

A GENERALISED TREATMENT OF THE USE OF DIALLEL CROSSES IN QUANTITATIVE INHERITANCE

BRUCE GRIFFING

Department of Genetics, Cambridge

Received 15.vii.55

IN problems of quantitative inheritance it has been customary to make genetic analysis of populations using covariances between relatives, the method outlined in the classical paper by Fisher (1918). Associated with this approach is the notion of partitioning the total genotypic variance into additive and non-additive genetic components. The importance of this treatment of quantitative inheritance to the genetical theory of natural selection has been made clear by Fisher (1930), and the practical applications of these concepts to plant and animal improvement have been increasingly realised.

In recent years techniques involving "diallel crosses" have been used in problems which concern quantitative inheritance. The methods used and problems attacked by this technique have been diverse. Sprague and Tatum (1942), Henderson (1948, 1952), and Griffing (1953) have defined and applied the notions of general and specific combining ability to plant and animal experimental material using a variety of diallel crossing methods, but without an exact generalised genetic interpretation of the combining ability effects and variances. Hull (1946, 1952), Griffing (1950), Jinks (1954) and Hayman (1954*a*, 1954*b*), again using diallel crossing systems, have given procedures for estimating other genetic parameters in terms of restricted gene models.

The present paper is an attempt to give, in terms of population genetics, a generalised treatment of the major problems which occur with the use of diallel crosses, to integrate wherever possible the already existing techniques, and to show the relationship of the diallel crossing method to Fisher's method of covariances between relatives as expressed in terms of additive and non-additive genetic variances.

We shall use the term "diallel crosses" to describe a procedure in which a set of p inbred lines are chosen and crosses among these lines are made. There are a maximum of p^2 possible crosses, which can be represented by a $p \times p$ matrix with elements x_{ij} such that x_{ii} represents the i^{th} inbred, $x_{ij}(i \pm j)$ represents the F_1 between the i^{th} and j^{th} inbreds, and x_{ji} represents its reciprocal. Thus, the p^2 combinations can be divided conveniently into three groups: (1) the p inbreds themselves; (2) one group of $\frac{p(p-1)}{2}$ F_1 's; (3) a group of $\frac{p(p-1)}{2}$ reciprocal F_1 's.

Experimental methods utilising diallel crosses may vary, depending

upon whether or not the inbreds and/or the reciprocal F_1 's are included. With this basis for classification, there are four possible methods: (1) inbreds, one set of F_1 's, and reciprocal F_1 's are included (all p^2 combinations); (2) inbreds and one set of F_1 's are included but reciprocal F_1 's are not ($\frac{p(p+1)}{2}$ combinations); (3) one set of F_1 's and reciprocals are included but not the inbreds ($p(p-1)$ combinations); (4) one set of F_1 's, but neither inbreds nor reciprocal F_1 's are included ($\frac{p(p-1)}{2}$ combinations).

Theoretically the important difference among the methods is whether or not the inbreds are included. Since the term "diallel" has in the past been associated by other authors with the methods which include inbreds, we shall use the term "modified diallel" to designate methods (3) and (4), which do not include the parents. It is with these methods that we shall be most concerned.

Once the experimental data has been obtained by use of one of the above methods, the problem is to estimate certain genetic parameters of the population from which the inbreds are derived. The problem may be considered in three major parts: (1) an examination of assumptions and conditions which are necessary if valid inductive inferences are to be made from the experimental material about the parent population; (2) a specification of estimable genetic parameters and their genetic interpretation using a generalised gene model; (3) a description of the available methods by which experimental data are used to test hypotheses and estimate population parameters.

I. ASSUMPTIONS AND CONDITIONS NECESSARY FOR VALID INDUCTIVE INFERENCE

When a set of lines are used in a diallel crossing system for the purpose of estimating genetic parameters of the population from which the lines were derived, a chain of assumptions which relates the experimental material to the original population must be made. Starting with the experimental material, we must assume that the set of inbreds is a random sample from a population of inbred lines which in turn were derived from the parent population by means of an inbreeding system free from forces which change gene frequency.

In addition to making these assumptions we must, before we can completely specify the assumptions and conditions necessary for valid inductive inferences, (1) determine the restrictions which must be placed on the parent population to ensure that it can be truly represented by diallel crossing among its derived inbreds, and (2) determine which of the four possible diallel methods gives unbiased estimates of the population parameters.

We shall investigate the first problem by considering the particular case in which the limiting frequencies of the homozygous genotypes

are derived by imposing a self-mating system on an arbitrary population. This population consists of genotypes involving two alleles at each of two loci which may or may not be linked, *i.e.* recombination value $a \leq \frac{1}{2}$. We shall start with arbitrary initial genotypic frequencies and obtain the limiting frequencies of homozygous genotypes using the methods described by Fisher (1949). The self-mating system is employed because it involves a minimum number of mating types and because the mating type frequencies are equivalent to the genotypic frequencies in any generation. We are generalising a solution given by Nelder (1952).

TABLE 1
*Generation matrix with self-fertilisation relating the genotypic frequencies in one generation with those in the next generation **

Genotype	Frequency	q_0	r_0	s_0	t_0	u_0	v_0	w_0	x_0	y_0	z_0
A_1B_1/A_1B_1	q_1	1	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{b^2}{4}$	$\frac{a^2}{4}$
A_2B_2/A_2B_2	r_1	0	1	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{b^2}{4}$	$\frac{a^2}{4}$
A_1B_2/A_1B_2	s_1	0	0	1	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0	$\frac{a^2}{4}$	$\frac{b^2}{4}$
A_2B_1/A_2B_1	t_1	0	0	0	1	0	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{a^2}{4}$	$\frac{b^2}{4}$
A_1B_2/A_1B_1	u_1	0	0	0	0	$\frac{1}{2}$	0	0	0	$\frac{ab}{2}$	$\frac{ab}{2}$
A_2B_1/A_2B_1	v_1	0	0	0	0	0	$\frac{1}{2}$	0	0	$\frac{ab}{2}$	$\frac{ab}{2}$
A_1B_2/A_2B_2	w_1	0	0	0	0	0	0	$\frac{1}{2}$	0	$\frac{ab}{2}$	$\frac{ab}{2}$
A_2B_2/A_2B_1	x_1	0	0	0	0	0	0	0	$\frac{1}{2}$	$\frac{ab}{2}$	$\frac{ab}{2}$
A_1B_1/A_2B_2	y_1	0	0	0	0	0	0	0	0	$\frac{b^2}{2}$	$\frac{a^2}{2}$
A_1B_2/A_2B_1	z_1	0	0	0	0	0	0	0	0	$\frac{a^2}{2}$	$\frac{b^2}{2}$

* "a" is the recombination fraction such that $a+b = 1$.

The generation matrix (denoted as A) relating the frequencies in one generation to those in the next is given in table 1. Solution of the determinantal equation $|A - \lambda I| = 0$ yields the latent roots and their corresponding principal components as given in table 2. Associated with each of the multiple roots $\lambda = 1, \frac{1}{2}$ are four pair wise orthogonal vectors which taken together form a 4×10 matrix of rank four. This matrix premultiplies its appropriate λ matrix to give a 4×10 zero matrix.

Using these latent roots and principal components we obtain the limiting frequencies for the homozygous genotypes as $n \rightarrow \infty$. These are given in table 3.

If the initial genotypic frequencies in the parent population are those that would occur under equilibrium conditions with random

mating, and if the gene frequencies for A_1 , A_2 , B_1 and B_2 are p_1 , p_2 , q_1 and q_2 respectively, then

$$\begin{aligned} &(\text{freq. of } A_1A_1B_1B_1) \xrightarrow{n \rightarrow \infty} p_1q_1, \quad (\text{freq. of } A_2A_2B_2B_2) \xrightarrow{n \rightarrow \infty} p_2q_2, \\ &(\text{freq. of } A_1A_1B_2B_2) \xrightarrow{n \rightarrow \infty} p_1q_2, \quad \text{and} \quad (\text{freq. of } A_2A_2B_1B_1) \xrightarrow{n \rightarrow \infty} p_2q_1. \end{aligned}$$

TABLE 2
Latent roots and their associated principal components

Latent roots	Principal components										
		q	r	s	t	u	v	w	x	y	z
$\lambda = 1$	$\left\{ \begin{array}{l} A = \\ B = \\ C = \\ D = \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 1 \\ 1 \\ (1+2a) \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ -1 \\ -1 \\ (1+2a) \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 1 \\ -1 \\ -(1+2a) \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ -1 \\ 1 \\ -(1+2a) \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 1 \\ 0 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 0 \\ 1 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 0 \\ -1 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ -1 \\ 0 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 0 \\ 0 \\ (b-a) \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 0 \\ 0 \\ -(b-a) \end{array} \right.$
$\lambda = \frac{1}{2}$	$\left\{ \begin{array}{l} E = \\ F = \\ G = \\ H = \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 1 \\ 0 \\ 1 \\ -1 \end{array} \right.$	$\left\{ \begin{array}{l} -1 \\ 0 \\ 1 \\ -1 \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ 1 \\ 1 \\ 1 \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ -1 \\ 1 \\ 1 \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ 0 \\ 2 \\ 0 \end{array} \right.$	$\left\{ \begin{array}{l} 0 \\ 0 \\ 2 \\ 0 \end{array} \right.$
$\lambda = \frac{1-2ab}{2}$	$I =$	0	0	0	0	0	0	0	0	1	1
$\lambda = \frac{b-a}{2}$	$J =$	0	0	0	0	0	0	0	0	1	-1

It is obvious that the genotypic frequencies occurring in the original population can be obtained by diallel crossing among the derived inbreds (if their limiting frequencies are taken into consideration).

TABLE 3
Limiting frequencies of homozygous genotypes

Genotype	Frequency
A_1B_1/A_1B_1	$q_n \xrightarrow{n \rightarrow \infty} q_0 + \frac{1}{2}u_0 + \frac{1}{2}v_0 + \left[\frac{1}{2(1+2a)} \right] y_0 + \left[\frac{a}{1+2a} \right] z_0$
A_2B_2/A_2B_2	$r_n \xrightarrow{n \rightarrow \infty} r_0 + \frac{1}{2}w_0 + \frac{1}{2}x_0 + \left[\frac{1}{2(1+2a)} \right] y_0 + \left[\frac{a}{1+2a} \right] z_0$
A_1B_2/A_1B_2	$s_n \xrightarrow{n \rightarrow \infty} s_0 + \frac{1}{2}u_0 + \frac{1}{2}w_0 + \left[\frac{a}{1+2a} \right] y_0 + \left[\frac{1}{2(1+2a)} \right] z_0$
A_2B_1/A_2B_1	$t_n \xrightarrow{n \rightarrow \infty} t_0 + \frac{1}{2}v_0 + \frac{1}{2}x_0 + \left[\frac{a}{1+2a} \right] y_0 + \left[\frac{1}{2(1+2a)} \right] z_0$

It is equally clear that with populations which cannot be considered in equilibrium with random mating, diallel crossing among the inbreds will not generate the original population, and thus valid inferences are not possible using this technique.

With regard to the second problem, the author is indebted to Sir Ronald Fisher for pointing out that when random samples are drawn from the population of inbreds it is necessary to employ the "modified diallel" crossing systems to obtain unbiased estimates of the population parameters. We must defer to the appendix a demonstration of this fact since we consider parameters not yet defined and estimating procedures not yet described.

2. SPECIFICATION AND INTERPRETATION OF GENETIC PARAMETERS

Since the diallel analysis is applicable only to random mating populations in equilibrium, the following discussion will be restricted to such populations. However, in other genetic respects we wish to consider a completely general situation. Thus, we want the genetic analysis to deal with an arbitrary number of alleles at each of an arbitrary number of loci. We want, in the final analysis, to interpret the genetic parameters in terms of a completely generalised gene model which will allow any magnitude of additive genetic, dominance, and epistatic effects.

The genetic parameters of interest are the additive and non-additive genetic components of the parent population genotypic variance. We assume that the population phenotypic variance is partitionable into various genotypic and environmental components as follows :

$$\sigma^2_P = \sigma^2_G + \sigma^2_E, \text{ and since } \sigma^2_G = \sigma^2_A + \sigma^2_{NA}, \text{ then } \sigma^2_P = \sigma^2_A + \sigma^2_{NA} + \sigma^2_E$$

where

- σ^2_P = population phenotypic variance.
- σ^2_G = population genotypic variance.
- σ^2_A = additive genetic variance.
- σ^2_{NA} = non-additive genetic variance.

σ^2_{NA} may be partitioned into dominance and epistatic variance, as will be shown later.

The concept of partitioning the total genotypic variance into various genetic compounds is due to Fisher (1918) and used by him in evaluating covariances between relatives. Fisher (1918) also originated the completely generalised genetic model which will be used by the present author in connection with these components. A model developed by Kempthorne (1955), which is adapted to random mating populations and which allows a complete orthogonal partitioning of the total epistatic variance, will also be used.

The estimation of the additive and non-additive genetic components will be made from the experimental material in terms of general and specific combining ability variances. Therefore, it is necessary to start with random mating populations and (1) review the definitions of additive and non-additive genetic variances, (2) define the combining ability variances, and (3) show the relationship

between the two sets of parameters in terms of a completely generalised gene model.

In this section we shall use the summation convention in which repeated suffixes (in those terms which involve gene or genotypic frequencies) imply summation over these suffixes. For example,

$$p_i p_j d_{ij} = \sum_i \sum_j p_i p_j d_{ij} \text{ and } p_i d_{ij} = \sum_i p_i d_{ij} \text{ etc.}$$

(a) Definition of additive and non-additive genetic effects and variances

The definitions will be reviewed first for genotypes at one locus only, and then generalised to include genotypes at two or more loci.

TABLE 4
Genotypic values and frequencies for a random mating population of genotypes derived from alleles at one locus *

	A ₁	A ₂	A ₃	—	—	—	A _m	Marginal frequencies
A ₁	d_{11}	d_{12}	d_{13}	—	—	—	d_{1m}	p_1
A ₂	d_{21}	d_{22}	d_{23}	—	—	—	d_{2m}	p_2
A ₃	d_{31}	d_{32}	d_{33}	—	—	—	d_{3m}	p_3
—	—	—	—	—	—	—	—	—
A _m	d_{m1}	d_{m2}	d_{m3}	—	—	—	d_{mm}	p_m
Marginal frequency	p_1	p_2	p_3	—	—	—	p_m	1

* Genotypic frequencies are the product of appropriate marginal frequencies; i.e. frequency of d_{ij} is $p_i p_j$.

A notation similar to that used by Kempthorne (1955) will be used.

I. DEFINITIONS FOR GENOTYPES AT ONE LOCUS. Consider m alleles A_1, A_2, \dots, A_m with frequencies p_1, p_2, \dots, p_m respectively. Genotypic values, measured as deviations from the weighted population mean, together with their associated frequencies are given in table 4. The genotypic value of $A_i A_j$ is defined to be d_{ij} , and such that $p_i p_j d_{ij} = 0$, $i, j = 1, 2, \dots, m$.

We define the additive genetic effect of A_i as $a_i = p_j d_{ij}$. These effects are subject to the restriction $p_i a_i = 0$. We define the non-additive (dominance) effect of the $A_i A_j$ genotype as $\delta_{ij} = d_{ij} - a_i - a_j$ and these effects are subject to the following restrictions: $\delta_{ij} = \delta_{ji}$ and $p_i \delta_{ij} = 0$ for all j . The δ_{ij} may be thought of as the interaction between the A_i and A_j alleles.

In this way we derive the following model which represents the genotypic value in terms of additive and non-additive genetic effects $d_{ij} = a_i + a_j + \delta_{ij}$. The elements in this model have mean zero and are uncorrelated in the population.

The total genotypic variance is partitionable into variances due to additive and non-additive genetic components as follows :

$$p_i p_j d_{ij}^2 = 2p_i a_i^2 + p_i p_j \delta_{ij}^2.$$

These variances may be represented symbolically as $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$ where $\sigma_G^2 = p_i p_j d_{ij}^2$ is the total genotypic variance, $\sigma_A^2 = 2p_i a_i^2$ is the additive genetic variance, and $\sigma_D^2 = p_i p_j \delta_{ij}^2$ is the non-additive (dominance) genetic variance.

We note that the a 's can be obtained by the least squares procedure. Assuming that the genotypic values can be expressed by the model $d_{ij} = a_i + a_j + \delta_{ij}$ we choose a 's such that the sum of the weighted squared deviations from the linear model ($a_i + a_j$) is a minimum. The a 's so defined, together with the associated restrictions, are identical to those defined above.

TABLE 5
Genotypic values and frequencies for a random mating population of genotypes derived from alleles at two loci *

	A ₁ B ₁	A ₁ B ₂	A ₁ B ₃	—	—	—	A _m B _n	
A ₁ B ₁	$1_1 d_{11}$	$1_1 d_{12}$	$1_1 d_{13}$	—	—	—	$1_m d_{1n}$	$p_1 q_1$
A ₁ B ₂	$1_1 d_{21}$	$1_1 d_{22}$	$1_1 d_{23}$	—	—	—	$1_m d_{2n}$	$p_1 q_2$
—	—	—	—	—	—	—	—	—
A _m B _n	$m_1 d_{n1}$	$m_1 d_{n2}$	$m_1 d_{n3}$	—	—	—	$m_m d_{nn}$	$p_m q_n$
Marginal frequency	$p_1 q_1$	$p_1 q_2$	$p_1 q_3$	—	—	—	$p_m q_n$	1

* Genotypic frequencies are the product of appropriate marginal frequencies; i.e. frequency of $_{ij}d_{kl}$ is $p_i p_j q_k q_l$.

2. DEFINITIONS FOR GENOTYPES AT TWO OR MORE LOCI. We shall consider two genetic models under this heading: (a) a model in which loci effects are additive, and (b) a model in which the loci effects are not additive.

(a) *Additive loci effects.* When the loci effects are additive the total genotypic variance is the sum of the separate genotypic variances for each locus. To illustrate, consider the two loci case in which there are alleles A₁, A₂, ... A_m with frequencies $p_1, p_2, \dots p_m$ at one locus, and alleles B₁, B₂, ... B_n with frequencies $q_1, q_2, \dots q_n$ at the other locus.

Genotypic values, measured as deviations from the weighted population mean, together with their associated frequencies are given in table 5. The genotypic value of A_iA_jB_kB_l is defined to be $_{ij}d_{kl}$ and such that $p_i p_j q_k q_l \text{ }_{ij}d_{kl} = 0$. In the following notation, suffixes placed before the constants pertain to the A locus and those after pertain to the B locus.

We define the effects at the A locus to be :

$$\begin{aligned} {}_i\alpha &= p_i q_k q_l {}_{ij}d_{kl} && = \text{additive genetic effect of } A_i, \\ {}_j\alpha &= p_i q_k q_l {}_{ij}d_{kl} && = \text{additive genetic effect of } A_j, \\ {}_{ij}\delta &= q_k q_l {}_{ij}d_{kl} - {}_i\alpha - {}_j\alpha && = \text{interaction (dominance) of } A_i \text{ and } A_j. \end{aligned}$$

The restrictions on these constants are

$$p_i {}_i\alpha = 0, {}_{ij}\delta = {}_{ji}\delta \text{ and } p_i {}_{ij}\delta = 0.$$

Similarly, we define the effects at the B locus to be :

$$\begin{aligned} \alpha_k &= p_i p_j q_l {}_{ij}d_{kl} && = \text{additive genetic effect of } B_k, \\ \alpha_l &= p_i p_j q_k {}_{ij}d_{kl} && = \text{additive genetic effect of } B_l, \\ \delta_{kl} &= p_i p_j {}_{ij}d_{kl} - \alpha_k - \alpha_l && = \text{interaction (dominance) of } B_k \text{ and } B_l. \end{aligned}$$

The restrictions on these constants are

$$q_k \alpha_k = 0, \delta_{kl} = \delta_{lk} \text{ and } q_k \delta_{kl} = 0.$$

The model is thus derived to represent the genotypic value of $A_i A_j B_k B_l$ in terms of additive and non-additive genetic effects and may be written

$${}_{ij}d_{kl} = {}_i\alpha + {}_j\alpha + {}_{ij}\delta + \alpha_k + \alpha_l + \delta_{kl}.$$

After the total genotypic variance has been partitioned into additive and non-additive components for each locus the loci components may be summed, *i.e.*

$$p_i p_j q_k q_l {}_{ij}d_{kl}^2 = 2p_i {}_i\alpha^2 + 2q_k \alpha_k^2 + p_i p_j {}_{ij}\delta^2 + q_k q_l \delta_{kl}^2,$$

or symbolically $\sigma_G^2 = \sigma_A^2 + \sigma_{NA}^2$ where

$$\begin{aligned} \sigma_G^2 &= p_i p_j q_k q_l {}_{ij}d_{kl}^2 = \text{total genotypic variance,} \\ \sigma_A^2 &= 2p_i {}_i\alpha^2 + 2q_k \alpha_k^2 = \text{total additive genetic variance, and} \\ \sigma_{NA}^2 &= \sigma_D^2 = p_i p_j {}_{ij}\delta^2 + q_k q_l \delta_{kl}^2 = \text{total non-additive genetic variance.} \end{aligned}$$

When loci effects are additive the extension to any number of loci each having an arbitrary number of alleles presents no difficulty.

(b) *Loci effects not additive.* Fisher (1918) introduced epistatic constants to account for the interaction of genotypes at one locus with those of a second locus. The epistatic constants and their frequencies for genotypes formed by alleles at each of two loci are given in table 6. The epistatic constants are defined as

$${}_{ij}\epsilon_{kl} = {}_{ij}d_{kl} - {}_i\alpha - {}_j\alpha - {}_{ij}\delta - \alpha_k - \alpha_l - \delta_{kl}$$

and by rewriting this equation we arrive at the representation of the genotypic value of $A_i A_j B_k B_l$ by the following completely general gene model :

$${}_{ij}d_{kl} = {}_i\alpha + {}_j\alpha + {}_{ij}\delta + \alpha_k + \alpha_l + \delta_{kl} + {}_{ij}\epsilon_{kl}.$$

The total genotypic variance is partitionable as follows :

$$p_i p_j q_k q_l {}_{ij}d_{kl}^2 = 2p_i {}_i\alpha^2 + 2q_k \alpha_k^2 + p_i p_j {}_{ij}\delta^2 + q_k q_l \delta_{kl}^2 + p_i p_j q_k q_l {}_{ij}\epsilon_{kl}^2$$

which may be written symbolically as

$$\sigma_G^2 = \sigma_A^2 + \sigma_{NA}^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2.$$

σ_{NA}^2 is broken down into variances due to two different types of interaction, namely :

$$\begin{aligned} \sigma_D^2 &= p_i p_j \delta^2 + q_k q_l \delta_{kl}^2 = \text{interaction between alleles at the same locus, and} \\ \sigma_I^2 &= p_i p_j q_k q_l \epsilon_{ijkl}^2 = \text{interaction between genotypes at different loci.} \end{aligned}$$

Imposing the restrictions

$$i j \epsilon_{k l} = j i \epsilon_{k l}, i i \epsilon_{k l} = i j \epsilon_{l k}, p_i p_j i i \epsilon_{k l} = 0, \text{ and } q_k q_l i i \epsilon_{k l} = 0$$

on the 16 constants in table 6, results in four independent constants

TABLE 6
*Epistatic constants and their frequencies for genotypes derived from two alleles at each of two loci **

	B ₁ B ₁	B ₁ B ₂	B ₂ B ₁	B ₂ B ₂	Marginal frequency
A ₁ A ₁	11 ϵ_{11}	11 ϵ_{12}	11 ϵ_{21}	11 ϵ_{22}	p_1^2
A ₁ A ₂	12 ϵ_{11}	12 ϵ_{12}	12 ϵ_{21}	12 ϵ_{22}	$p_1 p_2$
A ₂ A ₁	21 ϵ_{11}	21 ϵ_{12}	21 ϵ_{21}	21 ϵ_{22}	$p_2 p_1$
A ₂ A ₂	22 ϵ_{11}	22 ϵ_{12}	22 ϵ_{21}	22 ϵ_{22}	p_2^2
Marginal frequency	q_1^2	$q_1 q_2$	$q_2 q_1$	q_2^2	1

* The frequency of the constant $i j \epsilon_{k l}$ is $p_i p_j q_k q_l$.

which we arbitrarily choose to be $i i \epsilon_{k k}$. The genotypic variance may then be rewritten in the following form given by Fisher (1918).

$$\begin{aligned} p_i p_j q_k q_l i j d_{kl}^2 &= 2p_i i a^2 + 2q_k a_k^2 + p_i p_j i j \delta^2 + q_k q_l \delta_{kl}^2 \\ &+ \frac{1}{4p_1 p_2 q_1 q_2} \left[z^2 + 2p_1^3 q_1 q_2 (q_1 11\epsilon_{11} + q_2 11\epsilon_{22})^2 + 2p_1 p_2 q_1^3 (p_1 11\epsilon_{11} + p_2 22\epsilon_{11})^2 \right. \\ &\left. + 2p_1 p_2 q_2^3 (p_1 11\epsilon_{22} + p_2 22\epsilon_{22})^2 + 2p_2^3 q_1 q_2 (q_1 22\epsilon_{11} + q_2 22\epsilon_{22})^2 \right], \end{aligned}$$

where

$$z = (p_1^2 q_1^2 11\epsilon_{11} - p_1^2 q_2^2 11\epsilon_{22} - p_2^2 q_1^2 22\epsilon_{11} + p_2^2 q_2^2 22\epsilon_{22}).$$

(b) *Definition of general and specific combining ability effects and variances and their relationship to additive and non-additive genetic effects and variances*

In an analysis of diallel crossing experiments it is convenient to redefine the genetic model in terms of general combining ability (g.c.a.) and specific combining ability (s.c.a.) effects and variances.

To give a general definition of g.c.a. and s.c.a. effects we need only consider the representation of the zygotic array of a random mating population in equilibrium as the square of the gametic array.

This representation may take the form of a two-way multiplication table (for example table 5) in which the multiplying elements forming the row and column margins are the gametes together with their respective frequencies, and the products in the table represent the genotypes and their frequencies. We define, then, the g.c.a. effect associated with the i^{th} gamete, say, as the weighted mean value of genotypes in the i^{th} row of the multiplication table. Thus, if the value of the genotype in the ij^{th} cell of the multiplication table is d_{ij} (measured from the population mean) with frequency $f_i f_j$, the g.c.a. effect for the i^{th} gamete is $g_i = f_j d_{ij}$. The s.c.a. effect for the ij^{th} genotype is $s_{ij} = d_{ij} - g_i - g_j$.

By use of these effects we can represent the ij^{th} genotypic value as a linear function of the average value associated with each of the gametes which combined to form the genotype plus the interaction effect between these gametes, *i.e.* $d_{ij} = g_i + g_j + s_{ij}$.

We have defined the combining ability effects in terms of gametes. When homozygous lines are used in diallel crossing techniques the combining ability effects are often defined in terms of the inbred lines. In this case the terms are equivalent since there is a one-to-one correspondence between inbreds and gametes.

To show the exact relationships between combining ability and additive and non-additive genetic effects we shall first consider populations of genotypes involving only two loci and later generalise to populations of genotypes which involve any number of loci.

I. POPULATIONS OF GENOTYPES INVOLVING TWO LOCI. We shall again consider the two situations in one of which the loci effects are additive and in the other not additive.

(a) *Additive loci effects.* Let us consider a population of genotypes derivable from an arbitrary number of alleles at each of two loci as given in table 5.

We define the g.c.a. effect associated with the $A_i B_k$ gamete (or with the inbred $A_i A_i B_k B_k$) as ${}_i g_k = p_j q_l {}_{ij} d_{kl}$. By substituting

$${}_{ij} d_{kl} = {}_i \alpha + {}_j \alpha + {}_{ij} \delta + \alpha_k + \alpha_l + \delta_{kl}$$

in the equation for ${}_i g_k$ and employing the restrictions associated with the elements in the model, we are able to determine the exact relationships between g.c.a. effects and additive genetic effects, which are as follows :

$$\left. \begin{aligned} {}_i g_k &= {}_i \alpha + \alpha_k \\ {}_j g_l &= {}_j \alpha + \alpha_l \end{aligned} \right\} \text{with restrictions } p_v q_w {}_v g_w = 0.$$

The s.c.a. effects are defined as ${}_{ij} s_{kl} = {}_{ij} d_{kl} - {}_i g_k - {}_j g_l$ and when appropriate substitutions are made we find that the s.c.a. effects are a function of the dominance effects, *i.e.* ${}_{ij} s_{kl} = {}_{ij} \delta + \delta_{kl}$. The restrictions are ${}_{ii} s_{kl} = {}_{ii} s_{lk}$, ${}_{ij} s_{kk} = {}_{ii} s_{kk}$, ${}_{ij} s_{kl} = {}_{ji} s_{lk}$, and $p_j q_l {}_{ij} s_{kl} = 0$ for each ik .

The genotypic value of $A_i A_j B_k B_l$ in terms of g.c.a. and s.c.a.

effects is $_{ij}d_{kl} = _i g_k + _i g_l + _{ij} s_{kl}$. The total genotypic variance can be partitioned into g.c.a. and s.c.a. variances as follows :

$$p_i p_j q_k q_l \text{ } _{ij}d_{kl}^2 = 2p_i q_k \text{ } _i g_k^2 + p_i p_j q_k q_l \text{ } _{ij} s_{kl}^2.$$

Symbolically this can be represented as $\sigma^2_G = 2\sigma^2_{g.c.a.} + \sigma^2_{s.c.a.}$ and hence the following equivalence exists between combining ability variances on the one hand and additive and non-additive variances on the other : $\sigma^2_A = 2\sigma^2_{g.c.a.}$, and $\sigma^2_D = \sigma^2_{s.c.a.}$.

(b) *Non-additive loci effects.* Consider the more restricted population of genotypes derived from two alleles at each of two loci. The combining ability effects are defined as in (1a). However, the additional complication of dual epistacy necessitates the following genetic model : $_{ij}d_{kl} = _i a + _j a + _{ij} \delta + \alpha_k + \alpha_l + \delta_{kl} + _{ij} \epsilon_{kl}$, and, therefore, the g.c.a. effects now include a linear function of the epistatic effects as follows :

$$\begin{aligned} 1g_1 &= 1a + \alpha_1 + \frac{I}{4p_1q_1} z & 2g_1 &= 2a + \alpha_1 - \frac{I}{4p_2q_1} z \\ 1g_2 &= 1a + \alpha_2 - \frac{I}{4p_1q_2} z & 2g_2 &= 2a + \alpha_2 + \frac{I}{4p_2q_2} z \end{aligned}$$

where $z = (p_1^2 q_1^2 \text{ }_{11} \epsilon_{11} - p_1^2 q_2^2 \text{ }_{11} \epsilon_{22} - p_2^2 q_1^2 \text{ }_{22} \epsilon_{11} + p_2^2 q_2^2 \text{ }_{22} \epsilon_{22})$.

Twice the g.c.a. variance contains not only all of the additive genetic variance but also a portion of the epistatic variance, *i.e.*

$$2\sigma_{g.c.a.}^2 = 2p_i \text{ } _i a^2 + 2q_k \alpha_k^2 + \frac{I}{8p_1 p_2 q_1 q_2} z^2$$

(z is a function of the linear \times linear interaction contrast in a 3² factorial experiment—see Cockerham (1954)).

The s.c.a. variance (residual genotypic variance) includes all of the dominance and the remaining epistatic variance.

Thus, when dual epistacy is taken into consideration, the partitioning of the population genotypic variance is as follows :

$$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I \text{ OR } \sigma^2_G = 2\sigma_{g.c.a.}^2 + \sigma_{s.c.a.}^2.$$

However, $2\sigma_{g.c.a.}^2 = \sigma^2_A + \frac{1}{4}(\text{linear} \times \text{linear portion of } \sigma^2_I)$, and $\sigma_{s.c.a.}^2 = \sigma^2_D + (\text{residual } \sigma^2_I)$, where σ^2_I is the total epistatic variance.

2. POPULATIONS OF GENOTYPES INVOLVING MORE THAN TWO LOCI.
To generalise the analysis to any number of alleles per locus and any number of loci, it is convenient to consider the gene model developed by Kempthorne (1955) which allows a complete orthogonal partitioning of the total epistatic variance. For definitions and an elaboration of the use of the model we refer the reader to Kempthorne (1955). We shall briefly consider the results when this model is applied to the concept of general and specific combining ability.

In populations of genotypes derived from an arbitrary number of alleles at each of two loci the following holds :

(1) The genotypic value of $A_iA_jB_kB_l$ as measured from the population mean is expressed as :

$$ij^d_{kl} = {}_i\alpha + {}_j\alpha + {}_{ij}\delta + \alpha_k + \alpha_l + \delta_{kl} + {}_i(\alpha\alpha)_k + {}_i(\alpha\alpha)_l + {}_j(\alpha\alpha)_k \\ + {}_j(\alpha\alpha)_l + {}_i(\alpha\delta)_{kl} + {}_j(\alpha\delta)_{kl} + {}_{ij}(\delta\alpha)_k + {}_{ij}(\delta\alpha)_l + {}_{ij}(\delta\delta)_{kl}.$$

(2) The population genotypic variance is partitioned as :

$$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_{AA} + \sigma_{AD} + \sigma^2_{DD},$$

where σ^2_A = additive genetic variance, σ^2_D = dominance variance, σ^2_{AA} = additive \times additive epistatic variance, σ^2_{AD} = additive \times dominance epistatic variance, and σ^2_{DD} = dominance \times dominance epistatic variance.

(3) The g.c.a. effect associated with the A_iB_k gamete is

$${}_i g_k = {}_i\alpha + \alpha_k + {}_i(\alpha\alpha)_k,$$

where ${}_i\alpha$ and α_k are additive genetic effects, and ${}_i(\alpha\alpha)_k$ is an additive \times additive epistatic effect.

(4) The g.c.a. variance is $\sigma_{g.c.a.}^2 = p_i {}_i\alpha^2 + q_k \alpha_k^2 + p_i q_k {}_i(\alpha\alpha)_k^2$ which symbolically is $\sigma_{g.c.a.}^2 = \frac{1}{2}\sigma^2_A + \frac{1}{4}\sigma^2_{AA}$.

(5) The population genotypic variance can be partitioned in terms of the combining ability variances as follows :

$$2\sigma_{g.c.a.}^2 = \sigma^2_A + \frac{1}{2}\sigma^2_{AA} \\ \frac{\sigma_{s.c.a.}^2}{\sigma^2_G} = \frac{\sigma^2_D + \frac{1}{2}\sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DD}}{\sigma^2_A + \sigma^2_D + \sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DD}}$$

There is no difficulty in generalising to any number of loci each of which has an arbitrary number of alleles. The corresponding partitioning of the population genotypic variance can be written down as follows :

$$2\sigma_{g.c.a.}^2 = \sigma^2_A + \frac{1}{2}\sigma^2_{AA} + \frac{1}{4}\sigma^2_{AAA} + \text{etc.} \\ \frac{\sigma_{s.c.a.}^2}{\sigma^2_G} = \frac{\sigma^2_D + \frac{1}{2}\sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DD} + \frac{3}{4}\sigma^2_{AAA} + \sigma^2_{AAD} + \sigma^2_{ADD} + \sigma^2_{DDD} + \text{etc.}}{\sigma^2_A + \sigma^2_D + \sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DD} + \sigma^2_{AAA} + \sigma^2_{AAD} + \sigma^2_{ADD} + \sigma^2_{DDD} + \text{etc.}}$$

We note that $\sigma_{g.c.a.}^2$ is equal to the covariance between parent and offspring in a random mating population at equilibrium. This illustrates the genetic connection between the diallel method and the method involving covariances between relatives.

3. METHODS OF TESTING AND ESTIMATING GENETIC PARAMETERS FROM EXPERIMENTAL DATA

We assume that the experimental material represents a random sample from some specified random mating population about which inferences are to be made.

Analyses for the two different "modified diallel" experimental

methods will be given. For convenience we shall designate the analysis as follows :

Analysis (1) : For the experimental method in which $\frac{p(p-1)}{2}$ F_1 's and their reciprocals are included.

Analysis (2) : For the experimental method in which only one set of $\frac{p(p-1)}{2}$ F_1 's are included.

In both analyses we assume that a replicated experiment has been conducted which yields the necessary phenotypic mean values and a pooled error variance (with n degrees of freedom) associated with these mean values. We shall not consider the more complicated situation resulting from disproportionate sub-class numbers which has been excellently treated by Henderson (1948).

Analysis (1) (Involving $\frac{p(p-1)}{2}$ F_1 's and their reciprocals)

We assume that the phenotypic value of the ij^{th} observation can be represented by the model :

$$x_{ij} = \mu + g_i + g_j + s_{ij} + r_{ij} + \text{error},$$

where $i, j = 1, 2, \dots, p, (i \neq j)$, and,

μ = population mean.

$g_i(g_j)$ = g.c.a. effect of the $i^{th}(j^{th})$ inbred line.

s_{ij} = s.c.a. effect associated with the cross of the i^{th} and j^{th} inbreds and such that $s_{ij} = s_{ji}$.

r_{ij} = reciprocal effect associated with the crosses of the i^{th} and j^{th} inbreds and such that $r_{ij} = -r_{ji}$.

error = environmental error effect associated with the ij^{th} observation.

We assume that these error effects are not correlated with each other nor with the various genotypic effects and that they are normally distributed with mean zero and variance σ_e^2 . σ_e^2 is a pooled error mean square obtained from the replicated experiment and has n degrees of freedom.

The analysis of variance and expectations of mean squares are given in table 7. In determining the expectations of mean squares we assume that g_k, s_{ij} and r_{ij} are random variables with zero means and variances σ_g^2, σ_s^2 and σ_r^2 . The construction of this analysis of variance is given by Yates (1947) and Kempthorne (1952).

To test the hypothesis $\sigma_r^2 = 0$,

use $F = \frac{Mr}{Me}$ with $\left[\frac{p(p-2)}{2}, n \right]$ degrees of freedom.

To test $\sigma_s^2 = 0$,

use $F = \frac{Ms}{Me}$ with $\left[\frac{p(p-3)}{2}, n \right]$ degrees of freedom.

To test $\sigma_g^2 = 0$,

use $F = \frac{Mg}{Ms}$ with $\left[(p-1), \frac{p(p-3)}{2} \right]$ degrees of freedom.

The estimation of the components of variance is accomplished by equating the observed to the expected mean squares and solving. Thus,

$$\hat{\sigma}_r^2 = \frac{Mr - Me}{2}, \hat{\sigma}_s^2 = \frac{Ms - Me}{2} \text{ and } \hat{\sigma}_g^2 = \frac{Mg - Ms}{2(p-2)}$$

TABLE 7
Analysis of variance for the experimental method in which F_1 's and their reciprocals are included

Source	DF	Sum of squares *	Mean squares	Expectation of mean squares
R(g)	$p-1$	$\frac{1}{2(p-2)} \sum (X_{i.} + X_{.i})^2 - \frac{2}{p(p-2)} X_{..}^2$	Mg	$\sigma_e^2 + 2\sigma_g^2 + \frac{2}{2(p-2)} \sigma_g^2$
R(s)	$\frac{p(p-3)}{2}$	$\left\{ \begin{aligned} &\sum_{\substack{i,j \\ i < j}} \frac{1}{2} (x_{ij} + x_{ji})^2 - \frac{1}{2(p-2)} \sum (X_{i.} + X_{.i})^2 \\ &+ \frac{1}{(p-1)(p-2)} X_{..}^2 \end{aligned} \right\}$	Ms	$\sigma_e^2 + 2\sigma_s^2$
R(r)	$\frac{p(p-1)}{2}$	$\sum_{\substack{i,j \\ i < j}} \frac{1}{2} (x_{ij} - x_{ji})^2$	Mr	$\sigma_e^2 + 2\sigma_r^2$
Error from replicated experiment	n	...	Me	σ_e^2

* where $X_{..} = \sum_{\substack{i,j \\ i \neq j}} x_{ij}$.

TABLE 8
Analysis of variance for the experimental method in which one set of F_1 's only is included

Source	DF	Sum of squares *	Mean squares	Expectation of mean squares
R(g)	$p-1$	$\frac{1}{p-2} \sum X_i^2 - \frac{4}{p(p-2)} X_{..}^2$	Mg	$\sigma_e^2 + \sigma_g^2 + \frac{4}{(p-2)} \sigma_g^2$
R(s)	$\frac{p(p-3)}{2}$	$\sum_{\substack{i,j \\ i < j}} x_{ij}^2 - \frac{1}{p-2} \sum_i^2 X + \frac{2}{(p-1)(p-2)} X_{..}^2$	Ms	$\sigma_e^2 + \sigma_s^2$
Error from replicated experiment	n	...	Me	σ_e^2

* where $X_{..} = \sum_{\substack{i,j \\ i < j}} x_{ij}$.

Analysis (2) (Involving one set of $\frac{p(p-1)}{2} F_1$'s only)

We assume that the ij^{th} observation may be represented by

$$x_{ij} = \mu + g_i + g_j + s_{ij} + \text{error},$$

where $i < j = 1, 2, \dots, p$. The elements in the model may be defined in the same way as in analysis (1).

The analysis of variance which partitions the total sum of squares into orthogonal reductions is presented in table 8. This analysis is essentially the same as that first presented by Sprague and Tatum (1942) and later corrected by Federer (1951). The expectations of mean squares are found in the same way as in analysis (1).

To test the hypothesis $\sigma_s^2 = 0$,

$$\text{use } F = \frac{M_s}{M_e} \text{ with } \left[\frac{p(p-3)}{2}, n \right] \text{ degrees of freedom,}$$

and to test $\sigma_g^2 = 0$,

$$\text{use } F = \frac{M_g}{M_s} \text{ with } \left[(p-1), \frac{p(p-3)}{2} \right] \text{ degrees of freedom.}$$

The estimation of components of variance is accomplished as follows :

$$\hat{\sigma}_s^2 = M_s - M_e, \text{ and } \hat{\sigma}_g^2 = \frac{M_g - M_s}{p-2}.$$

4. SUMMARY

When a diallel crossing system is used in quantitative inheritance for the purpose of estimating genetic parameters of the population from which the inbreds were derived, the following assumptions and conditions must be met : (1) the parent population must be a random mating population in equilibrium ; (2) the experimental set of lines must be a random sample from a population of inbred lines which were derived from the parent population by the imposition of an inbreeding system free from forces which change gene frequencies ; (3) a " modified diallel " crossing system must be used in which the lines themselves are not included in the experimental set.

The additive and non-additive components of the parent genotypic variance are estimated by use of general and specific combining ability components of variance. The exact relationship existing between these two sets of parameters is given using a completely generalised gene model.

When interpreted in terms of the classical method of covariances between relatives, the method of diallel crosses yields estimates equivalent to those obtained by covariance between parents and offspring.

Analyses of variance are given which may be used for testing and estimating the population general and specific combining ability variances from experimental material.

5. APPENDIX

We wish to demonstrate by considering a particular case that use of the " modified diallel " crossing system gives rise to unbiased estimates of the population parameters. We shall start with a random mating parent population in equilibrium consisting of genotypes involving two alleles at each of two loci, which may or may not be linked.

It has been shown that if an inbreeding system is imposed on the entire population (in the absence of forces which change gene frequency) the population of derived inbreds will have the following distribution :

Genotype	Designation	Frequency
$A_1A_1B_1B_1$	L_1	$f_1 = p_1q_1$
$A_1A_1B_2B_2$	L_2	$f_2 = p_1q_2$
$A_2A_2B_1B_1$	L_3	$f_3 = p_2q_1$
$A_2A_2B_2B_2$	L_4	$f_4 = p_2q_2$

We assume : (1) the population of inbreds is indefinitely large and, therefore, the frequencies of the inbred lines are not changed by sampling without replacement ; (2) random samples of size p are drawn from the inbred population, allowing repetition of lines ; (3) for each sample of lines a " modified diallel " crossing system is imposed, and for convenience the method which involves one set of crosses but not reciprocals will be used.

Specifically it is necessary to show that (1) the frequency of genotypes generated by this experimental procedure, when averaged over all possible samples and when calculated after sample frequencies have been taken into consideration, will yield the genotypic frequencies of the original parent population, and (2) the expectations of $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ taken over all samples are the true population parameters σ_g^2 and σ_s^2 . (Symbolically this condition is indicated as $E_s(\hat{\sigma}_g^2) = \sigma_g^2$, and $E_s(\hat{\sigma}_s^2) = \sigma_s^2$).

TABLE 9

Frequencies of genotypes resulting from a modified diallel crossing system using the lines indicated in the column entitled " Typical sample."

Parti- tion	Sample type	No. of samples	Sample frequency	Typical sample	A_1B_1	A_1B_2	A_2B_1	A_2B_2	A_1B_1	A_1B_2	A_1B_1	A_1B_2	A_2B_1	A_2B_2
					A_1B_1	A_1B_2	A_2B_1	A_2B_2	A_1B_2	A_2B_1	A_2B_2	A_2B_1	A_2B_2	A_2B_1
(4)	S_{iiii}	4	f_i^4	S_{1111}	1	0	0	0	0	0	0	0	0	0
(31)	S_{iiij}	12	$4f_i^3f_j$	S_{1112}	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	0	0	0	0
(2 ²)	S_{iiij}	6	$6f_i^2f_j^2$	S_{1122}	$\frac{1}{6}$	$\frac{1}{6}$	0	0	$\frac{1}{6}$	0	0	0	0	0
(21 ²)	S_{iijk}	12	$12f_i^2f_jf_k$	S_{1123}	$\frac{1}{6}$	0	0	0	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	0	0
(1 ⁴)	S_{ijkl}	1	$24f_1f_2f_3f_4$	S_{1234}	0	0	0	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The particular illustration with which we are concerned necessitates making a complete enumeration of samples of size four (this is the smallest possible number) and deriving the expected values of the statistics in question over all such samples.

The sample frequency distribution may be obtained from the multinomial $(f_1L_1 + f_2L_2 + f_3L_3 + f_4L_4)$.⁴ We shall denote a sample having L_i, L_j, L_k and L_l as S_{ijkl} . It is convenient in tabulating frequencies to group the samples according to the partitions of the number four. Thus (4) corresponds to all samples S_{iiii} in which all four lines are L_i ($i = 1, 2, 3, \text{ or } 4$), (31) corresponds to all samples S_{iiij} in which three lines are L_i and one is L_j ($i, j = 1, 2, 3, \text{ or } 4$, but $i \neq j$), etc.

Table 9 gives the frequencies necessary to show that the " modified diallel " crossing system does yield genotypic frequencies which, when averaged over all samples, are identical to the genotypic frequencies occurring in the parent population. To verify this, one must multiply the genotypic frequencies within each

sample by the corresponding sample frequency and add over all samples. The frequency terms can be reduced from the fourth to the second degree if use is made of the identity $(\sum f_i)^2 = 1$.

To demonstrate that the "modified diallel" crossing system yields unbiased estimates of the population combining ability variances, it is necessary to perform an analysis of variance of the type given in table 8 on F_1 's derived from each of the 35 different possible samples. For these analyses it is assumed that a genotypic value, as measured from the population mean, may be represented by the model $d_{ij} = g_i + g_j + s_{ij}$, where d_{ij} is the genotypic value of the cross between the i^{th} and j^{th} inbreds, $g_i(g_j)$ is the g.c.a. effect of the $i^{th}(j^{th})$ inbred, and s_{ij} is the s.c.a. effect of the ij^{th} cross. We then determine the composition of $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ for each sample in terms of squares and cross-products of the elements in the model, and finally show that when all samples are taken into consideration the following is obtained: $E_s(\hat{\sigma}_g^2) = \sum f_i g_i^2 = \sigma_g^2$, and $E_s(\hat{\sigma}_s^2) = \sum f_{ij} s_{ij}^2 = \sigma_s^2$.

To illustrate the procedure of evaluating $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ for a given sample, let us consider the class of samples designated S_{iijj} ($i, j = 1, 2, 3, \text{ or } 4, \text{ but } i < j$). We need perform only one generalised analysis of variance for this class of samples and then obtain particular solutions for each sample by assigning appropriate values for i and j .

The mean values for F_1 's derived from the sample of inbreds (S_{iijj}) by the imposition of the "modified diallel" crossing system may be recorded as follows:

	i	i	j	j													
i	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%;"></td> <td style="width: 25%; text-align: center;">x_{ii}</td> <td style="width: 25%; text-align: center;">x_{ij}</td> <td style="width: 25%; text-align: center;">x_{ij}</td> </tr> <tr> <td></td> <td></td> <td style="text-align: center;">x_{ij}</td> <td style="text-align: center;">x_{ij}</td> </tr> <tr> <td></td> <td></td> <td></td> <td style="text-align: center;">x_{jj}</td> </tr> </table>					x_{ii}	x_{ij}	x_{ij}			x_{ij}	x_{ij}				x_{jj}	x_i
	x_{ii}	x_{ij}	x_{ij}														
		x_{ij}	x_{ij}														
			x_{jj}														
i					x_i												
j					x_j												
j					x_j												

where : $x_{ab} = g_a + g_b + s_{ab}, a, b, = i \text{ or } j$
 $x_i = 4g_i + 2g_j + s_{ii} + 2s_{ij}$
 $x_j = 2g_i + 4g_j + 2s_{ij} + s_{jj}$
 and $x.. = 6g_i + 6g_j + s_{ii} + 4s_{ij} + s_{jj}$

The problem is to determine the expressions for $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ in terms of the g_a and s_{ab} elements using the analysis of variance obtained from table 8 by putting $p = 4$.

The mean square for $R(g)$ is found to be

$$Mg = \frac{1}{3} \left(\frac{1}{2} \sum x_a^2 - \frac{X^2}{2} \right) = \frac{1}{6} [2(g_i - g_j) + (s_{ii} - s_{jj})]^2$$

The mean square for $R(s)$ is

$$Ms = \frac{1}{3} \left(\sum_{a < b} x_{ab}^2 - \frac{1}{2} \sum x_a^2 + \frac{X^2}{3} \right) = \frac{1}{6} (s_{ii} - 2s_{ij} + s_{jj})^2$$

Using these mean squares we determine $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ as follows:

$$\hat{\sigma}_g^2 = \frac{1}{2}(Mg - Ms) = \frac{1}{6} [(g_i - g_j)^2 + (s_{ii} - s_{ij})(s_{ij} - s_{jj}) + (g_i - g_j)(s_{ii} - s_{jj})]$$

and $\hat{\sigma}_s^2 = Ms = \frac{1}{6}(s_{ii} - 2s_{ij} + s_{jj})^2$.

Thus, for each of the six possible samples which can be represented by S_{iijj}

TABLE 10

Composition of $\hat{\sigma}_s^2$ in terms of squares and cross-products of specific combining ability elements for each class of samples corresponding to the partitions of the number four

Partition	Sample type	Number of samples	Sample frequency	Square and cross-product terms
(4)	S_{iiii}	4	f_i^4	None
(31)	S_{iiij}	12	$4f_i^3f_j$	None
(2 ²)	S_{ijij}	6	$6f_i^2f_j^2$	$\frac{1}{6}(s_{ii}-2s_{ij}+s_{jj})^2$
(21 ²)	S_{iijk}	12	$12f_i^2f_jf_k$	$\frac{1}{6}(s_{ii}-s_{ij}-s_{ik}+s_{jk})^2$
(1 ⁴)	S_{ijkl}	1	$24f_i f_j f_k f_l$	$\frac{1}{6}[(s_{ij}+s_{ik}+s_{il}+s_{jk}+s_{jl}+s_{kl})^2-3s_{ij}(s_{ik}+s_{il}+s_{jk}+s_{jl})-3s_{ik}(s_{il}+s_{jk}+s_{kl})-3s_{il}(s_{jl}+s_{kl})-3s_{jk}(s_{jl}+s_{kl})-3s_{jl}^2s_{kl}]$

TABLE 11

Composition of $\hat{\sigma}_g^2$ in terms of squares and cross-products of general and specific combining ability elements for each class of samples corresponding to the partitions of the number four

Partition	Sample type	Number of samples	Sample frequency	Square and cross-product terms
(4)	S_{iiii}	4	f_i^4	None
(31)	S_{iiij}	12	$4f_i^3f_j$	$\frac{1}{4}[(g_i-g_j)+(s_{ii}-s_{ij})]^2$
(2 ²)	S_{ijij}	6	$6f_i^2f_j^2$	$\frac{1}{6}[(g_i-g_j)^2+(s_{ii}-s_{ij})(s_{ij}-s_{jj})+(g_i-g_j)(s_{ij}-s_{jj})]$
(21 ²)	S_{iijk}	12	$12f_i^2f_jf_k$	$\frac{1}{12}\{4g_i(g_i-g_j-g_k)+3(g_j-g_k)^2+4g_jg_k+2s_{ii}(s_{ij}+s_{ik}-2s_{jk})+s_{ij}(s_{ij}-6s_{ik}+2s_{jk})+s_{ik}(s_{ik}+2s_{jk})+2[2g_i(s_{ii}-s_{jk})-g_i(s_{ii}-2s_{ij}+2s_{ik}-s_{jk})-g_k(s_{ii}+2s_{ij}-2s_{ik}-s_{jk})]\}$
(1 ⁴)	S_{ijkl}	1	$24f_i f_j f_k f_l$	$\frac{1}{12}\{3(g_i^2+g_j^2+g_k^2+g_l^2)-2(g_i g_j+g_i g_k+g_i g_l+g_j g_k+g_j g_l+g_k g_l)+s_{ij}(s_{ik}+s_{il}+s_{jk}+s_{jl})+s_{ik}(s_{il}+s_{jk}+s_{kl})+s_{il}(s_{il}+s_{kl})+s_{jk}(s_{jl}+s_{kl})+s_{jl}^2s_{kl}-4(s_{ij}^2s_{kl}+s_{ik}^2s_{jl}+s_{il}^2s_{jk})+2[g_i(s_{ij}+s_{ik}+s_{il}-s_{jk}-s_{jl}-s_{kl})+g_j(s_{ij}-s_{ik}-s_{il}+s_{jk}+s_{jl}-s_{kl})-g_k(s_{ij}-s_{ik}+s_{il}-s_{jk}+s_{jl}-s_{kl})-g_l(s_{ij}+s_{ik}-s_{il}+s_{jk}-s_{jl}-s_{kl})]\}$

(i.e. S_{1122} , S_{1133} , S_{1144} , S_{2233} , S_{2244} , and S_{3344}) we find that the values for $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ are obtained by substituting the appropriate values for i and j into the generalised expressions given above.

In a similar manner we can obtain generalised $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ expressions for each class of samples corresponding to the five partitions of the number four. These are given in tables 10 and 11.

To show that $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$ are unbiased estimates of the population parameters σ_g^2 and σ_s^2 (i.e. $E_s(\hat{\sigma}_g^2) = \sigma_g^2$ and $E_s(\hat{\sigma}_s^2) = \sigma_s^2$) we may (1) expand tables 10 and 11 to give a complete enumeration of all possible samples, (2) multiply the frequencies of square and cross-product terms within a given sample by the sample frequency, and (3) sum over all samples to obtain the weighted mean values for all square and cross-product terms.

We are required to demonstrate that these expressions for $E_s(\hat{\sigma}_g^2)$ and $E_s(\hat{\sigma}_s^2)$ which contain both square and cross-product terms can be transformed into expressions containing squared elements only, and, in fact, to prove that

$$E_s(\hat{\sigma}_g^2) = \sum_{i,j} f_i g_i^2 \text{ and } E_s(\hat{\sigma}_s^2) = \sum_{i,j} f_i f_j s_{ij}^2.$$

Let us consider $E_s(\hat{\sigma}_s^2)$ first. Since coefficients of terms involving g.c.a. effects only and coefficients of cross-product terms involving g.c.a. and s.c.a. effects are all zero, we need consider only square and cross-product terms involving s.c.a. effects. If to these terms we add the following five expressions each of which equals zero,

$$\left\{ \begin{array}{l} 2 f_i (\sum_j f_j s_{ij})^2, \text{ for } i = 1, 2, 3, \text{ and } 4 \\ \text{and } (-) (\sum_{i,j} f_i f_j s_{ij})^2 \end{array} \right\}$$

we find that $E_s(\hat{\sigma}_s^2) = \sum_{i,j} f_i f_j s_{ij}^2 = \sigma_s^2$.

To determine $E_s(\hat{\sigma}_g^2)$ we must consider square and cross-product terms of g.c.a. effects only, of s.c.a. effects only, and cross-product terms involving both g.c.a. and s.c.a. effects.

By use of the identity, $\sum_i f_i^2 g_i^2 (\sum_j f_j)^2 = -2 \sum_{i < j} f_i f_j g_i g_j$, it is possible to show that the square and cross-product terms involving the g.c.a. effects only, when summed over all samples, yield simply $\sum_i f_i g_i^2$ which is σ_g^2 .

The over-all contribution to $E_s(\hat{\sigma}_g^2)$ of square and cross-product terms involving only s.c.a. effects can be shown to be zero, if to the coefficients we add the following expressions each of which is equal to zero :

$$\left\{ \begin{array}{l} (-f_i) (\sum_j f_j s_{ij})^2 \text{ for } i = 1, 2, 3, \text{ and } 4 \\ \text{and } (\sum_{i,j} f_i f_j s_{ij})^2. \end{array} \right\}$$

Likewise, the over-all contribution to $E_s(\hat{\sigma}_g^2)$ of the cross-product terms involving both g.c.a. and s.c.a. effects can be shown to be zero, if to the existing coefficients we add the following expressions each of which is equal to zero :

$$\left\{ \begin{array}{l} (-2) f_i g_i \sum_j f_j s_{ij} \text{ for } i = 1, 2, 3, \text{ and } 4 \\ 2 (\sum_k f_k g_k) (\sum_{i,j} f_i f_j s_{ij}) \end{array} \right\}$$

When all terms are taken into consideration we find that $E_s(\hat{\sigma}_g^2) = \sum_i f_i g_i^2 = \sigma_g^2$.

This completes the demonstration that for inbred samples of size four, the "modified diallel" crossing system yields unbiased estimates of the population parameters.

Acknowledgment. The author is grateful to Dr Henry Bennett for reading the manuscript and making many valuable suggestions.

This study was undertaken at the Department of Genetics, Cambridge University where the author held a National Research Council (U.S.A.) Fellowship.

6. REFERENCES

- COCKERHAM, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics*, 39, 859-882.
- FEDERER, W. T. 1951. Evaluation of variance components from a group of experiments with multiple classifications. *Iowa Agr. Exp. Sta. Res. Bul.*, p. 380.
- FISHER, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh*, 52, 399-433.
- FISHER, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- FISHER, R. A. 1949. *The Theory of Inbreeding*. Oliver and Boyd, Edinburgh.
- GRIFFING, J. B. 1950. Analysis of quantitative gene action by constant parent regression and related techniques. *Genetics*, 35, 303-321.
- GRIFFING, J. B. 1953. An analysis of tomato yield components in terms of genotypic and environmental effects. *Iowa Agr. Exp. Sta. Res. Bul.*, p. 397.
- HAYMAN, B. I. 1954a. The analysis of variance of diallel tables. *Biometrics*, 10, 235-244.
- HAYMAN, B. I. 1954b. The theory and analysis of diallel crosses. *Genetics*, 39, 789-809.
- HENDERSON, C. R. 1948. Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine. Unpublished Ph.D. thesis, Iowa State College Library, Ames, Iowa.
- HENDERSON, C. R. 1952. Specific and general combining ability. In *Heterosis*, ed. Gowen, Iowa State College Press, Ames, Iowa.
- HULL, F. 1946. *Maize Genetics Cooperation News Letter*, 20.
- HULL, F. 1952. Recurrent selection and overdominance. In *Heterosis*, ed. Gowen, Iowa State College Press, Ames, Iowa.
- JINKS, J. E. 1954. The analysis of continuous variation in a diallel cross of *Nicotiana rustica* varieties. *Genetics*, 39, 767-788.
- KEMP THORNE, O. 1952. *Design and Analysis of Experiments*. John Wiley and Sons, Inc., New York.
- KEMP THORNE, O. 1955. The theoretical values of correlations between relatives in random mating populations. *Genetics*, 40, 153-167.
- NELDER, J. A. 1952. Some genotypic frequencies and variance components occurring in biometrical genetics. *Heredity*, 6, 387-394.
- SPRAGUE, G. F., AND TATUM, L. A. 1942. General versus specific combining ability in single crosses of corn. *Jour. Amer. Soc. Agron.*, 34, 923-932.
- YATES, F. 1947. The analysis of data from all possible reciprocal crosses between a set of parental lines. *Heredity*, 1, 287-301.