A METHOD FOR CALCULATING LINKAGE VALUES[1]

HUGO W. ALBERTS

Department of Agronomy, University of Illinois, Urbana, Illinois

Received August 7, 1925

TABLE OF CONTENTS

INTRODUCTION

The calculation of linkage values in many plants is confined to F_2 distributions, due to difficulties in making backcrosses. EMERSON (1916), HALDANE (1919), and other authors have proposed methods by which the gametic ratio can be obtained from the observed phenotypic zygotic proportions and from this ratio the percentage of crossing over can be calculated. WOODWORTH (1916), COLLINS (1924), BRUNSON (1924) and

[1] Contribution from the Division of Plant Breeding, Department of Agronomy, UNIVERSITY OF ILLINOIS. Published with the approval of the Director of the Station.

others have modified or extended EMERSON'S method for use in special cases, involving duplicate, complementary or supplementary factors.

A study of the theoretical gametic ratios with their corresponding phenotypic zygotic ratios in a dihybrid F_2 distribution (table 1) has revealed some interesting relationships. Some of these relationships have already been worked out by the authors cited above, but most of those contained in this paper have not been published so far as I know. It seems desirable, therefore, to present these relationships together with additional formulae for measuring linkage intensities.

TABLE 1

Giving a number of different gametic series together with their corresponding theoretical F_2 *phenotypic series.*

GAMETIC SERIES				F_2 PHENOTYPIC SERIES				TOTAL (n)	CONSTANT (k)
r	s	s	r	AB	Ab	aB	ab		
10	1	1	10	342	21	21	100	484	14.611
9	1	1	9	281	19	19	81	400	14.48
8	1	1	8	226	17	17	64	324	14.320
7	1	1	7	177	15	15	49	256	14.125
6	1	1	6	134	13	13	36	196	13.877
5	1	1	5	97	11	11	25	144	13.555
4	1	1	4	66	9	9	16	100	13.12
3	1	1	3	41	7	7	9	64	12.5
2	1	1	2	22	5	5	4	36	11.555
1	1	1	1	9	3	3	1	16	10
1	2	2	1	19	8	8	1	36	8.888
1	3	3	1	33	15	15	1	64	8.5
1	4	4	1	51	24	24	1	100	8.32
1	5	5	1	73	35	35	1	144	8.222
1	6	6	1	99	48	48	1	196	8.163
1	7	7	1	129	63	63	1	256	8.125
1	8	8	1	163	80	80	1	324	8.098
1	9	9	1	201	99	99	1	400	8.08
1	10	10	1	243	120	120	1	484	8.066

SYMBOLS USED

The following symbols will be used:

r, the first and fourth member of the gametic series ($r=1$ in the repulsion phase).

s, the second and third member of the gametic series ($s=1$ in the coupling phase).

AB, Ab, aB and ab, theoretical values of the members of the F_2 phenotypic zygotic series.

AB_0, Ab_0, aB_0 and ab_0 observed values of the members of the F_2 phenotypic zygotic series.

n, the sum of AB, Ab, aB and ab.

n_0, the sum of AB_0, Ab_0, aB_0 and ab_0.

k, constant for a given series obtained by dividing the sum of AB and ab by n (or by dividing the sum of AB_0 and ab_0 by n_0) and multiplying the quotient by 16, (formulae 1 and 2).

E_0, the sum of AB_0 and ab_0.

M_0, the sum of Ab_0 and aB_0.

c, coefficient obtained by dividing n_0 by n.

THE CONSTANT (k)

As is shown in table 1, the value of the constant k is dependent on the relationship existing between the genetic factors concerned. If there is independent inheritance, k has a value of 10; if A and B are coupled, the value of k varies from 10 to 16, depending on the strength of the coupling; and if A and b are linked (repulsion phase), k varies from 10 to 8 depending on the strength of the linkage. If the linkage in the coupling phase is complete, k has a value of 16, but in case of complete linkage in the repulsion phase, k has a value of 8.

Figure 1 shows that as the value of r (coupling phase) increases, the value of the constant k also increases, approaching 16 as a limit. Figure 2 shows that as the value of s (repulsion phase) increases, the value of the constant k decreases, approaching 8 as a limit. Table 2 shows that when large values for r and s are assumed, the constant k approaches 16 in case of coupling and 8 in case of repulsion. With such high values for r and s the linkage is practically complete.

TABLE 2

Showing the value of k when the values of r (coupling) and s (repulsion) are extremely high.

GAMETIC SERIES				TOTAL (n)	CONSTANT (k)
r	s	s	r		
3199	1	1	3199	40,960,000	15.99500078125
1	3199	3199	1	40,960,000	8.00000078125

FORMULAE

The formulae given below are applicable when deviations from a dihybrid Mendelian F_2 distribution are caused by linkage. When such disturbances as differential viability of gametes or failure of many seeds of certain types to germinate are operating, the results obtained by the application of the formulae to observed values will be misleading. The following formula shows the method of obtaining the value of k from the theoretical F_2 phenotypic series:

$$k=\frac{16(AB+ab)}{n} \tag{1}$$

The general formula also holds for obtaining the value of k from observed F_2 phenotypic ratios:

$$k=\frac{16(AB_0+ab_0)}{n_0} \tag{2}$$

Most of the formulae that follow indicate the relationship of the members of *theoretical* series. Several formulae are given to show the relationship when *observed* values are used. All formulae which are applicable to observed values are applicable also to theoretical values, but formulae which are applicable to theoretical values are applicable to observed values only when the value of c is unity. The theoretical values can be calculated from the observed values by means of the constant k.

Formulae for finding the value of n when the value of r, s, AB, Ab, aB, ab, or k is known

Coupling phase

$$n=4(r+1)^2 \tag{3}$$

From formulae (15) and (17) the following value for AB is derived,

$$AB=\frac{n}{2}+\left(\frac{\sqrt{n}}{2}-1\right)^2 \tag{4}$$

Hence,

$$n=\left[\frac{2\sqrt{3AB-2}+2}{3}\right]^2 \tag{5}$$

$$n=(Ab+1)^2 \text{ or } (aB+1)^2 \tag{6}$$

$$n=\left[2(\sqrt{ab}+1)\right]^2 \tag{7}$$

$$n=\left[\frac{4(\sqrt{2k-16}+4)}{16-k}\right]^2 \tag{8}$$

Repulsion phase

$$n=4(s+1)^2 \tag{9}$$

$$n=2(AB-1) \tag{10}$$

$$n=4(Ab+1) \tag{11}$$

$$n=\frac{32}{k-8} \tag{12}$$

Formulae for finding the value of any member of either gametic or F_2 phenotypic series when the value of n is known

Coupling phase

$$r = \frac{\sqrt{n}}{2} - 1 \tag{13}$$

$$AB = \frac{3n - 4(\sqrt{n} - 1)}{4} \quad \text{(derived from formula 5)} \tag{14}$$

$$AB = \frac{n}{2} + ab \tag{15}$$

$$Ab \text{ or } aB = \sqrt{n} - 1 \tag{16}$$

$$ab = \left[\frac{\sqrt{n}}{2} - 1\right]^2 \tag{17}$$

$$k = \frac{16\left[(\sqrt{n} - 1)^2 + 1\right]}{n} \tag{18}$$

Repulsion phase

$$s = \frac{\sqrt{n}}{2} - 1 \tag{19}$$

$$AB = \frac{n}{2} + 1 \tag{20}$$

$$Ab \text{ or } aB = \frac{n}{4} - 1 \tag{21}$$

$$k = \frac{8(n+4)}{n} \tag{22}$$

Formulae showing the relation of the different members of the phenotypic series to one another

Coupling phase

Substituting in formula 4, the value of n obtained in formula 6, we have

$$AB = \frac{(Ab+1)^2}{2} + \left[\frac{Ab+1}{2} - 1\right]^2$$

which reduces to

$$AB = \frac{3\,(Ab)^2 + 2Ab + 3}{4} \tag{23}$$

Substituting in formula 4, the value of n obtained in formula 7, we have

$$AB=\frac{[2(\sqrt{ab}+1)]^2}{2}+\left[\frac{2\sqrt{ab}+2}{2}-1\right]^2$$

which reduces to

$$AB=3ab+4\sqrt{ab}+2 \qquad (24)$$

$$Ab=\frac{2\sqrt{3AB-2}-1}{3} \text{ (derived from formula 23)} \qquad (25)$$

Substituting in formula 16, the value of n obtained in formula 7, we have

$$Ab=2\sqrt{ab}+1 \qquad (26)$$

Substituting in formula 17, the value of n obtained in formula 5, we have

$$ab=\left[\frac{\sqrt{3AB-2}+1}{3}-1\right]^2$$

which reduces to

$$ab=\frac{3AB-4\sqrt{3AB-2}+2}{9} \qquad (27)$$

Substituting in formula 17, the value of n obtained in formula 6, we have

$$ab=\left[\frac{Ab+1}{2}-1\right]^2$$

which reduces to

$$ab=\left[\frac{Ab-1}{2}\right]^2 \qquad (28)$$

Repulsion phase

$$AB=2Ab+3 \qquad (29)$$

Formulae for finding the percentage of crossing over when the value of r, s, AB, Ab, aB, ab, n or k is known

Coupling phase

$$\text{Percentage of crossing over}=100\left[\frac{1}{r+1}\right] \qquad (30)$$

From formulae 3 and 4 we derive

$$AB=\frac{4(r+1)^2}{2}+\left[\frac{2(r+1)}{2}-1\right]^2$$

From this equation we obtain

$$r=\frac{\sqrt{3AB-2}-2}{3} \qquad (31)$$

Substituting in formula 30, the value of r obtained in formula 31, we have

$$\text{Percentage of crossing over} = 100\left[\frac{3}{\sqrt{3AB-2}+1}\right] \tag{32}$$

$$r = \frac{Ab-1}{2} \quad \text{(derived from formulae 3 and 6)} \tag{33}$$

From formulae 30 and 33 we derive

$$\text{Percentage of crossing over} = 100\left[\frac{2}{Ab+1}\right] \tag{34}$$

$$r = \sqrt{ab} \quad \text{(derived from formulae 3 and 7)} \tag{35}$$

From formulae 30 and 35 we derive

$$\text{Percentage of crossing over} = 100\left[\frac{1}{\sqrt{ab}+1}\right] \tag{36}$$

$$\text{Percentage of crossing over} = 100\left[\frac{2}{\sqrt{n}}\right] \quad \text{(derived from formulae 3 and 30)} \tag{37}$$

From formula 8 we derive

$$r+1 = \frac{2(\sqrt{2(k-8)}+4)}{16-k} \tag{38}$$

Hence,

$$\text{Percentage of crossing over} = 100\left[\frac{16-k}{2\sqrt{2(k-8)}+8}\right] \tag{39}$$

Repulsion phase

$$\text{Percentage crossing over} = 100\left[\frac{2}{\sqrt{2(AB-1)}}\right] \tag{40}$$

$$\text{Percentage crossing over} = 100\left[\frac{1}{\sqrt{Ab+1}}\right] \tag{41}$$

$$\text{Percentage crossing over} = 100\left[\frac{2}{\sqrt{n}}\right] \tag{42}$$

$$\text{Percentage crossing over} = 100\left[\frac{k-8}{2\sqrt{2(k-8)}}\right] \tag{43}$$

Formulae for calculating the percentage of crossing over from observed values

Coupling phase

Substituting in formula 39, the value of k obtained in formula 2, we have

$$\text{Percentage crossing over} = 100\left[\frac{\left(16-\frac{16(AB_0+ab_0)}{n_0}\right)}{2\sqrt{2\left(\frac{16(AB_0+ab_0)}{n_0}\right)-16+8}}\right]$$

which is equivalent to

$$\text{Percentage crossing over} = 100\left[1-\sqrt{\frac{AB_0-Ab_0-aB_0+ab_0}{AB_0+Ab_0+aB_0+ab_0}}\right]$$

Substituting E_0 for the sum of AB_0 and ab_0, M_0 for the sum of Ab_0 and aB_0, and n_0 for the sum of AB_0, Ab_0, aB_0 and ab_0, we have,

$$\text{Percentage of crossing over} = 100\left[1-\sqrt{\frac{E_0-M_0}{n_0}}\right] \tag{44}$$

Repulsion phase

Substituting in formula 43, the value of k obtained in formula 2, we have

$$\text{Percentage crossing over} = 110\left[\frac{\frac{16(AB_0+ab_0)}{n_0}-8}{2\sqrt{2\left(\frac{16(AB_0+ab_0)}{n_0}-8\right)}}\right]$$

which is equivalent to

$$\text{Percentage crossing over} = 100\sqrt{\frac{AB_0-Ab_0-aB_0+ab_0}{AB_0+Ab_0+aB_0+ab_0}}$$

Substituting E_0 for the sum of AB_0 and ab_0, M_0 for the sum of Ab_0 and aB_0, and n_0 for the sum of AB_0, Ab_0, aB_0 and ab_0, we have

$$\text{Percentage crossing over} = 100\sqrt{\frac{E_0-M_0}{n_0}} \tag{45}$$

The percentage of crossing over may also be obtained by dividing E_0 by M_0 and referring to table 3. Fractional percentages can be obtained by interpolation. As an example, the percentage of crossing over may be calculated as follows from the first distribution in table 4:

$$E_0=631; \qquad M_0=50$$

$$\frac{E_0}{M_0}=12.6$$

By referring to table 3 we find that when the value of $E_0 \div M_0 = 13.8038$ the percentage of crossing over is 7 and when the value of $E_0 \div M_0 = 12.0208$ the percentage of crossing over is 8. The difference $= 1.7830$

$$13.8038 - 12.6 = 1.2038$$

$$\frac{1.2038}{1.7830} \text{ of 1 per cent } = 0.67 \text{ per cent}$$

$$7.0 \text{ per cent } + 0.67 \text{ per cent} = 7.67 \text{ per cent}$$

TABLE 3

Crossing-over percentages from 1 to 50 and the corresponding ratios of E_0 *to* M_0.

COUPLING PHASE				REPULSION PHASE			
Percent crossing over	$\frac{E_0}{M_0}$	Percent crossing over	$\frac{E_0}{M_0}$	Percent crossing over	$\frac{E_0}{M_0}$	Percent crossing over	$\frac{E_0}{M_0}$
1	99.5025	26	3.4209	1	1.0002	26	1.1450
2	49.5051	27	3.2817	2	1.0008	27	1.1573
3	32.8409	28	3.1528	3	1.0018	28	1.1701
4	24.5102	29	3.0331	4	1.0032	29	1.1836
5	19.5128	30	2.9215	5	1.0050	30	1.1978
6	16.1822	31	2.8175	6	1.0072	31	1.2126
7	13.8038	32	2.7202	7	1.0098	32	1.2282
8	12.0208	33	2.6291	8	1.0129	33	1.2444
9	10.6346	34	2.5436	9	1.0163	34	1.2614
10	9.5263	35	2.4631	10	1.0202	35	1.2792
11	8.6200	36	2.3875	11	1.0245	36	1.2978
12	7.8652	37	2.3162	12	1.0292	37	1.3172
13	7.2271	38	2.2488	13	1.0344	38	1.3375
14	6.6805	39	2.1852	14	1.0400	39	1.3588
15	6.2072	40	2.1250	15	1.0460	r0	1.3810
16	5.7935	41	2.0679	16	1.0525	41	1.4041
17	5.4288	42	2.0139	17	1.0595	42	1.4283
18	5.1050	43	1.9625	18	1.0670	43	1.4537
19	4.8157	44	1.9137	19	1.0749	44	1.4802
20	4.5556	45	1.8674	20	1.0833	45	1.5078
21	4.3206	46	1.8232	21	1.0923	46	1.5368
22	4.1073	47	1.7813	22	1.1017	47	1.5670
23	3.9128	48	1.7412	23	1.1117	48	1.5988
24	3.7349	49	1.7031	24	1.1222	49	1.6319
25	3.5714	50	1.6667	25	1.1333	50	1.6667

DISCUSSION

In order to illustrate the method of application of some of the formulae, table 4 was compiled. The data in this table are the same as those used by EMERSON (1916). Formula 2 was first used to obtain the value of k. From the value of k, the value of n was calculated by using formula 8 for the first two distributions and formula 12 for the third distribution. Finding the value of n in the coupling phase involves several computations because formula 8 is quite complex, but after n has once been obtained the calculation from n, of the other members of the zygotic series, is relatively simple. The value of k in the fourth distribution is 7.94. Since this value is less than 8, the repulsion formulae are not applicable. EMERSON (1916) found that the value of r for this distribution could not be determined by the formula which he proposed because $(1335+2)-(643+714)$ is a negative quantity (-20) and has no real root.

TABLE 4

Theoretical values of various terms obtained from observed values.

OBSERVED					THEORETICAL						
AB	Ab_0	aB_0	ab_0	n_0	k	$r:s$	AB	Ab	aB	ab	n
493	25	25	138	681	14.825	12.1 : 1	488.8	25.2	25.2	146.1	685.3
165	58	58	78	359	10.830	1.46 : 1	14.30	3.93	3.93	2.13	24.34
336	150	143	11	640	8.675	1:2.44	24.7	10.85	10.85	1	47.4
1335	643	714	2	2694	7.94	......	......	......	......	......	......

The formulae reported in this paper are applicable only when the value of k is between 10 and 8 (repulsion phase) or 10 and 16 (coupling phase). With a reduction of the same number of AB_0 or ab_0 types for a given value of r or s the limits for the application of the formulae are reached sooner in the repulsion phase than in the coupling phase as is indicated by the nature of the curves in figures 1 and 2. In the coupling phase the value of the points on the curve approaches 16 *gradually*, while in the repulsion phase it approaches 8 *rapidly* at first, then more slowly.

The relationship of E_0 and M_0 to r and s is shown in the following formulae:

$$\sqrt{\frac{E_0-M_0}{E_0+M_0}}=\frac{r}{r+1} \quad \text{(coupling phase)} \tag{46}$$

and

$$\sqrt{\frac{E_0-M_0}{E_0+M_0}}=\frac{1}{s+1} \quad \text{(repulsion phase)} \tag{47}$$

The relationship between E_0 and M_0 may also be expressed as follows:

$$E_0=M_0+4cr^2 \quad \text{(coupling phase)} \tag{48}$$

and

$$E_0=M_0+4c \quad \text{(repulsion phase)} \tag{49}$$

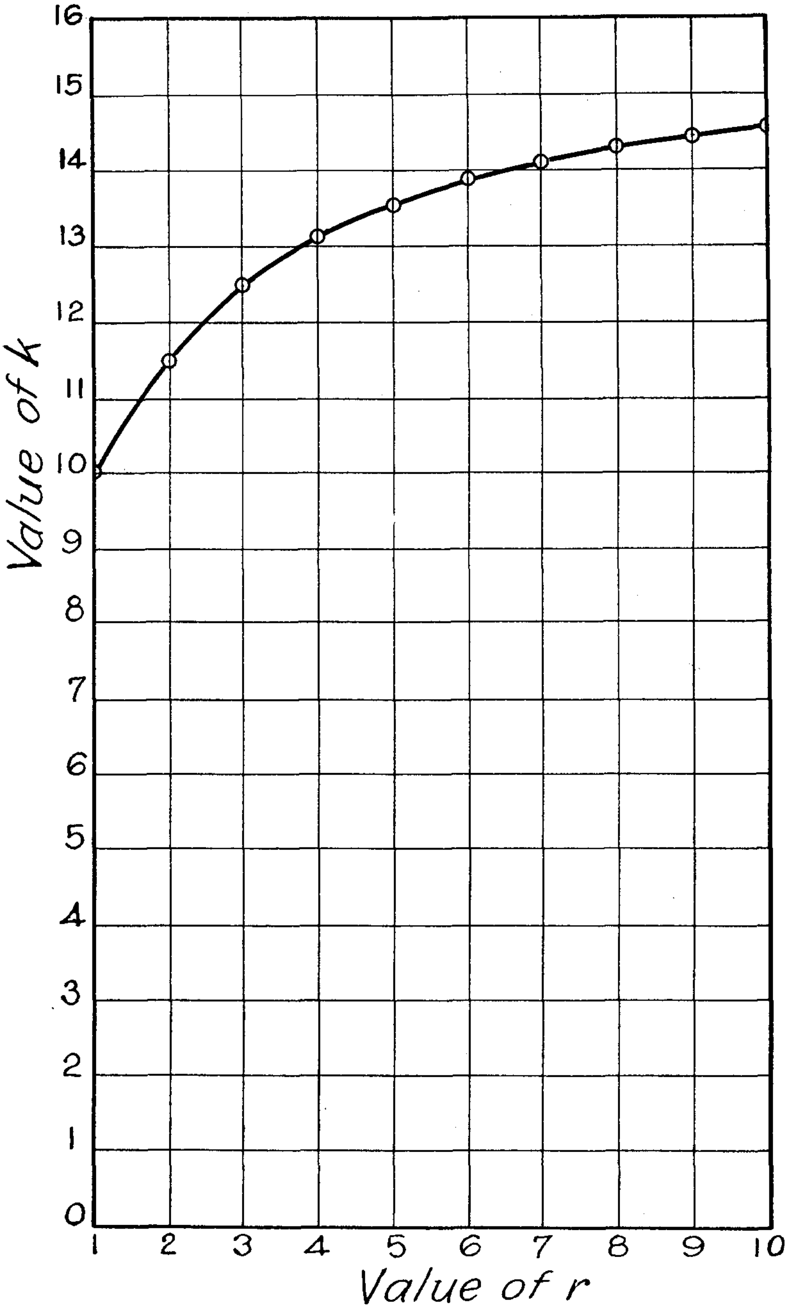

FIGURE 1.—The curve shows that when the value of r is increased, the value of k approaches 16 gradually.

A further relationship between E_0 and M_0 is indicated by formulae 50 and 51:

$$\frac{E_0}{M_0} = r + \frac{1}{2} + \frac{c}{M_0} \quad \text{(coupling phase)} \tag{50}$$

and

$$\frac{E_0}{M_0} = 1 + \frac{4c}{M_0} \quad \text{(repulsion phase)} \quad (51$$

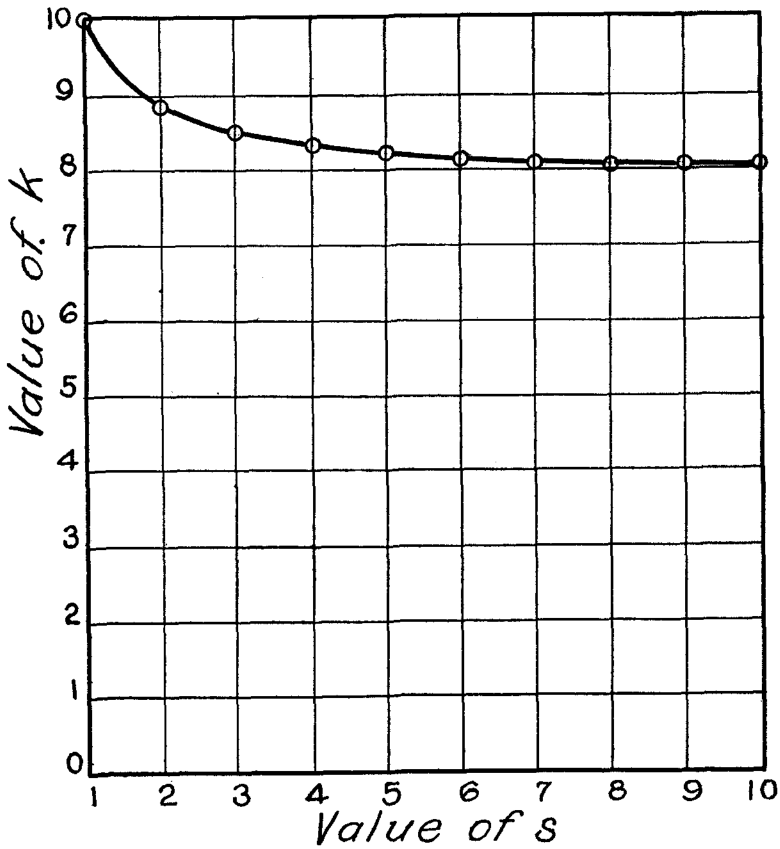

FIGURE 2.—The curve shows that when the value of s is increased, the value of k approaches 8, rapidly at first, then more slowly.

The value of E_0 divided by M_0 in the coupling phase is approximately equal to $r+\frac{1}{2}$ and the greater the magnitude of r the less is the magnitude of $\frac{c}{M_0}$ which approaches zero as the value of r approaches infinity. On the other hand, the value of E_0 divided by M_0 in the repulsion phase is approximately equal to 1, and the magnitude of $\frac{4c}{M_0}$ decreases as the value of s increases. The value of $\frac{4c}{M_0}$ approaches zero as the value of s approaches infinity. In other words, the ratio of E_0 to M_0 is approximately equal to the magnitude of $r+\frac{1}{2}$ in the coupling phase while in the repulsion phase E_0 is approximately equal to M_0. With this relationship in mind it is evident that a considerable reduction of AB_0 and ab_0 types, caused perhaps by differential viability or selective fertilization, is permissible in the coupling phase to meet the conditions necessary for the application of the formulae. The effect of such reductions, however, is not corrected by the formulae and thus an erroneous value for r is obtained. On the other hand, owing to the approximate equality of E_0 and M_0 in the repulsion phase, a slight reduction of the AB_0 and ab_0 types may cause the value of E_0 to be less than M_0. If the reduction of these types is greater than $4c$, then E_0 minus M_0 will have a negative value.

In order that the formulae may be usable, the observed values should deviate as little as possible from the result obtained by multiplying the theoretical distribution by c. If the observed distribution corresponds to this distribution, the quotient obtained by dividing the sum of the types AB_0 and Ab_0 by the types aB_0 and ab_0 is a constant, being equal to 3 in both coupling and repulsion phases.

$$\frac{AB_0+Ab_0}{aB_0+ab_0}=3 \tag{52}$$

SUMMARY

1. A constant k which shows the relation of observed values in linkage studies to independent inheritance is obtained.

2. The value of k indicates the following:
   - (a) When $k=10$, independent inheritance.
   - (b) When $k=$more than 10 but less than 16, coupling.
   - (c) When $k=$less than 10 but more than 8, repulsion.
   - (d) When $k=$less than 8, some factor other than chance resulting in marked deviations from theoretical ratios.

3. Formulae are given showing the relation of the members of the theoretical gametic and zygotic series to one another.

4. The relationship between the extreme terms and the middle terms of a distribution is shown.

5. Formulae and tables for determining the percentage of crossing over are presented.

ACKNOWLEDGMENT

The writer is indebted to Professor C. M. WOODWORTH for helpful suggestions and constructive criticisms given during the preparation of this paper and the constant interest shown throughout its progress.

LITERATURE CITED

BRUNSON, A. M., 1924 The inheritance of a lethal pale green seedling character in maize. Memoir Cornell Univ. 72. pp. 1-22.

COLLINS, G. N., 1924 Measurement of linkage values. Jour. Agric. Res. **27**: 881-891.

EMERSON, R. A., 1916 The calculation of linkage intensities. Amer. Nat. **50**: 411-420.

HALDANE, J. B. S., 1919 The probable errors of calculated linkage values and the most accurate method of determining gametic from certain zygotic series. Jour. Genetics **8**: 291-297.

WOODWORTH, C. M., 1923 Calculation of linkage intensities where duplicate factors are concerned. Genetics **8**: 106-115.