ANALYSIS OF QUANTITATIVE GENE ACTION BY CONSTANT PARENT REGRESSION AND RELATED TECHNIQUES[1]

BRUCE GRIFFING
Iowa State College, Ames, Iowa

Received November 14, 1949

ANALYSIS of multigenic or quantitative inheritance is one of the important problems of theoretical and applied genetics. Such analysis requires techniques of a different nature from those used in studying individual genes of large effect. The techniques thus far devised have not been wholly satisfactory, and for this reason new methods of analyzing quantitative inheritance are of considerable interest. A relatively new method, the constant parent regression technique, is reviewed and extended in this paper.

There is a considerable body of evidence to demonstrate that typical quantitative characters are controlled by a large number of gene pairs. It is assumed that the following statements are corollaries of this multigenic nature of inheritance: (1) linkage results because the large number of quantitative genes for a particular trait are located on a relatively small number of chromosome pairs, and (2) in general, the effect of each gene pair is relatively small. Because of this low order of effect of individual genes, it is necessary to study the action of these genes "en masse" by statistical techniques. This obviously results in inferences about the average properties of a set of quantitative genes.

To estimate the average gene action one may construct models involving different types of gene interaction and choose the model which best fits the experimental data. In thus reducing the total gene action entering into the expression of a complex characteristic to that of a simple model, it cannot be assumed that all genes behave in the manner prescribed by the chosen model. In fact, it has been suggested that quantitative genes probably have as diverse types of action as the so-called qualitative ones and differ from qualitative genes only in magnitude of effect. However, it has been shown in various quantitative traits that the system of genes involved does have average properties which are measurable. Estimation of these group genetic parameters is the objective of a statistical analysis, and these estimated parameters are associated with a gene model. The principal approach that has developed over the years involves, (1) a description of frequency distributions resulting from segregating populations by use of first, second, and third moments, and (2) a partitioning of variances into components. Most of the techniques have been summarized by MATHER (1949).

Use of segregating populations offers certain advantages but entails both theoretical and practical difficulties. First, with segregation, linkage effects are generated which contribute to the genotypic variance. This complex problem has been dealt with to some extent by MATHER (1949). Second, a large

[1] Journal Paper No. J-1719 of the IOWA AGRICULTURAL EXPERIMENT STATION, Ames, Iowa. Project No. 250.

number of individuals are needed to estimate properly the parameters of any one particular genetically segregating population. For example: justifiably to partition the variance of an F_2, the experimenter needs to include a fairly large number of F_3 or biparental families each with adequate number of individuals to estimate their own distributions. This leads to the physical difficulty of making observations on a vast number of individuals. If the characteristics involved are functions of indeterminate type of growth, such as yield of tomatoes, the amount of work becomes tremendous. Third, difficulties arise in designing experiments from which estimates of different populations are accurate and comparable. There is also the problem of confounding genetic and block effects in a replicated experiment.

By considering only the non-segregating generations (homozygous parents and F_1s) the constant parent regression method, evolved by Hull (1946, 1947), considerably reduces the previously mentioned sources of error. Since the P_1s and F_1s are not segregating, relatively few plants are needed to estimate the expected genotypic values. The actual number necessary depends on the heritability of the traits involved, and the accuracy may be increased to any practical desired level by merely increasing the number of plants. Variance within a line is assumed to be entirely environmental and thus, with appropriate sampling techniques, an analysis of the source of these environmental variations is provided. Experimental designing presents no problems and there is no genetic confounding. (There is, of course, a genotypic-environmental interaction which may be estimated.)

Linkage disturbs the constant parent regression analysis only in the sampling procedure of obtaining the set of parents used from all theoretical possible homozygous genotypes. In populations resulting from recent hybrid origin, linkages existing in the original parental types may reduce the probability of occurrence of some gene combinations and raise the probability of obtaining others. It can be demonstrated that coupling linkages do not affect the constant parent regression results whereas repulsion linkages may give rise to some pseudo-overdominance as measured by the second order regression coefficient. However, with random mating over a period of time the coupling and repulsion phases are expected to become equally frequent in the population, and thus, the effect of linkage in sampling parental lines disappears. The possibility of linkage effect should not be overlooked but in most cases is probably of little concern.

Experiments with segregating populations generally involve only two original parents. On the other hand, the P_1-F_1 test is conducted with a variety of different parents, thus providing a wider sampling of the available germplasm, and allowing broader inferences to be made from the result obtained. By using a number of parents, the experimenter interested in more than one characteristic should be able to choose lines which collectively give desired ranges of expression in all traits. This might be difficult to do when only two parents must be chosen.

The P_1-F_1 methodology also directly and easily yields estimates of herita-

bility and of phenotypic and genotypic correlations. These estimates may be used to develop discriminant function selection indices if desired (see GRIFFING 1948).

The main disadvantages of the P_1-F_1 technique are that estimation of gene number or number of effective factors (MATHER 1949), detection of major genes and examination of gene-chromosome relationships which could involve linkage and phase of linkage are not directly possible.

CONSTANT PARENT REGRESSION METHOD

The following presentation will consider the experimental situation in which a set of inbred (homozygous) lines are available together with all possible F_1 combinations of these lines. These restrictions are not entirely necessary in that not all F_1s need be present, different males and females may be used, and F_2s and later generations may be considered (HULL 1947a). However, for simplicity the above mentioned restrictions will be imposed. Generally, one set of F_1s is used and not the reciprocals, *i.e.* $F_{ij} = F_{ji}$ in table 1.

Notation is as follows:

$P_i = i^{th}$ parent
$F_{ij} = F_1$ (or hybrid) of i^{th} and j^{th} parents
$F_{ij} = F_{ji}$
c.p. = constant parent (homozygous)
c.p.r. = constant parent regression, also noted as bp.

F_1 and F_{ij} are used interchangeably. The term "F_1" is used in the general sense, and F_{ij} is used more specifically to indicate the F_1 of the i^{th} and j^{th} parents.

(1) "n" parents are crossed in all possible combinations of pairs to give

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} \text{ hybrids.}$$

TABLE 1

Combination of n inbred lines to give all possible F_1's.

	P_1	P_2	P_3 - - - - - - - - - - P_j - - - - P_n
P_1	P_1	F_{12}	F_{13} - - - - - - - - - - F_{1j} - - - - F_{1n}
P_2	F_{21}	P_2	F_{23} - - - - - - - - - - F_{2j} - - - - F_{2n}
P_3	F_{31}	F_{32}	P_3 - - - - - - - - - - F_{3j} - - - - F_{3n}
.	.	.	. . .
.	.	.	. . .
.	.	.	. . .
P_i	F_{i1}	F_{i2}	F_{i3} - - - - P_i - - - - F_{ij} - - - - F_{in}
.	.	.	. . .
.	.	.	. . .
.	.	.	. . .
P_n	F_{n1}	F_{n2}	F_{n3} - - - - F_{ni} - - - - F_{nj} - - - - P_n

(2) Since there are "n" parents there are "n" c.p. groups; the i^{th} c.p. group is defined as follows:
   (a) The i^{th} row constitutes the i^{th} c.p. group and has
      1. P_i as c.p.; P_j where $j=1, 2, \cdots, n, i \neq j$ as variable parents.
      2. $(n-1)$ hybrids ($F_{i1}, F_{i2}, \cdots, F_{in}$); *i.e.*, F_1s resulting from P_i crossed with all other parents.

With this scheme as a basis, HULL (1946) ingeniously devised a statistical approach to the problem of estimating the relative dominance effect for the simplest gene model in which between-loci-effects are additive (no between-loci-interactions). We shall consider this gene model as well as some aspects of an epistatic gene model in which between-loci interactions occur.

The statistical approach is made by evaluating the trend of c.p.r. coefficients. This involves a regression of F_1s on the variable parents within each c.p. group. A c.p.r. coefficient is calculated for each c.p. group. As will be shown later, the trend (increasing or decreasing) of c.p.r. coefficients relative to the value of the corresponding c.p. will give information regarding both direction and magnitude of dominance. This trend is measured by a "second order" regression, *i.e.*, the regression of c.p.r.'s on the c.p. values.

First, models involving only two-gene pairs will be considered, and later these will be extended to the "n" gene case, using models having three-gene pairs to point out difficulties encountered.

A. *Two-gene-pair models*

With two gene pairs, four c.p. groups are possible, each having three F_1s, as shown in table 2.

Two types of models will be discussed, one involving dominance alone with no epistasis and the other considering both dominance and epistasis.

1. Dominance but no epistasis

Assumptions:

1. Gene effects are the same for both gene pairs. The notation followed is a modification of FISHER (1918).

$A_1A_1=A_2A_2=2d$	where:	$h=0$	no dominance
		$h=\pm 1$	complete dominance
$A_1a_1=A_2a_2=d+dh$		$-1<h<0$	incomplete neg. dom.
		$0<h<+1$	incomplete pos. dom.
$a_1a_1=a_2a_2=0$		$h>+1$ or $h<-1$	super or over dom.

2. Between-loci effects are additive.

These algebraic values give rise to table 3.

If c.p.r. coefficients are calculated for each c.p. group, one will find that they take on the algebraic values found in table 4.

On examination of table 4, it is obvious that the direction of c.p.r. trend, with respect to parental values, indicates plus or minus dominance. Thus, with no dominance ($h=0$) there is no trend; all c.p.r. coefficients are equal to .5. Diagrammatically, all the F_1s lie at midparental values, and, considering values with no error, there will be no deviations from regression. With positive

TABLE 2

General arrangement of Ps and all possible F1s for two-gene models.

	P_1	P_2	P_3	P_4
	$a_1a_1a_2a_2$	$A_1A_1a_2a_2$	$a_1a_1A_2A_2$	$A_1A_1A_2A_2$
P_1	P_1	$A_1a_1a_2a_2$	$a_1a_1A_2a_2$	$A_1a_1A_2a_2$
P_2		P_2	$A_1a_1A_2a_2$	$A_1A_1A_2a_2$
P_3			P_3	$A_1a_1A_2A_2$
P_4				P_4

dominance a decreasing trend occurs for the c.p.r. coefficients as the parental values increase. This trend becomes more severe with an increase of the dominance value. With complete positive dominance the trend goes from +1.0 to 0, and with overdominance the extreme c.p.r. values are 1+a/2 and −a/2. In this way a c.p.r. exceeding one or less than zero is indicative of overdominance. The c.p.r. values together with their trend as measured by b_2 give an indirect method of evaluating the direction and magnitude of dominance. A test of significance for b_2 and a more direct method of estimating "h" will be discussed later.

TABLE 3

Algebraic values for P1s and F1s for a two-gene model.

		P_1 0	P_2 2d	P_3 2d	P_4 4d
P_1	0		d+dh	d+dh	2d+2dh
P_2	2d	d+dh		2d+2dh	3d+dh
P_3	2d	d+dh	2d+2dh		3d+dh
P_4	4d	2d+2dh	3d+dh	3d+dh	

TABLE 4

Algebraic values for c.p.r.'s, together with values for these regression coefficients with different values of "h."

CONSTANT PARENT	CONSTANT PARENT REGRESSION VALUES					
	ALGEBRAIC VALUES	h=0	h=+1	h>+1 OR h=1+a, a>0	h=−1	n<−1 OR h=a−1, a<0
$P_1=0$	$\frac{1+h}{2}$	.5	1.0	$1+\frac{a}{2}$	0	$+\frac{a}{2}$
$P_2=2d$	.5	.5	.5	.5	.5	.5
$P_3=2d$	.5	.5	.5	.5	.5	.5
$P_4=4d$	$\frac{1-h}{2}$	.5	0	$-\frac{a}{2}$	1.0	$1-\frac{a}{2}$

$b_2=-\frac{h}{4d}$ (second order regression coeff. of c.p.r. on c.p. value)

2. Both dominance and epistasis

In setting up this model we may regard diagrammatically epistasis as an interaction between loci analogous to dominance as an interaction between alleles at the same locus. The basic diagram is the following which shows the four parental values.

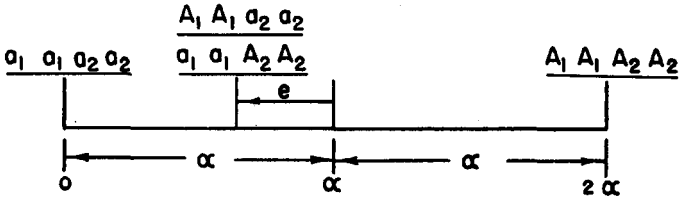

DIAGRAM 1

The genotypic values for the heterozygotes may be obtained by the fact that the F_1 is a dominance deviation away from the midpoint between its homozygotes. For example, the following diagram shows the derivation of $A_1a_1a_2a_2$.

$$A_1a_1a_2a_2 = 1/2(1 - e)(1 + h)\alpha.$$

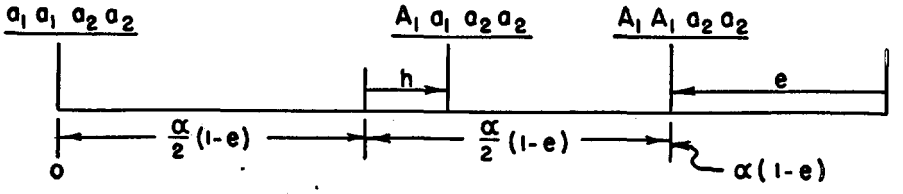

DIAGRAM 2

Following such a scheme all P_1 and F_1 values may be established, and the algebraic values are given in table 5.

TABLE 5

Algebraic values for P_1s and F_1s involving both dominance and epistasis.

	P_1 $a_1a_1a_2a_2$ 0	P_2 $A_1A_1a_2a_2$ $\alpha(1-e)$	P_3 $a_1a_1A_2A_2$ $\alpha(1-e)$	P_4 $A_1A_1A_2A_2$ 2α
$P_1=0$		$\frac{\alpha}{2}(1-e+h-eh)$	$\frac{\alpha}{2}(1-e+h-eh)$	$\frac{\alpha}{2}(2-e+2h+eh^2)$
$P_2=\alpha(1-e)$	$\frac{\alpha}{2}(1-e+h-eh)$		$\frac{\alpha}{2}(2-e+2h+eh^2)$	$\frac{\alpha}{2}(3-e+h+eh)$
$P_3=\alpha(1-e)$	$\frac{\alpha}{2}(1-e+h-eh)$	$\frac{\alpha}{2}(2-e+2h+eh^2)$		$\frac{\alpha}{2}(3-e+h+eh)$
$P_4=2\alpha$	$\frac{\alpha}{2}(2-e+2h+eh^2)$	$\frac{\alpha}{2}(3-e+h+eh)$	$\frac{\alpha}{2}(3-e+h+eh)$	

TABLE 6

Algebraic values for c.p.r. coefficients for both dominance and epistasis.

CONSTANT PARENT	C.P.R. COEFFICIENTS
$P_1=0$	$bp_1=\frac{1}{2}\frac{\{(1+h)+e[(1+h)^2+eh(1+h)]\}}{(1+e)^2}$
$P_2=\alpha(1-e)$	$bp_2=\frac{9-2eh(2eh-3)}{6(e^2+3)}$
$P_3=\alpha(1-e)$	$bp_3=\frac{9-2eh(2eh-3)}{6(e^2+3)}$
$P_4=2\alpha$	$bp_4=\frac{1}{2}\frac{\{(1-h)-e[(1-h)^2+eh(1-h)]\}}{(1-e)^2}$

Using these algebraic values, c.p.r. coefficients may be calculated for each c.p. group, and these are listed in table 6.

It may be noted from tables 5 and 6 that if "e" is equated to zero, then all algebraic values reduce to those in the first model (letting $\alpha=2d$).

A two-way classification table may be constructed to demonstrate the trends of c.p.r. coefficients for different numerical values of "e" and "h." These values are found in table 7.

The following characteristics of the "e" by "h" table may be noted.

(1) When $e=0$, $h=0$, $bp_i=.5$, as found earlier.

(2) When $e=0$, and "h" varied (first row), the same trend occurs as found in table 4, which is a steadily decreasing regression trend from bp_1 to bp_4. The relative decrease depends on the degree of dominance.

TABLE 7

Two-way classification of various values of "e" and "h" for constant parent regression coefficients.

EPISTASIS	DOMINANCE		
	h= 0	h=+.5	h=+1.0
e=0	$bp_1=$ .50	= .75	= 1.00
	$bp_2=$ .50	= .50	= .50
	$bp_3=$ .50	= .50	= .50
	$bp_4=$ .50	= .25	= 0
e=+.5	$bp_1=$ .33	= .63	= 1.00
	$bp_2=$ .46	= .53	= .56
	$bp_3=$ .46	= .53	= .56
	$bp_4=1.00$	= .63	= 0
e=+1.0	$bp_1=$ .25	= .56	= 1.00
	$bp_2=$ .38	= .46	= .46
	$bp_3=$ .38	= .46	= .46
	$bp_4=+\infty$	$=+\infty$	$=+\infty$

(3) When h=0, and "e" varied (first column), epistasis is considered alone without the confounding influence of dominance. The regression trend is opposite to that of dominance alone, as the regression trend steadily increases from bp_1 to bp_4. The relative increase depends on the degree of epistasis.

(4) When "h" and "e" both have values other than zero—*i.e.*, e=+.5, h=+.5—then conflicting regression trends result in a decrease from bp_1 to bp_2 and an increase from bp_3 to bp_4.

A consideration of RASMUSSON's (1933) interaction hypothesis can be made by merely assigning "e" negative values. The hypothesis may be summarized by stating that the phenotypic result of the addition of genes follows the law of diminishing returns. In other words, genes added when the character expression is low would have much greater effect than when the character expression is near a physiological limit.

The above two-gene pair model is only one of many possible approaches to between-loci interaction. It is an attempt to combine in one model both dominance and epistatic effects and still allow algebraic solution. An extension of this model to more gene pairs may be difficult and for an extended epistatic and dominance model the logarithmic case is chosen and will be discussed later.

HULL (1947a) has presented a regression equation which involves both dominance and complementary gene action.

BURDICK (1949) has developed an epistatic model for multiplicative effects of gene pairs assuming no dominance. The c.p.r.s yield an increasing trend in much the same manner as the epistatic model described above with h=0.

Testing for significance of b_2 (in both models) by an ordinary analysis of variance for the regression of c.p.r. coefficients on parental values is not valid because the c.p.r. coefficients are highly correlated. The problem can be approached by fitting constants to the following regression equation by least squares procedure:

$$F_{ij} = m + a_1(P_i + P_j) + a_2(P_i^2 + P_j^2) + a_3(P_iP_j) + e_{ij}$$

i.e. by estimating and testing a_1, a_2, and a_3. Because a_3 is equal to b_2 for the case of dominance alone, a test of significance of a_3 is equivalent to a test of significance for b_2 for dominance alone. Similarly since a_2 approximates b_2 for the case of epistasis alone (for the two-gene-pair model) a test of significance for a_2 is equivalent to a test for b_2 for epistasis alone. When "e" and "h" both have some value, the c.p.r. trend is curvilinear with respect to c.p. values so that the regression coefficient b_2 would have little meaning.

B. *Extension of two-gene-pair models*

1. Dominance but no epistasis

Model specifications are as follows:

1. There are "n" loci affecting the specified character.
2. Effect of gene pairs are identical. However, both d and h may vary.

$$A_1A_1 = A_2A_2 \cdots = A_nA_n = 2d$$
$$A_1a_1 = A_2a_2 \cdots = A_na_n = d + dh$$
$$a_1a_1 = a_2a_2 \cdots = a_na_n = 0.$$

3. Effects of genes at different loci are additive.

With "n" loci and the alternative of two alleles at any one locus, there are 2^n number of different possible homozygous parental lines. These inbreds may be denoted simply by their gamete arrangement, and all possible gamete arrangements may be indicated merely by the expansion of

$$\prod_{i=1}^{n} (A_i + a_i).$$

Parents of different genotypic constitution may have the same genotypic value, and thus F_1s with different genotypic values may result from crosses in which the parents have the same genotypic values. For example, from the three-gene-pair case as illustrated in table 8, one can see that parent No. 6 ($A_1a_2A_3$)

TABLE 8

Algebraic values for all possible parents and F_1s involved in the three-gene-pair model.

			P_1 $a_1a_2a_3$ 0	P_2 $A_1a_2a_3$ 2d	P_3 $a_1A_2a_3$ 2d	P_4 $a_1a_2A_3$ 2d	P_5 $A_1A_2a_3$ 4d	P_6 $A_1a_2A_3$ 4d	P_7 $a_1A_2A_3$ 4d	P_8 $A_1A_2A_3$ 6d
P_1	$a_1a_2a_3$	0	0	d+dh	d+dh	d+dh	2d+2dh	2d+2dh	2d+2dh	3d+3dh
P_2	$A_1a_2a_3$	2d	d+dh	2d	2d+2dh	2d+2dh	3d+dh	3d+dh	3d+3dh	4d+2dh
P_3	$a_1A_2a_3$	2d	d+dh	2d+2dh	2d	2d+2dh	3d+dh	3d+3dh	3d+dh	4d+2dh
P_4	$a_1a_2A_3$	2d	d+dh	2d+2dh	2d+2dh	2d	3d+3dh	3d+dh	3d+dh	4d+2dh
P_5	$A_1A_2a_3$	4d	2d+2dh	3d+dh	3d+dh	3d+3dh	4d	4d+2dh	4d+2dh	5d+dh
P_6	$A_1a_2A_3$	4d	2d+2dh	3d+dh	3d+3dh	3d+dh	4d+2dh	4d	4d+2dh	5d+dh
P_7	$a_1A_2A_3$	4d	2d+2dh	3d+3dh	3d+dh	3d+dh	4d+2dh	4d+2dh	4d	5d+dh
P_8	$A_1A_2A_3$	6d	3d+3dh	4d+2dh	4d+2dh	4d+2dh	5d+dh	5d+dh	5d+dh	6d

has a value of 4d, and parents No. 4 ($A_1a_2a_3$) and No. 3 ($a_1A_2a_3$) both have values of 2d. However, the F_1 $(6\times4)=3d+dh$, and the F_1 (6×3) has a different value namely, 3d+3dh. Therefore, it is impossible to determine uniquely the F_1 genotypic value with knowledge of only the parental genotypic values. In other words, in the three-dimensional coordinate system, where P_i, P_j and F_{ij} are the three axes, a three-dimensional constellation of points would be generated with "n" gene pairs rather than a simple surface (if the measurements used were the genotypic values).

HULL (1946, 1947) suggests a condensation process, which may be illustrated by the three-gene pair example, for collapsing the scatter of F_1s into a regression surface. In table 8, all rows having parents of like genotypic values are combined and similarly all columns of like genotypic values are combined. Table 8 then reduces to table 9, which has only one row and one column for

TABLE 9

Algebraic values for the condensed table involving the three-gene-pair model.

		P_1	P_2	P_3	P_4
		0	2d	4d	6d
P_1	0	0	d+dh	2d+2dh	3d+3dh
P_2	2d	d+dh	$2d+\frac{4}{3}dh$	$3d+\frac{5}{3}dh$	4d+2dh
P_3	4d	2d+2dh	$3d+\frac{5}{3}dh$	$4d+\frac{4}{3}dh$	4d+dh
P_4	6d	3d+3dh	4d+2dh	5d+dh	6d

each parental genotypic value. The F_1 values in table 9 are averages and may be obtained in two ways.

1. The F_1 values in table 8 may be averaged in the manner described above.
2. HULL (1946, 1947) suggests the following general scheme for determining the genotypic value for any particular F_1 cell of table 9.

Let u and w be the proportion of AA loci in P_i and P_j, respectively, and $(1-u)$ and $(1-w)$ be the proportion of aa loci. The average F_1 value may be determined as follows:

			P_i gametes		
			A	a	
			u	$1-u$	Parental coded values are:
	A	w	uw	$w(1-u)$	
P_j					$P_i = u$ 2nd
gametes					$P_j = w$ 2nd
	a	$1-w$	$u(1-w)$	$(1-u)(1-w)$	

Then

$$F_{ij} = uw(2nd) + \{w(1-u) + u(1-w)\}n(d + dh). \qquad (1)$$

(Parental and F_1 values are coded so that the completely recessive genotype, $a_1a_1a_2a_2 \cdots a_na_n$, has a genotypic value of zero.)

With HULL's theoretical procedure a regression surface is fitted to the three-dimensional scatter, and equation (1) may be used to predict the exact average F_1 values found in the cells of table 9.

Let us examine the situation more closely by considering separately the four constant parent groups possible from table 9. In figure 1, each constant parent group is plotted together with the regression of the F_1s on the variable parents. In all cases the F_1 values fall directly on the regression line, and the regression coefficients form a linear series so that there is a perfect, simple correlation between regression coefficients and the corresponding constant parent value.

To illustrate this point, one may first re-write equation (1) in terms of parental values and then partially differentiate the F_1 with respect to the variable parent.

$$F_{ij} = \frac{P_i + P_j}{2}(1 + h) - \frac{h}{2nd} P_i P_j.$$

The rate of change in the F_{ij} with change in P_j (variable parent) and considering P_i as the constant parent may be expressed as follows:

$$\frac{\partial F_{ij}}{\partial P_j} = \frac{1 + h}{2} - \frac{h}{2nd} P_i = bp_i \text{ (c.p.r. for } P_i).$$

The second order regression coefficient, which represents the rate of change in bp_i with change in the constant parent (P_i), may be represented as follows:

$$\frac{\partial bp_i}{\partial P_i} = -\frac{h}{2nd} = b_2.$$

However, in practice where, obviously, only a sample of all possible parental genotypes can be used one is not confronted with the idealized regression surface but rather the actual F_1 scatter. Under such conditions the c.p.r. scheme is not exact. This may be illustrated by forming different sets of parents in

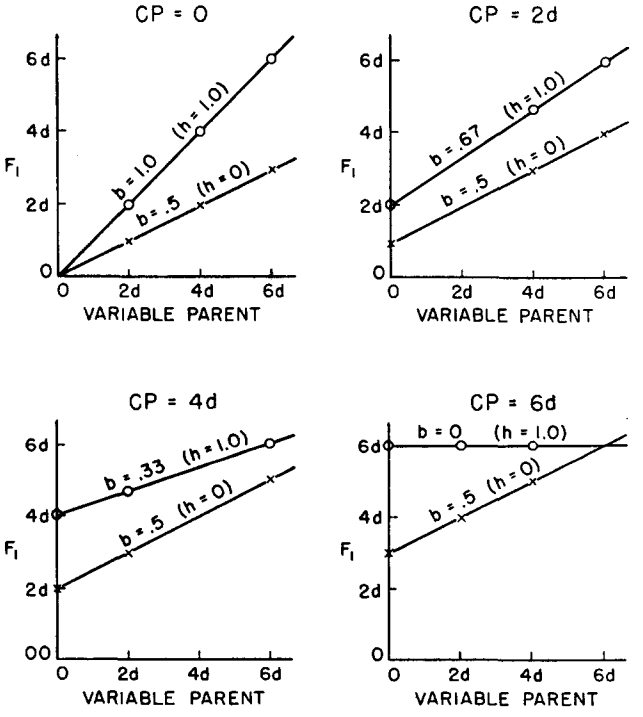

FIGURE 1.—Constant parent regressions for the four constant parent groups involved in the condensed set of F_1s.

which no one set contains all possible parents. From the three-gene-model, let us consider two parental sets ($S_1 = P_1$, P_2, P_5, P_8 and $S_2 = P_1$, P_4, P_5, P_8) chosen from table 8. Figures 2 and 3 give the c.p.r.'s for S_1 and S_2 respectively for the condition of complete dominance. The no-dominance ($h = 0$) regressions for each c.p. group are also included. Table 10 gives algebraic values for c.p.r. coefficients, variances of F_1s, and other statistics for the three sets, S_1, S_2, and the condensed (ideal) set. It is obvious that the regression trends in all three sets reflect the dominance effect when "h" takes on some value other than zero. Although the same "h" value may be assigned to all three sets, the correspond-

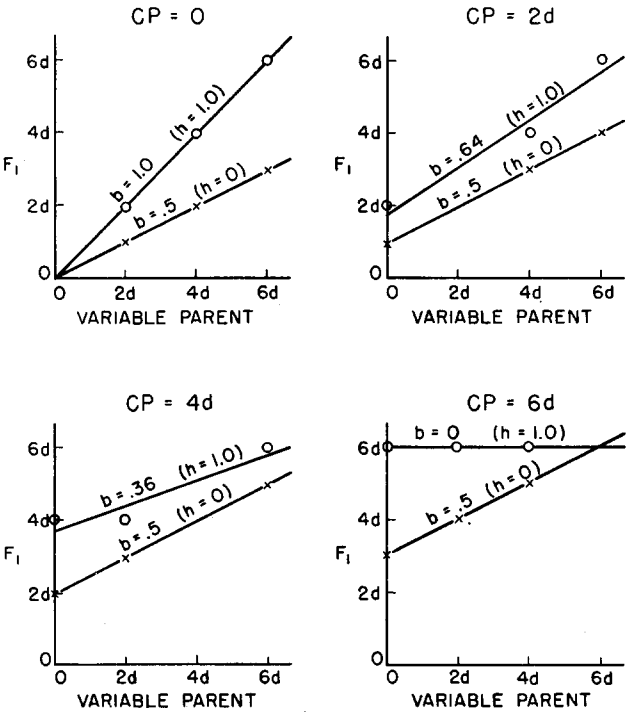

FIGURE 2.—Constant parent regressions for the four constant parent groups involved in set No. 1.

ing c.p.r. coefficients do not exactly agree. Thus, because of the sampling procedure operating in obtaining the set of parents, the c.p.r.'s do not exactly reflect the magnitude of dominance, although the inexactitude would probably be small if a wide range of character expression is exhibited among the parents. However, the fact that one has to deal with unpredictable points in the F_1 scatter, although causing some irregularities in the practical use of the c.p.r. technique, does offer another approach to the problem of dominance estimation.

This method of attack is to consider the deviations from regression and examine the components of F_1 variances. The general procedure due to FISHER (1918) of breaking down genotypic variances into additively genetic and non-additive portions is not of particular interest in this case. With

TABLE 10

Algebraic values for c.p.r. coefficients, variance among F_1s, and other statistics for the 4 c.p.groups in each of the 3 chosen sets, S_1, S_2, and condensed set.

	PARENTAL VALUE	c.p.r's.	VARIANCE AMONG F_1S	AMOUNT OF VARIANCE ATTRIBUTABLE TO REGRESSION $\frac{\{\sum xy\}^2}{\sum x^2}$	DEV'S FROM REG. $\sum d\overline{y \cdot x}^2$	CORRELATION BETWEEN DOM. & BASIC GEN EFFECTS ρ
Set No. 1	$P_1=0$	$\frac{1+h}{2}$	$\frac{1}{3}(2d^2+4d^2h+2d^2h^2)$	$\frac{1}{3}(2d^2+4d^2h+2d^2h^2)$	0	$+1$
	$P_2=2d$	$\frac{1}{2}+\frac{h}{7}$	$\frac{1}{3}\left(\frac{14}{3}d^2+\frac{8}{3}d^2h+\frac{2}{3}d^2h^2\right)$	$\frac{1}{3}\left(\frac{14}{3}d^2+\frac{8}{3}d^2h+\frac{8}{21}d^2h^2\right)$	$\frac{6}{63}d^2h^2$	$+\sqrt{\frac{4}{7}}$
	$P_3=4d$	$\frac{1}{2}-\frac{h}{7}$	$\frac{1}{3}\left(\frac{14}{3}d^2-\frac{8}{3}d^2h+\frac{2}{3}d^2h^2\right)$	$\frac{1}{3}\left(\frac{14}{3}d^2-\frac{8}{3}d^2h+\frac{8}{21}d^2h^2\right)$	$\frac{6}{63}d^2h^2$	$-\sqrt{\frac{4}{7}}$
	$P_4=6d$	$\frac{1-h}{2}$	$\frac{1}{3}(2d^2-4d^2h+2d^2h^2)$	$\frac{1}{3}(2d^2-4d^2h+2d^2h^2)$	0	-1
		$b_2=-\frac{23h}{140d}$				
Set No. 2	$P_1=0$	$\frac{1+h}{2}$	$\frac{1}{3}(2d^2+4d^2h+2d^2h^2)$	$\frac{1}{3}(2d^2+4d^2h+2d^2h^2)$	0	$+1$
	$P_2=2d$	$\frac{1}{2}+\frac{3}{14}h$	$\frac{1}{3}\left(\frac{14}{3}d^2+4d^2h+2d^2h^2\right)$	$\frac{1}{3}\left(\frac{14}{3}d^2+4d^2h+\frac{6}{7}d^2h^2\right)$	$\frac{24}{63}d^2h^2$	$+\sqrt{\frac{3}{7}}$
	$P_3=4d$	$\frac{1}{2}-\frac{3}{14}h$	$\frac{1}{3}\left(\frac{14}{3}d^2-4d^2h+2d^2h^2\right)$	$\frac{1}{3}\left(\frac{14}{3}d^2-4d^2h+\frac{6}{7}d^2h^2\right)$	$\frac{24}{63}d^2h^2$	$-\sqrt{\frac{3}{7}}$
	$P_4=6d$	$\frac{1-h}{2}$	$\frac{1}{3}(2d^2-4d^2h+2d^2h^2)$	$\frac{1}{3}(2d^2-4d^2h+2d^2h^2)$	0	-1
		$b_2=-\frac{6h}{35d}$				
Condensed Set	$P_1=0$	$\frac{1+h}{2}$	$\frac{1}{3}(2d^2+4d^2h+2d^2h^2)$	$\frac{1}{3}(2d^2+4d^2h+2d^2h^2)$	0	$+1$
	$P_2=2d$	$\frac{1}{2}+\frac{1}{6}h$	$\frac{1}{3}\left(\frac{14}{3}d^2+\frac{28}{9}d^2h+\frac{14}{27}d^2h^2\right)$	$\frac{1}{3}\left(\frac{14}{3}d^2+\frac{28}{9}d^2h+\frac{14}{27}d^2h^2\right)$	0	$+1$
	$P_3=4d$	$\frac{1}{2}-\frac{1}{6}h$	$\frac{1}{3}\left(\frac{14}{3}d^2-\frac{28}{9}d^2h+\frac{14}{27}d^2h^2\right)$	$\frac{1}{3}\left(\frac{14}{3}d^2-\frac{28}{9}d^2h+\frac{14}{27}d^2h^2\right)$	0	-1
	$P_4=6d$	$\frac{1-h}{2}$	$\frac{1}{3}(2d^2-4d^2h+2d^2h^2)$	$\frac{1}{3}(2d^2-4d^2h+2d^2h^2)$	0	-1
		$b_2=-\frac{h}{6d}$				

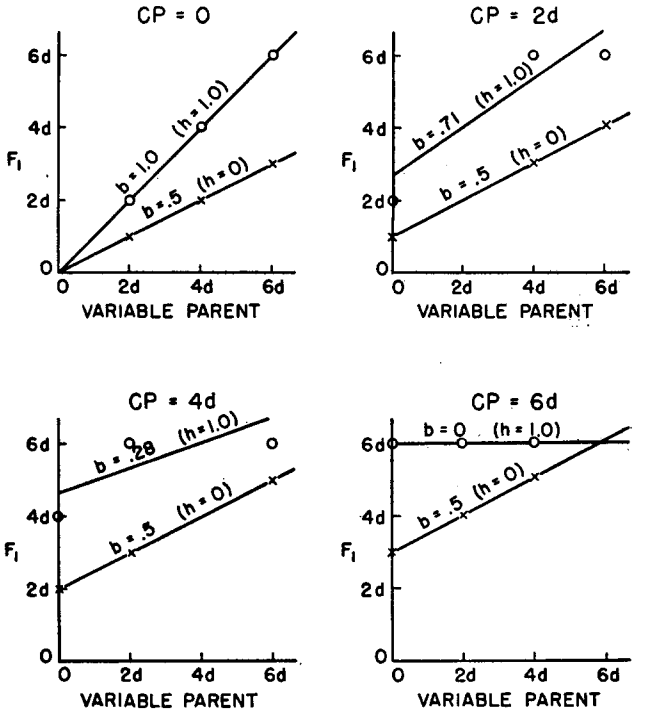

FIGURE 3.—Constant parent regressions for the four constant parent groups involved in set No. 2

homozygous parents and controlled matings a slightly different mathematical model may be used. The genotypic value of an F_1 is composed of two parts, a basic gene effect which is a function of "d," and a dominance effect which is a function of "dh." The variance among F_1s then involves a basic gene variance, a dominance variance, and a covariance, as these two effects may be correlated (see table 10). The partitioning of variances will be discussed after the next section.

2. Both dominance and epistasis (logarithmic case)

There are undoubtedly many ways that interaction between loci can be handled. The only general scheme that will be mentioned here will be that resulting from logarithmic gene action. When characters controlled by logarithmic or geometric gene action are measured in arithmetic values, it is obvious that gene interaction exists. In other words, the addition of a gene has an exponential or multiplicative effect, the value of which depends not only on its own activity but also on the presence of other genes in the genotype. By working out simple numerical examples, it may be readily observed that this general type of interaction as expressed on the arithmetic scale is detected by an increasing regression trend similar to that in the epistatic two-gene-pair model.

When, however, the data are transformed to logarithms, the gene action

itself is transformed to the additive scheme and the interaction disappears. This type of interaction is termed metrical bias and results when variations are measured on a different scale than is actually involved in the growth and expression of the character.

The problem then is first to choose whether the arithmetic or logarithmic scale will yield the best fitting model. This is taken up in the last section which discusses the general procedure for fitting gene models.

PARTITIONING OF VARIANCE OF F_1S WITHIN A CONSTANT PARENT GROUP

Let us assume that the proper transformation has been made so that between-loci effects as well as environmental effects are additive and independent. Consider, then, the i^{th} constant parent group, which has CP_i as the constant parent, VP_js where $j=1, 2, \cdots, n$ $(i\neq j)$ as variable parents, and F_{ij}s are F_1s resulting by crossing CP_i on all other variable parents.

The mathematical models involved are as follows:

$$CP_i = g_i + e_i$$

$$VP_j = g_j + e_j$$

$$F_{ij} = \frac{g_i}{2} + \frac{g_j}{2} + h_{ij} + e_{ij}$$

$$MP_{ij} = \frac{g_i}{2} + \frac{g_j}{2} + \frac{e_i}{2} + \frac{e_j}{2}$$

(where MP_{ij} is the midparent between CP_i and VP_j)

CP_i, VP_j, F_{ij} and MP_{ij} represent the phenotypic measurements for the particular trait considered on the i^{th} constant parent (CP_i), the j^{th} variable parent (VP_j), etc. g_i is the basic gene effect, which, in the gene model, is a function of "d" only. e_i, e_{ij} are individual errors. h_{ij} is the dominance effect for the F_{ij} and is a function of "dh" only.

Using the above mentioned models the F_{ij}s of the i^{th} constant parent group, associated variable parents and midparents may be represented as follows:

F_{ij}	VP_j	MP_{ij}
$F_{i1} = \frac{g_i}{2} + \frac{g_1}{2} + h_{i1} + e_{i1}$	$VP_1 = g_1 + e_1$	$MP_{i1} = \frac{g_i}{2} + \frac{g_1}{2} + \frac{e_i}{2} + \frac{e_1}{2}$
$F_{i2} = \frac{g_i}{2} + \frac{g_2}{2} + h_{i2} + e_{i2}$	$VP_2 = g_2 + e_2$	$MP_{i2} = \frac{g_i}{2} + \frac{g_2}{2} + \frac{e_i}{2} + \frac{e_2}{2}$
.	.	.
.	.	.
.	.	.
$F_{in} = \frac{g_i}{2} + \frac{g_n}{2} + h_{in} + e_{in}$	$VP_n = g_n + e_n$	$MP_{in} = \frac{g_i}{2} + \frac{g_n}{2} + \frac{e_i}{2} + \frac{e_n}{2}$

where $i\neq j$.

The following assumptions are used to obtain expected values of variances and covariances.

$E(g_k - \bar{g}.) = E(h_{ij} - \bar{h}_{i.}) = E(e_m) = 0$ where k = i or j; m = i, j or ij

$E(g_k - \bar{g}.)^2 = \sigma_g^2$; $E(h_{ij} - \bar{h}_{i.})^2 = \sigma_h^2$; $E(e_m^2) = \sigma_e^2$

$E(g_k - \bar{g}.)(h_{ik} - \bar{h}_{i.}) = \rho\sigma_g\sigma_h$, expected values of all other cross-products are zero.

$\bar{g}.$ is the mean genotypic value of the parents.

$\bar{h}_{i.}$ is the mean dominance value of the F_1s in the i^{th} c.p. group.

The problem is to partition the variance of F_1s within each constant parent group and estimate the variances σ_g^2 and σ_h^2, and the correlation ρ. The variances and covariances are denoted as follows:

$V_{(F_{ij})}$ = variance of the F_1s within the i^{th} c.p. group.

$V_{(VP)}$ = variance of the variable parent values in the i^{th} c.p. group.

$V_{(MP)}$ = variance of midparental values of the variable parents and the constant parent.

$V_{(F_1-MP)}$ = variance of the differences between the F_1s and their corresponding midparents.

$CV_{(F_1 \times VP)}$ = covariance of the F_1s and their corresponding variable parents.

The expected values of these variances and covariances are as follows:

$$E\{V_{(F_{ij})}\} = (\tfrac{1}{4})\sigma_g^2 + \sigma_h^2 + \rho\sigma_g\sigma_h + \sigma_e^2 \qquad E\{V_{(F_1-MP)}\} = \sigma_h^2 + (\tfrac{5}{4})\sigma_e^2$$

$$E\{V_{(VP)}\} = \sigma_g^2 + \sigma_e^2$$

$$E\{V_{(MP)}\} = (\tfrac{1}{4})\sigma_g^2 + (\tfrac{1}{4})\sigma_e^2 \qquad E\{CV_{(F_1 \times VP)}\} = (\tfrac{1}{2})\sigma_g^2 + \rho\sigma_g\sigma_h.$$

Estimates of the genotypic components are obtained by equating observed and expected values.

$$(\tfrac{1}{4})\hat{\sigma}_g^2 = 1/4\{(V_{VP} - \hat{\sigma}_e^2)\} \qquad \widehat{\rho\sigma_g\sigma_h} = V_{(F_{ij})} - (\tfrac{1}{4})\hat{\sigma}_g^2 - \hat{\sigma}_h^2 - \hat{\sigma}_e^2$$

$$\hat{\sigma}_h^2 = V_{(F_1-MP)} - (\tfrac{5}{4})\hat{\sigma}_e^2.$$

The error component σ_e^2 is directly estimated from the analysis of variance of the replicated experiment and is the variance of parental and F_1 means (experimental error variance divided by the number of observations per parent or F_1).

The genotypic c.p.r. coefficient is:

$$b = \frac{(\tfrac{1}{2})\sigma_g^2 + \rho\sigma_g\sigma_h}{\sigma_g^2} = 1/2 + \rho\frac{\sigma_h}{\sigma_g} = 1/2 + \beta_{h \times g}$$

where $\beta_{h \times g}$ is the regression of the dominance contributions on the parental values.

These expected c.p.r. coefficients are estimated by:

$$\hat{b} = \frac{CV_{(F_1 \times VP)}}{V_{(VP)} - \hat{\sigma}_e^2}.$$

In general, we may assume that the best gene model is that which yields the most accurate prediction of F_1 values from those of its parents. This means that within a constant parent group the most desirable model or scale of measurement would be that which would minimize the mean square of deviations

from the constant parent regression. Thus, the separation of the F_1 variance into (1) the portion attributable to regression and (2) the portion due to deviations from regression, is of considerable importance as an aid in identifying the best fitting model. It is also of interest to point out the relationship existing between these portions and the components of the F_1 variance which have previously been discussed, namely σ_g^2, σ_h^2 and $\rho\sigma_g\sigma_h$. In terms of components the expected genotypic fraction of an F_1 variance attributable to the c.p.r. is as follows:

$$\frac{\left[(\tfrac{1}{2})\sigma_g^2 + \rho\sigma_g\sigma_h\right]^2}{\sigma_g^2} = (\tfrac{1}{4})\sigma_g^2 + \rho\sigma_g\sigma_h + \rho^2\sigma_h^2,$$

and the expected fraction due to deviations from regression is:

$$(1 - \rho^2)\sigma_h^2.$$

It is obvious that if $\rho = 0$ then the basic gene variance and the mean square attributable to regression are equivalent; also the dominance component and deviations from regression are identical. If $\rho \neq 0$ this relationship does not hold.

In the gene model in which dominance exists but no between-loci-interaction, and all of the dominance effects are in the same direction of character expression, one can use variance components to estimate the value of "h." This necessitates theoretically the inclusion among the set of parents, of a completely recessive or completely dominant inbred or both. Then with their c.p. groups the dominance value to be associated with the gene model may be estimated as follows:

$$\hat{h} = \sqrt{\frac{\hat{\sigma}_h^2}{\hat{\sigma}_g^2}}.$$

Overdominance is indicated when $\hat{\sigma}_h^2$ is greater than $\hat{\sigma}_g^2$.

Tests of significance for the dominance component may be obtained by the following F test;

$$F = \frac{V_{(F_1-MP)}}{(\tfrac{5}{4})\hat{\sigma}_e^2}$$

with $(n-1)$ degrees of freedom for the numerator where "n" is the number of F_1s in the c.p. group, and degrees of freedom for the denominator equal the degrees of freedom on which the experimental variance is based.

USE OF POTENCE RATIOS FOR CALCULATING APPROXIMATE MAGNITUDE OF DOMINANCE

In addition to the use of regression trends and variance component analyses, one can approximate a dominance value from the means of the non-segregating populations.

The proper procedure is to determine the appropriate scale which will transform between-loci gene action into the additive scheme and then compare

the F_1 means with mid-parental values to determine a "potence" value (h_p) as follows:

$$h_p = \frac{F_1 - M_p}{|P - M_p|}$$

where: M_p = midparent
P = parent (extreme)

This potence value will sufficiently approximate an average dominance value for the models discussed if one of the parents is at either extreme of character expression.

GENERAL PROCEDURE FOR FITTING GENE MODELS

The following steps are listed as the general procedure which may be used to fit the data to a specific gene model.

1. From arithmetic P_1 and F_1 means various statistics are calculated, such as (a) c.p.r. coefficients, (b) variance components, (c) deviations from the c.p.r., and (d) h_p values for the F_1s.
2. If the regression trend is decreasing, then positive dominance is assumed with arithmetic gene action. Amount of dominance is estimated by severity of regression trend, components, and h_p values.
3. If the regression trend is increasing, then the first problem arising is to distinguish between, (1) arithmetically cumulative action with negative dominance, and (2) logarithmically cumulative gene action with or without dominance.

This may be done by transforming the data to logarithms and comparing the various statistical values with the arithmetic analysis. Then,

(1) If on transformation to logarithms the c.p.r. trend is drastically reduced in comparison with the trend of the arithmetic c.p.r.s, this indicates that the arithmetic trend is due to logarithmic interaction (metrical bias) which can be removed by transformation.

(2) If on transformation to logarithms the deviations from regression are greatly reduced, this indicates that the deviations from regression are due mostly to metrical bias which disappears with transformation, and the best model would involve the logarithmic scale.

(3) If h_p values are irregular with arithmetic data and on transformation to logarithms become much more uniform, then this indicates that the better gene model is logarithmic.

(4) If in arithmetic analysis the c.p.r.s are all positive but increase sharply with largest valued c.p.r. considerably over $+1.0$, then this indicates logarithmic gene action. With negative dominance the c.p.r. coefficients should never exceed $+1.0$ (disregarding errors associated with c.p.r.'s) except in the case of overdominance, and usually one c.p.r. should have a negative value.

These four criteria may be used effectively to differentiate logarithmic from arithmetic gene action. When the proper basic gene action is decided upon, then estimates of amount of dominance are obtained by using the appropriate data (arithmetic or logarithmic) from the following statistics, (a) b_2, severity of regression trend, (b) components of variance, and (c) h_p values.

The development of the c.p.r. technique is obviously in its preliminary stages. However, when the trends of c.p.r. coefficients are used in combination with other approaches, a satisfactory analysis of the average gene action may be obtained. It should be emphasized again that as wide a range of character expression as possible is desirable. At least one extreme or both should be included among the parents so that regression trends, potence and component ratios are more reliable.

The real test of a new technique is in its practical application. HULL (1948) has reported some results in corn. Probably the most extensive use has been in the analysis of some ten characteristics in tomatoes (GRIFFING 1948). Here considerable success was obtained in fitting gene models to the data by the techniques outlined in this paper.

SUMMARY

An approach to the analysis of quantitative inheritance by use of the constant parent regression method is outlined. Variance components are also considered as an aid in understanding the relation between constant parent regressions and dominance deviations, and in evaluating the magnitude of dominance.

The approach is essentially that of constructing different gene models and then choosing the model that fits the data most accurately as determined by various statistics. As an introduction to the techniques, two-gene-pair models are considered first involving (1) additive gene action with dominance and no epistasis, and (2) gene-interaction involving both dominance and epistasis. The models are extended to "n" gene pairs with logarithmic gene action being the only type of gene interaction discussed. A procedure for differentiating the two basic models, arithmetic and logarithmic, is outlined together with an estimation of the amount of dominance on the closest fitting scale.

ACKNOWLEDGEMENTS

The author is indebted to DRS. JOHN W. GOWEN, SEWALL WRIGHT, and FRED H. HULL for constructive criticisms of the manuscript and to PROFESSOR OSCAR KEMPTHORNE for valuable aid in the statistical aspects of the paper.

LITERATURE CITED

BURDICK, A. B., 1949 Heterosis in the tomato; estimation of the relative contribution of epistasis and dominance. Unpublished Ph.D Dissertation, University of California.

FISHER, R. A., 1918 The correlation between relatives, on the supposition of Mendelian inheritance. Trans. roy. Soc. Edinb. **52:** 399–433.

GRIFFING, J. B., 1948 Statistical analysis of gene action in components of tomato yield. Unpublished Ph.D Dissertation, Ames, Iowa State College, Ames, Iowa.

HULL, F. H., 1946 Maize Genetics Cooperation News Letter 20.

1947a Theoretical regression of F_n on homozygous parents with additive or complementary gene action. Genetics **32:** 90–91.

1947b Maize Genetics Cooperation News Letter 21.

1948 Maize Genetics Cooperation News Letter 22.

MATHER, K., 1949 Biometrical Genetics, pp. 82–97. Dover Publ. Inc., London.

RASMUSSON, J., 1933 A contribution to the theory of quantitative character inheritance. Hereditas **18:** 245–261.