THE THEORY AND ANALYSIS OF DIALLEL CROSSES*

B. I. HAYMAN

A.R.C. Unit of Biometrical Genetics, Department of Genetics, University of Birmingham

Received January 28, 1954

1. INTRODUCTION

THE diallel cross method of investigating the genetical properties of a group of homozygous lines has recently received much attention. HULL (1945) has considered some aspects of the method. A short summary of a more general approach by JINKS and HAYMAN (1953) and its application to several published sets of maize data has also appeared. In another paper JINKS (1954) has described experiments on inbred lines of *Nicotina rustica*, and has given an account of some of the associated statistics together with a discussion of the results. In this paper we apply a genetic algebra to the theory of the diallel cross, not only to re-establish the formulae of JINKS, but also to investigate more complex genetical systems. We will show how to measure additive and dominance variation, how to describe the relative dominance properties of the parental lines and how to detect non-allelic genic interaction. A worked example illustrates the theory.

The following definitions will be used. A *diallel cross* is the set of all possible matings between several genotypes. The genotypes may be defined as individuals, clones, homozygous lines, etc., and, if there are n of them, there are n^2 mating combinations, counting reciprocals separately. A *diallel table* is an arrangement in a square of n^2 measurements corresponding one-to-one to the mating combinations of a diallel cross, each row and column of the square corresponding to offspring with a common parental genotype. This general definition is necessary because a diallel table need not be restricted to containing measurements on the progeny of a diallel cross, but may be used for later generations obtained by selfing these progeny or backcrossing them to their parents. We shall investigate the diallel cross consisting of the progeny of n selfed homozygous lines and their $n^2 - n$ crosses.

2. A SIMPLE GENETICAL SYSTEM

2.1. Hypotheses. Certain of the hypotheses listed below hold in many genetical systems; the others are useful simplifications and the effects of their failures are discussed in the 4th section. We assume;

(i) Diploid segregation,
(ii) No difference between reciprocal crosses,

* Part of the cost of the accompanying mathematical formulae and tables has been paid by the GALTON AND MENDEL MEMORIAL FUND.

(iii) Independent action of non-allelic genes, and in the diallel cross:
(iv) No multiple allelism
(v) Homozygous parents
(vi) Genes independently distributed between the parents.

2.2. Algebra. Consider a metrical character controlled by k genes, each with two alleles, A & a, . . . I & i, ... MATHER (1949) has discussed the case where genes at non-homologous loci influence the character independently and has investigated various mating systems. We consider a method which can be extended beyond sets of independent genes to those in which the genes at non-homologous loci interact. This consists of expressing the (metrical) phenotype as an algebraic function of a suitable representation of the genotype and then, using Mendel's laws to establish the relation between the genotypic representations of parent and progeny, obtaining statistics such as means, variances and covariances by the usual algebraic procedures.

In the notation of MATHER the genotypes II, ii and Ii at the ith locus have phenotypes $c + d_i$, $c - d_i$ and $c + h_i$ respectively where c is constant, $d_i > 0$ and h_i may take either sign. Let us represent the genotype by a variable θ_i which takes the values 1, -1 and 0 respectively, so that the phenotype is the polynomial $c + d_i\theta_i + h_i(1 - \theta_i^2)$. When the genotype controlling the character is the set, $\Theta = (\theta_1, \theta_2, \ldots \theta_k)$, and when genes at non-homologous loci act independently, the phenotype is

$$\sum_{i=1}^{k} \{d_i\theta_i + h_i(1 - \theta_i^2)\}$$

(omitting the constant which does not appear in differences and moments).

Some general results on this representation are:
(i) Since $\theta_i^3 = \theta_i$, $\sum_i \{d_i\theta_i + h_i(1 - \theta_i^2)\}$ is the most general polynomial involving $\theta_1, \theta_2, \ldots \theta_k$ independently, i.e. excluding products like $\theta_i\theta_j$.
(ii) The individual with genotype Θ produces gametes containing I and i in the frequencies $\frac{1}{2}(1 + \theta_i)$ and $\frac{1}{2}(1 - \theta_i)$.
(iii) The cross $\Theta' \times \Theta''$ (which may be either reciprocal cross) produces progeny II, ii and Ii in the frequencies $\frac{1}{4}(1 + \theta_i')(1 + \theta_i'')$, $\frac{1}{4}(1 - \theta_i')(1 - \theta_i'')$ and $\frac{1}{2}(1 - \theta_i'\theta_i'')$, i.e. these are the probabilities that $\theta_i = 1, -1$ or 0 in the progeny.
(iv) In the progeny of the cross $\Theta' \times \Theta''$ the expectations of θ_i and $1 - \theta_i^2$ are $\frac{1}{2}(\theta_i' + \theta_i'')$ and $\frac{1}{2}(1 - \theta_i'\theta_i'')$. The expected mean phenotype of these progeny is thus $\frac{1}{2}\sum_i \{d_i(\theta_i' + \theta_i'') + h_i(1 - \theta_i'\theta_i'')\}$.

2.3. The statistics of the diallel table. In this section, only genetical variation is considered, and the phenotypes are taken to be exactly

$$\sum_i \{d_i\theta_i + h_i(1 - \theta_i^2)\}.$$

Environmental variation is discussed in 3.1.

The genotypes of the offspring in the n × n diallel cross are determined

by the parental genotypes which are also the genotypes corresponding to the leading diagonal of the diallel table. The parents are assumed to be homozygous with u_i and v_i ($u_i + v_i = 1$) as the frequencies of parents with positive and negative homozygotes at the ith locus. Thus $\theta_i = 1$ in nu_i parents and $\theta_i = -1$ in nv_i parents. The mean of θ_i is $u_i - v_i = w_i$. Also $\theta_i^2 = 1$ and var $\theta_i = 1 - w_i^2 = 4u_iv_i$. Since the genes are assumed distributed independently in the parents cov $(\theta_i,\theta_j) = 0$ $(i \neq j)$.

Let the genotypes of the n parents be $\Theta_r = (\theta_{r1},\theta_{r2}, \ldots \theta_{rk})(r = 1, - -n)$ so that their phenotypes are $y_r = \sum_i d_i\theta_{ri}$ with mean $m_{L0} = \sum_i d_iw_i$. The progeny of $\Theta_r \times \Theta_s$ (or of $\Theta_s \times \Theta_r$) all have the phenotype $y_{rs} = \frac{1}{2}\sum_i \{d_i(\theta_{ri} + \theta_{si}) + h_i(1 - \theta_{ri}\theta_{si})\} = \frac{1}{2}\sum_i \{(d_i - h_i\theta_{si})\theta_{ri} + d_i\theta_{si} + h_i\}$. The mean of all the progeny of Θ_r (i.e. including Θ_r itself) is

$$\bar{y}_r = \tfrac{1}{2}\sum_i \{(d_i - h_iw_i)\theta_{ri} + d_iw_i + h_i\}$$

and the mean of the whole n^2 progeny is $m_{L1} = \sum_i \{d_iw_i + \frac{1}{2}h_i(1 - w_i^2)\}$. The difference between the mean of the parents and the mean of their n^2 progeny is $m_{L1} - m_{L0} = \frac{1}{2}\sum_i h_i(1 - w_i^2)$.

Consider the set of parents, the rth array (complete row or column) and the set of array means of the diallel table. The variance of the parents,

$$V_{0L0} = \operatorname*{var}_r \sum_i d_i\theta_{ri} = \sum_i d_i^2 \operatorname*{var}_r \theta_{ri} = \sum_i d_i^2(1 - w_i^2).$$

The covariance between the parents and their offspring in the rth array,

$$\begin{aligned} W_{01(r)L01} &= \operatorname*{cov}_s [\sum_i d_i\theta_{si}, \tfrac{1}{2}\sum_i \{(d_i - h_i\theta_{ri})\theta_{si} + d_i\theta_{ri} + h_i\}] \\ &= \tfrac{1}{2}\sum_i d_i(d_i - h_i\theta_{ri}) \operatorname*{var}_s \theta_{si} \\ &= \tfrac{1}{2}\sum_i d_i(d_i - h_i\theta_{ri})(1 - w_i^2) \end{aligned}$$

The variance of the rth array,

$$\begin{aligned} V_{1(r)L1} &= \tfrac{1}{4} \operatorname*{var}_s \sum_i \{(d_i - h_i\theta_{ri})\theta_{si} + d_i\theta_{ri} + h_i\} \\ &= \tfrac{1}{4}\sum (d_i - h_i\theta_{ri})^2(1 - w_i^2). \end{aligned}$$

The covariance between the array means and the rth array,

$$W_{01(r)L1} = \tfrac{1}{4}\sum (d_i - h_i\theta_{ri})(d_i - h_iw_i)(1 - w_i^2).$$

The means of the last three statistics are

$$W_{0L01} = \tfrac{1}{2}\sum d_i(d_i - h_iw_i)(1 - w_i^2)$$

which is also the covariance between the parents and the means of their offspring,

$$V_{1L1} = \tfrac{1}{4}\sum (d_i^2 - 2d_ih_iw_i + h_i^2)(1 - w_i^2),$$

and

$$V_{0L1} = \tfrac{1}{4}\sum (d_i - h_iw_i)^2(1 - w_i^2)$$

which is also the variance of the means of the arrays.

Notation. The suffix L refers to the diallel cross mating system and its extension by selfing. The subsequent figure(s), commencing with zero for the parents, refer to the generation(s) under consideration. In variances of individual measurements the preceding figure not in brackets is the same as that following; in variances of means and in covariances the preceding figure(s) refer to the generation(s) of the common parent(s) from which these means are descended. The bracketed figure(s) occur in variances and covariances of arrays and their omission indicates averaging over all arrays. In the L2 generation obtained by selfing the L1 individuals (which are the progeny of the original diallel cross) the variances within progenies are $V_{2(rs)L2}$ and these may be averaged, firstly to give $V_{2(r)L2}$ for each array, and secondly to give V_{2L2}. The extension of the notation over any number of generations of selfing parallels those of MATHER and VINES (1952) for continued selfing or sib-mating from a single cross, except that here we are also interested in statistics from parts of the diallel table. For convenience $W_{01(r)L01}$ and $V_{1(r)L1}$ will be abbreviated to W_r and V_r.

2.4. Genetical components. The statistics in the previous section (with the omission of the here unimportant $W_{01(r)L1}$) may be written in a form similar to that used by MATHER (l.c.) as

$$V_{0L0} = D$$
$$W_r = \tfrac{1}{2}D - \tfrac{1}{4}F_r$$
$$W_{0L01} = \tfrac{1}{2}D - \tfrac{1}{4}F$$
$$V_r = \tfrac{1}{4}D - \tfrac{1}{4}F_r + \tfrac{1}{4}H_1$$
$$V_{1L1} = \tfrac{1}{4}D - \tfrac{1}{4}F + \tfrac{1}{4}H_1$$
$$V_{0L1} = \tfrac{1}{4}D - \tfrac{1}{4}F + \tfrac{1}{4}H_1 - \tfrac{1}{4}H_2$$

and
$$m_{L1} - m_{L0} = \tfrac{1}{2}h$$

where
$$D = \Sigma d_i^2(1 - w_i^2)$$
$$F_r = 2\Sigma d_i h_i \theta_{ri}(1 - w_i^2)$$
$$F = 2\Sigma d_i h_i w_i(1 - w_i^2)$$
$$H_1 = \Sigma h_i^2(1 - w_i^2)$$
$$H_2 = \Sigma h_i^2(1 - w_i^2)^2$$

and
$$h = \Sigma h_i(1 - w_i^2)$$

Those equations independent of r may be solved directly for D, F, H_1, H_2 and h. $D - F + H_1 - H_2 (4V_{0L1})$ and H_2 are MATHER'S (l.c.p.75) random mating D and H. When $w_i = 0$, D and H_1 (or H_2) become MATHER'S (l.c.P.56) selfing and sib-mating D and H.

2.5. Distribution of alleles. $H_1 - H_2 = \Sigma h_i^2 w_i^2(1 - w_i^2) \geqslant 0$. Now $w_i^2 \neq 1$. Hence, if some loci exhibit dominance ($H_1 \neq 0$), the vanishing of $H_1 - H_2$ means that $w_i = 0$ ($u_i = v_i$) at these loci, while $H_1 > H_2$ means that $u_i \neq v_i$, i.e. the positive and negative alleles at these loci are not in equal proportions in the parents. Since $H_1 - H_2$ depends on the square of w_i it is not possible to decide in the latter case whether the positive or negative alleles are in excess. Now $w_i^2(1 - w_i^2)$ vanishes at $w_i = 0$, increases slowly with w_i^2 to a maximum at $w_i^2 = \frac{1}{2}$(u_i:v_i = 0·85:0·15 or 0·15:0·85) and then decreases to zero at $w_i^2 = 1$. Hence $H_1 - H_2$ fails to detect weak

asymmetry in general and extreme asymmetry in large diallel crosses. A test of significance of $H_2 - H_1$ is referred to in 3.3. An estimator of the mean value of u_iv_i at loci exhibiting dominance is $H_2/4H_1$. This is, however, biassed towards the larger values of u_iv_i.

2.6. Dominance. In qualitative definitions of dominance the heterozygote of the gene is taken to have the same phenotype as one of the homozygotes—the dominant homozygote—so that parents which contain the greatest number of dominant homozygotes of the genes controlling the character in question produce offspring with the least variation among themselves and with the least covariation with their other and more recessive parents in that character.

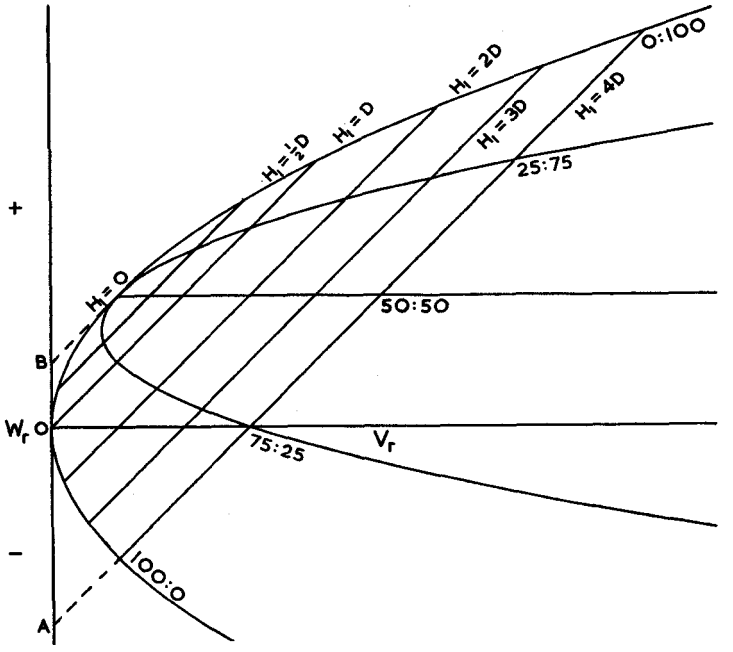

FIGURE 1.—Diallel cross dominance relationships in terms of V_r, the variance of all the offspring of the rth parent, and W_r, the covariance between these offspring and their non-recurrent parents, environmental variation being neglected.

For a diallel cross with a certain value of H_1/D, the points (V_r, W_r) are distributed along a corresponding straight line of unit slope inside the limiting parabola, $W_r^2 = V_r V_{0L0}$. This is one of the full sloping lines in the diagram. If the continued line cuts the W_r-axis in A, and if the parallel tangent to the parabola cuts it in B, then the line is determined by AB/OB = H_1/D. The line marked A in the diagram corresponds to a diallel cross with $H_1/D = 4$.

The position of (V_r, W_r) on the line reveals the relative proportions of dominant and recessive genes in the rth parent. For any diallel cross, the point corresponding to a parent containing p% dominants and q% recessives lies on the curve labelled p:q. Completely recessive parents correspond to points at the upper ends of the sloping lines on the part of the limiting parabola labelled 0:100, and completely dominant parents to points at the lower ends on the part labelled 100:0.

In an experiment with no dominance all the points coincide at $(\frac{1}{4}D, \frac{1}{2}D)$ ($H_1 = 0$ in the diagram).

Quantitatively we may consider any degree of dominance (measured by $|h_i|/d_i$ at the ith locus), the dominant homozygote deviating from the mid-homozygote in the same direction as the heterozygote so that $h_i\theta_i = |h_i|$. Similarly $h_i\theta_i = -|h_i|$ for the recessive homozygote. ($|h_i|$ means the positive value of h_i.)

In the diallel cross an overall measure of dominance is provided by $(H_1/D)^{\frac{1}{2}}$, the square root of the ratio of weighted means of h_i^2 and d_i^2. This ratio may be obtained from the graph of W_r against V_r. From 2.4, $W_r - V_r = \frac{1}{4}(D - H_1) = W_{0L01} - V_{1L1}$, so that the points (V_r,W_r) lie on a straight line of unit slope through their mean point (V_{1L1},W_{0L01}). The statistical inequality $W_r^2 \leqslant V_rV_{0L0}$ means that points can only lie on that part of the line inside the parabola $W_r^2 = V_rV_{0L0}$. Let the line cut the OW axis in A and let the parallel tangent to this limiting parabola cut the same axis in B. Then $AB/OB = H_1/D$ and this is identical with the value given by the equations in 2.4. In figure 1 the line $H_1/D = 4$ has been labelled A. When there is no dominance ($H_1 = 0$) the line is tangent to the limiting parabola and all the points (V_r,W_r) coincide at the point of contact, $(\frac{1}{4}D,\frac{1}{2}D)$. With complete dominance the line passes through the origin (0,0) while with partial dominance it lies above and with overdominance below the origin. In the latter case W_r may be negative.

With the aid of this graph we can make a detailed study of the relative dominance properties of the parents. From 2.4,

$$W_r = \tfrac{1}{2}D - \tfrac{1}{4}F_r$$

and

$$V_r = \tfrac{1}{4}D - \tfrac{1}{4}F_r + \tfrac{1}{4}H_1$$

where $F_r = 2\Sigma d_ih_i\theta_{ri}(1 - w_i^2)$. Since $h_i\theta_{ri}$ is positive for a dominant homozygote and negative for a recessive the greater values of F_r correspond to the more dominant parents and the lesser values to recessive parents. Thus, in the (V_r,W_r) graph, points with lower values of V_r and W_r correspond to dominant parents and points with higher values to recessive parents.

If we define the completely dominant parent to be the (possibly fictitious) parent carrying the dominant homozygotes of all the genes (which may have either positive or negative effects) and the completely recessive parent that carrying all the recessive homozygotes then, for the complete dominant,

$$V_D = \tfrac{1}{4}\Sigma(d_i - |h_i|)^2(1 - w_i^2)$$

and for the complete recessive

$$V_R = \tfrac{1}{4}\Sigma(d_i + |h_i|)^2(1 - w_i^2).$$

Now, unless the degree of dominance, $|h_i|/d_i$, of every gene is the same, (V_D,W_D) and (V_R,W_R) lie just inside the parabola. However, assuming that the points where the straight line cuts the parabola correspond to the completely dominant and recessive parents, it is found that

$$(V_D,W_D) = (V_{0L0}x_1^2,V_{0L0}x_1)$$

and

$$(V_R,W_R) = (V_{0L0}x_2^2,V_{0L0}x_2)$$

where x_1 and x_2 are the roots of $V_{0L0}x^2 - V_{0L0}x + W_{0L01} - V_{1L1} = 0$.

Suppose that the parent Θ_r contains k_{rD} dominant genes and k_{rR} recessive genes. Then, with certain restrictions about equality of gene effects,

$\frac{k_{rD}}{k_{rR}} = \frac{V_R - V_r}{V_r - V_D}$ = the ratio of the lengths of the two segments into which the point (V_r,W_r) divides the chord joining (V_R,W_R) to (V_D,W_D). (See fig. 1). The ratio of the total numbers of dominant to recessive genes in all the parents is

$$\frac{k_D}{k_R} = \frac{V_R - V_{1L1}}{V_{1L1} - V_D} = \frac{(4DH_1)^{\frac{1}{2}} + F}{(4DH_1)^{\frac{1}{2}} - F}.$$

2.7. Dominance by regression methods. With no dominance the regression of progeny with one parent in common on their non-recurrent parents is a straight line of slope $\frac{1}{2}$. When dominance is present consider the measurements, y_{rs}, of the progeny of Θ_r—the rth array—and the parental measurements, $y_s (s = 1, \ldots n)$. We have $y_{rs} - \frac{1}{2}y_s = \frac{1}{2}\sum_i \{d_i\theta_{ri} + h_i(1 - \theta_{ri}\theta_{si})\}$, which is not now independent of s, and in fact a linear relationship no longer exists between y_{rs} and y_s for given r. A best fitting regression line may, of course, be found for the offspring of each parent and its slope, W_r/V_{0L0}, will vary from parent to parent. HULL (1945, 1952) uses this variation in slope to detect and measure dominance.

His method is to fit the regression surface

$$y_{rs} = c + \tfrac{1}{2}b_1(y_r + y_s) - b_2y_ry_s$$

to the diallel table. b_2 is also the regression of regression slopes onto the corresponding parents and its existence indicates the presence of dominance. HULL's estimator of the degree of dominance reduces to $\{(1 - b_1)^2 + 4cb_2\}^{\frac{1}{2}}$. Fitting his regression surface to our more general genetical model we find that this estimator becomes

$$\{(V_{0L0} - 2W_{0L01})^2 + 4\mathrm{Cov}(W_r,y_r)(m_{L0} - m_{L1})\}^{\frac{1}{2}}/V_{0L0}$$
$$= \{(\Sigma d_ih_iw_i(1 - w_i^2))^2 + \Sigma d_i^2h_i(1 - w_i^2)^2 \cdot \Sigma h_i(1 - w_i^2)\}^{\frac{1}{2}}/\Sigma d_i^2(1 - w_i^2)$$
$$= (\Sigma d_i^2h_i \cdot \Sigma h_i)^{\frac{1}{2}}/\Sigma d_i^2, \text{ in the special case when all } w_i = 0.$$

Since HULL, unlike us, omits the diagonal of the diallel table in computing parent-offspring covariances these are not exact representations of his estimator but they suffice to reveal the main differences from our estimator, $(H_1/D)^{\frac{1}{2}} = (\Sigma h_i^2/\Sigma d_i^2)^{\frac{1}{2}}$ when all $w_i = 0$.

Clearly HULL's estimator measures mean dominance and not mean square dominance and must underestimate the average degree of dominance in any case but that of unidirectional dominance (for which it was admittedly designed). The regression coefficient, b_2, which provides the test of significance of dominance suffers from the same disadvantage. Furthermore, the third degree statistic, $\mathrm{Cov}(W_r,y_r)$, is a greater source of sampling error than the second degree statistics used in our estimator. It is interesting to note that, perhaps contrary to expectation, our more general theory has enabled us to obtain the simpler and more reliable estimator of the degree of dominance.

2.8. Number of genes. An estimate of the summed value of h_i is (2.4) $2(m_{L1} - m_{L0}) = h = \Sigma h_i(1 - w_i^2)$. In general this underestimates mean dominance because positive and negative values of h_i cancel out, but its sign does show whether positive or negative dominants are in the majority, or which exhibit the greatest degree of dominance.

Let k_+ and k_- be the number of groups of genes distributed independently in the parents for which the dominance is respectively positive and negative. Then $\frac{h^2}{H_2} = \frac{(k_+ - k_-)^2}{(k_+ + k_-)}$ again with certain restrictions about equality of gene effects. Now, if either k_+ or k_- is zero, i.e. all dominants have the same sign, this ratio estimates the number of groups which control the character and exhibit dominance to some degree. Usually, however, it underestimates this number, and it provides no information about groups of genes exhibiting little or no dominance. These groups of genes are not to be confused with MATHER'S (l.c.) effective factors.

It is important to distinguish the different discussions of the gene distribution in the previous sub-sections. The proportions in all the parents of positive and negative homozygotes at each locus which exhibits dominance are considered in 2.5; 2.6 gives the relative proportions of dominant and recessive homozygotes in each parent; here a lower bound is found to the number of genes exhibiting dominance in the parents.

2.9. Dominance and size. The sign of h (2.8) gives the mean direction of dominance. A measure of association between the signs of dominant genes is the correlation between parental size and parental order of dominance. The parental measurement, y_r, is closely correlated with the number of positive homozygotes in the parent while $(W_r + V_r)$ bears the same relation to the number of recessive homozygotes (3.2). When the correlation, ρ, between y_r and $(W_r + V_r)$ is nearly one the recessive genes must be mostly positive; when ρ is minus one the dominant genes are positive; when ρ is small equal proportions of the dominant genes are positive and negative.

When ρ^2 is nearly unity, the regression of y_r on $(W_r + V_r)$ exists, and the substitution of $(W_R + V_R)$ and $(W_D + V_D)$ in the regression equation predicts the measurements of the completely dominant and recessive parents. These must also be predictions of the possible limits of selection from amongst the genes exhibiting dominance, but, as before, we have no information about possible limits of selection from amongst the other genes.

3. COMPONENTS OF VARIATION

3.1. Environmental variation. Interaction between environmental fluctuations and the genotypes in a diallel cross is revealed by heterogeneity of the variances within (or between duplicate) parental and F_1 families. Such heterogeneity may be handled in at least three cases.

(i). When the sole source of heterogeneity is a difference between parental and F_1 variances the environmental variances of y_r and y_{rs} $(r \neq s)$ may be denoted by E and $\frac{1}{2}E'$ respectively. E is estimated from differences between

duplicate plots, E′ is estimated from the same source or from reciprocal differences, and the factor of $\frac{1}{2}$ compensates for the replacement of each pair of measurements of reciprocals by their common mean, which is done before evaluating statistics from the diallel table. With the environmental expectations included the equations of 2.4 become

$$\begin{aligned}
V_{0L0} &= D + E \\
W_r &= \tfrac{1}{2}D - \tfrac{1}{4}F_r + E/n \\
W_{0L01} &= \tfrac{1}{2}D - \tfrac{1}{4}F + E/n \\
V_r &= \tfrac{1}{4}D - \tfrac{1}{4}F_r + \tfrac{1}{4}H_1 + (E + \tfrac{1}{2}(n-1)E')/n \\
V_{1L1} &= \tfrac{1}{4}D - \tfrac{1}{4}F + \tfrac{1}{4}H_1 + (E + \tfrac{1}{2}(n-1)E')/n \\
V_{0L1} &= \tfrac{1}{4}D - \tfrac{1}{4}F + \tfrac{1}{4}H_1 - \tfrac{1}{4}H_2 + (E + \tfrac{1}{2}(n-2)E')/n^2 \\
(m_{L1} - m_{L0})^2 &= \tfrac{1}{4}h^2 + (n-1)((n-1)E + E')/n^3.
\end{aligned}$$

Even W_r and W_{0L01} have environmental expectations because each array has a term in common with the parental array.

(ii). When each F_1 variance can be expressed as the sum of two components corresponding to its parents those of the above equations which are independent of r hold with E = E′ = overall mean family variance.

(iii). When a trend exists between all the L_1 family means and variances the genotype-environment interaction may be removed by rescaling. WRIGHT (1952) describes the standard method.

When the influence of the environment is independent of genotype all the above equations hold with E′ = E. We use them in this form in the next sub-section.

3.2. Accuracy of the components. Those of the equations in 3.1(i) which are independent of r, together with an estimate of E, furnish an exact solution for D, F, H_1, H_2 and h^2, but no estimate of their accuracy. However, we have observed (2.6) that there are n estimates, $W_r - V_r$, of $\frac{1}{4}(D - H_1)$, and the sampling variation in these may be used to provide approximate standard errors of the genetical and environmental components. In replicated experiments, block differences supply a further estimate of error.

In obtaining the least squares solution of the above equations we omit W_{0L01} and V_{1L1} as superfluous (because of W_r and V_r) and weight with a factor $n^{\frac{1}{2}}$ the equations for V_{0L1} and $(m_{L1} - m_{L0})^2$ and the estimate of E since, unlike the others, these three statistics depend on all the measurements of the diallel table. The solution is

$$\begin{aligned}
\hat{D} &= V_{0L0} - \hat{E} \\
\hat{F} &= 2V_{0L0} - 4W_{0L01} - 2(n-2)\hat{E}/n \\
\hat{H}_1 &= V_{0L0} - 4W_{0L01} + 4V_{1L1} - (3n-2)\hat{E}/n \\
\hat{H}_2 &= 4V_{1L1} - 4V_{0L1} - 2\hat{E} \\
\hat{h}^2 &= 4(m_{L1} - m_{L0})^2 - 4(n-1)\hat{E}/n^2 \\
\hat{F}_r &= 2(V_{0L0} - W_{0L01} + V_{1L1} - W_r - V_r) - 2(n-2)\hat{E}/n
\end{aligned}$$

where we have used W_{0L01} and V_{1L1} for the means of W_r and V_r. The estimates of D, F, H_1, H_2 and h^2 are the same as in the exact solution. Only the

TABLE 1

Covariance Matrix

	D	F	H_1	H_2	h^2	E
D	$n^5 + n^4$	$2n^5 + 2n^4 - 4n^3$	$n^5 + 3n^4 - 2n^3$	$2n^4$	$4n^3 - 4n^2$	$-n^4$
F	$2n^5 + 2n^4 - 4n^3$	$4n^5 + 20n^4 - 16n^3 + 16n^2$	$2n^5 + 22n^4 - 16n^3 + 8n^2$	$4n^4 - 8n^3$	$8n^3 - 24n^2 + 16n$	$-2n^4 + 4n^3$
H_1	$n^5 + 3n^4 - 2n^3$	$2n^5 + 20n^4 - 16n^3 + 8n^2$	$n^5 + 41n^4 - 12n^3 + 4n^2$	$22n^4 - 4n^3$	$12n^3 - 20n^2 + 8n$	$-3n^4 + 2n^3$
H_2	$2n^4$	$4n^4 - 8n^3$	$22n^4 - 4n^3$	$36n^4$	$8n^3 - 8n^2$	$-2n^4$
h^2	$4n^3 - 4n^2$	$8n^3 - 24n^2 + 16n$	$12n^3 - 20n^2 + 8n$	$8n^3 - 8n^2$	$16n^4 + 16n^2 - 32n + 16$	$-4n^3 + 4n^2$
E	$-n^4$	$-2n^4 + 4n^2$	$-3n^4 + 2n^3$	$-2n^4$	$-4n^3 + 4n^2$	n^4

with common multiplier s^2/n^5

Also $\mathrm{Var}F_r = 4(3n^3 + 3n^2 - 4n + 4)s^2/n^3$

$\mathrm{Cov}(F_r,F_s) = 4(n^3 + 3n^2 - 4n + 4)s^2/n^3 \quad (r \neq s)$

$\mathrm{Var}(D - H_1) = 4(9n^2 - 2n + 1)s^2/n^3$

estimate of F_r is new, and it shows that $(W_r + V_r)$, and not just V_r or W_r, provides the better measure of dominance order (2.6).

The expected values of the statistics, derived from the estimates of the components, are identical with the observed values for V_{0L0}, V_{0L1}, $(m_{L1} - m_{L0})^2$ and E, but

$\hat{W}_r = \frac{1}{2}(W_{0L01} - V_{1L1} + W_r + V_r)$
and $\hat{V}_r = \frac{1}{2}(-W_{0L01} + V_{1L1} + W_r + V_r)$ so that the residual sum of squares
$= \frac{1}{2}\{\Sigma(W_r - V_r)^2 - n(W_{0L01} - V_{1L1})^2\}$ with $n - 1$ degrees of freedom. The mean square, $s^2 = \frac{1}{2}\mathrm{Var}(W_r - V_r)$. The covariance matrix of $\hat{D}$, $\hat{F}_r$, $\hat{H}_1$, $\hat{H}_2$, $\hat{h}^2$ and $\hat{E}$ is the inverse of the matrix of the coefficients of these components in the least squares equations and, as the important components are D, F, H_1, H_2, h^2 and E, the covariance matrix may be contracted to refer only to these quantities (see table 1).

The sampling correlation between $\hat{D}$ and $\hat{H}_1$ is positive so that diallel crossing shares with repeated backcrossing from the F_1 of a single cross the advantage over continued selfing or sib-mating of enabling $D - H_1$ to be estimated with the greater accuracy. In fact, if we solely desire to test the deviation of $\hat{D} - \hat{H}_1$ from zero we may use the direct estimate, $\mathrm{Var}(\hat{D} - \hat{H}_1) = 16\mathrm{Var}(W_r - V_r)/n$. Combined with the test of the significance of H_2 (3.3), this quick test at once classifies the experiment into one of four categories as exhibiting no dominance, partial dominance, complete dominance or overdominance.

3.3. Analysis of variance. In another paper HAYMAN (1954) has constructed an analysis of variance of the diallel table to test additive and dominance variation in the multiple allele case. Such an analysis provides statistically sound tests of significance of some of the components discussed here, viz., $(D - F + H_1 - H_2)$ (i.e. V_{0L1}), H_2, $(H_1 - H_2)$ and h^2 as well as differences between reciprocal crosses. See table 3 of that paper and 6.3 in this paper. The error mean square from this analysis of variance may be used as an estimate of E.

4. GENERAL GENETICAL SYSTEMS

4.1. Testing the hypotheses. Before the theory of the previous sections can be applied to a diallel table it is necessary to show that the corresponding diallel cross conforms to the hypotheses postulated in 2.1. A consequence of those hypotheses was that $W_r - V_r$ was constant, i.e. independent of r (2.6). We therefore expect that failures of the hypotheses may upset this constancy, and this is borne out in the investigations below. Heterogeneity of $W_r - V_r$ is thus a good indication of such failures. Homogeneity of $W_r - V_r$, while always implied by the validity of the hypotheses, may also be attained in certain cases of balanced failure. Two tests for heterogeneity of $W_r - V_r$ are available.

(i). When the experiment is replicated the variance of $W_r - V_r$ may be analysed for line and block differences. A significant line effect indicates fail-

ure of the hypotheses. The main statistical fault in this test is that the n values of $W_r - V_r$ are correlated. This paragraph, with 3.2, reveals the importance of the variance of $W_r - V_r$ in the analysis of diallel crosses.

(ii). A test which is useful when the experiment is not replicated depends on the (V_r,W_r) graph. This is not a line of unit slope if $W_r - V_r$ varies. To provide a test which gives equal weight to both W_r and V_r the axes of the graph are rotated through 45° so that the coordinates of points become proportional to $W_r + V_r$ and $W_r - V_r$. The t testing the significance of regression in the new axes is given by

$$t^2 = \frac{n-2}{4} \cdot \frac{(\mathrm{Var}V_r - \mathrm{Var}W_r)^2}{\mathrm{Var}V_r\mathrm{Var}W_r - \mathrm{Cov}^2(V_r,W_r)} \text{ with } n-2 \text{ degrees of freedom.}$$

Significance indicates failure of the hypotheses. The weakness of this test is that it only detects variation in $W_r - V_r$ which is correlated with the dominance order of the parents. Variation which merely increases the scatter of points about the regression line without altering its slope can only be detected by the first test.

Table 2 lists, for both tests, the probabilities of the hypotheses holding for three characters from a diallel cross of *Nicotiana rustica* varieties which was repeated for three years. Asterisks indicate failure of the hypotheses. Evidently experiments should be replicated to ensure detection of such failures.

TABLE 2

Significance tests applied to the hypotheses of 2.1 *for height, flowering time, and leaf length by the two tests described. Asterisks indicate failure of the hypothesis.*

Character		Test (i)	Test (ii)
Height	1951*	.01–.001	.20–.10
	1952*	.01–.001	.10–.05
	1953*	.01–.001	.10–.05
Flowering time	1951	.05–.01	> .90
	1952*	.01–.001	.01–.001
	1953	.05–.01	.70–.60
Leaf length	1951	.10–.05	.80–.70
	1952	> .20	.80–.70

When failure of the hypotheses has been demonstrated the simple theory of section 2 is no longer applicable; a more complex genetical system must be postulated and new parameters introduced to represent it. In the absence of data from later generations the components will outnumber the statistics so that a solution for them cannot be found. However, it is still possible to draw the (V_r,W_r) graph and to make estimates of D, F, H_1, H_2 and h^2 from the formulae in 3.2. We will now discuss what genetical interpretations can be ascribed to such components and what forms the graph may take under failure of the various hypotheses. Let OW be the vertical axis and OV the horizontal axis.

4.2. Reciprocal differences. When reciprocal differences exist, two values of each statistic are obtained from the table, one by using the columns, and the other by using the rows, as arrays. This ambiguity may be removed by replacing all entries in the table by their mean reciprocals. If the differences were independent of genotype then estimates of the genetical components are now correct as long as E contains the variation between reciprocals. If the differences depended on genotype then the genetical components are only averages in some way over maternal effects. 3.3 refers to a test for reciprocal differences.

4.3. Residual heterozygosity in the parents. A full treatment of diallel crosses between partially inbred material will require a separate paper and only the pertinent results are mentioned here. Firstly, the (V_r,W_r) graph is a scatter about a line of unit slope, points above the line corresponding to heterozygous parents and points below the line to inbred parents. Secondly, the formulae of 3.2 produce underestimates of H_1/D and $H_2/4H_1$ and an overestimate of F so that the degree of dominance is underestimated, asymmetry of the gene distribution exaggerated and the proportion of dominants overestimated.

4.4. Correlated gene distributions. When the genes at different loci are uncorrelated $\underset{r}{\mathrm{Cov}}(\theta_{ri},\theta_{rj}) = 0 (i \neq j)$ but in the general case we must take $\underset{r}{\mathrm{Cov}}(\theta_{ri},\theta_{rj}) = c_{ij}$. Then, neglecting E, the formulae of 3.2 give

$$\begin{aligned}
\hat{D} &= \sum_i d_i^2(1 - w_i^2) + \sum_{i \neq j} d_i d_j c_{ij} \\
\hat{F} &= 2\Sigma d_i h_i w_i(1 - w_i^2) + 2\Sigma d_i h_j w_j c_{ij} \\
\hat{H}_1 &= \Sigma h_i^2(1 - w_i^2) + \Sigma h_i h_j c_{ij}(w_i w_j + c_{ij}) \\
\hat{H}_2 &= \Sigma h_i^2(1 - w_i^2)^2 + \Sigma h_i h_j c_{ij}^2 \\
\hat{h} &= \Sigma h_i(1 - w_i^2).
\end{aligned}$$

Taking all $w_i = 0$ for simplicity we get

$$\begin{aligned}
\hat{D} &= \Sigma d_i^2 + \Sigma d_i d_j c_{ij} \\
\hat{F} &= 0 \\
\hat{H}_1 &= \Sigma h_i^2 + \Sigma h_i h_j c_{ij}^2 = \hat{H}_2 \\
\hat{h} &= \Sigma h_i
\end{aligned}$$

Since $\hat{F} = 0$ and $\hat{H}_1 = \hat{H}_2$ the measures of gene frequency are unaffected by the correlation. $\hat{H}_1$ (or $\hat{H}_2$) is greater than Σh_i^2 when the dominance is unidirectional, and slightly less if the h_i are randomly distributed in sign and magnitude. The estimate of gene number, $\hat{h}^2/\hat{H}_2$, is therefore usually depressed by the occurrence of correlation. $\hat{D}$ $(= V_{0L0})$ is zero when each parent has the same phenotype, and this occurs if each parent contains a suitable combination of positive and negative genes (dispersion). When genes of like effect are together in each parent (association) most $c_{ij} > 0$ and $\hat{D}$ is greater than Σd_i^2. Thus the measure of degree of dominance, $\hat{H}_1/\hat{D}$, may be either increased or decreased by the combined effects of correlation on $\hat{H}_1$ and $\hat{D}$ but the particular combination of dispersion and unidirectional dominance causes serious inflation and may easily turn partial dominance into apparent overdominance.

The effect on the (V_r,W_r) graph is interesting. When the c_{ij} are small the values of V_r still depict the order of dominance. Further

$$W_r - V_r = \tfrac{1}{4}\Sigma_i(d_i^2 - h_i^2)(1 - w_i^2) + \tfrac{1}{4}\Sigma_{i\neq j}(d_id_j - h_ih_j\theta_{ri}\theta_{rj})c_{ij}.$$

The completely dominant and recessive parents correspond to $W_D - V_D = W_R - V_R = \tfrac{1}{4}\Sigma(d_i^2 - h_i^2)(1 - w_i^2) + \tfrac{1}{4}\Sigma(d_id_j - |h_ih_j|)c_{ij}$. This is on the average the minimum value of $W_r - V_r$ for association and the maximum value for dispersion. Parents with intermediate proportions of dominants and recessives correspond to the maximum and minimum values respectively. Hence with association the (V_r,W_r) curve is convex upwards and with dispersion convex downwards.

4.5. Non-allelic gene interaction. Like 4.3, this heading provides scope for a separate paper. However, we can mention here that, now that a perfectly general representation of gene effect has been developed (HAYMAN and MATHER 1955), genic interaction presents no insuperable problems and we can give some of the pertinent results. A complementary type of interaction distorts the (V_r,W_r) graph, inflates $\hat{H}_1/\hat{D}$, depresses $\hat{h}^2/\hat{H}_2$ but has little effect on the estimators of gene frequency. A duplicate type of interaction depresses $\hat{h}^2/\hat{H}_2$, increases the apparent proportion of dominants but leaves $\hat{H}_1/\hat{D}$ and $\hat{H}_2/4\hat{H}_1$ and the (V_r,W_r) graph almost unaltered.

4.6. Scaling. Interaction between genes at non-homologous loci in a single cross may often be eliminated by a suitable change of the scale of measurement, but when interaction has been shown to be present in a diallel cross it may well be impossible to find one scale on which every individual cross exhibits no interaction. This can, nevertheless, be accomplished when a trend exists between $W_r - V_r$ on the one hand and both the corresponding parental and array mean values, y_r and $\bar{y}_r$, on the other hand. The method is given by WRIGHT (l.c.).

4.7. Multiple allelism. In the absence of segregation this is equivalent to polygenic biallelism exhibiting genic interaction and distributional correlation as illustrated here. A gene with four alleles forms four homozygotes and is thus equivalent to two genes with two alleles each. There are 9 independent comparisons between ten genotypes, whether derived from the four multiple alleles of one gene, or from two biallelic genes. However, whereas in the first case these comparisons correspond to additive and dominance effects, in the second case only 4 of the comparisons have this property, the others corresponding one to position effect and 4 to genic interaction. Any set of 2^p multiple alleles is equivalent in a like manner to p pairs of alleles. A set of q ($\neq 2^p$) multiple alleles requires p pairs of alleles to represent it, where $2^{p-1} < q < 2^p$, but there must also be some genic correlation to reduce the number of homozygotes to q. Evidently the effects of multiple allelism can be extremely complicated.

4.8. When there is no true dominance (all $h_i = 0$) the extended discussion of the previous five sub-sections may be reduced to the simple statement that any failure of the hypotheses which can be detected in the (V_r,W_r) graph causes the components estimated by the formulae of 3.2 to

exhibit spurious dominance. This follows directly from the concept of the limiting parabola in the (V_r,W_r) diagram. When there is no dominance all the points should coincide at $(\frac{1}{4}D,\frac{1}{2}D)$ on the limiting parabola. Any failure of the hypotheses which scatters the points (V_r,W_r) causes their mean (V_{1L1},W_{0L01}) to lie inside and not on the limiting parabola. Therefore, from 2.6, $\hat{H}_1 > 0$.

4.9. Analysis when the hypotheses fail. The approach to the problem of a diallel table exhibiting genic interaction (or indeed any failure of the hypotheses) has been to show how the (V_r,W_r) graph is affected. This is all that can be done with data from one generation because rescaling (4.6) is not possible if the scale has already been fixed to minimise genotype-environment interaction (3.1). Certainly no exact analysis can be undertaken while the genetical system fails to satisfy the hypotheses of section 2.

The way to further progress is to find some sub-table of the diallel table which does satisfy all the hypotheses and from which valid conclusions may be drawn as to the degree of dominance, asymmetry of gene distribution, etc., in the corresponding sub-group of non-interacting lines. Interacting lines usually have extreme values of $W_r - V_r$, i.e. lie well off the line of unit slope through (V_{1L1},W_{0L01}). A surer method of discovering which line to eliminate is to remove the measurements on the offspring of each line in turn and to test for heterogeneity of $W_r - V_r$ in each of the resulting $(n - 1) \times (n - 1)$ diallel tables. If any of these heterogeneities is not significant then, as far as this test is concerned, the corresponding sub-table satisfies the hypotheses and the lines remaining in it may be analysed as in sections 2 and 3.

If the interaction is still significant whichever line is removed, the next step is to remove every possible pair of lines in turn and test the remaining sub-tables for heterogeneity of $W_r - V_r$, then, if necessary, all triples of lines, etc., until a diallel table satisfying the hypotheses is obtained. In practice, if the removal of no single line can eliminate the interaction, it is usually sufficient, and far less laborious, instead of removing every possible pair of lines, to remove that line whose omission minimises the heterogeneity and then to remove each of the remaining lines in turn.

When only two values of $W_r - V_r$ (for r = s and t) show failure of the hypotheses by deviating markedly from the common value of the others, the cause may be interaction in the single cross $\Theta_s \times \Theta_t$. This is the case when it is possible to adjust the value of this phenotype to make $W_s - V_s$ and $W_t - V_t$ revert to the common value of the other $W_r - V_r$. Theoretically this should be done by minimising the heterogeneity of $W_r - V_r$ but in practice a missing plot fit based on the analysis of variance of the diallel table (3.3) seems to be good enough and is simpler. If y_{ts} and y_{st} are fitted values then

$$(n - 2)(n - 3)(y_{st} + y_{ts}) = (n - 1)(y_{s\cdot} + y_{\cdot s} + y_{t\cdot} + y_{\cdot t} - 2y_s - 2y_t) - 2y_{\cdot\cdot} + 2y_{\cdot}$$

and $(n - 2)(y_{st} - y_{ts}) = y_{s\cdot} - y_{\cdot s} - y_{t\cdot} + y_{\cdot t}$ where the sums $y_{s\cdot}$, etc.,

exclude the missing values. If this adjustment eliminates the interaction the whole table may be analysed without having to omit all the progeny either of Θ_s or of Θ_t.

This procedure of selecting certain lines which conform to the hypotheses is open to the criticism that some sub-group of lines is likely to do so merely through sampling variation. Replication of the experiment in time and space is necessary to validate such selection.

5. SUMMARY OF THE METHOD OF ANALYSIS

(i). Test the variances within families of parents and F_1's for genotype-environment interaction. This should lie within one of the categories in 3.1 for further analysis to be possible.

(ii). Form the diallel table of reciprocal means, compute V_r and W_r ($r = 1, \ldots n$) and test $W_r - V_r$ for heterogeneity (4.1). If this is significant, plot $W_r - V_r$ against $\bar{y}_r$ to decide if rescaling would be useful (4.6) and, if this fails, determine the interacting lines or crosses by inspection or otherwise and remove or adjust them (4.9).

(iii). When a diallel table with uniform $W_r - V_r$ has been obtained analyse its variance to find the significance of some of the genetical components of variation and to provide the estimate of E (3.3).

(iv). From V_{0L0}, W_{0L01}, V_{1L1}, V_{0L1}, $(m_{L1} - m_{L0})^2$ and the estimate of E compute $\hat{D}$, $\hat{F}$, $\hat{H}_1$, $\hat{H}_2$ and $\hat{h}^2$ and find their standard errors (3.2).

(v). Evaluate and interpret $\frac{\hat{H}_1}{\hat{D}}$ (2.6), $\frac{\hat{H}_2}{4\hat{H}_1}$ (2.5), $\frac{(4\hat{D}\hat{H}_1)^{\frac{1}{2}} + \hat{F}}{(4\hat{D}\hat{H}_1)^{\frac{1}{2}} - \hat{F}}$ (2.6), and $\frac{\hat{h}^2}{\hat{H}_2}$ (2.8) when the relevant components are significant.

(vi). Note the order of dominance of the parents (3.2). Find the correlation between $W_r + V_r$ and y_r and, if it is significant, predict the limits of selection (2.9).

A quicker and more superficial analysis would omit the analysis of variance of the diallel table and estimate E from differences between reciprocals.

6. NUMERICAL EXAMPLE

6.1. The data used to illustrate the analysis of a diallel table were kindly supplied by DR. JINKS. They are the flowering times, in days from a date in 1952, of *Nicotiana rustica* plants from a diallel cross of eight inbred varieties. These plants were grown in two blocks each containing 64 plots; each cross or self was represented by 10 progeny, grown in two plots of 5, with one plot in each block. Table 3 contains the mean flowering times per plot of the progeny. The variances within plots do vary significantly, but in the manner described in 3.1(ii) so that it is legitimate to use an average value for E.

6.2. Table 3 also contains the variances, V_r, of arrays and the covariances, W_r, between the arrays and the parental array, calculated for each block from the diallel table of mean reciprocals. $W_r - V_r$ is reasonably constant except for $r = 1$ and 3; the table shows exceptional and consistent heterosis

TABLE 3

Mean flowering time per plot of the progeny of a diallel cross of eight varieties.

Block I

		♀ 1	2	3	4	5	6	7	8	W_r	V_r	$W_r - V_r$
	1	22.8	14.4	27.2	17.2	18.3	16.2	18.6	16.4	18.29	24.46	−6.18
	2	15.4	17.2	14.8	18.6	15.2	17.0	14.4	10.8	7.44	5.55	1.89
	3	31.8	21.0	24.8	24.6	19.2	29.8	12.8	13.0	23.17	30.59	−7.42
	4	16.2	11.4	16.8	18.4	12.4	16.8	12.6	9.6	10.37	7.48	2.89
♂	5	14.6	12.2	15.2	15.2	15.2	18.0	10.4	13.4	4.41	2.53	1.88
	6	20.2	14.2	18.6	22.2	14.3	20.2	9.0	11.8	15.38	14.10	1.28
	7	14.0	12.2	13.6	13.8	15.6	15.6	11.4	13.0	3.72	1.96	1.76
	8	15.2	10.0	17.0	20.8	20.0	17.0	13.0	14.0	2.67	3.76	−1.08

Block II

		♀ 1	2	3	4	5	6	7	8	W_r	V_r	$W_r - V_r$
	1	24.2	16.2	30.8	27.0	20.2	16.8	14.4	16.0	14.02	25.39	−11.37
	2	16.5	18.8	14.6	18.6	15.3	15.2	14.8	13.2	7.18	6.61	0.56
	3	30.4	23.0	21.2	25.4	20.0	28.4	14.2	14.4	17.91	22.84	−4.93
	4	17.8	13.0	16.3	18.0	14.2	14.8	12.2	11.2	11.10	10.19	0.91
♂	5	18.8	13.6	15.4	13.8	15.2	16.0	12.2	20.0	5.73	4.82	0.91
	6	23.4	14.0	14.8	17.0	17.3	22.6	10.2	12.8	16.57	18.35	−1.79
	7	16.6	9.2	16.2	14.4	15.6	11.0	10.6	9.8	4.84	4.79	0.05
	8	17.2	11.6	18.2	20.8	17.4	12.6	9.8	15.8	4.04	8.33	−4.29

in the progeny of the corresponding crosses. The tests of 4.1 confirm heterogeneity of $W_r - V_r$.

(i). The analysis of the variance of $W_r - V_r$ is

	Mean square	Df	P
Lines	31.67	7	.01–.001
Blocks	13.98	1	.05–.01
Error	2.46	7	

(ii). In the test of the deviation of the slope of the (V_r, W_r) line from unity, $t_{14} = 4.068$ and P = .01–.001, which is highly significant.

The significant variation of $W_r - V_r$ from line to line means that these diallel tables do not satisfy the hypotheses of section 2. The graph of $W_r - V_r$ against $_r$ reveals no trend so that a change of scale is not suggested. The source of the trouble is probably complementary gene action in the cross $\Theta_1 \times \Theta_3$, and an adjustment of the plot means of the progeny of this cross should make the diallel cross conform to the hypotheses.

The formulae of 4.9 provide estimates of the plot means in the diallel table of means over the two blocks. The estimate of the difference between any two corresponding plot means in the two blocks = (BI total − BII total)/$(n^2 - 2)$ which minimises the block interaction sum of squares in the analysis of variance of two diallel tables. (The totals, as in 4.9, exclude the means of the progeny of $\Theta_1 \times \Theta_3$). The estimated plot means are

Block	♀1 × ♂3	♀3 × ♂1
I	23.3	19.2
II	23.6	19.6

which are about two-thirds of the observed values. The other new entries in Table 3 are

Block	W_1	V_1	$W_1 - V_1$	W_3	V_3	$W_3 - V_3$
I	10.27	7.82	2.45	17.51	17.44	1.37
II	10.29	9.94	0.34	10.33	6.44	3.89

The tests now show $W_r - V_r$ to be uniform

(i). In the analysis of variance of $W_r - V_r$ the probabilities of both line and block mean squares are .20–.10.

(ii). The slope of the (V_r, W_r) line does not differ significantly from unity ($P > .90$).

6.3. We are now free to analyse the variance of the diallel tables and to

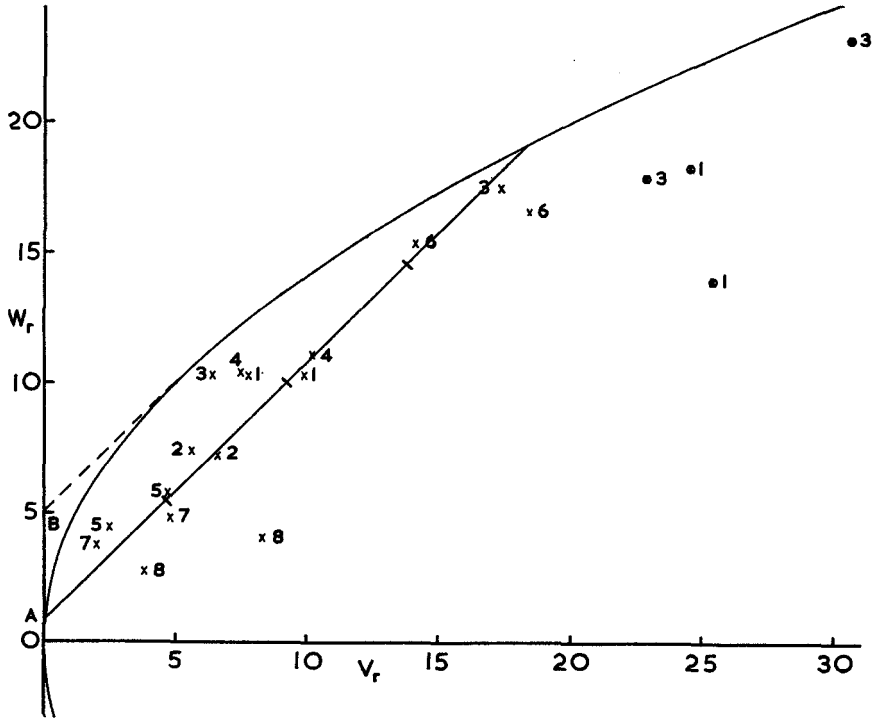

FIGURE 2.—(V_r, W_r) graph for flowering time in a replicated cross of eight inbred varieties of *Nicotiana rustica*. After adjusting the measurements of the progeny of the crosses of varieties 1 and 3, the points lie near a straight line of unit slope inside the limiting parabola $W_r^2 = 19.97V_r$. The unadjusted points (circled) lie well off this line. The three points of division and the ends of the line correspond to parents with 100%, 75%, 50%, 25% and 0% dominants (reading up from the bottom). The two blocks agree, except for line 3. AB/OB estimates H_1/D.

find the components of variation, remembering, of course, that our results provide no information about the cross $\Theta_1 \times \Theta_3$. Figure 2 depicts the points (V_r, W_r), the limiting parabola $W_r^2 = V_{0L0}V_r$, and the line of unit slope through the adjusted mean point (V_{1L1}, W_{0L01}). The unadjusted points lie well off this line.

The analysis of variance is derived from the adjusted table 3 in the way described by HAYMAN (l.c.). Table 4 contains the significance levels of the

TABLE 4

Significance levels of the components of variance derived from the data in table 3.

Item	Probability
(a) $D - F + H_1 - H_2$	$<.001$
(b_1) h^2	$<.001$
(b_2) $H_1 - H_2$	$>.20$
(b_3)	$<.001$
(b) H_2	$<.001$
(c) Reciprocal	$<.001$
(d) differences	$<.001$
(B) Blocks	$>.20$

components. The significance of (b) shows that dominance is present while (b_1) shows that it is largely unidirectional. From the sign of h we see that the progeny mean is less than the parental mean so that this dominance is in the direction of early flowering time. Asymmetry is not significant (b_2) at loci exhibiting dominance. This also implies that (a) detects purely additive variation and this is highly significant. The significant differences between reciprocal families are an unfortunate systematic effect due to faulty lay-out of seed-boxes in the greenhouse so that the use of mean reciprocals in all other computations is justified. The overall block difference is not significant. The block interaction mean square, which is the estimate of E, is 3.37.

6.4. The values of the statistics of 3.2 are

Block	V_{0L0}	W_{0L01}	V_{1L1}	V_{0L1}	$(m_{L1} - m_{L0})^2$
I	20.33	8.97	7.58	4.20	3.63
II	19.60	8.76	8.68	5.13	3.41

and the mean estimates of the components of variation with their standard errors are

$\hat{D}$	F	$\hat{H}_1$	$\hat{H}_2$	$\hat{h}^2$	$\hat{E}$	$D - \hat{H}_1$
16.59	−0.59	7.76	7.11	12.60	3.37	8.83
±1.79	±4.24	±4.12	±3.59	±2.40	±0.60	±3.54

The sum of the squares of deviations of observed from expected values of $V_{0L}{}^0$, V_{1L1}, V_r and W_r $(r = 1, \ldots n)$ is 142.84. The two diallel tables supply 37 statistics, and 12 constants are fitted to them, leaving 25 degrees of freedom for error. Hence the mean squared deviation is 5.71. Now, from the top left corner of the matrix in table 1, $\mathrm{Var}\hat{D} = \frac{1}{2}(n + 1)s^2/n$, the factor of $\frac{1}{2}$ allowing for the use in the estimation of means of statistics from two blocks. Since $n = 8$ and $s^2 = 5.71$, $\mathrm{Var}\hat{D} = 3.21$. Similarly, from the

next term down the leading diagonal of the matrix, $\text{Var}\hat{F} = \frac{1}{2} \times 6.28 \times 5.71 = 17.94$, etc. The square roots of these variances give the standard errors above. Their large values are due to the variations in V_3 and W_3 between blocks which suggest that this variety is unstable to environmental variations. Apparently $\hat{H}_1$ and $\hat{H}_2$ are not quite significant but the more reliable analysis of variance in 6.3 shows that $\hat{H}_2$ is significant. Here $\hat{D} - \hat{H}_1$ is just significantly different from zero; it is highly so if we use the simple formula for $\text{Var}(\hat{D} - \hat{H}_1)$ at the end of 3.2 which gives a standard error of 1.96 for $\hat{D} - \hat{H}_1$. Evidently dominance is present but it is definitely not complete dominance.

6.5. $(\hat{H}_1/\hat{D})^{\frac{1}{2}} = 0.68$ is an estimate of the mean degree of dominance over all loci. In figure 2, AB/OB estimates H_1/D without allowing for E. If the formulae of 3.2 are applied to the original diallel tables (i.e. without adjusting the cross $\Theta_1 \times \Theta_3$) then $(\hat{H}_1/\hat{D})^{\frac{1}{2}} = 1.02$. This illustrates the extent to which nonallelic genic interaction can inflate this estimate of the degree of dominance and emphasises the importance of the preliminary survey with the (V_r, W_r) graph.

The estimate of uv is $\hat{H}_2/4\hat{H}_1 = 0.23$ agreeing with the result in the analysis of variance (6.3) that H_1 is not significantly different from H_2.

$\{(4\hat{D}\hat{H}_1)^{\frac{1}{2}} + \hat{F}\}/\{(4\hat{D}\hat{H}_1)^{\frac{1}{2}} - \hat{F}\} = 0.95$ which is near enough to unity, implying equality between the numbers of dominant and recessive alleles in the parents. This is, of course, a necessary consequence of the foregoing result that $u_i = v_i = \frac{1}{2}$.

$\hat{h}^2/\hat{H}_2 = 1.77$ so that at least two of the genes controlling flowering time exhibit some degree of dominance.

6.6. The order of dominance of the parents, determined by $W_r + V_r$, is 75821436 and the order of flowering time is 78524631, parent 7 being the earliest and carrying the most dominants. The correlation between y_r and $W_r + V_r$ is 0.80 corroborating our result above (6.3) that early flowering is partially dominant.

6.7. This example has been worked in detail more to illustrate our methods and some of the difficulties which arise than to present a set of results conforming closely to our theories. Measurements of height and leaf length from the same plants were more consistent than flowering time over the two blocks and showed less or no reciprocal differences and little genotype-environment interaction. Further, while height exhibited genic interaction, leaf length satisfied our hypotheses completely.

SUMMARY

Experiments with diallel crosses provide a powerful method of investigating polygenic systems. The theory of a diallel cross between homozygous lines is discussed in terms of components of variation similar to MATHER'S (l.c.) D and H components. Assuming a simple underlying genetical system, we show that the various statistics obtained from measurements on the progeny provide estimates of the overall degree of dominance, of the rela-

tive dominance properties of the parents, and of the symmetry or otherwise of the gene distribution in the lines. The dominance relations are exhibited graphically. The effects of complications such as genic interaction are also considered. A genetic algebra is used to simplify the mathematical computations. The final sections are a summary of the method of analysis and a worked example.

LITERATURE CITED

HAYMAN, B. I., 1954 The analysis of variance of diallel crosses. Biometrics **10:** 235–244.

HAYMAN, B. I. and K. MATHER, 1955 The description of genic interaction in continuous variation. Biometrics (in press).

HULL, F. H., 1945 Regression analyses of yields of hybrid corn and inbred parent lines. Maize Genetics News Letter **19:** 21–27.

1952 Overdominance and recurrent selection. Heterosis, chap. 28, Iowa State College Press, Ames, Iowa.

JINKS, J. L., 1954 The analysis of heritable variation in a diallel cross of *Nicotiana rustica* varieties. Genetics **39:** 767–788.

JINKS, J. L. and B. I. HAYMAN, 1953 The analysis of diallel crosses. Maize Genetics News Letter **27:** 48–54.

MATHER, K., 1949 Biometrical Genetics. London, Methuen and Co.

MATHER, K. and A. VINES, 1952 The inheritance of height and flowering time in a cross of *Nicotiana rustica*. Quantitative Inheritance, 49–80, H.M.S.O., London.

WRIGHT, S., 1952 The genetics of quantitative variability. Quantitative Inheritance, 5–41, H.M.S.O., London.