THE ANALYSIS OF CONTINUOUS VARIATION IN A DIALLEL CROSS OF NICOTIANA RUSTICA VARIETIES

J. L. JINKS

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

Received January 28, 1954

I. THE ANALYSIS OF F_1 DATA

INTRODUCTION

STATISTICAL methods for the analysis of continuous variation have been developed using second degree statistics such as variance and covariance and successfully applied to crosses between two inbred lines. These have allowed the recognition of the more familiar phenomena associated with Mendelian genetics although they appear in the analyses in a new and less obvious form (MATHER 1949; BATEMAN and MATHER 1951; MATHER and VINES 1953). The methods can be used over the whole range of mating systems employed in plant and animal breeding, selfing, sib mating, assortative mating, and can also be adapted to randomly mating groups.

In the present paper this approach has been extended to the analysis of data from a diallel cross between a number of inbred lines. The application of this method to a diallel cross in *Nicotiana rustica* and published results of similar crossing programs in maize (JINKS and HAYMAN 1953) show it to be an efficient means of obtaining a rapid, overall picture of the genetical control of a character in a number of inbred lines; at the same time throwing considerable light on the genetical basis of heterosis in the F_1 progeny of these lines.

MATERIAL

The inbred lines used in the diallel cross were varieties of *N. rustica* collected by PROFESSOR K. MATHER while at the John Innes Horticultural Institution. The advantages of these are of a technical nature, e.g., ease of selfing and emasculating. The most important advantage, however, from the point of view of the present experiment is the great range of continuous variation within the species. A number of major gene differences have been recorded mainly as a check on pollination and segregation but also in the hope that linkage studies between polygenically and major genically controlled differences may be followed in later generations. These include the genes for anthocyanin and type of inflorescence (*A-a* and *M-m* respectively) already described by MATHER and VINES in addition to duplicate factors which control the color of ovary and flower (*yo*) and a gene responsible for seed color (*B-b*) i.e., brown-black or yellow.

The eight varieties used in these experiments are 2, 5, 7, 12, 14, 29, 38 and 41, using the numbers ascribed to them by MATHER and VINES.

EXPERIMENTAL PLAN

In 1950 crosses were made between the 8 varieties in all possible combinations, including reciprocals. (Although 9 lines were originally included in the experiment, the crosses in one line, variety 31, differed from the rest in showing a significant difference between reciprocals. In consequence, variety 31 and its crosses have been excluded from the main diallel experiment and will not be discussed further here.) The 56 F_1's and the 8 parental lines were grown in 1951. Plant height, flowering time, leaf length, width and distance of maximum width from the leaf apex were recorded. A further character, the total number of seed capsules per plant was also recorded for the four parental lines 5, 12, 14 and 41 and for the 12 F_1's between them.

The families were started in a glasshouse and were later planted out in a randomized layout of two blocks in the field. Each block consisted of 80 plots (i.e., 9 lines and all possible F_1's, except that 2×41 was omitted as it gave no viable seed). Each plot consisted of 5 plants in a row. Taking both blocks into account, 10 plants were grown of each family. All rows were set 2 feet apart and plants 1 foot apart within the rows. The whole experiment was surrounded by a guard row of plants which were not scored so as to avoid edge effects. The two blocks were identical except that the plots within them were randomized separately.

The F_1's grown in 1951 were selfed to give F_2 seed and also backcrossed reciprocally to their respective parental lines. The diallel cross between the parental lines was also repeated. In 1952 the layout again consisted of two

TABLE 1

The means and variances of the diallel crosses averaged over blocks and reciprocals for height 1951. The means occupy the upper right half of the table and the variances the lower left.

		Means							
Parents		2	5	7	12	14	29	38	41
Variances	2	38.5 6.0	51.3	40.4	67.1	40.9	40.3	41.8	38.8
	5	4.1	39.3 4.1	44.6	50.0	45.6	50.5	42.4	42.7
	7	8.6	3.2	37.2 4.7	56.0	39.1	41.7	33.3	37.1
	12	7.9	5.4	13.3	52.9 9.1	54.3	60.7	49.1	52.9
	14	10.7	11.6	5.7	11.7	38.3 15.6	40.3	33.8	30.7
	29	8.4	17.4	4.2	9.0	8.9	44.3 4.4	34.9	40.6
	38	7.4	6.5	8.0	19.9	4.2	5.8	30.7 1.5	33.1
	41	7.7	10.5	3.9	11.3	5.9	25.4	5.1	37.1 25.9

TABLE 2

The means and variances of the diallel crosses averaged over blocks and reciprocals for height 1952. The means occupy the upper right half of the table and the variances the lower left.

Parents		Means							
		2	5	7	12	14	29	38	41
Variances	2	33.5 1.6	51.0	39.1	62.9	40.2	36.1	34.6	37.3
	5	9.3	38.3 10.4	39.2	51.8	44.7	48.6	40.8	43.3
	7	5.2	5.5	27.4 4.4	55.6	39.5	36.7	35.4	35.2
	12	28.9	5.9	7.0	51.3 10.6	53.2	57.2	50.6	52.5
	14	4.3	9.4	3.9	4.9	35.8 5.8	41.0	36.4	37.7
	29	3.3	7.3	2.8	10.2	4.6	39.5 8.2	35.3	37.6
	38	8.4	8.7	8.5	8.3	3.9	6.5	35.9 1.0	33.7
	41	5.2	14.0	10.3	9.0	4.6	6.8	12.5	33.3 3.5

blocks, but with 287 plots of 5 plants per block. Since line 31 and its crosses had not been included, these 287 plots were made up of 8 parental lines and 55 F_1's (the family 2×14 being omitted as no viable seed was obtained) of 1 plot per block, 56 F_2's of 2 plots per block and 108 backcrosses of 1 plot per block.

ANALYSIS OF MEANS

The method of recording height and flowering time was the same as in MATHER and VINES' experiment. In the current experiment the time of flowering is expressed in days after July 1st in 1951 and after June 20th in 1952. These dates are chosen quite arbitrarily, the earlier date being used in 1952 so that all the figures would be positive. This in no way affects the subsequent analyses, or their results.

The leaf characters were measured to the nearest half centimeter. The largest leaf on each plant, usually one of the first four leaves to be formed in these varieties, was chosen and the length from apex to the insertion of the petiole, the maximum width and the distance of this point of maximum width from the leaf apex were measured.

Because of the large numbers involved (100–400 capsules per plant) and the time taken to measure such a character, it was only recorded for 4 parental lines and the 12 F_1's between them.

The mean heights for the 8 parental lines and the 56 F_1 combinations for the seasons 1951 and 1952 are given in tables 1 and 2. These results are averaged over two blocks and reciprocal families within each block.

(i) *Combination and correlation of characters*

In order to investigate the possibility of obtaining a discriminant function for the three leaf measurements, the variances within and between families were calculated for each of the three characters in 1951.

The ratio of these two variances was almost uniform for the three leaf characters. Furthermore these characters are highly correlated, e.g., the correlation coefficient for leaf length and width is 0.8819. It appears, therefore, that the three measurements are contributing much the same information so that little is to be gained by combining them into a discriminant function.

The correlation coefficients between height and flowering time and height and leaf length for 1951 are not significant. So far as these data go, these characters are independent of one another. Height and number of capsules per plant are, however, highly correlated. Capsule number, leaf width and position of maximum width, were consequently not scored in 1952.

(ii) *Reciprocal differences*

A comparison of the means of reciprocal F_1 families has been made for all the characters both in 1951 and 1952 using the analysis described by YATES (1947). In 1951 there were no significant reciprocal differences after excluding the progenies of parent 31 (mentioned earlier). Even when the progenies of this parent were not included, however, flowering time in 1952 gave significant reciprocal differences ($P = 0.01$–0.001), while height and leaf length were significant at the 5% level. Unlike the reciprocal differences encountered in the progenies of parent 31 in 1951, those of 1952 could not be traced to the progenies of one or a small number of parental lines: they were sporadically distributed throughout the F_1 progenies of all the parental lines. They could be traced, however, with some confidence to the history of the plants prior to planting in the field. Owing to the peculiarities of the 1952 season, the young plants were kept under glass in boxes up to a much later stage than is normal. Reciprocal families being in different boxes suffered to different extents from the resulting overcrowding and eventual setback when they were ultimately planted in the field. Furthermore, flowering commenced within a few weeks of planting out, as compared with the usual period of about two months. As a result, duplicate plots in the two blocks, grown from plants from the same box, although more widely separated in the field than reciprocal families within the same block, were more alike, so giving rise, one assumes, to the significant reciprocal differences. Flowering time, scored before the plants recovered from the setback, was affected most, while height and leaf length, which are scored at the end of the growing season, were less upset, and indeed show barely significant reciprocal differences ($P = 0.05$).

In the absence of significant reciprocal differences, the mean sum of squares of reciprocal differences is our estimate of E_2 (MATHER 1949), i.e., the environmental component of the variances of family means. In the presence of reciprocal differences of the type found for the 1952 flowering time, E_2 can be obtained as the variance of differences of duplicate plots between the two

blocks. When this course is followed, the analysis to be described later must be carried out on the means of reciprocal families within each block.

(iii) *Heterosis*

An indication of the overall direction of the deviation of the F_1 means from their corresponding mid-parents can be obtained by comparing the mean of all parental lines with that of all F_1's. In all characters except number of capsules and flowering time the F_1 mean is larger. This difference can be tested for significance using an empirical error variance derived from the variation of the deviations of each F_1 family mean from its corresponding mid-parent, taking sign into account, around the overall mean deviation. This method in all cases gives high significance for the mean deviations (table 3).

It must be emphasized that these results only represent the overall direction of heterosis for any character. The significant positive heterosis for height, for example, does not exclude the possibility that individual F_1 families may show significant negative heterosis or no heterosis. Indeed examination of the results show that such families exist, although less frequently than families that show positive heterosis.

TABLE 3

Overall parental and F_1 means and the significance of the deviation (F_1 mean − Parental mean) for the characters scored over the two years.

	Height		Flowering time		Leaf length	
	1951	1952	1951	1952	1951	1952
Parental mean	39.62	36.85	30.39	18.15	20.69	21.16
F_1 mean	43.50	42.55	26.19	16.54	21.95	22.38
(F_1 − P) mean	+3.88	+5.70	−4.20	−1.61	+1.26	+1.22
Significance	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

ENVIRONMENT-GENOTYPE INTERACTION

The variance of the five plants within a plot around the plot mean has been obtained for each of the characters scored over both years. These variances, averaged over blocks and reciprocals, are given for height in tables 1 and 2. In a small number of cases these variances are based on less than 4 degrees of freedom due to the death of an occasional plant. In no case, however, are they based on less than 2 degrees of freedom, the average being 3.9. In so far as the parental lines are homozygous (and there is nothing to lead us to believe that they are not) these variances are due entirely to non-heritable effects, i.e., the E_1 component of the variance of segregating families (MATHER 1949).

The STEVENS test of homogeneity of variance (FABERGÉ 1936) when applied to these variances gives high significance ($P < 0.001$ in all cases) for heterogeneity for all the characters scored over both years. The result is the same whether duplicate families in the two blocks are averaged or analyzed independently. The non-heritable variances are thus not uniform over families.

Since there was the possibility that the heterogeneity of the variances could be removed by a suitable change of scale, the relationship between the variances and means of families were investigated to see whether low means and variances and high means and variances were associated. Scatter diagrams of the means and variances showed no obvious relationship, and the regression, which was calculated in the case of height in 1951, proved to be non-significant (P = 0.70–0.60). No simple transformation of the data would thus make the variances homogeneous.

To see whether the heterogeneity could be traced to each genotype having its own specific environmental interaction, an analysis of variance was carried out on the non-heritable variances as illustrated below for flowering time 1951. (This analysis includes the progenies of parent 31).

		SS	DF	V	VR	P
genotype	♂ arrays	8,231.7710	8	1,028.9714	3.4	0.01–0.001
	♀ arrays	9,136.7459	8	1,142.0932	3.8	0.01–0.001
	blocks	6.5214	1	6.5214	0.02	Not sig.
	♂ × ♀ arrays	27,112.8553	64	423.6384	1.4	Not sig.
	blocks × ♂ arrays	2,706.3355	8	338.2919	1.1	Not sig.
	blocks × ♀ arrays	1,311.4206	8	163.9276	0.5	Not sig.
	blocks × ♂ × ♀ arrays (error)	19,398.3828	64	303.0997		
	Total	67,904.0325	161			

Analysis of the other characters gave essentially the same results, i.e., only the male and female lines used for the F_1's have any significant effect on the size of the error variance. One could, therefore, predict the error variance of each F_1 as the sum of two constants characteristic of its two parents. The variation, or heterogeneity in the non-heritable variances is thus attributable to the varying reaction of the different genotypes to environmental differences.

Certain differences between the flowering time as just analyzed and the other characters are worth noting. For height in 1951 the male and female array sums of squares were only significant if pooled to give a genotype sum of squares for 16 degrees of freedom (P = 0.05–0.01). The block × female sum of squares was also significant (P = 0.05–0.01), but was no longer significant if pooled with the block × male sum of squares to give block × genotype sum of squares (P = 0.10–0.05). In height 1952 the female array sum of squares was significant ($P < 0.001$), but the male array sum of squares was not. Jointly, the genotype sum of squares for 14 degrees of freedom was significant ($P < 0.001$). Neither of the block interactions was significant (P = 0.10–0.05). In the leaf analysis 1951 the male array sum of squares was significant (P = 0.05–0.01), but the female array sum of squares was not, neither were any of the interactions. The summed male and female array sum of squares was also not significant (P = 0.10). This is almost certainly the result of a larger relative error in choosing and measuring the largest leaf inflating the error sum of squares. The number of capsules gave significance for the

TABLE 4

The variance within plots of all parents and all F_1's in 1951 and 1952.

Character	Parents		F_1's		Averages of both years	
	1951	1952	1951	1952	Parents	F_1's
Height	6.8200	5.6656	6.0734	7.7583	6.2428	6.9160
Flowering time	19.0806	16.6375	10.4790	14.1219	17.8591	12.2705
Leaf length	3.3066	2.2907	2.1495	3.8967	2.7987	3.0231

male array sum of squares (P = 0.01) and the joint genotype sum of squares (P = 0.01–0.001) but not for the female array sum of squares (P = 0.1–0.05).

Despite these minor disturbances there can be little doubt that different genotypes show different amounts of non-heritable variation; or, to put in another way, different genotypes respond in different ways to given changes in environmental conditions.

In table 4 are given the mean variance within plots of all parents and all F_1's for the characters scored over the two years.

Only in the case of flowering time does the difference in variability in parents and F_1's approach significance and then only at the 5% level. For the other characters the difference is neither significant nor consistent in sign over the two years. Overall there appears to be very little difference in the response of the homozygous parents and heterozygous F_1's to differences in the environment.

SECOND DEGREE STATISTICS

The expected variances and covariances obtainable from the parents and progenies of a diallel cross can be found by the methods of MATHER (1949). Let us consider one locus represented by two allelomorphs A and a in the parental lines such that a proportion u_a of the parental lines are AA and $v_a(= 1 - u_a)$ are aa. The array of gametes will then be $u_a A$ and $v_a a$. If the overall mid-parent is taken as 0 and A adds on $+ d_a$ and a, $- d_a$, while the heterozygote deviates by h_a from the mid parent, then in a diallel set of crosses the contribution of this locus to the means of the families will be as follows.

Parents	Genotype	AA	aa	
				array
Males	Frequency	u_a	v_a	
				means
Females	Mean	d_a	$-d_a$	
AA u_a d_a		AA u_a^2 d_a	Aa $u_a v_a$ h_a	$u_a d_a + v_a h_a$
aa v_a $-d_a$		Aa $u_a v_a$ h_a	aa v_a^2 $-d_a$	$-v_a d_a + u_a h_a$

Parental mean $(u_a - v_a)d_a$ Overall mean of progenies $(u_a - v_a)d_a + 2u_a v_a h_a$.

The mean variance of the F_1's in an array, i.e., all the progenies of a common parent, around the array mean is

$$\text{mean variance of an array} = V_{1L1} = u_a[u_a d_a^2 + v_a h_a^2 - (u_a d_a + v_a h_a)^2] + v_a[u_a h_a^2 + v_a d_a^2 - (u_a h_a - v_a d_a)^2]$$

This is similar to the expectations for the variance of a random mating population given by MATHER (1949), which reduces to

$$u_a v_a[d_a + (v_a - u_a)h_a]^2 + 4u_a^2 v_a^2 h_a^2.$$

For a number of independent genes, i.e., genes showing neither interaction nor non-random distribution in the parental lines, this becomes

$$V_{1L1} = \tfrac{1}{4}D + \tfrac{1}{4}H + E_2.$$

Where $D = 4\Sigma uv(d + (v-u)h)^2$ and $H = 16\Sigma u^2 v^2 h^2$ as defined by MATHER 1949 for random mating systems. The component E_2 is included to allow for the non-heritable variance of means due to effects of the environment.

Similarly the contribution of the gene *A-a* to the variance of array means around the overall progeny mean is given by

$$u_a(u_a d_a + v_a h_a)^2 + v_a(u_a h_a - v_a d_a)^2 - [(u_a - v_a)d_a + 2u_a v_a h_a]^2$$

and summing over all independent genes we find $V_{0L1} = \tfrac{1}{4}D$. Now this will have a non-heritable component E_3 which is $1/n\ E_2$, where n is the number of families in the array, i.e., the number of parental lines.

The overall variance of progeny means around the overall progeny mean is, of course, the sum of the above two variances, i.e., $\tfrac{1}{2}D + \tfrac{1}{4}H$, the non-heritable factor again being E_2. When $u = v = \tfrac{1}{2}$, the overall variance of the diallel crosses is the same as the variance of an F_2 family from a cross between two inbred lines.

In addition to these variances one can also obtain the covariance of offspring on parents. For a diallel cross there are two methods of obtaining this covariance, viz., the covariance of array means on the common parent of the array and the mean covariance of members of an array on the non-common parents, both of which reduce to

$$2u_a v_a d_a[d_a + (v_a - u_a)h_a].$$

Unlike the variances obtainable from the diallel cross it is impossible to express this covariance in terms of D and H appropriate to the variances and covariances of a random mating population. It is necessary, therefore, to redefine our D for the covariances as $D_W = \Sigma 4uvd[d + (v-u)h]$. Each term of this D is the geometric mean of the corresponding terms of the random mating D and $D(= \Sigma d^2)$ of a one by one cross between inbred lines, though of course this property does not hold for the D's themselves. Covariances normally have no non-heritable component (E) since the pairs of measurements, which give the cross-products from which covariances are calculated, are as likely to be affected in opposite ways as the same way by non-heritable agencies. In diallel crosses, however, each parental value occurs also as a member

of its own progeny array, hence the covariance of offspring on parents contains square terms as well as cross products, the former having a non-heritable component that does not cancel out. In an $n \times n$ diallel cross, of the n^2 progenies, n are of this type, so that the covariance has a non-heritable component $1/n^{th}$ of the mean variance of arrays, i.e., E_3.

When reciprocal F_1 families are grown two methods of analyzing the data are available in the absence of any significant reciprocal differences. Firstly, all the statistics may be calculated for both male and female arrays or their array means and then averaged. In this case the non-heritable components have the coefficients given above. Alternatively, reciprocal F_1 families may be averaged prior to the calculations, this method being essential when significant reciprocal differences are present. This necessitates adjustments in the coefficients of some of the non-heritable components as each F_1 cross is the mean of two plots while each parental mean is represented by only one plot. If n equals the number of parental lines in the diallel cross, the E components of the variance and covariance are:—

$$V_{1L1}, \quad \frac{n+1}{2n}E_2 \qquad V_{0L1}, \quad \frac{1}{2n}E_2 \qquad \text{and} \qquad W_{0L01}, \quad \frac{1}{n}E_2$$

Table 5 gives the estimates of these three statistics, corrected for the appropriate E components, along with the estimate of E_2 from which these corrections have been calculated. These estimates have been found separately for each block and then averaged.

TABLE 5

Estimates of the second degree statistics obtainable from a diallel cross for the N. rustica experiment.

Character		Statistic						
		V_{1L1}	V_{0L1}	W_{0L01}	D	D_W	H	E_2
Height	1951	49.3107	31.5812	34.0085	126.3248	68.0170	70.9180	2.4885
	1952	48.0110	28.5165	31.5141	114.0660	63.0282	77.9780	7.7086
Flowering time	1951	22.0089	13.0203	34.7186	52.0812	69.4372	27.9544	2.5609
	1952	13.2525	7.8853	10.0018	31.5412	20.0036	21.4688	3.3738
Leaf length	1951	5.5397	4.3739	7.4065	17.4956	14.8130	4.6632	2.1706
	1952	3.7602	3.8592	6.8291	15.4368	13.6582	−0.3960	2.8814

No estimates of error are given as only exact fit to observation can be obtained from the number of statistics available, and the estimate from block differences, if obtained, would only be based on one degree of freedom.

Comparison of H with either D or D_W gives little idea of the average degree of dominance since, when $u \neq v \neq \frac{1}{2}$, D includes some effect of h, and H is correspondingly less than the summed effects of all the squared h deviations. (The relationship between D and Σd^2 has been worked out by MATHER (1949) for a range of gene frequencies in the parental lines.) A value of H,

however, of the same order as D and D_W does suggest that dominance is operative in these progenies, e.g., in height.

A better idea of the average degree of dominance can be obtained from a comparison of V_{1L1} and W_{0L01} after subtracting the environmental components. Thus V_{1L1} can be written as

$$\Sigma uv[d^2 + h^2 + 2(v - u)dh]$$

while

$$W_{0L01} = \Sigma uv[2d^2 + 2(v - u)dh]$$

Therefore if $d = h$ then $V_{1L1}/W_{0L01} = 1$, while if $d > h$ $V_{1L1}/W_{0L01} < 1$, or if $d < h$ $V_{1L1}/W_{0L01} > 1$.

For independent genes with $u = v = \frac{1}{2}$ for all the genes the graph connecting V_{1L1}/W_{0L01} and h^2/d^2 is a straight line. If $u \neq v \neq \frac{1}{2}$ the line becomes a curve but always passes through the two points ($h^2/d^2 = 0$, $V_{1L1}/w_{0L01} = \frac{1}{2}$) and ($h^2/d^2 = 1$, $V_{1L1}/w_{0L01} = 1$). This estimate of the average degree of dominance has one advantage over the other estimates in that it does not suffer from errors arising from partitioning D and H which are correlated in all the available statistics. The ratio of V_{1L1}/W_{0L01} is given for the various characters in tables 6 and 8. Discussion of these ratios will be reserved until later when other estimates of dominance are given.

A further useful second degree statistic can be obtained from diallel crosses, namely the variance of the parental means around the mid parent. The expectations for this statistic in terms of a single gene are

$$\begin{aligned} V_{0L0} &= u_a d_a^2 + v_a d_a^2 - [(u_a - v_a)d_a]^2 \\ &= 4u_a v_a d_a^2. \end{aligned}$$

Summing over a number of independent genes this becomes

$$\Sigma\, 4uvd^2.$$

This of course has a non-heritable component E_2.

Now terms in $\Sigma\, uvd^2$ are present in all the available statistics. Furthermore, this new statistic provides sufficient equations for obtaining an independent estimate of $\Sigma\, dh$ terms, leaving terms in $\Sigma\, uvd^2$ and $\Sigma\, uvh^2$, which, having similar coefficients, are directly comparable. D can now be defined to cover both variances and covariances as $4\Sigma\, uvd^2$, H_1 as $4\Sigma\, uvh^2$, $H_2 (= H)$ as $16\Sigma\, u^2v^2h^2$ and F as $8\Sigma\, uv(u - v)dh$, the last being the only statistic that can take sign.

The compositions of the variances and covariance in terms of the newly defined D, H_1, H_2 and F are

$$\begin{aligned} V_{0L0} &= D \\ V_{1L1} &= \tfrac{1}{4}D + \tfrac{1}{4}H_1 - \tfrac{1}{4}F \\ V_{0L1} &= \tfrac{1}{4}D + \tfrac{1}{4}H_1 - \tfrac{1}{4}H_2 - \tfrac{1}{4}F \\ W_{0L01} &= \tfrac{1}{2}D - \tfrac{1}{4}F. \end{aligned}$$

If $u = v = \frac{1}{2}$, D becomes Σd^2 and $H_1 = H_2 = \Sigma h^2$ as obtained for a cross between two inbred lines (MATHER 1949).

The estimates of these parameters are given in table 6 for height 1951 and 1952 along with the estimates of dominance H_1/D and V_{1L1}/W_{0L01}.

Again there is no useful estimate of error, since five parameters (including E_2) are estimated from five equations. Nevertheless, the consistency of all the statistics over years and over blocks, within years (not given in table) allows one to interpret these results with confidence even in the absence of standard errors of the components.

Both the ratios H_1/D and V_{1L1}/W_{0L01} suggest that overdominance is present, i.e., H_1/D and $V_{1L1}/W_{0L01} > 1$, but it must be remembered that these ratios are measuring dominance on different scales, although they can be converted to the same scale when $u = v = \frac{1}{2}$. H_1/D then becomes $\Sigma h^2/\Sigma d^2$ while V_{1L1}/W_{0L01} becomes $\Sigma h^2/\Sigma 2d^2 + \frac{1}{2}$. Thus subtracting half from the ratio V_{1L1}/W_{0L01} and doubling the remainder should lead to a value approaching H_1/D. In the absence of information concerning u and v this transforma-

TABLE 6

Estimates of the parameters D, H_1, H_2 and F and dominance H_1/D and V_{1L1}/W_{0L01} for height 1951 and 1952, including and excluding arrays 1, 2 and 4.

	Year	Parameters					
		D	H_1	H_2	F	H_1/D	V_{1L1}/W_{0L01}
All arrays	1951	42.3092	103.5180	70.9178	−51.4151	2.4467	1.45
	1952	43.0078	108.9958	77.9780	−40.0402	2.5343	1.68
Omitting arrays 1, 2 and 4	1951	19.5548	11.8878	11.1522	− 6.1518	0.6079	0.8306
	1952	22.0921	19.0100	9.7440	+18.0896	0.8605	0.8716

tion has in all cases given results that agree more closely with H_1/D, e.g., height 1951, $H_1/D = 2.45$, $2(V_{1L1}/W_{0L01} - \frac{1}{2}) = 1.90$, 1952, $H_1/D = 2.53$, $2(V_{1L1}/W_{0L01} - \frac{1}{2}) = 2.36$.

RELATIONSHIP BETWEEN ARRAY VARIANCE AND ARRAY COVARIANCE

Returning to our consideration of a single gene we find that the variance of the *AA* array (V_{AA}) is $u_a v_a (d_a - h_a)^2$ and the covariance (W_{AA}), $2u_a v_a d_a (d_a - h_a)$. Similarly for the *aa* array, $V_{aa} = u_a v_a (d_a + h_a)^2$, while $W_{aa} = 2u_a v_a d_a (d_a + h_a)$. Thus substituting *aa* for *AA* changes both V and W by the same amount, $4u_a v_a d_a h_a$; or to put it in another $V_r - W_r$ is constant over arrays. Extending this to a number of independent genes gives the same result, viz., that $V_r - W_r$ is constant. The regression of array covariance and array variance should therefore give a straight line of slope 1. The only limiting condition is that each gene should be independent of all others in its action. It should therefore be possible to detect genic interaction by deviations from this expected regression line.

A further property of this regression line can be utilized in the study of

the dominance relations. When the variance of an array is zero, its covariance is $\frac{1}{4}D - \frac{1}{4}H_1$, i.e., $W_r - V_r = \frac{1}{4}D - \frac{1}{4}H_1$. Now we have a test of significance to see whether the regression line passes through the origin, i.e., whether D is greater than, equal to, or less than H_1 in value, so that we can test for significant overdominance, complete dominance or underdominance. Furthermore, if there is significantly less than complete dominance, it is possible to test whether H_1 is significantly different from zero, since if $H_1 = 0$, there is no significant regression, i.e., the variances and covariances of all arrays are identical, within experimental error, being estimates of a single point where $W_r/V_r = 2$ (fig. 1).

Before proceeding to analyze the data we will first examine the ways in which genic interaction might be expected to result in deviations from a uniform regression of unit slope. Interaction affecting a proportion of the

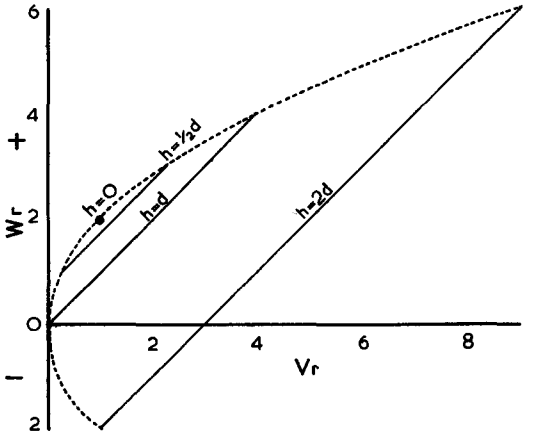

FIGURE 1.—The theoretical regressions of W_r on V_r for various degrees of dominance, i.e., h/d ratios. The curve (broken line) joins the points of the arrays whose common parents contain all the dominant or all the recessive allelomorphs.

members of an array will increase the variance of that array while the parent offspring covariance will fall. Examination of figure 1 shows that any increase in variance relative to covariance will move the W_r, V_r graph to the right giving an apparent increase in dominance. At the same time, this increase in V_r relative to W_r will have a proportionately greater effect as we move from the point of origin along the regression line. As a result not only will the mean of the regression move towards the right (indicating higher dominance) but its slope will fall below the expected value of one.

The extent to which this happens and hence the ease with which it can be detected, from its effect on the slope, will depend on the ratio of interacting to non-interacting members in the arrays. When this ratio is high, no difficulty, of course, will be met with. If, however, it is low, the deviation of the few array points affected by interaction may serve only to inflate the error variance of the regression line to such an extent as to obscure the deviation

from unit slope. The latter case may be dealt with by omitting those arrays which lead to an improvement of the average slope and at the same time reduce the standard error of the regression coefficient.

The 1951 height data gave a regression coefficient of 0.4079 ± 0.04611 and the 1952, 0.5963 ± 0.1282, both of which differ significantly from a slope of one (fig. 2). There was no block × regression interaction, neither were there differences between block means in either sets of data. Examination of the graph of W_r against V_r showed that in both years the arrays fell into two groups. Most of the points fell on a straight line of slope 0.6–0.7, while three points for arrays 1, 2 and 4 (common parents, 2, 5 and 12) deviated more than the rest from a slope of one. This is compatible with an incidence of interaction highest in arrays 1, 2 and 4 but also occurring to a smaller extent

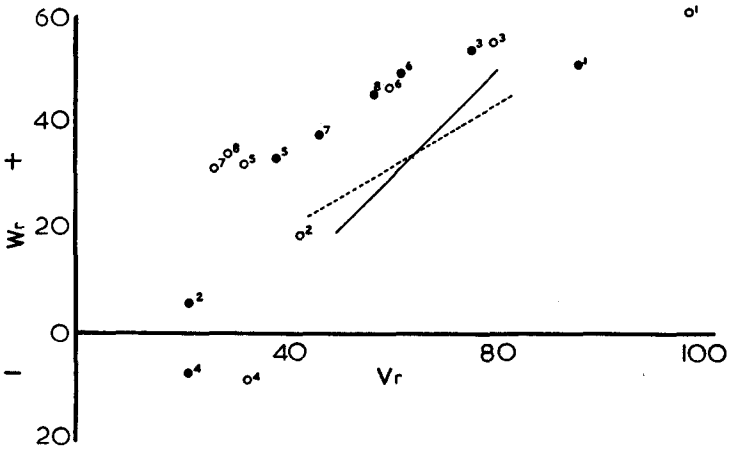

FIGURE 2.—The regression of W_r on V_r for height 1952. The best fitting regression line has a slope of 0.5963 ± 0.1282 (broken line). The complete line is the theoretical slope of 1 expected if there is no genic interaction. The points for arrays 1, 2 and 4 obviously deviate most from the best fitting line. The points for the duplicate blocks are differentiated by using full and open circles.

in the other arrays, possibly being confined to the families in these arrays having either 1, 2 or 4 as one parent. If this is correct then the removal of the progenies of 1, 2 and 4 from the data should result in the remainder fitting a regression coefficient of 1.

In the 1952 data this proved to be the case (fig. 3). Removal of the progenies of parents 1, 2 and 4, and only these three, gave a regression coefficient of 0.7806 ± 0.2217, which is not significantly different from a regression coefficient of 1. In the 1951 data, however, removal of the progenies of parents 1, 2 and 4 still did not account for all the genic interaction, the regression still being less than 1 ($P = 0.02$–0.01). Examination of the graph of W_r against V_r after removing the progenies of 1, 2 and 4 showed that apart from array 8 (common parent 41) all the other arrays fell on a straight line of approximately unit slope. This remaining interaction could be traced to the F_1 family

5×8 since the removal of the progenies of either 5 or 8, while not affecting the other array points, brought the alternative array into line. Furthermore omitting the array points of 5 and 8 but leaving their progenies in the other arrays gave a regression coefficient that was not significantly different from 1 ($b = 0.6878 \pm 0.2790$). The 1951 results agree therefore quite well with those of 1952.

It must be borne in mind that although removal of the progenies of arrays 1, 2 and 4 removes all the genic interaction in 1952 and most of it in 1951, it must obviously be possible to remove interaction by omitting the progenies of the parents with which 1, 2 and 4 interact in the F_1's. Since, however, the intensity of interaction is highest in these three arrays, removal of the interaction by the alternative means would presumably require the omission of a larger number of arrays.

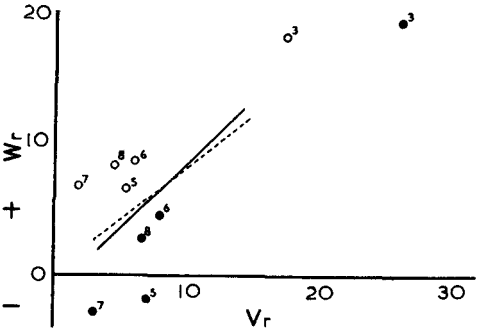

FIGURE 3.—The regression of W_r on V_r for height 1952 after removing the progenies in arrays 1, 2 and 4, i.e., the offspring of parents 2, 5 and 12. The best fitting line now has a slope of 0.7805 ± 0.2217, which does not differ significantly from the theoretical slope of 1 (the complete line). The array points are arranged along this line in the order of the number of dominant allelomorphs in their common parent, array 3 having least and array 7 the most. The points for the duplicate blocks are differentiated by using full and open circles. The slopes for the two blocks have significantly different means, i.e., different degrees of dominance in the two blocks.

We can now examine in the residual progenies the means and second degree statistics for height free from complications arising from the presence of interaction in the data. Unfortunately, in removing the interacting arrays the number of families has fallen from 64 to 25, hence the analysis is not so comprehensive nor the agreement between years so good as in the complete data. Nevertheless, certain consistencies are obvious over the two years. Firstly examination of the tables of progeny means (tables 1 and 2) shows that the removal of arrays 1, 2 and 4 (common parents 2, 5 and 12) removes all the families whose means are higher than their own better parent mean. Furthermore, parents 5 and 12 give good F_1 families with the same other parents but not with each other, while parent 2 gives high means with both 5 and 12 and with a number of other parents with which 5 and 12 give mediocre F_1's. Thus most of the parents give good F_1 families with either 2 or 5 and 12.

This agrees remarkably well with the regression analysis if we accept the view that the high means are the result of genic interaction of a type comparable with the complementary genic action of classical genetics. It would appear, therefore, that in these crosses heterosis is not the result of overdominance but of complementary gene interaction.

Secondly if we now examine the estimates of the second degree parameters D and H_1 from the residual progenies we find that the " overdominance " has disappeared and that we now have slightly less than complete dominance in both years (table 6).

If the deviation of $\frac{1}{4}D - \frac{1}{4}H_1$ from 0 is tested by the regression method we find that in both years there is no significant deviation from complete dominance. While admitting the possibility that the three arrays which were removed as the main source of genic interaction might well be the only arrays that also show overdominance, it would appear to be more reasonable to assume that the drop in apparent dominance is due almost entirely to the removal of the interaction.

A further test of interaction, though one not so clear as that afforded by the regression of W_r on V_r because of other disturbances it also detects, has been devised in collaboration with PROFESSOR K. MATHER. If we subtract half the deviation of the common parent of an array from the overall mid-parent, from the array mean the following expectations are obtained:

Array	array mean − ½deviation of common parent from parental mean
AA	$u_a d_a - v_a h_a - \frac{1}{2}u_a d_a = \frac{1}{2}u_a d_a + v_a h_a$
aa	$-v_a d_a + u_a h_a + \frac{1}{2}v_a d_a = \frac{1}{2}u_a d_a + u_a h_a$

If $u_a = v_a = \frac{1}{2}$, then these expectations will not differ significantly from one another unless there is genic interaction. Under these conditions the values will be homogeneous when tested against an error variance found as $E_2(1 + 1/n)$, where n = number of families in each array, as a χ^2 for n degrees of freedom. Homogeneity may break down when $u \neq v$, but even then the removal of any interaction which is present should at least lead to a significant lowering of the heterogeneity sum of squares, unless of course removing the interaction seriously disturbs the u, v distribution.

Applying this F_1 scaling test to the 1951 height data gives a heterogeneity SS of 86.3792 for 8 degrees of freedom which is significant when compared with the error MS of 1.0790. After removing the progenies of arrays 1, 2 and 4 the heterogeneity SS drops to 4.3538 and the error MS becomes 1.2450, so that the remaining progenies are homogeneous ($P = 0.8$–0.7). Hence this method, within its limitations, detects the same sources of interaction or heterogeneity as the regression analysis.

THE RELATIVE FREQUENCIES AND DISTRIBUTION OF ALLELOMORPHS IN THE PARENTAL LINES

An estimate of the average value of the product uv for all the gene differences in the parental lines can be obtained from the ratio $\frac{1}{4}H_2/H_1 =$

$4\Sigma u^2v^2h^2/4\Sigma uvh^2$. This, of course, will be subject to an error depending on the standard errors of H_1 and H_2 which are unknown. Since this estimate of $\overline{uv}$ is obtained solely from terms in Σh^2 it will of necessity only cover the frequencies of the allemorphs of genes that are exhibiting some degree of dominance. Thus it provides no evidence about the distribution of allelomorphic pairs exhibiting no dominance. The results for height 1951 and 1952 are $\overline{uv} = 0.1713$ and 0.1788 respectively.

The sign of $F[\Sigma 8uv(u-v)dh]$ depends on the sign of $(u-v)h$ which is positive for an excess of dominant allelomorphs and negative for an excess of recessives. The sign of F therefore is an indicator of the relative frequencies of dominant and recessive allelomorphs. For height in both 1951 and 1952 F was negative, hence there is greater frequency of recessive than dominant allelomorphs in the parental lines.

If we examine figure 1 we find that the ratio of W_r/V_r for the complete dominant and complete recessive array is characteristic of the degree of dominance (table 7). By complete dominant and complete recessive array we

TABLE 7

The relationship between the covariance and variance of the complete dominant and complete recessive arrays for different degrees of dominance.

Degree of dominance h/d	W_r/V_r	
	Complete dominant array	Complete recessive array
0	2.0	2.0
½	4.0	1.3
1	0.0	1.0
2	−2.0	0.6

mean arrays whose common parents carry all the dominant and recessive allelomorphs respectively, whether these are in the direction of increase or decrease in the size of the character under consideration.

Thus given the degree of dominance from the H_1/D, V_{1L1}/W_{0L01} or $2(V_{1L1}/W_{0L01} - ½)$ it is possible to see whether the complete dominant and complete recessive arrays are present in the data. Further, if we assume that all the gene differences are equal, i.e., $d_a = d_b = d_c \ldots d_n$, then the other arrays will be scattered along the regression line, connecting array covariance with array variance, between the complete dominant and recessive arrays, in proportion to the relative numbers of dominant and recessive allelomorphs in the common parent of the array (figs. 2, 3 and 4).

NUMBER OF EFFECTIVE FACTORS

So far this discussion has been in terms of individual genes but this terminology has been solely for ease of description. In biometrical genetics we are, of necessity, forced to work in terms of the average effects of all the gene differences affecting the character under consideration. These may, however, be resolved into smaller groups of genes whose effects are statistically separa-

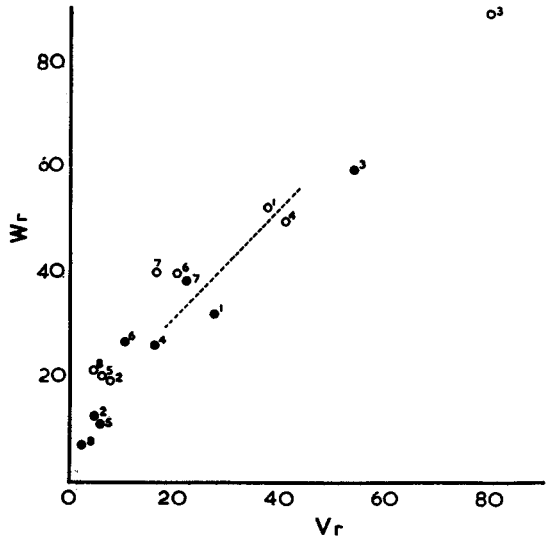

FIGURE 4.—The regression of W_r on V_r for flowering time 1951. This is almost a perfect fit with the regression slope of 1 expected in the absence of genic interaction. The array points are distributed along the regression line in the order of the number of dominant allelomorphs carried by their common parent, array 3 having least and array 8 the most. The points for the duplicate blocks are differentiated by using full and open circles.

ble on certain assumptions and whose numbers are thus capable of estimation. These smallest units of hereditary material recognizable in biometrical genetics are referred to as effective factors, which may or may not be synonymous with single genes in any particular case (MATHER 1949). In the estimations of the number of effective factors from crosses between two inbred lines described by MATHER tightly linked genes appear as one effective factor while independently segregating genes appear as separate effective factors. In the estimates from F_1 data described below there is a comparable disturbance resulting from the association of genes in the parental lines. Thus if two genes appear in all the parental lines either in reinforcement (the two d's with the same sign) or opposition (the d's taking opposite signs) then they appear as one effective factor. If, however, genes appear in all combinations in the parental lines then the apparent number of effective factors will depend on the relative frequencies of the reinforcing and opposing combinations, the maximum value being obtained when they are present in equal frequencies.

The estimate of the number of effective factors employed here is essentially the same as that described by MATHER (1949), but is based on the H, rather than the D, statistics. The difference between the overall progeny mean and the parental mean is 2Σ uvh. Therefore, the square of this difference divided by $\frac{1}{4}H_2(=4\Sigma u^2v^2h^2)$ gives an estimate of the number of effective factors

$$\text{i.e.,} \quad K = \frac{[2\Sigma uvh]^2}{4\Sigma u^2v^2h^2}.$$

Now this will be a minimal estimate for the same reasons as those given by MATHER (1949), i.e., it assumes equality and absence of oppositions in the h's for the various genes; both assumptions being unlikely to apply in practice. Unfortunately, there is no obvious way of obtaining an estimate from the D statistic which is only based on the first assumption and hence is likely to be more reliable. As in the case of the $\overline{uv}$ estimates, these estimates from the H statistics tell us nothing about genes that do not show some degree of dominance, i.e., genes that do not contribute to the H statistics.

The values of K for height are 0.85 for 1951 and 1.67 for 1952. These estimates are disappointingly small, but they are nevertheless of the same order as those obtained by MATHER (1949) from the analysis of a wide range of material. It would appear from these results that the basic assumptions on which this estimation is based are far from being realized in the present data and for that matter in any other data so far analyzed by these methods.

ANALYSIS OF THE OTHER CHARACTERS OF *N. rustica* STUDIED

A summary of the results of these analyses when applied to flowering time, leaf length and number of capsules per plant is given in table 8.

In the 1952 diallel experiment the cross 1×3 behaved anomalously for flowering time so that in the analysis it appeared as genic interaction. When the observed flowering time for this one family was discarded and a new value estimated by the missing plot technique the interaction disappeared, giving the results shown in table 8.

Although flowering time agrees in the degree of dominance and absence of genic interaction for the two years the value of D, and hence of F, differs considerably. When one considers, however, the effect of weather conditions on the rate of flowering, this year to year variation is perhaps not so surpris-

TABLE 8

Analysis of flowering time, leaf length and number of capsules in 1951 and 1952.

Statistic	Flowering time		Leaf length		No. of capsules, 1951
	1951	1952	1951	1952	
D	102.5686	16.5925	14.9831	11.3944	12,597.7664
H_1	51.7294	8.3950	5.5164	−1.7628	6,913.8168
H_2	35.9544	7.7408	4.6632	−0.7924	4,548.6752
F	+16.5657	−0.3950	−1.6590	−4.5276	−8,197.7008
H_1/D	0.5043	0.5052	0.3682	....	0.5488
V_{1L1}/W_{0L01}	0.6339	0.7555	0.7480	0.5506	0.8361
$2(V_{1L1}/W_{0L01} - \frac{1}{2})$	0.2678	0.5110	0.4960	0.1101	0.6722
F_1 scaling test	Homogeneous both years (P = 0.10–0.05)		Heterogeneous both years (P < 0.001)		Heterogeneous (P < 0.001)
b W_r/V_r	1.0520 ± 0.0903	0.9052 ± 0.1078	0.6871 ± 0.1820	0.7559 ± 0.1792	Not sig.
Dev. from b = 1	P = 0.6–0.5	P = 0.4–0.3	P = 0.2–0.1	P = 0.2–0.1	....
Sig. $\frac{1}{4}D - \frac{1}{4}H_1$	D > H_1 P = 0.01–0.001		D > H_1 P = 0.01–0.001	D > H_1 P < 0.001	....
$\overline{uv}$	0.1738	0.2309	0.2113	....	0.1645
K	1.96	1.77	0.055	....	0.095

ing. The genetic picture is essentially the same for the two years. Apart from the consistencies in degree of dominance and absence of genic interaction, mentioned above, we also find that the order of the array points along the V_r, W_r regression line is identical over the two years. The regression analysis further shows that $\frac{1}{4}D - \frac{1}{4}H_1$ is significantly positive for both years, so that D is significantly greater than H_1. Thus for flowering time dominance is incomplete: H_1 is shown to be significantly greater than 0 by the significant regression coefficients for W_r against V_r.

The ratio of W_r/V_r for the individual arrays suggests that the complete recessive parent is present. Thus array 3 gives $W_r/V_r = 1.12$. Taking $H_1/D = 0.5$ (table 8) then the complete recessive array should have a W_r/V_r of 1.18.

The number of effective factors is of the same order as for height.

Although the D's are of the same order for leaf length in both years the H terms show considerable divergence. Since, however, a negative H_1 and H_2 is theoretically impossible (both terms being quadratic in h), the negative values for 1952 could only arise by error variation in its effects on the negative correlation between D and H_1. The V_{1L1}/W_{0L01} and $2(V_{1L1}/W_{0L01} - \frac{1}{2})$ estimates of the degree of dominance do not suffer from such disturbances but they are nevertheless lower than the similar estimates for 1952, so suggesting that there is a genuine lower degree of dominance in the 1952 data. In both years D is significantly greater than H_1; dominance is real but incomplete.

The information from the regression analysis is remarkably consistent over the two years. Both regression coefficients fail to depart significantly from one, but examination of the graphs shows that although seven of the arrays fall on a line of slope $\simeq 0.9$, the array of parent 2 deviates considerably from this line resulting in an overall regression coefficient of $\simeq 0.7$. This suggests a high intensity of interaction confined to a small proportion of the arrays, so that the presence of the interaction is masked by the high standard error of the regression coefficient, resulting from the wide scattering of the points for the arrays showing the interaction. In this case, however, there is no difficulty in detecting the presence of genic interaction or tracing it to its source. As in the case of height, the array of parent 2 contains all the families whose leaf lengths are greater than their own best parent.

Since array 1 contributed a large proportion of the genic interaction for height, it would appear to be likely that the interaction in leaf length results from the action of the same gene complex. If this is true, then the intensity of the interaction is less marked for leaf length since it shows in only one array and not three as in the case of height.

Although the number of capsules per plant was recorded only in the 1951 experiment and then only for the four arrays 2, 4, 5 and 8 it is of interest in that this character also shows indications of genic interaction. The regression of array covariance on variance was not itself significant. The graph showed, however, that arrays 5 and 8 (parents 14 and 41) from both blocks fell on a straight line of slope approximately 0.7, while the array points for parents 5 and 12 deviated to the lower side of this line and occupied the same relative

positions as in the similar graph for height. Since only four arrays were involved, this could not be pursued further to see if the removal of the progenies in arrays 1 and 4 resulted in the remaining arrays giving a slope of 1. This similarity to height bears out the earlier observation that the means of the two characters were highly correlated.

DISCUSSION

Having tested the diallel analysis on a variety of material the results suggest that it may prove to be a powerful method for obtaining a rapid, overall picture of the genetical structure of a large number of parental lines. Although it does not allow an estimate of the standard errors of the individual statistics in its present form, the regression analysis at least makes it possible to test the significance of H_1/D, the statistic in which we are primarily interested. Furthermore, by raising the F_2 and subsequent generations by selfing each F_1 family and their progenies one can obtain sufficient statistics to allow estimates of the standard errors by the least squares method (MATHER 1949).

The regression analysis provides a simple method of detecting genic interaction which, as far as we have gone, shows good agreement with the standard scaling tests which detect non-additivity of gene action (MATHER 1949). Furthermore, this analysis has proved the presence of genic interaction in data where its presence had not been previously demonstrated although analyzed as far as already existing methods would allow (JINKS and HAYMAN 1953; JINKS 1954).

The conclusions of this analysis lend considerable support to the view that genic interaction is responsible for a large proportion of the observed heterosis in the F_1 progeny of crosses between inbred lines. It is impossible from these analyses to assess the relative contribution of other causes of heterosis, namely overdominance and the bringing together in one plant of complementary sets of genes in opposition which individually show only complete dominance or even lower degrees of dominance. One can say, however, that in the data analyzed the most outstanding F_1 families occur where genic interaction is superimposed on the other causes to give what has been termed "high combining ability."

The question of the number of effective factors is not quite so satisfactory. Firstly, our estimates must of necessity come from the H statistics which are undoubtedly inferior to those obtained from the D statistics. Secondly, any deviation from a random association of gene differences throughout the parental lines leads to a further minimizing of the estimate. In the *N. rustica* experiment the estimates never exceed two effective factors. This is undoubtedly a gross underestimate of the true position. At least two gene complexes must be postulated to explain the presence of genic interaction in height yet genic interaction is only a small proportion of the total genetical variation present as the analysis after excluding the interaction clearly shows.

Although the analysis has been applied only to diallel experiments involving plants it is equally applicable to animal breeding programs of this type. Fur-

thermore, it probably holds more advantages for the animal breeder than the plant breeder, with whom economic considerations in planning the programs do not play so important a part. The most important advantage, of course, is the time required: estimates can be obtained solely from the F_1 generations, and if confirmation of the results is required it can be supplied either from an F_2 or backcross generation. This would automatically reduce the time required to obtain estimates which is of great importance in slow maturing animals. Relieved of the burden of keeping experimental stocks for long periods and an ever expanding crossing program, the breeder can afford to use larger numbers of stocks in the initial diallel cross. In animals, of course, one would have to substitute herds for inbred lines of plants. These are less likely to be homozygous than the plant material but this would not be serious if the variation within stocks was of a much lower magnitude than that between them. The residual variation would be presumably randomly distributed throughout a herd and hence it could be investigated by using the random mating analysis (MATHER 1949). This differs from the diallel analysis only in the composition of the covariance. Herds are, however, presumably homozygous for the genes controlling the character that distinguish them and hence could be investigated by the diallel cross analysis.

As attention is increasingly turning to the possibilities of utilizing the F_1 heterosis from crosses between inbred lines as a means of increasing the yield of crop plants and animals, it is essential to gain some idea of the genetical structure of heterosis, not only from the theoretical viewpoint but also from considerations of the efficacy of choosing and testing suitable parental lines for their ability to give this heterosis. The diallel analysis described here is an attempt to provide an analytical method based on Mendelian concepts for use as a tool in investigations of this type.

SUMMARY

A method of analyzing quantitative data from diallel crosses based on the partitioning of second degree statistics, such as variances and covariances, has been developed, by extending the system described by MATHER (1949) for crosses between two inbred lines. This analysis has been successfully applied to a diallel cross between *Nicotiana rustica* varieties.

The method allows estimates of D and H_1 which are weighted terms in Σd^2 and Σh^2 respectively (MATHER 1949), so that H_1/D gives a direct estimate of the degree of dominance. The ratio of positive to negative allelomorphs in the parental lines can be obtained, as well as the ratio of dominant to recessive allelomorphs, from two further statistics, H_2 and F.

The regression of array covariance on variance has an expected slope of 1. Agreement with this expectation is found for flowering time in the *N. rustica* data. The remaining data all show varying degrees of deviation from this expected slope, the cause of which has been traced to genic interaction in some of the F_1 progenies. Removal of the progenies responsible in all cases gave the expected slope of 1.

When $V_r = 0$, then $W_r = \frac{1}{4}D - \frac{1}{4}H_1$, hence the regression analysis gives a test of significance for the level of dominance, i.e., whether W_r is equal to, greater than or less than 0 when $V_r = 0$.

Using this analysis it has been shown that genic interaction is responsible for all the apparent overdominance and heterosis in height and leaf length in *N. rustica*.

ACKNOWLEDGMENTS

I am indebted to PROFESSOR K. MATHER and MR. E. L. BREEZE for helpful discussion and criticism throughout the course of the work described here. This work was carried out while in receipt of a Research Studentship of the Agricultural Research Council.

LITERATURE CITED

BATEMAN, A. J., and K. MATHER, 1951 The progress of inbreeding in barley. Heredity **5**: 321–348.

FABERGÉ, A. C., 1936 The physiological consequences of polyploidy. J. Gen. **33**: 365–382.

JINKS, J. L., 1954 The genetical basis of heterosis. Maize Gen. Coop. News Let. **28**: 47–50.

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

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

MATHER, K., and A. VINES, 1953 The inheritance of height and flowering time in a cross of *Nicotiana rustica*. Quantitative Inheritance, H.M.S.O. London, 49–79.

YATES, F., 1947 Analysis of data from all possible reciprocal crosses between a set of parental lines. Heredity **1**: 287–301.