344 - .024q$). The simple empirical formula $\Delta q = -.35q(1-q)$ is adequate for most purposes.

The theoretical formula for selection in the simple case of an autosomal semi-dominant gene in a large population is $\Delta q = sq(1-q)$. The approach to this form in the present case in spite of the recessiveness of selective disadvantage in females ($U_{++} = U_{B+} = 1$, $U_{BB} = .37$) is due in part to the higher gene frequency in males than in females for most values of q which throws the point of maximum selection in males at values of

q less than .50, and to some extent to the falling off of selection in males at high values of q due to the presence of B/B females.

The variance of gene frequencies takes the form $q(1-q)/2N_e$ where N_e is the effective size of population. Table 10 shows the ratios $q(1-q)/\sigma_{\delta q}^2$ as estimates of $2N_e$. There is considerable irregularity but no well defined trend. The weighted average is 9.74 to be compared with the theoretical value 13.53 in populations of 4 females, 4 males (with sex linkage, no selection and fluctuation due merely to accidents of sampling). The value indicated by the rate of fixation of forked and its type allele, with very little selection, was 11.22.

The distribution of gene frequencies

It is of interest to compare the observed distribution of gene frequencies after attainment of equilibrium of form with that expected under the observed selection pressure and observed effective size of population. We may safely use the model of a monoecious diploid population.

The class with gene frequency q contributes to the classes of the next generation according to the expansion of

$$[(1 - q - \Delta q)(B^+) + (q + \Delta q)(B)]^{2N_e} f(q)$$

The solution for selection pressures of the fairly general form $\Delta q = (s + tq) \times q(1 - q)$, which allows for any degree of dominance, has been given in powers of s and t jointly, for very small k (Wright, 1942). It would be very tedious to solve it in this way for nonnegligible k since the value of k in relation to s and t must be established by repeated trial. Fortunately for our present purpose, the case, $\Delta q = spq$, is much simpler. The method used in a previous paper (Wright, 1931) leads at once to a solution in the form

$$\varphi(q) = e^{2Nsq}(a_0 + a_1q + a_2q^2 + a_3q^3 \dots)$$

in which s appears in the a 's only in even powers, after dropping terms in which the exponent of N is less than that of s . Thus the value of the power series remains the same on reversing the sign of the selection. As the distribution must remain the same in the case $\Delta q = sq(1 - q)$ on reversing the sign of s and replacing q by $(1 - q)$, it is evident that a series of the form

$$\varphi(q) = Ce^{2Nsq}(1 + C_1pq + C_2(pq)^2 + C_3(pq)^3 \dots)$$

can be used to advantage. In this case the coefficients came out as follows to a sufficient approximation.

$$\begin{aligned} C_1 &= (1 - 2Nk) \\ C_2 &= \frac{1}{3}[6 - 2Nk]C_1 + 2N^2s^2 \\ C_3 &= \frac{1}{6}[(15 - 2Nk)C_2 + 2N^2s^2C_1] \\ C_n &= \frac{2}{n(n + 1)} \{ [n(2n - 1) - 2Nk] \\ &\quad \times C_{n-1} + 2N^2s^2C_{n-2} \} \end{aligned}$$

The same result can be obtained from the Focker-Planck equation

$$\begin{aligned} \frac{1}{2} \frac{\partial^2}{\partial q^2} [\sigma_{sq}^2 \varphi(q, t)] \\ - \frac{\partial}{\partial q} [\Delta q \varphi(q, t)] = \frac{\partial \varphi(q, t)}{\partial t} \end{aligned}$$

in the form appropriate to constancy of form (excluding the fixed classes) with

decay at the rate k per generation (Wright 1945).

$$\begin{aligned} \frac{1}{2} \frac{d^2}{dq^2} [\sigma_{sq}^2 \varphi(q)] \\ - \frac{d}{dq} [\Delta q \varphi(q)] + k\varphi(q) = 0 \end{aligned}$$

For $\sigma_{sq}^2 = q(1 - q)/2N$, $\Delta q = sq(1 - q)$ this reduces to

$$\begin{aligned} \frac{d^2}{dq^2} [q(1 - q)\varphi(q)] \\ - 4N \frac{d}{dq} [q(1 - q)\varphi(q)] + 4Nk\varphi(q) \end{aligned}$$

On substituting $\varphi(q) = e^{2Nsq}\chi(q)$ we get after some reduction

$$\frac{d^2\chi(q)}{dq^2} - \left[4N^2s^2 - \frac{4Nk}{q(1 - q)} \right] \chi(q) = 0$$

If

$$\begin{aligned} \chi(q) = [1 + C_1q(1 - q) \\ + C_2q^2(1 - q)^2 \dots] \end{aligned}$$

as assumed above, we are led to the same relations among the coefficients as by the other method in this case without dropping small terms, as approximation is already introduced in the use of the Focker-Planck equation.

After $n(2n - 1)$ has become greater than $2Nk$, the first term in C_n has the same sign as C_{n-1} . The second always has the sign of C_{n-2} . Thus if two successive coefficients (after this point) ever have the same sign all later ones must have the same sign and will soon begin to increase. The series of coefficients becomes either divergent positive or negative unless there is a relation of $2Nk$ to $2N^2s^2$ at which the coefficients continue to alternate in sign and approach zero. The rapidity of divergence, one way or the other, is so great that it is easy to locate to any required number of decimal places the desired value of $2Nk$. This may be taken to be the limiting rate of fixation under the assumed selection. The final steps in the present case were as follows:

Trial values of $2Nk$

	2.13536800	2.13536802	2.13536810
C_1	-1.13536800	-1.13536802	-1.13536810
C_2	+ .46407350	+ .46407348	+ .46407341
C_3	- .09871538	- .09871544	- .09871568
C_4	+ .01291079	+ .01291063	+ .01290997
C_5	- .0011439	- .0011444	- .0011464
C_6	+ .0000747	+ .0000732	+ .0000674
C_7	+ .0000011	- .0000039	- .0000239
C_8	+ .0000154	- .0000012	- .0000674

A number of trials led to the value 2.135368 for $2Nk$. Using this, the coefficients alternate in sign to C_6 but thereafter all are positive and are divergent (after C_7). This shows that this estimate of $2Nk$ is too small. On trying 2.1353681 the signs alternate down to C_7 but are all negative thereafter, showing that this estimate is too large. Interpolation suggests 2.13536802. This also turns out to be too large but is correct for 9 significant figures.

It has been found empirically that for values of $2Ns$ up to 1, $2Nk$ is given to 11 significant figures by

$$2Nk = 1 + \frac{(2Ns)^2}{10} - \frac{(2Ns)^4}{7000} - \frac{(2Ns)^6}{1,050,000} - .000,000,004(2Ns)^8$$

This is accurate to 6 significant figures up to $2Ns = 4$, to 3 significant figures for $2Ns = 6$.

It is desirable to determine whether the value of $2Nk$, determined above, is consistent with the rates of fixation of the two alleles for given $2N$, deducible from the terminal ordinates of the distribution.

In the absence of selection, the formula for the distribution curve reduces at once to $\varphi(q) = 1$ or $f(q) = 1/(2N)$. The chance of fixation in a population with r representatives of the specified allele is e^{-r} (from the Poisson Law). The total amount of fixation of this allele should thus be

$$\frac{1}{2N} (e^{-1} + e^{-2} + e^{-3} \dots) = .582/(2N)$$

if all classes have the frequency $1/(2N)$ (Wright, 1931). As the actual rate is exactly $1/(4N)$ (in the ideal case of completely random union among $2N$ gametes) the above approximate formula for the curve must be considerably modified near the terminal ordinates. The correction to be applied to the approximate formula $\varphi(q) = 1$ was determined by Fisher (1930) on inspection of the above paper in manuscript: .8182 for the class with one of the genes, .9168 for that with two, .9449 for that with three, .9583 for that with four and thereafter $1 - [1/(6r)]$.

If selection is present and $2N$ is indefinitely large, the frequencies from the approximate formula

$$f(q) = \frac{C e^{2Nsq} (1 + C_1 pq + C_2 (pq)^2 \dots)}{(2N)}$$

take the value $C/(2N)$ in the neighborhood of $q = 0$ and the value $C e^{2Ns}/(2N)$ in the neighborhood of $q = 1$. Because of the approximate uniformity in each of these neighborhoods, corrections closely similar to those determined by Fisher for the case of no selection should be applicable.

Fixation occurs practically from only a very few of the classes in these terminal regions. Again because of the approximate uniformity of the frequencies in each region,

$$\begin{aligned} k_0 &= f(0)/(2N) = C/(4N) \\ k_1 &= f(1)/(2N) = C e^{2Ns}/(4N) \\ k &= C(1 + e^{2Ns})/(4N) \end{aligned}$$

Thus

$$\begin{aligned} C &= 4Nk/(1 + e^{2Ns}) \\ f(q) &= \frac{2k e^{2Nsq} (1 + C_1 pq + C_2 (pq)^2 \dots)}{(1 + e^{2Ns})} \end{aligned}$$

Let $y = e^{2Nsq} (1 + C_1 pq + C_2 (pq)^2 \dots)$ represent the relative value of an ordinate calculated from $2Ns$ and the values of the C 's given above. The ordinates of the curve of unit area can be obtained from the relation $\varphi(q) = y/\int_0^1 y dq$. If values of y are calculated at uniform intervals from $q = 0$ to $q = 1$, the area

of the curve can be estimated from a quadrature formula. The following were obtained from ordinates at intervals 0.1.

Quadrature rule

	Trapezoidal	Parabolic (Simpson's)	Gregory's	Correct values
2Nk ₀	2.0353	2.0659	2.06639	2.066,406
2Nk ₁	.0679	.0689	.06896	.068,962
2Nk	2.1032	2.1347	2.13535	2.135,368

It may be seen that the trapezoidal quadrature gives estimates of 2Nk that agree with the actual values to two significant figures, Simpson's rule to four and Gregory's to at least five. The estimate of 2Nk derived from the criterion of convergence is thus consistent with the values of the terminal ordinates.

As the values $k_0 = k_1 = 1/(4N)$ apply exactly in the absence of selection for all values of 2N, however small (in the ideal case of completely random union of gametes), it may be surmised that the formulae given above as applying in the presence of selection to very large populations will give fairly good approxima-

tions for small 2N. In the case $2N_e = 10$, we obtain $k_0 = .2066$, $k_1 = .0069$ in comparison with .2019 and .0075 respectively, the figures from direct determination of the stable distribution for $2N = 10$, $2N_s = -3.4$. The estimates for the frequencies of the unfixed classes (by adjusting the relative ordinates at intervals of 0.1 to give the total, $1 - k_0 - k_1$) are more seriously in error (table 11) and need correction of the type referred to.

The agreement in form for $2N = 10$ (direct) and $2N = \infty$ (no correction), both for $2N_s = -3.4$, is, however, sufficiently close that it may be surmised that the form for $2N = 12$, $2N_s = -3.4$, determined directly, should be very close to that for $2N = 10$ and the same $2N_s$ and thus will permit good estimates for the unfixed classes (1/12 Bar to 11/12 Bar) even though the effective value of 2N is about 10 instead of 12. The actual values are given in the second column of table 12 in comparison with uncorrected estimates from the appropriate ordinates for $2N = \infty$. We note again that the rates of fixation agree

TABLE 12. *The theoretical distribution of gene frequencies, including newly fixed populations, after stability of form has been attained.* Column 2: exact determinations for $2N = 12$, $2N_s = -3.4$. Column 3: estimates of same from limiting curve. Column 4: newly fixed populations from $2N = 10$, $2N_s = -3.4$, unfixed populations rated down from Column 2. Column 5: estimates of same from limiting curve. Column 6: observed populations. Column 7: expectations calculated from Column 4. Column 8: difference ($\chi^2 = 8.9$). Column 9: expectations calculated for Column 5. Column 10: difference ($\chi^2 = 11.7$).

	2N _s = -3.4									
	2N _e = 12		2N _e = 10		Obs.	Calc. 2N _e = 10 (Direct)	Diff.	Calc. 2N _e = 10 from 2N = ∞	Diff.	
	Direct	Estimate from 2N = ∞	Direct	Estimate from 2N = ∞						
0	.1689	.1722	(.2019)	(.2066)	65	59.6	+5.4	61.0	+4.0	
1	.2059	.2346	.1973	.2244	57	58.2	-1.2	66.2	-9.2	
2	.1647	.1642	.1579	.1571	42	46.6	-4.6	46.3	-4.3	
3	.1214	.1167	.1163	.1116	35	34.3	+0.7	32.9	+2.1	
4	.0894	.0842	.0857	.0806	27	25.3	+1.7	23.8	+3.2	
5	.0668	.0619	.0640	.0592	20	18.9	+1.1	17.5	+2.5	
6	.0507	.0462	.0486	.0442	19	14.3	+4.7	13.0	+6.0	
7	.0391	.0351	.0375	.0336	13	11.1	+1.9	9.9	+3.1	
8	.0306	.0271	.0293	.0259	5	8.6	-3.6	7.7	-2.7	
9	.0241	.0213	.0231	.0204	7	6.8	+0.2	6.0	+1.0	
10	.0188	.0170	.0180	.0163	1	5.3	-4.3	4.8	-3.8	
11	.0134	.0138	.0129	.0132	2	3.8	-1.8	3.9	-1.9	
12	.0062	.0057	(.0075)	(.0069)	2	2.2	-0.2	2.0	0.0	
	1.0000	1.0000	1.0000	1.0000	295	295.0	0.0	295.0	0.0	

much better than the frequencies of the unfixed classes.

The forms of these distributions can be compared by plotting the frequencies for $2N = 12$ on a percentage scale in which the units are 20% larger than those used in plotting the distribution for

$2N = 10$ (abscissas 0-12 in the former, 0-10 in the latter) and by plotting the curve for $2N = \infty$ in such a way that it has a unit area over the base $2N$ on the appropriate scale of percentages whether $2N = 10$ or 12. It may be seen that the forms for $2N = 10$ and $2N = 12$

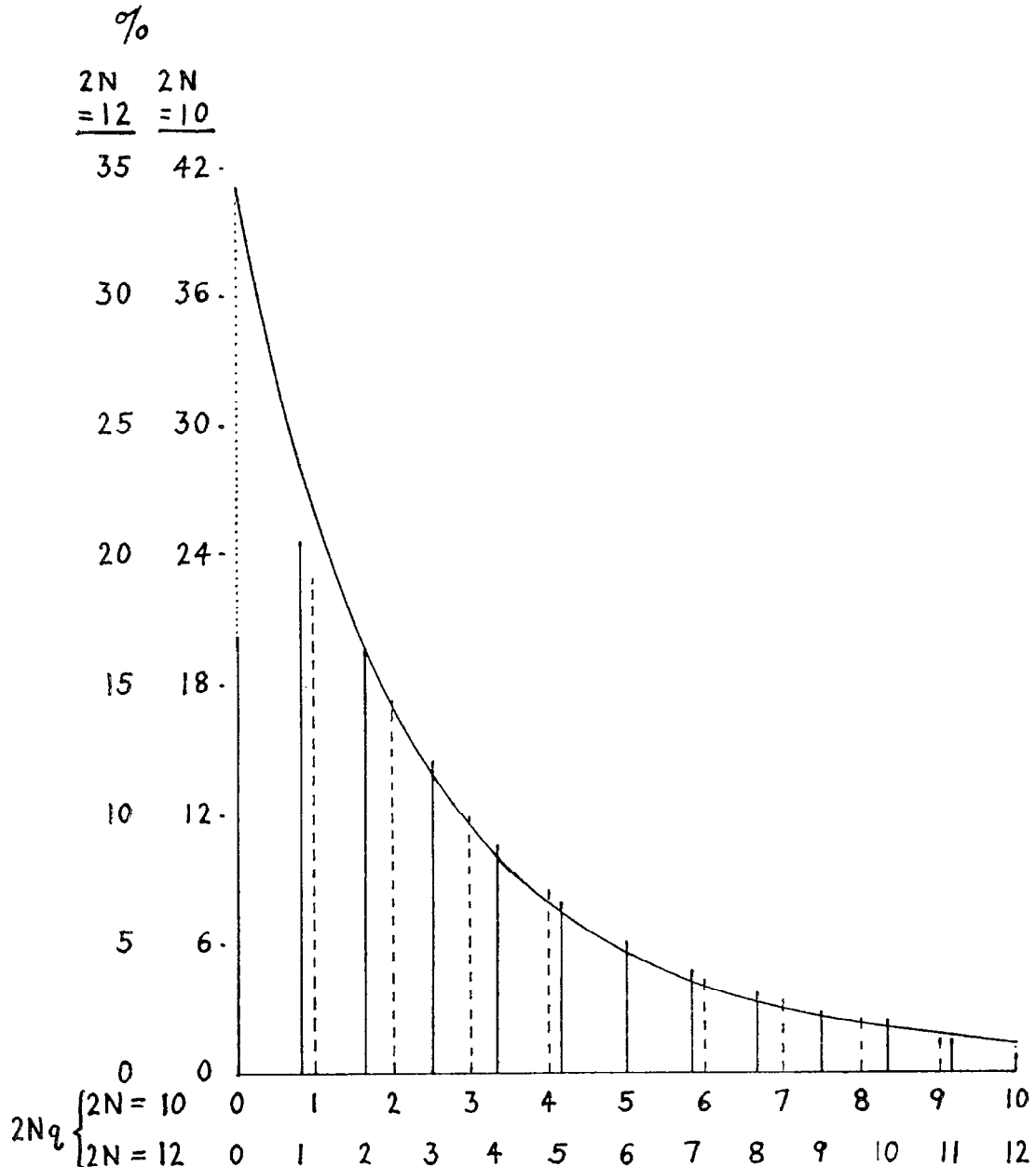


FIG. 2. The theoretical distributions of gene frequencies, including the classes of newly fixed genes, after attainment of stability of form, among populations in which $2Ns = -3.4$, $2N = 12$ (solid lines), $2N = 10$ (broken lines, except at $q = 0, .5$ and 1.0 at which there is agreement with the values for $2N = 12$ on the scales used). The scale unit for percentage frequencies is 20% greater for $2N = 12$ than for $2N = 10$ to bring out the similarity of form. The distribution for $2N = \infty$ is shown as a smooth curve of unit area (base 10 or 12, each with the appropriate scale of ordinates).

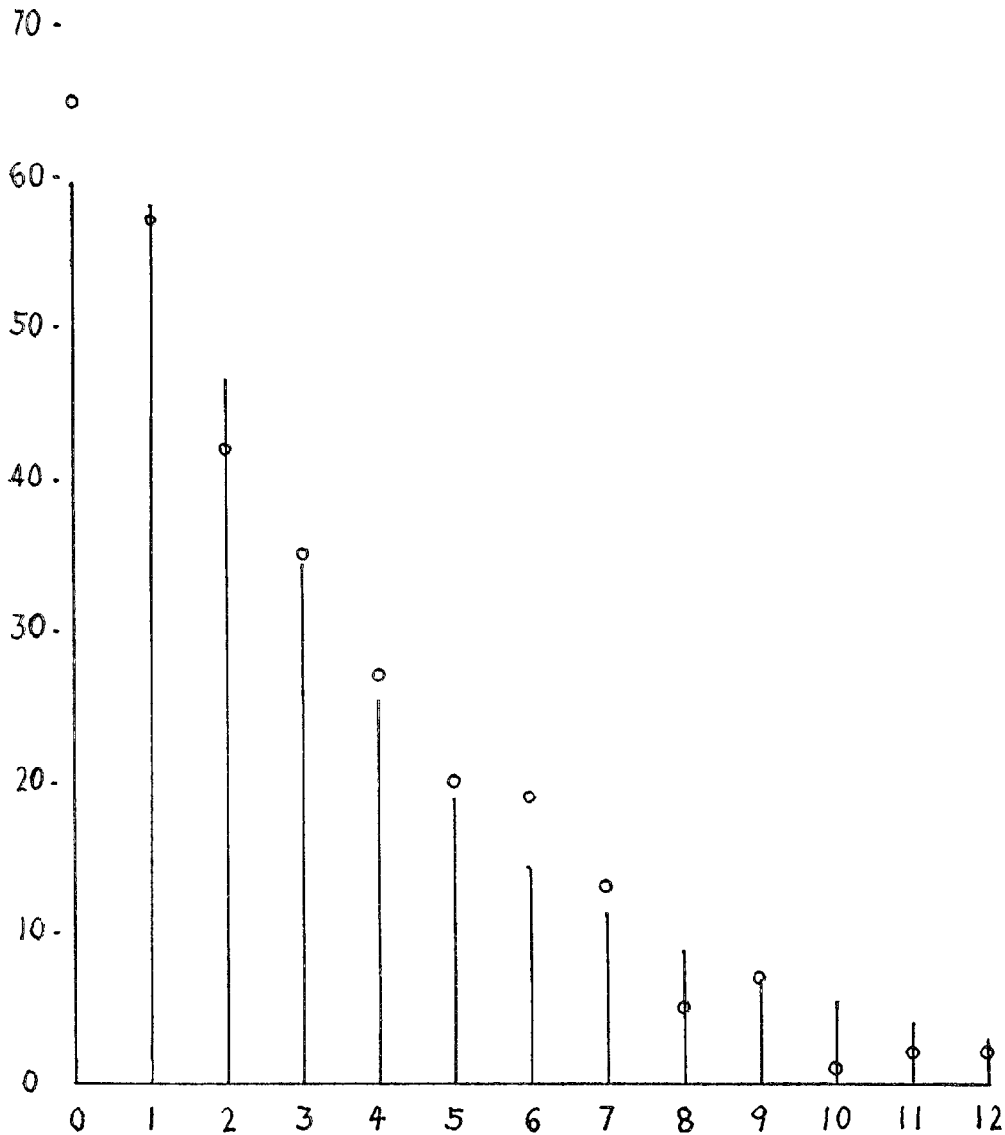


FIG. 3. The observed number of populations containing one to eleven Bar genes or newly fixed with respect to type (0) or Bar (12) in generations 4 to 10 (small circles). The expected numbers are indicated by vertical lines. Those for the newly fixed classes are based on the theoretical percentages of fixation for $2N = 10$, $2N_s = -3.4$. Those for the unfixed classes apportion the remainder according to the theoretical frequencies for $2N = 12$, $2N_s = -3.4$.

agree very closely, that for $2N = \infty$ less closely.

To test the goodness of fit of the experimental frequencies, the expected numbers of newly fixed classes are deduced from the distribution for $2N = 10$, while those for the unfixed are deduced from the relative frequencies of those for $2N = 12$. The expected proportions are shown in the fourth column of table 12. Similar estimates from the curve $2N = \infty$ (uncorrected) are given in the next col-

umn. The observed distribution of newly fixed and unfixed populations in generations 4 to 10, combined, are shown in column 6 and the expected values, rated up to the same total, 295, in column 7 and 9. The differences are given in columns 8 and 10. Calculation of χ^2 for the observed and direct estimates gives 8.9 with 10 degrees of freedom, probability .50-.70. Even the uncorrected estimates from the ordinates of the curve for $2N = \infty$ gives a χ^2 of only

11.7, probability about .30. The estimated rates of fixation from the latter, taking $2N_e = 9.74$ instead of 10 come out 21.2% for k_0 and 0.7% for k_1 in comparison with observed 22.0% and 0.7% (cf. fig. 3).

The calculation of the exact distributions for values of $2N$ and $2N_e$ much larger than 12 and 10 respectively would be a very tedious matter. Estimates from the curve for $2N = \infty$, observed $2N_s$, become increasingly satisfactory as the observed $2N$ and $2N_e$ become larger. From the goodness of fit in the present case, it appears that the method should give comparisons of observation and theory for larger populations, that are adequate for most purposes, especially if the corrections, derived from the case of no selection, are applied to the subterminal classes.

SUMMARY

Experiments with 108 very small populations of *Drosophila melanogaster* involving the mutation Bar and its allele (four females and four males as parents of each generation), carried to fixation of one allele or to the tenth generation, indicated little selective mortality, but severe selection against Bar from low productivity of homozygous Bar females (37% of that of B/+ or +/+) and Bar males (26% of the +/0 in the absence of B/B females, 59% in their presence).

Starting from 50% Bar genes in each case, the distribution of frequencies soon reached approximate stability of form (about four generations) as type came to be fixed at a rate of 22.0% per generation and Bar at a rate of 0.7% per generation. After generation 10, type had

been fixed in 95 lines, Bar in three and 10 were still unfixed.

The form of the distribution agreed well with that expected in a population of effective size 72% of the actual size and an empirically determined rate of change of the frequency (q) of the Bar gene, $\Delta q = -.35q(1 - q)$. The limiting form of the distribution curve for indefinitely large populations, but the observed value of $2N_s$, is derived, and shown to give a basis for fairly good estimates even with such small populations as in these experiments.

LITERATURE CITED

- EDGEWORTH, F. Y. 1908. On the probable error of frequency constants. Jour. Roy. Stat. Soc. London, **71**: 651-678.
- FISHER, R. A. 1922. On the mathematical foundations of theoretical statistics. Phil. Trans. Roy. Soc. London, Series A, **222**: 309-368.
- . 1930. The Genetical Theory of Natural Selection. Oxford, Clarendon Press, 272 p.
- KERR, W. E., AND S. WRIGHT. 1954. Experimental studies of the distribution of gene frequencies in very small populations of *Drosophila melanogaster*. I. Forked.
- L'HÉRITIER, PH., AND G. TEISSIER. 1937. Elimination des formes mutantes dans des populations de Drosophiles "Bar." C. R. Soc. Biol., **124**: 880-882.
- NEYMAN, J. 1951. Fisher's collected papers. Scientific Monthly, **62**: 406-408.
- PETIT, CLAUDINE. 1951. Le rôle de l'isolement sexuel dans l'évolution des populations de *Drosophila melanogaster*. Bull. Biol., **85**: 397-418.
- WRIGHT, S. 1931. Evolution in Mendelian populations. Genetics, **16**: 97-159.
- . 1942. Statistical genetics and evolution. Bull. Amer. Math. Soc., **48**: 223-246.
- . 1945. The differential equation of the distribution of gene frequencies. Proc. Nat. Acad. Sci., **31**: 382-389.