HEREDITY VOLUME 10 PART 1 APRIL 1956

THE F_2 AND BACKCROSS GENERATIONS FROM A SET OF DIALLEL CROSSES

J. L. JINKS

Agricultural Research Council Unit of Biometrical Genetics, Department of Genetics, University of Birmingham

Received 20.viii.55

1. INTRODUCTION

THE theory and analysis of F_1 and parental data from a set of diallel crosses has been described and illustrated using the results of our own crosses between eight varieties of *N. rustica* and published work on other species (Jinks and Hayman, 1953; Jinks, 1954*a*, 1954*b*, 1955; Hayman, 1954). The method, which is an extension of that developed by Mather (1949), permits the estimation of parameters for additive, dominance and environmental effects and allows the recognition of non-allelic interaction. The analysis has now been extended to cover the F_2 and backcross generations of a diallel set of crosses and is illustrated by the same diallel cross, the F_1 analysis of which was described in the earlier paper (Jinks, 1954*b*).

Before proceeding with the analyses of the later generations we will briefly restate the conclusions reached from the analysis of the parents and F_1s. Of the two main characters followed, the time of flowering showed incomplete but significant dominance and a complete absence of non-allelic interaction for the two seasons 1951 and 1952. The other character, final height, appeared to show overdominance but this was traced to a spurious inflation of the dominance component by non-allelic interaction of a type comparable with the complementary genes of classical genetics. After omitting from the analyses the crosses showing significant non-allelic interaction, all of which occurred in arrays 1, 2 and 4, re-analysis revealed only complete dominance.

2. MATERIAL AND METHOD

The 1951 and 1952 experimental layout has already been described in detail. In including F_2s and backcrosses as well as parents and F_1s the 1953 experiment was identical in design with that of 1952 so that we have the parents and F_1s of the diallel cross for three consecutive seasons and the F_2 and backcross generations for the last two.

3. THE ANALYSIS OF MEANS

(i) *Genotype-environment interaction*

In the first paper of this series (Jinks, 1954*b*) one method of detecting genotype-environment interaction was discussed, namely the within plot variance of non-segregating parental and F_1 families, *i.e.* E_1. The analysis showed that for the two seasons 1951 and 1952

there was no significant difference between the mean E_1 for parents and that for the F_1s. Furthermore, the significant heterogeneity of the E_1s over families, *i.e.* both parents and F_1s, was attributable to the varying reaction of the different genotypes to environmental differences. Since then estimates of the E_1s for 1953 have been obtained and a second genotype-environment interaction, the variance of family means over the three seasons, has been investigated. The analysis of these two statistics has been presented and discussed by Jinks and Mather (1955) in relation to the stability in development of homozygotes and heterozygotes.

(ii) *Scaling tests*

The expected means of each generation following a cross between two inbred lines in the absence and presence of non-allelic interaction are given by Mather (1949) and Hayman and Mather (1955) respectively. In the absence of interaction the following relationships hold :—

$$\bar{F}_2 = \tfrac{1}{4}\bar{P}_1 + \tfrac{1}{4}\bar{P}_2 + \tfrac{1}{2}\bar{F}_1$$
$$\bar{B}_1 = \tfrac{1}{2}\bar{P}_1 + \tfrac{1}{2}\bar{F}_1$$
$$\bar{B}_2 = \tfrac{1}{2}\bar{P}_2 + \tfrac{1}{2}\bar{F}_1$$

In the presence of non-allelic interaction, however, these equalities no longer hold. Although this test can be made on the three equalities independently as described by Mather (1949) and Mather and Vines (1952), a more convenient method which combines these three tests has been developed by Cavalli (1952) and used for the detection of non-allelic interaction by Jinks (1954*a*, 1955). This test consists of estimating by weighted least squares the three parameters Σd, Σh and M (the mid-parent) taking as weights the reciprocals of the squared standard errors of the generation means. The expectations of the generations in terms of these three parameters are as follows :

$\bar{P}_1 = M + \Sigma d$
$\bar{P}_2 = M - \Sigma d$
$\bar{F}_1 = M + \Sigma h$
$\bar{F}_2 = M + \tfrac{1}{2}\Sigma h$
$\bar{B}_1 = M + \tfrac{1}{2}\Sigma d + \tfrac{1}{2}\Sigma h$
$\bar{B}_2 = M - \tfrac{1}{2}\Sigma d + \tfrac{1}{2}\Sigma h$

(Σd and Σh here refer to the balance of genes in opposition, *e.g.* Σd is the net balance of the alleles with a positive contribution to the mean (+d) minus the contribution of the alleles with a negative effect (−d)).

The estimated parameters can be tested for consistency over generations by comparing the observed and expected generation means, in this particular case as a χ^2 for three degrees of freedom.

(a) *Height.* Our earlier analyses of height using the regression of array covariance on array variance suggested that we should find non-allelic interaction in arrays 1, 2 and 4, *i.e.* crosses involving lines 2, 5 and 12 as one parent. They further suggest that while parent 2 interacts with 5 and 12 the latter do not interact with one another. The scaling tests fully confirm these expectations. For both the 1952 and 1953 seasons all crosses, apart from the cross 3×8, exhibiting significant interaction on the scaling test, have one of these three lines as a parent (tables 1*a* and 1*b*).

One difference in detail over the two seasons is that two crosses 5×2 and 5×4 which show no significant interaction in 1952 (P = 0·1-0·05 and 0·7-0·5 respectively) do so in 1953.

Having separated all the crosses into two groups, the interacters and non-interacters, we can now return to the individual scaling tests to examine the ways in which the three equalities given earlier have failed. For this purpose we have

TABLE 1

The incidence of significant non-allelic interaction for height in the 8×8 diallel in 1952 and 1953 and flowering time in 1953 as revealed by the joint scaling test

	Height														Flowering time						
	1a 1952							*1b* 1953							*1c* 1953						
Arrays	2	3	4	5	6	7	8	2	3	4	5	6	7	8	2	3	4	5	6	7	8
1	+	−	+	−	−	−	+	+	−	+	−	−	−	+	−	+	−	+	+	−	+
2		+	−	−	+	−	+		+	−	+	+	−	+		+	+	−	+	−	+
3			+	−	−	−	+			+	−	−	−	+			−	−	+	−	+
4				−	+	−	+				+	+	−	+				−	−	+	−
5					−	−	−					−	−	−					+	−	−
6						−	−						−	−						+	+
7							−							−							−

independently pooled all interacting and non-interacting crosses over reciprocals, blocks and seasons for each generation (table 2).

The expected values for the F_2 and backcross generation means have been calculated according to the three equalities given earlier. Examination of the observed-expected deviations in table 2 shows that the most marked difference

TABLE 2

The three scaling tests carried out independently for the non-interacting and interacting crosses pooled over reciprocals, blocks and seasons for each generation for the character height

	Generation / Item	$\bar{P}_1$	$\bar{P}_2$	$\bar{F}_1$	$\bar{F}_2$	$\bar{B}_1$	$\bar{B}_2$
Pooled interacting crosses	Observed	40·18	51·94	60·98	53·23	53·10	53·32
	Expected	...	...	...	53·52	50·58	56·46
	Deviation	...	...	...	*−0·29*	*+2·52*	*−3·14*
Pooled non-interacting crosses	Observed	38·76	45·01	45·93	43·47	42·54	44·28
	Expected	...	...	...	43·91	42·35	45·47
	Deviation	...	...	...	*−0·44*	*+0·19*	*+1·19*

between the crosses showing significant interaction and those showing no interaction lies in the two backcross means (B_1 and B_2). Thus the means of the backcross families showing non-allelic interaction are virtually the same irrespective of whether the F_1 is backcrossed to the smaller or the larger parent, *i.e.* B_1 and B_2 respectively.

(b) *Flowering time.* For flowering time the scaling tests revealed no crosses showing non-allelic interaction for the season 1952. This includes the cross 1×3 which gave rise to a deviation from the F_1 Wr/Vr regression in a manner suggesting non-allelic interaction. It will be recalled that the latter interpretation of this deviation was discounted at the time and an alternative explanation based on the

pre-planting treatment in this season was put forward. Contrasting with this we find a number of crosses showing significant interaction in the scaling tests for the 1953 season as shown in table 1*c*.

This raises a problem that we have not so far discussed, namely, that certain types of interaction may escape detection in the F_1 Wr/Vr regression analysis. Duplicate genes, for example, might well escape detection by this method and yet be picked up by the scaling tests. Further discussion of this matter will be postponed, however, until the results of other methods of detecting interactions have been described.

One thing is certain, even at this stage, namely, that the non-allelic interactions for flowering time in 1953 are a distinct system from those that appear in the inheritance of height and, as we shall see later, for leaf length. Apart from the differences in behaviour in their respective F_1 Wr/Vr regression, they also have a different distribution amongst the families and give rise to different types of upsets in the individual scaling tests. Thus of 12 crosses showing interaction for height

TABLE 3

The three scaling tests carried out independently for the interacting and non-interacting crosses pooled over reciprocals and blocks for leaf length

	Generation / Item	$\bar{P}_1$	$\bar{P}_2$	$\bar{F}_1$	$\bar{F}_2$	$\bar{B}_1$	$\bar{B}_2$
Pooled interacting crosses	Observed	18·60	24·42	23·81	21·60	22·29	22·98
	Expected	...	...	...	22·66	21·21	24·11
	Deviation	...	...	...	*−1·06*	*+1·08*	*−1·13*
Pooled non-interacting crosses	Observed	18·19	23·73	22·87	22·02	20·90	23·87
	Expected	...	...	...	21·92	20·53	23·30
	Deviation	...	...	...	*+0·10*	*+0·37*	*+0·57*

in 1953 and the 14 crosses showing interaction for flowering time in the same season only 5 show interaction for both characters—the expectation on a random basis being 6. Furthermore the effect of the interaction in flowering time is in all cases to give F_2 and backcross families which flower earlier (*i.e.* have a lower mean) than one would expect from their parental and F_1 flowering times.

(c) *Leaf length.* While the character leaf length failed to show any significant non-allelic interaction on the F_1 Wr/Vr regression analysis for the two seasons over which it was tested, there were some indications of interaction. For example, the regression coefficient was always lower than expected for no interaction ($\doteq$0·7) but not significantly so because of the large standard error. Examination of the graphs of Wr against Vr shows that the large error and low regression coefficient are in all cases due to the point for array 1, all other points being a good fit with the expected slope of one. The joint scaling test for 1952 (leaf length was not scored in 1953) shows that non-allelic interaction is indeed present and is mainly concentrated in array 1. Two crosses show significant interaction at the 1 per cent. level, namely 1×2 and 3×4, while a further six crosses, three of which occur in array 1, are significant at the 5 per cent. level. All crosses showing non-allelic interaction for leaf length also show interaction for height at a higher level of significance. It would seem, therefore, that we are dealing with the same group of interacting loci with two different levels of effect.

The observed and expected generation means for the averages of interacting and non-interacting crosses are given in table 3.

As with height, the backcross generation means for leaf length for the backcross to the smaller and larger parent are more alike than expected in the crosses exhibiting

non-allelic interaction. Unlike the height means, however, the deviation of the observed F_2 from its expectation is of the same order of magnitude as that of the backcrosses.

The existence of non-allelic interaction in certain crosses raises the problem of rescaling. This has been investigated not only in the *N. rustica* data but also in re-analyses of published maize diallels (Jinks, 1955). Suitable changes of scale such as log transformations will remove interaction from all crosses exhibiting this phenomenon. Unfortunately in a diallel set of crosses where only a proportion of the crosses show non-allelic interaction, any change of scale that successfully removes this interaction results in the appearance of significant disturbances in previously non-interacting crosses.

(iii) *Heterosis*

The relationship between the parental and F_1 means is the same for all three seasons. Thus the overall F_1 mean for height is significantly greater than that of the overall mid-parent, while the F_1 mean flowering time is significantly earlier than that of the average parent. An analysis of variance to test the consistency of the magnitude of the difference between the F_1 mean and the parental mean over seasons is given in table 4.

TABLE 4

The analysis of variance of the relationship of the overall parental and F_1 mean over seasons

Character	Flowering time		Height
Item	N	MSS	MSS
P v F_1	1	599·7	2151·4
P v F_1 × seasons	2	38·3	152·5
Duplicate error	192	3·9	5·9
Reciprocal differences	84	24·1	14·9

Two error variances are available for testing the significance of the parental and F_1 differences and their consistency over seasons. These are the duplicate error derived from the SS of differences between identical parental and F_1 families in the two blocks for the three seasons and the SS of differences between reciprocal F_1 families in the three seasons. The latter SS is significant for both characters against the duplicate error ; but this significance must be interpreted with caution. As pointed out in previous papers, the design of the experiment was such that while reciprocal families were independently randomised prior to sowing the seed, blocks were not independently randomised until planting out into the field. This inadequacy of the design has now been remedied, but for the present results the reciprocal difference SS must be regarded as the better estimate of error in the experiment as a whole.

For flowering time the F_1 heterosis is significant and consistent in magnitude over seasons. For height, on the other hand, there is a significant overall heterosis over the three seasons, but the magnitude of this heterosis varies significantly over seasons.

4. SECOND DEGREE STATISTICS

(i) *Analysis of F_1s and parents for* 1953

Before proceeding to the F_2 and backcross analyses we will first consider the analysis of the 1953 F_1 and parental means and compare them with those given earlier for 1951 and 1952.

In all essential details the 1953 results agree with the earlier ones (table 5).

The character height again shows significant overdominance $\left(\frac{H_1}{D} > 1\right)$ and significant non-allelic interaction ($b_{Wr/Vr} > 1$) while flowering time again shows incomplete but significant dominance $\left(0 < \frac{H_1}{D} < 1\right)$ and no indication of non-allelic interaction ($b_{Wr/Vr} = 1$). Analysis of array 7 for the character height, *i.e.* the only array showing no significant interaction in the scaling test (table 1*b*) gave only complete dominance. As in previous seasons, therefore, the high dominance ratio can be related to an inflation due to non-allelic interaction.

TABLE 5

The analysis of height and flowering time for 1953

Statistic	Height		Flowering time
	Complete	Array 7	
D	64·6635	64·6635	147·9608
H_1	231·0404	76·4384	83·0117
H_2	158·6446	...	69·2976
F	−133·5268	−139·4684	+40·4673
$\frac{H_1}{D}$	3·7551	1·1821	0·5619
$\overline{uv}$	0·1726	...	0·2084
b Wr/Vr	0·5626±0·1134	...	0·9154±0·0560

Other points of similarity over seasons are the product $\overline{uv}$ for height, which has never varied from 0·17, and the distribution of the array points on the Wr/Vr graphs for both characters. As the latter comparison has been analysed in detail by Allard (in preparation), it will not be discussed further here except in reference to the F_2 and backcross Wr/Vr graphs.

Certain statistics show differences over the three seasons. Thus the genetical components D, H_1, H_2 and F are greater in magnitude for both characters in 1953 than in either of the previous seasons. For flowering time this change has apparently not influenced the dominance ratio, but for height the ratio is definitely higher than in previous seasons. On the whole the height statistics have remained much more constant over seasons than those for flowering time. Further discussion of this subject will, however, be delayed until least squares estimates of the statistics have been obtained.

(ii) *The expectations for the F_2 and backcross analyses*

The expected statistics for the F_2 generation are of the same general form as those of the F_1 except that the contribution of h is halved by the one generation of inbreeding. For this reason the coefficients of H_1 and H_2 are $\frac{1}{4}$ of those of the F_1 statistics, while the coefficient of F is halved, being second and first degree statistics in h respectively (table 6).

TABLE 6

F_2 expected means and variances

Original parent lines Female / Male	Genotype . . Frequency . Mean . .	AA u_a d_a	aa v_a $-d_a$	Mean of array
AA. u_a. d_a.	Genotype of F_2 . Frequency . Mean . . Variance . .	AA u_a^2 d_a o	$\frac{1}{4}$AA : $\frac{1}{2}$Aa : $\frac{1}{4}$aa u_av_a $\frac{1}{2}h_a$ $\frac{1}{2}d_a^2+\frac{1}{4}h_a^2$	$u_ad_a+\frac{1}{2}v_ah_a$ $\frac{1}{2}v_ad_a^2+\frac{1}{4}v_ah_a^2$
aa. u_a. $-d_a$.	Genotype of F_2 . Frequency . Mean . . Variance . .	$\frac{1}{4}$AA : $\frac{1}{2}$Aa : $\frac{1}{4}$aa u_av_a $\frac{1}{2}h_a$ $\frac{1}{2}d_a^2+\frac{1}{4}h_a^2$	aa u_a^2 $-d_a$ o	$-v_ad_a+\frac{1}{2}u_ah_a$ $\frac{1}{2}u_ad_a^2+\frac{1}{4}u_ah_a^2$

Parental mean $(u_a-v_a)d_a$ Overall mean of progenies $(u_a-v_a)d_a+u_av_ah_a$

Overall mean variance of progenies $u_av_ad_a^2+\frac{1}{2}u_av_ah_a^2$.

The composition of the F_2 variances and covariances are as follows :—

F_2 statistic	One gene	Many independent genes
Mean variance of arrays	$= u_av_ad_a^2+\frac{1}{4}u_av_ah_a^2-u_av_a(u_a-v_a)d_ah_a$	$= \frac{1}{4}D+\frac{1}{16}H_1-\frac{1}{8}F+E_2$
Mean covariance of arrays	$= 2u_av_ad_a^2-u_av_a(u_av_a(u_a-v_a)d_ah_a$	$= \frac{1}{2}D-\frac{1}{8}F+\frac{1}{n}E_2$
Variance of array means	$= u_av_ad_a^2+\frac{1}{4}u_av_ah_a^2-u_a^2v_a^2h_a^2-u_av_a(u_a-v_a)d_ah_a$	$= \frac{1}{2}D+\frac{1}{16}H_1-\frac{1}{16}H_2-\frac{1}{8}F+\frac{1}{n}E_2$
Mean family variance	$= u_av_ad_a^2+\frac{1}{2}u_av_ah_a^2$	$= \frac{1}{2}D+\frac{1}{8}H_1+E_1$

The analysis of the means of reciprocal backcross families is divided into two parts. The first makes use of the fact that in any cross of a diallel set of crosses the following relationship holds between the F_2 family mean and the means of the two reciprocal backcrosses provided non-allelic interaction is not present. $\overline{F}_2 = \frac{1}{2}(\overline{B}_1+\overline{B}_2)$ (table 7).

Thus the expected statistics for the analysis of the means of reciprocal backcross families of a diallel set of crosses are identical with those given for the analysis of F_2 family means.

The second part of the analysis of backcrosses utilises the statistics obtained from the differences between the means of reciprocal backcross families. These terms are entirely composed of the additive component of variation, *i.e.* d (table 7).

The expected variance and covariance for the analysis of the differences between reciprocal backcross family means are as follows :—

Statistic		One gene		Many independent genes
Mean variance of arrays	=	$u_a v_a d_a^2$	=	$\frac{1}{4}D+E_2$
Mean covariance of arrays	=	$\pm 2u_a v_a d_a^2$	=	$\pm\frac{1}{2}D+\frac{I}{n}E_2$
Variance of array means	=	$u_a v_a d_a^2$	=	$\frac{1}{4}D+\frac{I}{n}E_2$

The mean covariance of arrays will be positive or negative according to whether the analysis has been arranged such that B_1 is the backcross of the F_1 to the larger (or the smaller) of the two parents.

One further statistic is available from the backcross generation, namely the mean summed reciprocal backcross family variance. For the one gene model the expectation of this variance is :—

$$u_a v_a d_a^2 + u_a v_a h_a^2$$

which for many independent genes becomes

$$\tfrac{1}{4}D+\tfrac{1}{4}H_1+2E_1.$$

(iii) *Non-allelic interaction*

The contribution of non-allelic interaction to the family means differs over the generations under consideration here (Hayman and Mather, 1955). Since the contribution of non-allelic interaction is taken up by the four genetic parameters D, H_1, H_2 and F, and as this contribution does not appear explicitly in our analyses, these parameters will be homogeneous over generations only in the absence of interactions (Mather, 1949 ; Mather and Vines, 1952). The homogeneity of least squares estimates of the four parameters over generations will therefore provide us with a further test of non-allelic interaction.

(iv) *Linkage*

Linkage, though not affecting family means in the absence of non-allelic interaction and hence the second degree variances and covariances derived from them, shows its effect in the within family variances of segregating generations (Mather, 1949).

The simplest case of two linked loci, A-a and B-b, independently distributed in the parental lines but showing a recombination value

TABLE 7

Backcross expected means and variances

Original parent lines Males Females							
	Genotype . .	AA	...	...	aa	...	...
	Frequency . .	u_a	...	...	v_a	...	...
	Mean . .	d_a	...	...	$-d_a$	...	...
			$\frac{1}{2}(B_1+B_2)$	(B_1-B_2)		$\frac{1}{2}(B_1+B_2)$	(B_1-B_2)
AA. u_a. d_a.	Frequency . .	u_a^2	...	...	u_av_a	...	...
	Mean B_1 . .	d_a	d_a	o	$\frac{1}{2}d_a+\frac{1}{2}h_a$	$\frac{1}{2}h_a$	d
	Mean B_2 . .	d_a			$-\frac{1}{2}d_a+\frac{1}{2}h_a$		
	Variance B_1 .	o	...	...	$\frac{1}{4}(d_a-h_a)^2$	...	...
	Variance B_2 .	o	...	...	$\frac{1}{4}(d_a+h_a)^2$	...	...
aa. v_a. $-d_a$.	Frequency . .	u_av_a	...	...	v_a^2	...	...
	Mean B_1 . .	$-\frac{1}{2}d_a+\frac{1}{2}h_a$	$\frac{1}{2}h_a$	$-d_a$	$-d_a$	$-d_a$	o
	Mean B_2 . .	$\frac{1}{2}d_a+\frac{1}{2}h_a$			$-d_a$		
	Variance B_1 .	$\frac{1}{4}(d_a+h_a)^2$	...	...	o	...	...
	Variance B_2 .	$\frac{1}{4}(d_a-h_a)^2$	...	...	o	...	...
Parental mean $(u_a-v_a)d_a$							
Mean of arrays	Mean . .	...	$u_ad_a+\frac{1}{2}v_ah_a$	$-v_ad_a$	...	$-v_ad_a+\frac{1}{2}u_ah_a$	$+u_ad_a$
	Variance . .	$\frac{1}{2}v_ad_a^2+\frac{1}{2}v_ah_a^2$	...	...	$\frac{1}{2}u_ad_a^2+\frac{1}{2}u_ah_a^2$	...	...

Overall mean of reciprocal backcrosses . . . . . $(u_a-v_a)d_a+u_av_ah_a$
Overall difference of reciprocal backcrosses . . . . . o
Overall mean summed reciprocal backcrosses family variance . . . . $u_av_ad_a^2+u_av_ah_a^2$

$p(=1-q)$, makes the following contribution to the mean variance of F_2 and summed reciprocal backcross families from a set of diallel crosses. Mean family variance of F_2

$$= u_a v_a d_a^2 + u_b v_b d_b^2 + \tfrac{1}{2} u_a v_a h_a^2 + \tfrac{1}{2} u_b v_b h_b^2 + 2 u_a v_a u_b v_b (1-2p)^2 h_a h_b$$

Mean summed reciprocal backcross variance

$$= u_a v_a d_a^2 + u_b v_b d_b^2 + u_a v_a h_a^2 + u_b v_b h_b^2 + 4 u_a v_a u_b v_b (1-2p) h_a h_b$$

If we define two linkage parameters such that

$$H_3 = 16 \Sigma u_a v_a u_b v_b (1-2p)^2 h_a h_b$$
$$\text{and } H_4 = 16 \Sigma u_a v_a u_b v_b (1-2p) h_a h_b$$

Then for many genes, some of which show linkage, the expectations for the two mean family variances are

F_2	$\tfrac{1}{4}D + \tfrac{1}{8}H_1 + \tfrac{1}{8}H_3 + E_1$
backcross	$\tfrac{1}{4}D + \tfrac{1}{4}H_1 + \tfrac{1}{4}H_4 + 2E_1$

Both linkage parameters can take sign. Thus if the linked genes show reinforcing dominance, *i.e.* the dominance deviations have the same sign, then H_3 and H_4 will be positive. If, on the other hand, dominance is in opposition, *i.e.* the dominance deviations have opposite signs, then H_3 and H_4 will be negative. Linkage between genes showing reinforcing dominance will, therefore, lead to an inflation of the within family variance, while linkage between genes showing opposing dominance will lead to a deflation.

Although complete specification of simultaneous linkage and non-allelic interaction is now possible (Hayman and Mather, 1955) the limited number of statistics available in the present analyses is not sufficient for their estimation. The question now arises as to how far tests of heterogeneity of the specified components D, H_1, H_2, F and E_1, over statistics, can distinguish between the two sources of disturbance. In so far as components for non-allelic interaction appear in all our statistics (except for the mean variance of F_2 families), it follows that non-allelic interaction can lead to heterogeneity of components whether these be derived from an analysis of family means or family variances of segregating generations of a diallel set of crosses. Linkage, on the other hand, effects only family variances and so can lead to heterogeneity only if this type of statistic is included in the estimation of components. Any heterogeneity of components over the statistics excluding family variances must, therefore, be ascribable to non-allelic interaction. Should such heterogeneity exist, then any further heterogeneity introduced by including the family variances in the estimations can obviously no longer be unambiguously ascribed to linkage. Our linkage test, therefore, is only valid where non-allelic interaction plays no significant part in the inheritance of the character under consideration.

(v) *Non-random distribution of alleles in the parent lines*

One further source of disturbance may lead to heterogeneity of the components of variation as estimated from the analysis of family means and those obtained from the mean variance of families of segregating generations. Such disturbances may arise from non-random distribution of alleles in the original parental lines. Since these do not affect the homogeneity of the components of variation obtained from the diallel analyses of family means, they will not be confounded with disturbances arising from non-allelic interactions. They will, however, be confounded with linkage. A full account of the possibilities and magnitudes of such confounding is outside the scope of this paper and only pertinent results will be mentioned here.

Association of alleles combined with reinforcing dominance will inflate the components of variation obtained from family means but will not affect those obtained from the mean family variances. It will thus mimic linkage between loci exhibiting opposing dominance. The alternative situation, *i.e.* dispersion of alleles and opposing dominance, in so far as it will deflate the components of variation from family means while not affecting those from mean family variances, will mimic linkage between loci exhibiting reinforcing dominance. Combinations of association and opposing dominance or dispersion and reinforcing dominance will also mimic linkage, but the magnitude of the effect will be much less than that arising from the previous two combinations, while the linkage phase it mimics will depend on the relative magnitude of the deviation from random distribution and the dominance ratio.

Two considerations make the confounding between correlated gene distribution and linkage less serious than it might appear at first. Firstly, non-random distribution of alleles can be detected before the estimation of the effects due to linkage is undertaken. The detection depends once again on the fact that the regression of Wr on Vr is a line of slope one only if our hypothesis of independence of the genes is true. It can therefore be used to detect not only non-allelic interactions but also non-random distribution of the alleles. Thus association gives a curve which is convex upwards while dispersion gives a curve which is convex downwards. Although a regular curvature of the line is unlikely to be mistaken for non-allelic interaction we have at least two methods of discrimination should such difficulty arise, namely, the scaling test and the test of homogeneity of the components of variation over statistics derived from family means, both of which detect only disturbances arising from non-allelic interactions. In the one case where difficulty may arise in interpreting the Wr/Vr regressions, *i.e.* where both sources of disturbance are present, the estimation of linkage is in any case invalidated by the presence of non-allelic interactions.

Secondly, in some cases, for example where the inbred lines used in the diallel crosses are derived from a common mating pool, any

non-random distribution of alleles will probably be due to the linked genes themselves. In such cases confounding between the two, while it might disturb the magnitude of the linkage effect, will not lead to the drawing of wrong conclusions regarding the presence or absence of linkage.

In our own data, where a valid estimate of linkage effects is confined to the 1952 flowering time results, there is no suggestion of a downward convexity of the Wr/Vr regressions, *i.e.* dispersion of alleles, which would give rise to the significant reinforcing linkage (see later).

(vi) *The regression of array covariance on array variance*

The relationship between array covariance and array variance described for the F_1 analyses also holds, within the same limitation of independence of the genes, in the F_2 and averaged reciprocal backcross analyses, *i.e.* the regression of array covariance (Wr) on array variance (Vr) has a slope of one. There is one difference, however, in that the point of interaction of the regression line with the Wr axis is no longer $\overline{Wr}-\overline{Vr} = \frac{1}{4}(D-H_1)$ but $\frac{1}{4}(D-\frac{1}{4}H_1)$.

TABLE 8

The regression coefficients of Wr on Vr combined over blocks and seasons for the three generations

Character	Height		Flowering time
Generation	Complete data	2, 4 and 7 diallel	Complete data
F_1 . . .	0·598±0·173	1·057±0·081	0·943±0·020
F_2 . . .	0·802±0·054	1·017±0·062	1·083±0·155
Backcross .	0·677±0·219	1·004±0·088	0·999±0·050

Interaction between non-allelic genes may lead to deviation from the expected slope in all three generations, although in our own data the deviation is less marked in the F_2s and backcrosses. The regression coefficients for the three generations obtained from a combined analysis over blocks and seasons are given in table 8. Also included for comparison are the regression coefficients of a 3×3 diallel extracted from the height data which contains no cross showing non-allelic interaction (table 1*a*, 1*b*). This small diallel consists of the three parents of arrays 2, 4 and 7 and their F_1, F_2 and backcross progeny means.

For flowering time and the 3×3 diallel extracted from the height data the regression coefficients do not differ significantly from one, so that there is no suggestion of non-allelic interaction in any of the three generations for either season. The regression coefficients for the complete height data, on the other hand, always give a regression coefficient less than one, but it is only significantly so for the F_1 generation.

In the absence of interaction, the Wr/Vr graphs for the F_2 and backcross generations should be identical, but in the presence of interaction this is not necessarily true. For both characters we find that the two regressions do not differ significantly either in slope or mean. While the joint regression for flowering time (fig. 1) does not differ significantly from a slope of one, the joint regression for height is significantly less than one ($b = 0{\cdot}722 \pm 0{\cdot}130$), indicating significant non-allelic interaction for height in these two generations.

A further relationship exists between the regressions of Wr on Vr for the F_1 and F_2 (or backcross) generations. If for each array we draw a line through the Wr/Vr coordinates for its F_1 and F_2 array

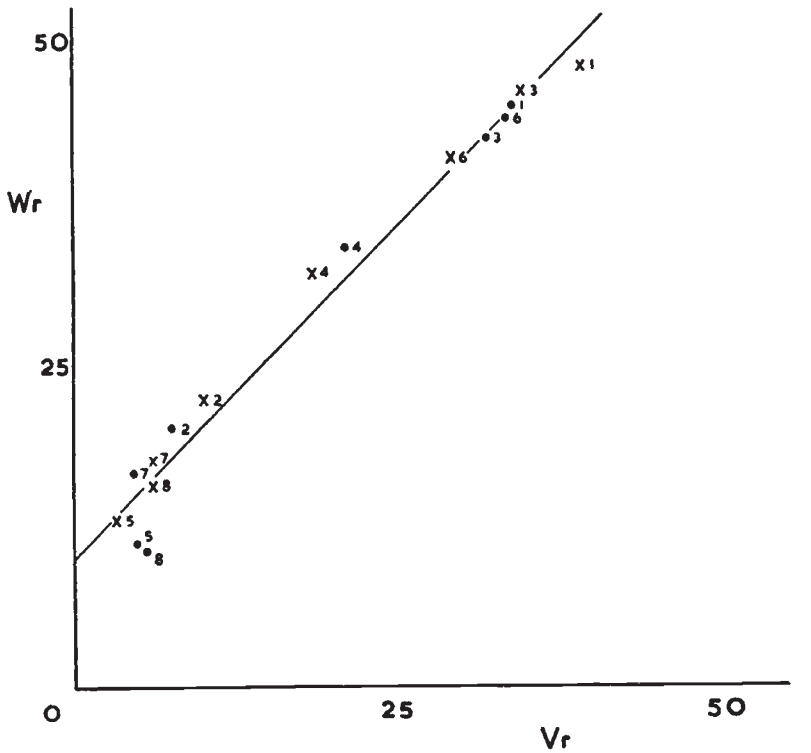

FIG. 1.—The regression of array covariance (Wr) on array variance (Vr) for the F_2 (full circles) and backcross (crosses) generations for flowering time. The array points of both generations fall on the same straight line of unit slope. This indicates absence of non-allelic interactions but presence of dominance. The 8 arrays fall in the same order along this line in both generations. This order reflects the varying proportions of dominant and recessive alleles in the common parents of the arrays. Those with most dominant alleles have low variances and covariances, *e.g.* 5, 8 and 7, while those with most recessive alleles have high variances and covariances, *e.g.* 1, 3 and 6.

points, they will all converge on a common point of intersection whose own coordinates vary characteristically with the degree of dominance. The Wr coordinate for the point of intersection is independent of the dominance relations as it contains only terms in d,

being $\frac{1}{2}D = \Sigma 2uvd^2$. The Vr coordinate, however, does vary with the degree of dominance and for a one gene model is given by the expression

$$u_a v_a (d_a - h_a)^2 + \frac{u_a v_a d_a h_a - \frac{3}{4} u_a v_a h_a^2}{u_a v_a d_a h_a} \cdot 2 u_a v_a d_a h_a.$$

This expression has a maximal value of $\frac{1}{4}D$ when $h = 0$.

The point of intersection for other generations, *e.g.* F_2 and F_3, can be obtained by halving the contribution of dominance in the above expression to allow for the additional generation of inbreeding. Thus for the n^{th} and $(n+1)^{th}$ generation of inbreeding the coordinates of intersection will be

$$Wr = 2u_a v_a d_a^2 \text{ and } Vr = u_a v_a \left(d_a - \frac{1}{n} h_a\right)^2 + \frac{\frac{1}{n} u_a v_a d_a h_a - \frac{3}{4} u_a v_a \left(\frac{1}{n} h_a\right)^2}{\frac{1}{n} u_a v_a d_a h_a} \cdot \frac{2}{n} u_a v_a d_a h_a$$

where n is large the Vr coordinate approaches the value of

$$\tfrac{1}{4}D = \Sigma u_a v_a d_a^2.$$

Therefore with a low initial degree of dominance or a large number of generations of selfing the point of intersection approaches the coordinates for the Wr/Vr array points of an F_1 diallel showing no dominance.

These relationships break down in the presence of non-allelic interaction in such a way as to allow us to detect the interacting arrays. This may be illustrated by the following two gene models. The first model, which shows no interaction, consists of four parental lines AABB, AAbb, aaBB and aabb such that $d_a = d_b = h_a = h_b = 2$. The second model is the same except that the allele A in the presence of B contributes 4 units instead of 2 to the mean. In both these models the points on the Wr/Vr graph for arrays AAbb and aaBB are identical leaving us with only three distinct points on the graphs. These are given in figs. 2 and 3. For the first model the lines shown through the F_1 and F_2 points of each array intersect at the point expected for a system showing complete dominance, namely $Wr = \frac{1}{2}D$, $Vr = \frac{1}{8}D$. In the second model, the line drawn through the array points for arrays AAbb and aaBB passes below the expected point of intersection while the other two arrays pass slightly above it but intersect very close to it.

Illustrations of both types of model can be found in the *N. rustica* data. For this purpose blocks and seasons have been combined, and the F_2 and backcross Wr/Vr points pooled to give the graphs of

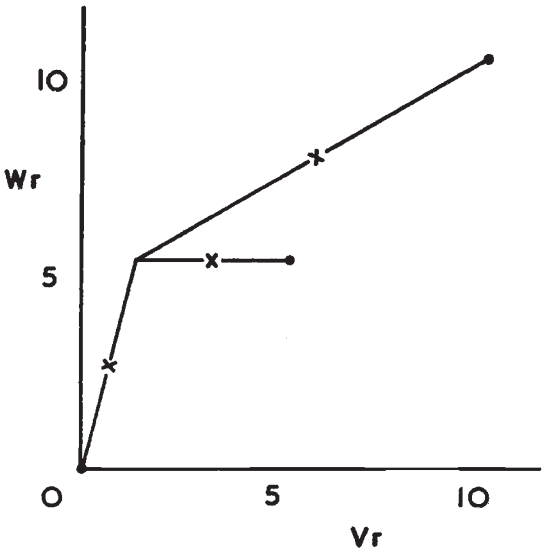

FIG. 2.—The above graph shows the F_1 (full circles) and F_2 (crosses) points of the Wr/Vr regression for a two gene model showing complete dominance. The lines drawn through the F_1 and F_2 array points for each array intersect at a common point. The Wr coordinate of this point is $\frac{1}{2}D$ and is independent of the degree of dominance. The Vr coordinate on the other hand increases with decreasing dominance to a maximum value of $\frac{1}{4}D$. For the complete dominance model $Vr = \frac{1}{8}D$.

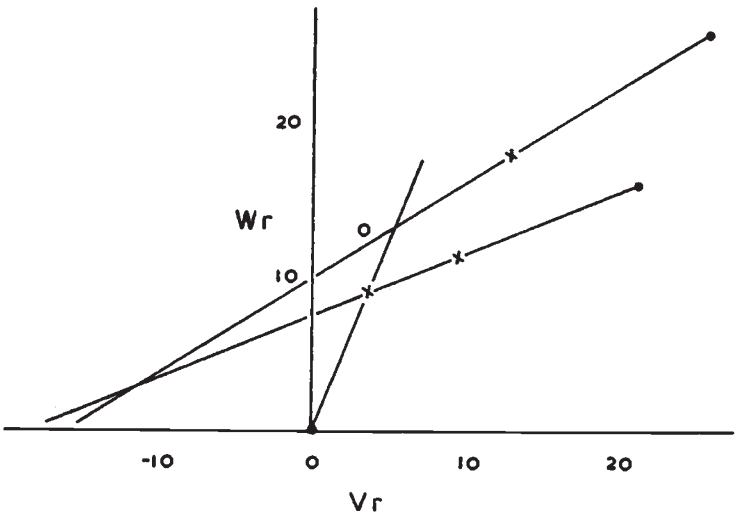

FIG. 3.—The above graph shows the F_1 (full circles) and F_2 (crosses) points of the Wr/Vr regression for a two gene model showing a complementary type of non-allelic interaction. There is no common point of intersection when lines are drawn through the array points for the F_1 and F_2 generations for each array. This absence of a common point of intersection provides a means of detecting non-allelic interaction. The Wr/Vr regressions, of course, have a slope of less than 1.

figs. 4 and 5. For the complete height analysis there is no common point of intersection, as would be expected from the high incidence of non-allelic interaction. The lines drawn through the array points for the F_1 and combined F_2 and backcross generations fall roughly into three groups (fig. 4). The first group consists solely of array 7, which, it will be recalled, has never given any indication of the presence of non-allelic interaction in any of the tests so far applied. This line passes through the coordinates Wr $= \frac{1}{2}$D, Vr $= \frac{1}{8}$D, which are those for a system showing complete dominance. The second group consists of arrays 2, 4 and 8 which intersect at a point whose coordinates are approximately Wr $= \frac{1}{5}$D, Vr $= \frac{1}{4}$D. These three arrays have certain interaction properties in common. Thus they all show significant non-allelic interaction in their crosses with the common

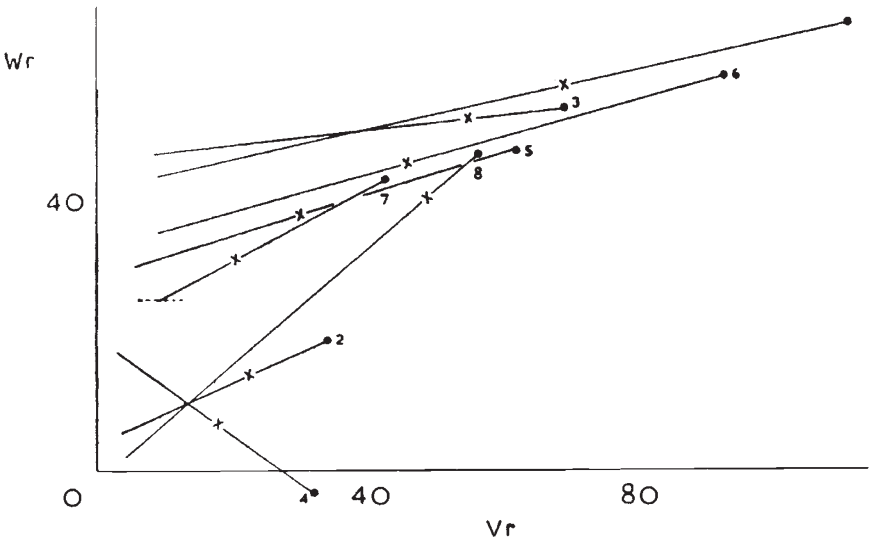

FIG. 4.—Illustrates the application of the method of detecting non-allelic interaction shown in figs. 2 and 3 for our height data. The F_1 array points are indicated by full circles and the F_2 and backcross generations have been combined to give the array points indicated by the crosses. The presence of non-allelic interactions leads to the absence of a common point of intersection. Only array 7, which contains no interacting crosses, intersects the Wr coordinate of the expected point of intersection (indicated by the dotted line). The remaining arrays fall roughly into two groups consisting of arrays 2, 4 and 8 and 1, 3, 5 and 6, a grouping that is confirmed by the scaling tests (tables 1*a* and 1*b*). The two Wr/Vr regression coefficients are less than 1.

parents of arrays 1, 3 and 6 (table 1*a*, 1*b*). The remaining four arrays 1, 3, 5 and 6, each of which shows interaction with two or more of the common parents of arrays 2, 4 and 8, all lie above the expected point of intersection and only two of them, 1 and 3, intersect with one another where Wr $\simeq$ D and Vr $\simeq \frac{4}{5}$D.

If we now analyse the small 3×3 diallel, involving the interaction-free crosses of arrays 2, 4 and 7, in the same way we obtain a picture similar to our first model (fig. 5). There is only one point of intersection and its coordinates are approximately Wr $= \frac{1}{2}$D and Vr $= \frac{3}{16}$D.

After allowing for the environmental components of the Wr's and Vr's, this situation is compatible with a system showing just under complete dominance and absence of non-allelic interaction.

When we apply the same test to the 1953 flowering time data we obtain unambiguous evidence of widespread non-allelic interaction as suggested by the earlier scaling tests. If we use the array points for the F_1 and joint F_2 and backcross Wr/Vr regressions we find that the lines drawn through only two arrays, 4 and 7, pass anywhere near the point of intersection expected from a genetical system showing partial dominance. This is in agreement with the scaling test which

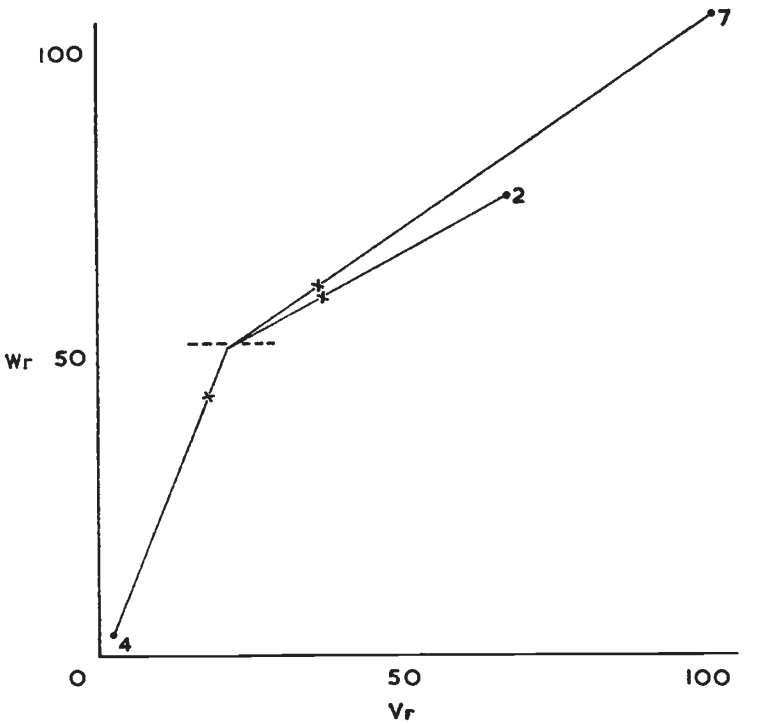

FIG. 5.—Illustrates the application of the method of detecting non-allelic interaction shown in figs. 2 and 3 to a portion of the height data known to contain no crosses showing non-allelic interaction. This is identical with the model of fig. 2. There is only one point of intersection which falls on the expected Wr coordinate, *i.e.* Wr = ½D (indicated by dotted line), while the Vr coordinate is approximately that expected of a system showing almost complete dominance, The F_1 Wr/Vr array points are again indicated by full circles and the combined F_2 and backcross points by crosses. The two Wr/Vr regressions have a slope of 1.

suggests that these two arrays exhibit only a low incidence of crosses showing non-allelic interaction. The lines for the remaining arrays, while not showing a common point of intersection, all converge on a small area whose centre is approximately at the point Wr = ⅓D, Vr = ¼D. In the 1952 analysis we find no serious evidence of interaction on this test, which is again in complete agreement with the scaling test and also the Wr/Vr regression analyses. Although three of the lines drawn through the array points, namely those for arrays 5,

8 and 6, do not pass through the expected point of intersection, they do pass close enough for any deviation to be regarded as the consequence of error variation.

While this method of detecting non-allelic interaction is not independent of the Wr/Vr regression method it does allow us to detect and in some cases classify the interacting arrays in a manner approaching the accuracy of the scaling test.

5. ESTIMATION OF THE COMPONENTS OF VARIATION

The process of estimation has been dealt with in detail by Mather (1949) and Mather and Vines (1952). In these analyses we depart from this method in only one respect, namely the treatment of the environmental components of variation. Of these only E_1 is independent of the design of the experiment, the rest, which are derivatives of E_2, being dependent on the number of parental lines employed. Because of the obvious advantages of a general least squares solution applicable to all diallel crosses, it was decided to correct the observed statistics for all environmental components dependent on " n " prior to the estimation of the remaining parameters.

E_1, which is included in the least squares estimation, and is only used in connection with the family variances of the segregating generations, is obtained as the mean within plot variance of the non-segregating parental and F_1 generations as described by Mather (1949). The E_2 component, however, is derived independently for each generation, as the mean variance of duplicate or reciprocal plot means, and is used only in connection with the generation from which it is derived. This is necessary because the variances and covariances of the diallel analysis of segregating generations contain both an environmental and a sampling component. By using an E_2 derived from the same generation, both these components are accommodated simultaneously, while an E_2 predicted from parents and F_1s, as described above for the E_1s, would depend only on the environmental portion. In many cases, the sampling error contributes the larger proportion of the non-inherited variation. This may be illustrated by reference to the E_2 components for the height analysis. For this purpose all E_2s have been expressed on a single plot basis and averaged over blocks and seasons (table 9).

The first column of table 9 gives the E_2s calculated from the mean variance of duplicate and reciprocal plot means for each generation. The second column contains the expected E_2s for the F_2 and backcross generations assuming no sampling error as $\frac{1}{2}\bar{P}_{E2}+\frac{1}{2}\bar{F}_{1E2}$. The excess of the observed E_2s over this predicted E_2, which is ascribable to sampling error, is contained in the third column. For comparison are included the expected sampling error calculated as $\frac{1}{n}$ (mean plot variance) where n is the number of plants per plot, *i.e.* five.

Having corrected the observed statistics for all these non-inherited components of variation except E_1 we can now proceed with the least squares estimation.

TABLE 9

The environmental and sampling components of variation for height

Generation	Observed E_2	Estimated environmental component	Estimated sampling component	Expected sampling component
P . . .	4·0538	4·0538	0·0000	...
F_1 . . .	8·1960	8·1960	0·0000	...
F_2 . . .	11·1847	6·1249	5·0598	7·9403
B_1 and B_2 .	17·5213	6·1249	11·3964	12·5843

(i) *Inclusive analysis*

Without making allowances for the possible effects of linkage, *i.e.* the inclusive analysis of Mather, five components of variation are involved, D, H_1, H_2, F and E_1. The sixteen basic equations available

TABLE 10

Inclusive c matrix

CDD	0·643932	CH_1D	0·615684	CH_2D	−0·257111	CFD	1·558234	CE_1D	−0·144625
CDH_1	0·615684	CH_1H_1	19·840244	CH_2H_1	10·411285	CFH_1	9·224883	CE_1H_1	−2·143653
CDH_2	−0·257111	CH_1H_2	10·411285	CH_2H_2	26·573565	CFH_2	−2·054586	CE_1H_2	−1·052370
CDF	1·558234	CH_1F	9·224883	CH_2F	−2·054586	CFF	12·021402	CE_1F	−1·155705
CDE_1	−0·144625	CH_1E_1	−2·143653	CH_2E_1	−1·052370	CFE_1	−1·155705	CE_1E_1	0·408042

for their estimation may be combined to give five equations yielding least squares estimates of the five components as described by Mather (1949.) The solution of these five sets of equations leads to a matrix of multipliers as shown in table 10.

TABLE 11

Exclusive c matrix

CDD	0·68508	CH_1D	1·14917	CH_2D	0·00000	CFD	1·85635	CE_1D	0·00000
CDH_1	1·14917	CH_1H_1	27·61940	CH_2H_1	14·22222	CFH_1	13·43646	CE_1H_1	0·00000
CDH_2	0·00000	CH_1H_2	14·22222	CH_2H_2	28·44444	CFH_2	0·00000	CE_1H_2	0·00000
CDF	1·85635	CH_1F	13·43646	CH_2F	0·00000	CFF	14·32044	CE_1F	0·00000
CDE_1	0·00000	CH_1E_1	0·00000	CH_2E_1	0·00000	CFE_1	0·00000	CE_1E_1	1·00000
CDH_3	−2·51934	CH_1H_3	−29·91774	CH_2H_3	−14·22222	CFH_3	−17·14917	CE_1H_3	−8·00000
CDH_4	−1·83425	CH_1H_4	−28·76857	CH_2H_4	−14·22222	CFH_4	−15·29282	CE_1H_4	−8·00000

CH_3D	−2·51934	CH_4D	−1·83425
CH_3H_1	−29·91774	CH_4H_1	−28·76857
CH_3H_2	−14·22222	CH_4H_2	−14·22222
CH_3F	−17·14917	CH_4F	−15·29282
CH_3E_1	−8·00000	CH_4E_1	−8·00000
CH_3H_3	162·95641	CH_4H_3	96·43708
CH_3H_4	96·43708	CH_4H_4	110·60283

(ii) *Exclusive analysis*

The analysis allowing for the presence of linkage, *i.e.* the exclusive analysis of Mather, involves two further components of variation, H_3 and H_4, making seven in all. We have the same sixteen equations for their estimation and we can again obtain a matrix of multipliers (table 11). As three of the components H_3, H_4 and E_1 are confined

to three equations, the latter must be a perfect fit with expectations. This provides an alternative method of exclusive analysis, whereby least square estimates of the four components D, H_1, H_2 and F are obtained from the remaining thirteen equations. The other three components are then obtained by direct estimation from the observed statistics using the least squares estimates of D and H_1. Both methods give identical results, but the seven by seven matrix of multipliers has the added advantage that the standard errors of all components can be obtained directly from it.

From the observed statistics we can now find the least squares estimates of the components of variation in both the inclusive and exclusive analyses for each of the two blocks within the two seasons.

TABLE 12

Analysis of variance of the height and flowering time data

Item		Height		Flowering time	
		N	MSS	N	MSS
Linkage		...	...	2	134·66
Residual interaction		11	150·39	9	28·68
Heterogeneity	*Between seasons*—				
	Linkage	...	...	2	170·29
	Residual interaction	11	31·41	9	88·94
	Components	5	261·77	5	5953·10
	Within seasons				
	Linkage	...	...	4	14·96
	Residual interaction	22	9·14	18	4·59
	Components	10	111·92	10	43·36

By substituting these estimates in the sixteen basic equations we can arrive at least squares estimates of the expected values of the observed statistics. The analysis of variance of the sum of squares of deviations of observed from expected and the consistency of these deviations over blocks and seasons have been dealt with in detail by Mather and Mather and Vines, so that only the final analyses are given here (table 12).

As pointed out earlier, significant non-allelic interaction invalidates our test for linkage, and therefore the linkage items are not given for the height analysis although they have been estimated and found to be non-significant.

(iii) *Consistency of components over seasons and blocks*

The above analysis reveals a marked difference in the stability of the components of variation over blocks and seasons for the two characters height and flowering time. Thus the components of

variation for height are not significantly more variable between seasons than they are between blocks in the same season. The components of variation for flowering time, on the other hand, are considerably more variable over seasons than they are within seasons. This one might expect if the change in the components in the former case had been brought about in response to differences in soil fertility presented by different blocks, while in the latter case they were induced by differences in the weather conditions in the two seasons. This explanation is not only adequate to explain our present analyses but has the added advantage of falling into line with the findings of Mather and Vines, in respect of the character height, and with current views on the physiology of the two characters we are investigating here.

TABLE 13

The components of variation of the exclusive analysis for flowering time 1952 and inclusive analysis for 1953

Component	Season	
	1952	1953
D	20·0288± 3·1603	131·9830± 3·9340
H_1	4·4254±20·0658	75·0777±21·8369
H_2	2·1550±20·3633	65·5492±25·2722
F	5·2862±14·4487	31·0838±16·9979
E_1	13·5321± 3·8181	22·5840± 3·1316
H_3	110·2602±48·7401	...
H_4	132·8985±40·1544	...

The above analysis of variance does not allow us to partition the heterogeneity of components over seasons between the additive and dominance components of variation. We can, however, arrive at a partial solution of this problem by comparing the least square estimates of the various components over the two seasons. The components for flowering time obtained from the exclusive analysis in 1952 and the inclusive analysis in 1953 are given in table 13.

Since the analyses of the separate seasons reveals no significant disturbances after allowing for linkage in 1952, the standard errors of the components for this season have been calculated from the sum of squares of block differences of the observed statistics and so are based on sixteen degrees of freedom.

In 1953 there are other significant disturbances which will be discussed later and for this reason the standard errors of the components include the significant sum of squares for heterogeneity over statistics within blocks (*i.e.* residual interaction) and are therefore based on a total of twenty-seven degrees of freedom. For the same reason the linkage components are not given as their estimation is invalided.

Of the 1952 components of variation for flowering time H_1, H_2 and

F are not significant, while only F is non-significant in 1953. That dominance plays a part in the inheritance of flowering time in 1952 is, however, beyond dispute. Apart from the significant linkage components which are both terms in Σh, we have the significant Wr/Vr regressions for all three generations, and the significant heterosis for early flowering. But what is of more importance to our present discussion is that all components of variation apart from F are significantly different in the two seasons. Furthermore the dominance components H_1 and H_2 have changed by a greater factor than any of the other components of variation. This greater change, while large and consistent, is not significant. While, therefore, there is a suggestion that the dominance components are somewhat more variable over seasons than the additive components, in the present analyses they are not significantly so.

TABLE 14

The components of variation of the inclusive analysis of height

Component	Season			
	1952	1953		
		block I	block II	combined blocks
D	37·1331± 3·7209	44·4929± 5·2323	66·0078± 5·9605	55·2504± 5·0740
H_1	76·7428±20·6538	217·4862±29·0433	228·4762±33·0852	222·9812±28·1648
H_2	61·7516±23·9030	155·6763±33·6122	181·2954±38·2900	168·4859±32·5956
F	−60·9226±16·0770	−149·3846±22·6074	−167·2167±25·7536	−158·3007±21·9236
E_1	7·2307± 2·9620	5·2820± 4·1651	10·7260± 4·7447	8·0040± 4·0391

We find essentially the same situation for the components of variation of height. Owing to the presence of significant non-allelic interaction the standard errors of the components are based on the pooled sums of squares of block differences and residual interaction for a total of twenty-seven degrees of freedom. The components themselves, of course, are variously biased by the non-specified effects of the non-allelic interaction. They do serve, however, to illustrate the present discussion.

Apart from E_1 all other components of variation are significantly higher in 1953 than in 1952. This increase in magnitude is again more marked in the case of the dominance components H_1 and H_2 than in the additive component D, but again there is no significant difference in this respect between the two types of components. Since the components show almost as large a difference between blocks within seasons as they do between seasons, the components for 1953 have been separated into the two blocks. The standard errors of these components have been derived from the heterogeneity of components over the observed statistics within each block, and are hence

based on eleven degrees of freedom. Although these standard errors are inflated by disturbances arising from significant non-allelic interaction, they are the only ones available and will therefore have to serve for our present purpose. The only component that shows a significant difference in the two blocks is the additive component D. We have therefore examples where the dominance components are apparently less stable over environmental fluctuations and others where the additive component seems the less stable. In conclusion we can only say that on the whole the two components of variation, the additive and the dominance, are equally susceptible to changes in the environment, the latter in the case of height being predominantly relatable to the soil conditions.

TABLE 15

The analysis of variance of flowering time in 1952 and 1953

Item		1952		1953	
		N	MSS	N	MSS
Linkage		2	168·71	2	136·21
Residual interaction		9	15·87	9	101·76
Heterogeneity	*Between blocks*—				
	Linkage	2	13·35	2	16·58
	Residual interaction	9	3·88	9	5·31
	Components	5	80·98	5	5·63

(iv) *The linkage components*

The analysis of the two seasons separately shows that linkage plays a significant role in the inheritance of flowering time in both seasons (table 15).

Not only is this linkage apparently significantly variable over seasons (tables 12 and 13) but in 1953 there is also significant residual interaction which, while confirming the findings of the scaling tests, etc., invalidates both the test for linkage and the least squares estimates of the linkage components. Further discussion of linkage must therefore be confined to the 1952 analyses where no further interactive disturbances have been revealed by any of the available tests.

In 1952 the linkage components H_3 and H_4 are both positive and significantly different from zero. This suggests a preponderance of linkage between loci exhibiting reinforcing dominance. Since we only estimate the balance of the two types of dominance relations of the linked loci we cannot exclude the possibility that a proportion of the linked loci controlling flowering time shows opposing dominance relations.

From the ratio of the least squares estimates of the two linkage components H_3 and H_4 we can obtain an estimate of the linkage relations of the linked loci.

Thus
$$\frac{H_3}{H_4} = \frac{16\Sigma u_a v_a u_b v_b (1-2p)^2 h_a h_b}{16\Sigma u_a v_a u_b v_b (1-2p) h_a h_b}$$

which when $h_a = h_b = h_n$ and $u_a = u_b = u_n$ becomes

$$\frac{H_3}{H_4} = 1-2\bar{p}$$

where $\bar{p}$ is the mean recombination frequency of the linked genes. If these conditions are not fulfilled then the mean recombination frequency will be weighted in favour of linked genes with larger

TABLE 16

The fall ratios

	Fall ratios (Mather, 1949)	Our comparable estimate	Observed *fall ratios* for 1952. Flowering time
1	$\frac{(d^2)}{D_{F_2}}$	$\frac{H_1}{H_1+2H_4}$	0·0197±0·0896
2	$\frac{(h^2)}{H_{F_2}}$	$\frac{H_1}{H_1+2H_3}$	0·0164±0·0777
3	$\frac{D_{F_2}-D_{F_3}}{D_{F_2}}$	$\frac{2H_3-2H_4}{H_1+2H_4}$	0·1676±0·4702

dominance effects and equal allele frequencies. For the 1952 flowering time data we find that $\bar{p} = 0{\cdot}0852 \pm 0{\cdot}1277$. Owing to the relatively large standard error of this estimate it tells us very little except that there is significant linkage, *i.e.* $\bar{p}$ is significantly smaller than 0·50.

With preponderant reinforcing linkage such as we have found in the 1952 flowering time analyses we can use the relative magnitude of H_1 and the two linkage components H_3 and H_4 to obtain some idea of the number of linked genes (Mather, 1949). The difference between these three components depends on the relative magnitude of $h_a h_b$ etc., on the recombination frequencies and the number of genes involved. The estimation of the number of genes assumes equality of the $h_a h_b$ etc. and equal spacing along the genetical chromosome. Failure of either or both assumptions leads to an underestimation of the number of genes as does also the presence of a proportion of genes showing opposition linkage. The process of estimation depends on the observation that the maximum *fall ratio* is characteristic of the number of linked genes within the limitations of the above assumptions. Three *fall ratios* described by Mather can be adapted to our present components (table 16).

In equating our *fall ratios* to those of Mather we are making one further assumption, namely that for each pair of linked genes $4u_av_a+4u_bv_b = 32u_av_au_bv_b$. Non-equality will lead to a further reduction in the apparent number of linked genes, as calculated by all the above *fall ratios*.

All three observed fall ratios are smaller than their own standard errors. In fact the third *fall ratio* in table 15 gives no evidence of linkage at all. The other two, in so far as they are both significantly smaller than one, show that there is at least significant linkage in the flowering time data. Furthermore they are both significantly smaller than the maximum possible *fall ratios* for four linked factors which are 0·40 and 0·54 respectively (P = 0·001).

(v) *Disturbances due to interaction*

As pointed out earlier in this account, interaction, though not appearing explicitly as components of variation, may nevertheless be detected by their effects in causing heterogeneity of the specified

TABLE 17

Exclusive matrix for single array analysis

CDD	0·716050	CH_1D	1·185185	CF_r	−1·925926
CDH_1	1·185185	CH_1H_1	28·444444	CF_rH_1	−14·222222
CDF_r	−1·925926	CH_1F_r	−14·222222	CF_rF_r	15·111043

components. In the case of height there is significant residual interaction in both seasons. The joint analysis of 1952 and 1953 (table 12) confirms the significant residual interaction and, while showing significant heterogeneity of interaction over seasons, this heterogeneity is significantly smaller than the seasonal heterogeneity of the specified components of variation. In fact the residual interaction items are more stable over seasons and blocks within seasons than any other component of variation.

A constant feature of the analyses of these experiments by the Wr/Vr regression method has been that, following the omission of all crosses showing non-allelic interaction from the data, we have never found a dominance ratio significantly greater than one. We can now test this finding on a joint analysis over generations and seasons. Two fragments of the height data have been consistently free from interaction on all the tests made so far, namely array 7 and the 3×3 diallel between the common parents of arrays 2, 4 and 7. The latter can be analysed by the method already employed for the complete diallel, but the analysis of the single array requires a new matrix of multipliers. The equations appropriate for the analysis of single arrays have been given elsewhere (Jinks, 1955) and they provide nine statistics for the estimation of three components of variation, D, H_1 and F_r ($F_r = 8\Sigma uvdh$). The appropriate matrix of multipliers for the general solution, *i.e.* one omitting the environmental components that depend on the design of the experiment, is given in table 17.

We can now calculate the least squares estimates of the three parameters in the usual way and from them predict the expectations of the nine equations. The analysis of variance of the sums of squares of deviations of observed from expected is given in table 18.

The only significant item in the above analysis of variance is the heterogeneity of components, *i.e.* D, H_1 and F_r, over seasons. There

TABLE 18

Analysis of array 7 for height (exclusive)

Item		N	MSS
Residual interaction . . .		6	80·61
Heterogeneity	*Between seasons*—		
	Residual interaction . .	6	121·22
	Components . . .	3	2315·87
	Within seasons—		
	Residual interaction . .	12	42·19
	Components . . .	6	136·52

is no significant residual interaction and the heterogeneity of residual interaction over seasons, which borders on significance, is probably merely a reflection of the significant heterogeneity of components. As the analysis of seasons separately reveals no significant disturbances, the standard errors of the components in the two seasons are based on the sum of squares of block differences of the observed statistics for nine degrees of freedom (table 19).

TABLE 19

The components of variation for height in array 7 in 1952 and 1953

Component	1952	1953
D . . .	44·5907± 3·2035	61·1961± 5·1225
H_1 . . .	38·1536±20·1910	94·4135±32·2854
F_r . . .	39·3485±14·7165	102·3600±23·5314

Apart from H_1 in 1952, which is not significantly different from zero at the 5-10 per cent. level of probability, all other components of variation are significant. Furthermore, there is no evidence of overdominance in this non-interacting array, *i.e.* D-H_1 is not significantly different from zero. This is in marked contrast to the analysis of the complete height data where there is significant interaction and significant overdominance. The only other point of note is that all the components of variation are greater in 1953 than in 1952 but only D and F_r are significantly so.

The analysis of the non-interacting 3×3 diallel does not differ from the above in any important detail. Thus the dominance ratio is again not significantly different from complete dominance, *i.e.* $H_1/D = 0{\cdot}90$, and there is no significant residual interaction.

Flowering time in 1952 shows no significant residual interaction (table 15) when tested against its own heterogeneity over duplicate blocks or against the pooled heterogeneities of all items. In 1953, on the other hand, we have highly significant residual interaction. This difference over seasons is confirmed by the joint analysis of the two seasons (table 12), where the heterogeneity of residual interaction is also significant. Thus these findings confirm those obtained from the scaling tests in both seasons. We have, therefore, four tests which agree in detecting no interaction in 1952, and three tests (that is all except the Wr/Vr regression), which have detected interaction in 1953. Their non-detection by the Wr/Vr regression is, however, a valuable pointer to the type of interaction involved. There can be no doubt that it is of a different type from that found in height, which is comparable with the complementary gene interactions of classical genetics. It in no way appears to lead to spurious inflations or any other changes in the dominance ratio. Thus the unit slopes of the Wr/Vr regressions show that the estimated dominance ratio is constant over arrays, *i.e.* $Wr-Vr = \frac{1}{4}(D-H_1)$ for the F_1 generation and $= \frac{1}{4}(D-\frac{1}{4}H)$ for the F_2 and backcross generations, is constant, and hence it must be relatively uninfluenced by the non-allelic interactions whose incidence varies amongst the arrays (table 1*c*). The only comparable type of interaction in classical genetics which would fit these observations are duplicate genes. Why these interactions should be confined to the 1953 season, or alternatively, why they should remain undetected by all available methods in 1952 we cannot say as yet. But it is possible that the smaller total variation for this character in 1952, which is approximately $\frac{1}{5}$th of that in 1951 and 1953 is at least partly responsible for this failure. On the other hand there may be a genuine absence of interaction in 1952, which must then be ascribed to genotype-environmental interactions which differ in the two seasons.

6. CONCLUSIONS

The present paper in the series on diallel crosses has two primary aims. Firstly, it aims to extend the theory and illustrate its application to F_2 and backcross generations derived from a set of diallel crosses ; and secondly, it aims to see how far conclusions and predictions from the analysis of parental and F_1 means are borne out by the fuller analyses made possible by the inclusion of these later generations.

The conclusions of these investigations into the character height are easily summarised—in all essential details there is complete agreement between the assessment based on parents and F_1s alone and that based on the analysis when extended to F_2s and backcrosses

This close agreement is nowhere more striking than in relation to the genetical control of heterosis for this character. Every test designed to detect non-allelic interaction has given a positive result with the height data. Furthermore, the tests which allow one to trace the origin of the non-allelic interaction to particular crosses or arrays, *e.g.* the Wr/Vr regressions and the scaling tests, have in all cases picked out the same combinations of parental lines as those responsible for giving rise to the interactions. We may further note that the inclusion of F_2s and backcrosses in these analyses has in no way altered our conclusions concerning the role of overdominance. In the joint analyses in both seasons we have found no evidence of overdominance after omitting from the analyses all crosses in which we have detected non-allelic interactions by means of the scaling tests and the F_1 Wr/Vr regressions. We can take it, therefore, that all the heterosis is the result of interactions of a complementary type between alleles at different loci. With such consistency over generations and seasons the diallel analysis of parental and F_1 means provides an accurate and rapid method of assessing the potentialities of the various F_1 combinations as well as providing valuable information about the genetical control of heterosis. The ready availability of such information cannot but seriously affect future methods of utilising this heterosis for economic purposes.

Turning to flowering time, we again find good agreement between the various methods of analysis, but for this character the agreement is confined to within seasons, there being significant differences between seasons for the genetical control of this character. Thus while all tests agree that non-allelic interaction is absent in 1952, the same tests consistently show its presence in 1953. The F_1 Wr/Vr regression alone of all the available methods of detecting non-allelic interaction has failed to detect non-allelic interaction for flowering time in 1953. Since this was the only test available in 1951 there is the possibility that non-allelic interaction played a significant role in the inheritance of flowering time in that season also. This failure of the F_1 Wr/Vr method has been traced to its inherent insensitivity to the type and magnitude of the non-allelic interaction found in 1953. Thus while this method is highly sensitive to the complementary type of interaction, as found in our own height data, it is not so sensitive to duplicate gene interactions which we have now found in the flowering time data in 1953. For flowering time, therefore, the diallel analysis of parental and F_1 means provides an incomplete assessment of the potentialities of the various crosses. Furthermore, the high degree of instability of the genetical control of this character over seasons makes an assessment based on any one season of doubtful value, although, of course, the ability to detect these seasonal differences and further to partition the changes in genetical control amongst the components of variation is an important contribution of this method of analysis.

7. SUMMARY

1. The theory of the diallel analysis of parental and F_1 means has now been extended to the F_2 and backcross generations derived from a diallel set of crosses.

2. The joint analysis provides sufficient statistics to give least squares estimates of the components of variation, *i.e.* the additive, dominance and environmental effects, and their standard errors.

3. A number of methods of detecting non-allelic interaction are given, including the regression of array covariance on array variance, which is applicable to any generation, the joint scaling tests and the homogeneity of the least squares estimates of the components of variation over statistics.

4. If the within-family variances of the segregating generations are also included in the analyses we can detect linkage and estimate its effect.

Illustrations of the analyses are drawn from our own data from *Nicotiana rustica*. This comprises an 8×8 diallel, the parents and F_1s of which have been grown in three consecutive seasons and the F_2s and backcrosses in the last two.

These analyses have shown that :

(i) All tests agree that heterosis for the character height is the result of a complementary type of non-allelic interaction.

(ii) The differences in magnitude between the components of variation for height in different seasons and different blocks within seasons is a genotype-environment interaction dependent on soil differences.

(iii) The analysis of parents and F_1s alone in any one season provides a satisfactory assessment of the genetical control of the character height that is completely borne out by subsequent generations and seasons.

(iv) There are significant differences in the genetical control of flowering time in the two seasons 1952 and 1953, primarily relatable to differences in the weather. These differences involve not only variation in the magnitude of the components of variation but also differences involving the presence of duplicate gene interactions in 1953 and their absence in 1952.

(v) Linkage between alleles which exhibit reinforcing dominance relations is detectable in the flowering time data of 1952 and involves at least four factors.

8. REFERENCES

CAVALLI, L. L. 1952. An analysis of linkage in quantitative inheritance. *Quantitative Inheritance*. H.M.S.O. London, 135-144.

HAYMAN, B. I. 1954. The theory and analysis of diallel crosses. *Genetics, 39*, 789-809.

HAYMAN, B. I., AND MATHER, K. 1955. The description of genic interactions in continuous variation. *Biometrics, 11*, 69-82.

JINKS, J. L. 1954*a*. The genetical basis of heterosis. *Maize Genetics*. Cooperation News Letter, Dept. of Plant Breeding, Cornell Univ., Ithaca, N.Y., *28*, 47-50.

JINKS, J. L. 1954*b*. The analysis of continuous variation in a diallel cross of *Nicotiana rustica* varieties. *Genetics*, *39*, 767-788.

JINKS, J. L. 1955. A survey of the genetical basis of heterosis in a variety of diallel crosses. *Heredity*, *9*, 223-238.

JINKS, J. L., AND HAYMAN, B. I. 1953. Analysis of diallel crosses. *Maize Genetics*. Cooperation News Letter, *27*, 48-54.

JINKS, J. L., AND MATHER, K. 1955. Stability in development of heterozygotes and homozygotes. *P.R.S.*, B, *143*, 561-578.

MATHER, K. 1949. *Biometrical Genetics*. Methuen, London.

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