COMPONENTS OF CONTINUOUS VARIATION IN PRESENCE OF EPISTASIS AND GENOTYPE × ENVIRONMENT INTERACTIONS*

BY T. C. S. R. SARMA§ AND PREM NARAIN

Institute of Agricultural Research Statistics, New Delhi

INTRODUCTION

THE manner in which the continuous variation may be partitioned into various components and the way in which these may be apportioned to fit certain experimental populations are of considerable interest to the plant or animal breeder. When two true breeding strains (or strains sufficiently inbred to be taken as true breeding) differing significantly in respect of the character under study are available, experiments based on different mating systems can be designed to yield a great multiplicity of statistics for the estimation of various components of continuous variation. The biometrical analysis in such a case, as suggested by Mather (1949), depends on the availability of a scale on which the genic effects are simply additive on the average and on which the contribution made by the environmental effects are independent of the genotype. Allowance is, however, made for the lack of additiveness due to dominance action of the genes by the inclusion of a special term in the analysis but the effects of epistatic and genotype × environment interactions are assumed to have been eliminated or minimised by proper scaling. The partitioning of the total observed variation is then into three components, *viz.*, (i) additive genetic (or simply genic), (ii) dominance and (iii) environmental. It may not, however, always be possible to find a scale which wholly removes the non-additiveness in effect due to epistatic and genotype × environment interactions. It is, therefore, necessary to take into account these effects by including further terms in the analysis. The partitioning of the total variation would then lead to two additional

components, *viz.*, (i) epistatic and (ii) genotype × environmental, the latter being further capable of breaking down into three components corresponding to three partitions of the genotype, *viz.*, genic, dominance and epistasis. Hayman and Mather (1955) studied the case when effects due to epistasis are taken into account but the effects due to genotype × environment interactions are left out whereas Mather and Jones (1958) studied the case when effects due to genotype × environment interactions are considered but the effects due to epistasis are assumed to be absent. In this paper both these effects are considered simultaneously and the resulting components of the continuous variation are described.

2. DESCRIPTION OF THE PARAMETERS

With two gene pairs *A–a* and *B–b*, there are nine genotypes out of which four are homozygotes, four single heterozygotes and one double heterozygote. If we consider a F_2 derived from the initial cross between two true breeding strains observed under two environments *P* and *Q* there are eighteen situations possible and seventeen parameters will be necessary to describe completely the differences among the eighteen phenotypes. Four of these parameters are the usual genetic quantities d_a, d_b, h_a and h_b. As defined by Mather (1949) d_a measures the effect of phenotypic difference between the homozygotes and h_a measures the phenotypic comparison of heterozygote with the mean of the two homozygotes for the gene pair *A–a* when the other gene pair is averaged out. Similarly, d_b and h_b can be defined. Four other parameters are $i_{ab/}$, $j_{a/b}$, $j_{b/a}$ and $l_{/ab}$ corresponding to digenic interactions as defined by Hayman and Mather (1955). The ninth parameter is e_1 representing the difference between the two environments, being the amount by which the phenotype in environment *P* when averaged over genotypes exceeds and that in environment *Q* similarly falls short of the overall mean. The interactions of the eight parameters with e_1 comparison can be represented by the remaining eight parameters $g_{d_{a_1}}$, $g_{d_{b_1}}$, $g_{h_{a_1}}$, $g_{h_{b_1}}$, $g_{i_{ab/_1}}$, $g_{j_{a/b_1}}$, $g_{j_{b/a_1}}$, g_{l/ab_1}. The two sets of eight parameters are independent of one another and within a set also the parameters are independent of each other. With these parameters, the eighteen phenotypes of F_2 observed in two environments *P* and *Q* can be described as shown in Table I.

The above description can be extended to cover cases of more than two gene pairs, provided digenic interactions are only considered as a first approximation to trigenic and other higher order interactions. Another gene difference represented by *C–c* when added brings in

4

TABLE I

Eighteen phenotypes from nine genotypes in two environments

Geno-type	Environment	P — Q

Genotype	P — Q
AABB	$[d_a + d_b + i_{ab/} - \frac{1}{2}(j_{a/b} + j_{b/a}) + \frac{1}{4}l_{/ab}]$ $\pm [g_{d_{a_1}} + g_{d_{b_1}} + g_{iab/_1} - \frac{1}{2}(g_{ja/b_1} + g_{jb/a_1}) + \frac{1}{4}g_{l/ab_1} + e_1]$
AABb	$[d_a + h_b + \frac{1}{2}j_{a/b} - \frac{1}{4}l_{/ab}]$ $\pm [g_{d_{a_1}} + g_{h_{b_1}} + \frac{1}{2}g_{ja/b_1} - \frac{1}{4}g_{l/ab_1} + e_1]$
AAbb	$[d_a - d_b - i_{ab/} - \frac{1}{2}(j_{a/b} - j_{b/a}) + \frac{1}{4}l_{/ab}]$ $\pm [g_{d_{a_1}} - g_{d_{b_1}} - g_{iab/_1} - \frac{1}{2}(g_{ja/b_1} - g_{jb/a_1}) + \frac{1}{4}g_{l/ab_1} + e_1]$
AaBB	$[h_a + d_b + \frac{1}{2}j_{b/a} - \frac{1}{4}l_{/ab}]$ $\pm [g_{h_{a_1}} + g_{d_{b_1}} + \frac{1}{2}g_{jb/a_1} - \frac{1}{4}g_{l/ab_1} + e_1]$
AaBb	$[h_a + h_b + \frac{1}{4}l_{/ab}]$ $\pm [g_{h_{a_1}} + g_{h_{b_1}} + \frac{1}{4}g_{l/ab_1} + e_1]$
Aabb	$[h_a - d_b - \frac{1}{2}j_{b/a} - \frac{1}{4}l_{/ab}]$ $\pm [g_{h_{a_1}} - g_{d_{b_1}} - \frac{1}{2}g_{jb/a_1} - \frac{1}{4}g_{l/ab_1} + e_1]$
aaBB	$[-d_a + d_b - i_{ab/} + \frac{1}{2}(j_{a/b} - j_{b/a}) + \frac{1}{4}l_{/ab}]$ $\pm [-g_{d_{a_1}} + g_{d_{b_1}} - g_{iab/_1} + \frac{1}{2}(g_{ja/b_1} - g_{jb/a_1}) + \frac{1}{4}g_{l/ab_1} + e_1]$
aaBb	$[-d_a + h_b - \frac{1}{2}j_{a/b} - \frac{1}{4}l_{/ab}]$ $\pm [-g_{d_{a_1}} + g_{h_{b_1}} - \frac{1}{2}g_{ja/b_1} - \frac{1}{4}g_{l/ab_1} + e_1]$
aabb	$[-d_a - d_b + i_{ab/} + \frac{1}{2}(j_{a/b} + j_{b/a}) + \frac{1}{4}l_{/ab}]$ $\pm [-g_{d_{a_1}} - g_{d_{b_1}} + g_{iab/_1} + \frac{1}{2}(g_{ja/b_1} + g_{ib/a_1}) + \frac{1}{4}g_{l/ab_1} + e_1]$
F_2 Mean	$[\frac{1}{2}h_a + \frac{1}{2}h_b] \pm [\frac{1}{2}g_{h_{a_1}} + \frac{1}{2}g_{h_{b_1}} + e_1]$

Note.—Environment P will be having + ve sign whereas Q will be having − ve sign.

a further d_c, h_c, $i_{ac/}$, $j_{a/c}$, $j_{c/a}$, $l_{/ac}$, $i_{bc/}$, $j_{b/c}$, $j_{c/b}$, $l_{/bc}$. To accommodate their interactions with the environment, a further set of g's, *i.e.*, $g_{d_{c_1}}$, $g_{h_{c_1}}$, $g_{iac/_1}$, g_{ja/c_1}, g_{jc/a_1}, g_{l/ac_1}, $g_{ibc/_1}$, g_{jb/c_1}, g_{jc/b_1}, g_{l/bc_1}, can be introduced. These

g's will be as independent of one another in their contribution to variances as are the various d's, h's, i's, j's and l's.

Turning to the case of more than two environments it can be seen that, whatever may be their number, their differences are expressible by a series of orthogonal comparisons equal in number to the degrees of freedom between them. In case of two environments there is only one comparison as seen above and represented by e_1. Three environments will yield two parameters e_1 and e_2 and so on for more than three environments. The average phenotypes of the different environments can be combined in a variety of ways giving appropriate sets of orthogonal comparisons. As pointed out by Mather and Jones (1958), we may choose any one set of comparisons whichever is the most convenient for representing environmental differences. As long as the comparisons in a set are orthogonal to one another, the sum of squares will necessarily be equal to the sum of squares of deviations of the environmental means from the general mean. The number of g's will increase as the number of environments is increased. With two environmental comparisons e_1 and e_2, the interaction parameters with these will be represented by $g_{d_{a_1}}$, $g_{h_{a_1}}$, $g_{d_{b_1}}$, $g_{h_{b_1}}$, $g_{i_{ab/_1}}$, $g_{j_{a/b_1}}$, $g_{j_{b/a_1}}$, g_{l/ab_1}, $g_{d_{a_2}}$ $g_{h_{a_2}}$, $g_{d_{b_2}}$, $g_{h_{b_2}}$, $g_{i_{ab/_2}}$, $g_{j_{a/b_2}}$, $g_{j_{b/a_2}}$ and g_{l/ab_2}. It is thus seen that with more environments, the total number of parameters will increase manifold. For instance, when the nine genotypes of a F_2 are observed in four environments there will be thirty-six phenotypes and in all there will be thirty-five parameters representing differences among these phenotypes.

3. CONTRIBUTION OF INTERACTION TO VARIATION

When the chances of occurrence of the various genotypes are the same in all environments the overall mean phenotype will be independent of the genotype × environment interactions represented by g items. The effects of the interaction on the phenotypic variances can be illustrated from the case of two gene differences in two environments P and Q as given in Table I.

The variance of a line pure for $AABB$ genotype over two environments is given by:

$$[e_1 + g_{d_{a_1}} + g_{d_{b_1}} + g_{i_{ab/_1}} - \tfrac{1}{2}g_{j_{a/b_1}} - \tfrac{1}{2}g_{j_{b/a_1}} + \tfrac{1}{4}g_{l/ab_1}]^2$$
$$+ [d_a + d_b + i_{ab/} - \tfrac{1}{2}j_{a/b} - \tfrac{1}{2}j_{b/a} + \tfrac{1}{4}l_{/ab}]^2$$

when taken around zero. In a similar way the variances of various homozygotes and heterozygotes can be obtained. These variances are given in Table II.

TABLE II

Variances of homozygotes and heterozygotes

Genotype	Variances when taken about zero
$AABB$	$[d_a+d_b+i_{ab/}-\frac{1}{2}(j_{a/b}+j_{b/a})+\frac{1}{4}l_{/ab}]^2$ $+[e_1+g_{d_{a_1}}+g_{d_{b_1}}+g_{i_{ab/_1}}-\frac{1}{2}(g_{j_{a/b_1}}+g_{j_{b/a_1}})+\frac{1}{4}g_{l/_{ab_1}}]^2$
$AABb$	$[d_a+h_b+\frac{1}{2}j_{a/b}-\frac{1}{4}l_{/ab}]^2+[e_1+g_{d_{a_1}}+g_{h_{b_1}}+\frac{1}{2}g_{j_{a/b}}-\frac{1}{4}g_{l/_{ab_1}}]^2$
$AAbb$	$[d_a-d_b-i_{ab/}-\frac{1}{2}(j_{a/b}-j_{b/a})+\frac{1}{4}l_{/ab}]^2$ $+[e_1+g_{d_{a_1}}-g_{d_{b_1}}-g_{i_{ab/_1}}-\frac{1}{2}(g_{j_{a/b_1}}-g_{j_{b/a_1}})+\frac{1}{4}g_{l/_{ab_1}}]^2$
$AaBB$	$[h_a+d_b+\frac{1}{2}j_{b/a}-\frac{1}{4}l_{/ab}]^2+[e_1+g_{h_{a_1}}+g_{d_{b_1}}+\frac{1}{2}g_{j_{b/a_1}}-\frac{1}{4}g_{l/_{ab_1}}]^2$
$AaBb$	$[h_a+h_b+\frac{1}{4}l_{/ab}]^2+[e_1+g_{h_{a_1}}+g_{h_{b_1}}+\frac{1}{4}g_{l/_{ab_1}}]^2$
$Aabb$	$[h_a-d_b-\frac{1}{2}j_{b/a}-\frac{1}{4}l_{/ab}]^2+[e_1+g_{h_{a_1}}-g_{d_{b_1}}-\frac{1}{2}g_{j_{b/a}}-\frac{1}{4}g_{l/_{ab_1}}]^2$
$aaBB$	$[-d_a+d_b-i_{ab/}+\frac{1}{2}(j_{a/b}-j_{b/a})+\frac{1}{4}l_{/ab}]^2$ $+[e_1-g_{d_{a_1}}+g_{d_{b_1}}-g_{i_{a/b_1}}+\frac{1}{2}(g_{j_{a/b_1}}-g_{j_{b/a_1}})+\frac{1}{4}g_{l/_{ab_1}}]^2$
$aaBb$	$[-d_a+h_b-\frac{1}{2}j_{a/b}-\frac{1}{4}l_{/ab}]^2+[e_1-g_{d_{a_1}}+g_{h_{b_1}}-\frac{1}{2}g_{j_{a/b}}-\frac{1}{4}g_{l/_{ab_1}}]^2$
$aabb$	$[-d_a-d_b+i_{ab/}+\frac{1}{2}(j_{a/b}+j_{b/a})+\frac{1}{4}l_{/ab}]^2$ $+[e_1-g_{d_{a_1}}-g_{d_{b_1}}+g_{i_{ab/_1}}+\frac{1}{2}(g_{j/_{ab_1}}+g_{j_{b/a_1}})+\frac{1}{4}g_{l/_{ab_1}}]^2$

A familiar type of experiment used for the estimation of components of variation is to cross two true breeding strains and make intercrossing between the resulting F_1's (or self F_1 in case of plants practising self-fertilization) which leads to F_2 generation. This F_2 generation can be selfed to give F_3 generations in case of plants practising self-fertilization or it can be crossed by taking pairs of individuals at random from it giving rise to biparental progenies (BIP). In case of plants it can also give maternal progenies (MAT) obtained by exposing mother plants to the pollen of other plants. Such an experiment leads to multiplicity of second degree statistics containing information about the various components of variation. In particular, we may derive expressions for (*a*) variance of parents considered as

genetically pure, (b) variance of F_1, (c) variance of F_2, (d) variances in F_3 generation, (e) variances in BIP generation, and (f) variances in MAT generation.

(a) *Mean variance of true breeding parents*: $\{\frac{1}{2}(V_{P_1} + V_{P_2})\}$:—With two gene differences, there can be two types of parental crosses (a) 'Associated', *i.e.*, $AABB \times aabb$ and (b) 'Dispersed', *i.e.*, $AAbb \times aaBB$. In the former case, using the results given in Table II, the variance is given by

$$(e_1 + g_{i_{a/b_1}} + \tfrac{1}{4}g_{l/ab_1})^2 + [g_{d_{a_1}} + g_{d_{b_1}} - \tfrac{1}{2}(g_{j_{a/b_1}} + g_{j_{b/a_1}})]^2,$$

whereas in the latter case it is given by

$$(e_1 - g_{i_{a/b_1}} + \tfrac{1}{4}g_{l/ba_1})^2 + [g_{d_{a_1}} - g_{d_{b_1}} - \tfrac{1}{2}(g_{j_{a/b_1}} - g_{j_{b/a_1}})]^2,$$

the variances having been taken around their own means.

The two types of situation enumerated above can, however, be described in one way if parameters θ_a and θ_b are introduced to indicate the association of the genes in the parents. θ_a and θ_b are positive unity when the first parent contains the positive homozygote of the corresponding gene and negative unity, when the second parent contains the positive homozygote. The mean variance of the parents can then be expressed as

$$(e_1 + g_{i_{ab/_1}}\theta_a\theta_b + \tfrac{1}{4}g_{l/ab_1})^2 + [g_{d_{a_1}}\theta_a + g_{d_{b_1}}\theta_b - \tfrac{1}{2}(g_{j_{a/b_1}}\theta_a + g_{j_{b/a_1}}\theta_b)]^2.$$

With more than two gene differences this expression would take the form

$$[e_1 + g_{i_1} + \tfrac{1}{4}g_{l_1}]^2 + (g_{d_1} - g_{j_1})^2$$

where g_{d_1}, g_{i_1}, g_{j_1}, and g_{l_1} are pooled effects of corresponding interactions as defined below:

$$g_{d_1} = \sum_a g_{d_{a_1}}\theta_a, \qquad\qquad g_{i_1} = \sum_{a<b} g_{i_{ab/_1}}\theta_a\theta_b$$

$$g_{j_1} = \tfrac{1}{2}\sum_{a<b}[g_{j_{a/b_1}}\theta_a + g_{j_{b/a_1}}\theta_b], \quad g_{l_1} = \sum_{a<b} g_{l/ab_1}.$$

Bringing in more than two environments would lead to the general expression.

$$\tfrac{1}{2}(V_{P_1} + V_{P_2}) = \sum_k (e_k + g_{i_k} + \tfrac{1}{4}g_{l_k})^2 + \sum_k (g_{d_k} - g_{j_k})^2$$

$$= V_{(e+g_i+\frac{1}{4}g_l)} + V_{(g_d-g_j)} \tag{1}$$

since

$$\sum_k (e_k + g_{i_k} + \tfrac{1}{4} g_{l_k}) = \sum_k (g_{d_k} - g_{j_k}) = 0,$$

summation being over different environmental comparisons.

(b) *Variance of F_1*: (V_{F_1}):—The F_1 would contain uniformly $AaBb$ genotypes with two gene differences. The variance, according to Table II, is given by $[e_1 + g_{h_{a1}} + g_{h_{b1}} + \tfrac{1}{4} g_{l/ab_1}]^2$ when taken about its own mean over the environments. With several gene differences, this variance would take the form

$$[e_1 + g_{h_1} + \tfrac{1}{4} g_{l_1}]^2$$

where

$$g_{h_1} = \sum_a g_{h_{a_1}} \; : g_{l_1} = \sum_{a<b} g_{l/ab_1}$$

Taking into account several environments we get

$$V_{F_1} = \sum_k (e_k + g_{h_k} + \tfrac{1}{4} g_{l_k}]^2$$

$$= V_{(e+g_h+\frac{1}{4}gl)} \tag{2}$$

(c) *Variance of F_2*: (V_{F_2}):—In F_2, the nine genotypes listed in Table II occur in the proportion $1:2:1:2:4:2:1:2:1$. The contribution of two gene differences when two environments are only considered can, therefore, be obtained by multiplying the expressions for variances about the origin zero given in Table II by respective proportions and subtracting the square of overall F_2 mean, i.e., $(\tfrac{1}{2}h_a + \tfrac{1}{2}h_b)$ from it. When simplified this contribution is found to be

$$\tfrac{1}{2}(d_a{}^2 + d_b{}^2) + \tfrac{1}{4}(h_a{}^2 + h_b{}^2) + \tfrac{1}{4} i^2{}_{ab/} + \tfrac{1}{8}(j^2{}_{a/b} + j^2{}_{b/a})$$

$$+ \tfrac{1}{16} l^2{}_{/ab} + \tfrac{1}{2}(g_{d_a}{}^2 + g_{d_b}{}^2) + \tfrac{1}{4}(g_{h_{a_1}}{}^2 + g_{h_{b_1}}{}^2)$$

$$+ \tfrac{1}{4} g^2{}_{l_{ab/_1}} + \tfrac{1}{8}(g^2{}_{j_{a/b_1}} + g^2{}_{j_{b/a_1}}) + \tfrac{1}{16} g^2{}_{l/ab_1}$$

$$+ (e_1 + \tfrac{1}{2}g_{h_{a_1}} + \tfrac{1}{2}g_{h_{b_1}})^2.$$

When more than two environments are considered, this takes the form

$$\tfrac{1}{2}(d_a{}^2 + d_b{}^2) + \tfrac{1}{4}(h_a{}^2 + h_b{}^2) + \tfrac{1}{4} i^2{}_{ab/} + \tfrac{1}{8}(j^2{}_{a/b} + j^2{}_{b/a})$$

$$+ \tfrac{1}{16} l^2{}_{/ab} + \tfrac{1}{2}(V_{g d_a} + V_{v d_b}) + \tfrac{1}{4}(V_{g h_a} + V_{g h_b})$$

$$+ \tfrac{1}{4} V_{g i_{ab/}} + \tfrac{1}{8}(V_{g j_{a/b}} + V_{g j l/a}) + \tfrac{1}{16} V_{g l/ab}$$

$$+ V_{(e+\frac{1}{2}g h_a + \frac{1}{2}g h_b)}$$

where

$$V_x = \sum_k x_k{}^2 \quad \text{since} \quad \sum_k x_k = 0$$

summation being over different environmental comparisons.

With several gene differences, V_{F_2} is given by

$$V_{F_2} = \tfrac{1}{2}D + \tfrac{1}{4}H + \tfrac{1}{4}I + \tfrac{1}{8}J + \tfrac{1}{16}L$$
$$+ \tfrac{1}{2}G_D + \tfrac{1}{4}G_H + \tfrac{1}{4}G_I + \tfrac{1}{8}G_J + \tfrac{1}{16}G_L + E_{F_2}, \qquad (3)$$

where

$$D = \sum_a d_a{}^2 \;;\; H = \sum_a h_a{}^2 \;;\; I = \sum_{a<b} i^2{}_{ab/}$$

$$J = \sum_{a<b} (j^2{}_{a/b} + j^2{}_{b/a}) \;\;;\; L = \sum_{a<b} l^2{}_{/ab}$$

$$G_D = \sum_a V_{g d_a} \;\;;\; G_H = \sum_a V_{g h_a} \;\;;\; G_I = \sum_{a<b} V_{g i_{ab/}}$$

$$G_J = \sum_{a<b} (V_{g i_{a/b}} + V_{g j_{b/a}}) \;\;;\; G_L = \sum_{a<b} V_{g l/ab} \quad \text{and}$$

$$E_{F_2} = V_{\left(e + \frac{1}{2} \sum_a g_{h_a}\right)} \cdot$$

(d) *Variance of F_3 generation*: (V_{F_3}):—In a F_3 population obtained by selfing F_2 in case of plants capable of self-fertilization, the nine genotypes, with two gene differences, would occur in the proportions $9:6:9:6:4:6:9:6:9$. As before, the contribution to V_{F_3} is obtained by multiplying the expressions for variances around zero given in Table II with the respective proportions and subtracting the square of overall F_3 mean, *i.e.*, $(\tfrac{1}{4}h_a + \tfrac{1}{4}h_b + \tfrac{1}{16}l_{/ab})$ from it. After simplification, the resulting expression for any number of environments and several gene differences is given by

$$V_{F_3} = \tfrac{3}{4}D' + \tfrac{3}{16}H' + \tfrac{9}{16}I + \tfrac{9}{64}J + \tfrac{9}{256}L + \tfrac{3}{4}G_{D'}$$
$$+ \tfrac{3}{16}G_{H'} + \tfrac{9}{16}G_I + \tfrac{9}{64}G_J + \tfrac{9}{256}G_L + E_{F_3} \qquad (4)$$

where

$$D' = \sum_{a \neq b} [d_a - \tfrac{1}{4}\Sigma j_{a/b}]^2 \;\;:\; H' = \sum_{a<b} (h_a - \tfrac{1}{4}\Sigma l_{/ab}]^2$$

$$G_{D'} = \sum_e V'_{g d_a}, \qquad G_{H'} = \sum_a V'_{g h_a}$$

$$V'_{g d_a} = \sum_k (g_{d_{a_k}} + \tfrac{1}{4} g_{j_{a/b_k}}]^2$$

$$V'_{g h_a} = \sum_k [g_{h_{a_k}} - \tfrac{1}{4} g_{l/ab_k}]^2, \qquad \text{and}$$

$$E_{F_3} = V_{[e + \frac{1}{4} \sum_a g_{h_a} + \frac{1}{16} \sum_{a < b} g_{l/ab}]}$$

I, J, L, G_I, G_J and G_L being defined as in (3).

The total F_3 variance is, however, composed of two parts V_{1F_3} (variance of F_3 means) and V_{2F_3} (mean variance of F_3 families). If each F_3 family is consisting of a large number of individuals and the families are distributed randomly over the environments, the variance of F_3 families would be the variance of families averaged over environments. With such a design of experiment the families would have means as shown in Table III taking for illustration only two environments P and Q.

These families occur in the proportion $1:2:1:2:4:2:1:2:1$ and therefore the contribution of two gene differences to V_{1F_3} would be obtained by multiplying the squares of the expressions for means over the two environments given in Table III with the respective proportions and subtracting the square of the overall F_3 mean from it. This, on simplification, leads to

$$\tfrac{1}{2} (d_a - \tfrac{1}{4} j_{a/b})^2 + \tfrac{1}{2} (d_b - \tfrac{1}{4} j_{b/a})^2 + \tfrac{1}{16} (h_a - \tfrac{1}{4} l_{/ab})^2$$
$$+ \tfrac{1}{16} (h_b - \tfrac{1}{4} l_{/ab})^2 + \tfrac{1}{4} i^2_{ab/} + \tfrac{1}{32} (j^2_{a/b} + j^2_{b/a}) +$$
$$+ \tfrac{1}{256} l^2_{/ab}.$$

With several gene differences this is expressed in the form

$$V_{1F_3} = \tfrac{1}{2} D' + \tfrac{1}{16} H' + \tfrac{1}{4} I + \tfrac{1}{32} J + \tfrac{1}{256} L \tag{5}$$

which is same as that obtained by Hayman & Mather (1955) without taking into account the genotype × environment interactions.

Using

$$V_{F_3} = V_{1F_3} + V_{2F_3}, \quad V_{2F_3} \text{ is then given by}$$

$$V_{2F_3} = \tfrac{1}{4} D' + \tfrac{1}{8} H' + \tfrac{5}{16} I + \tfrac{7}{64} J + \tfrac{1}{32} L$$
$$+ \tfrac{3}{4} G_{D'} + \tfrac{3}{16} G_{H'} + \tfrac{9}{16} G_I + \tfrac{9}{64} G_J + \tfrac{9}{256} G_L + E_{F_3} \tag{6}$$

Where the families are small in size or the individuals of each family are kept together in an experimental design and not distributed randomly over the environments, V_{1F_3} would be inflated partly by sampling variation and partly by the environmental and associated interactive differences between their sites or plots. It may be noted

!

TABLE III

Phenotypic F$_3$ family means in two environments

Geno-type	Environment	$+P$ $-Q$	Mean
AABB	..	$[d_a+d_b+i_{ab/}-\frac{1}{2}(j_{a/b}+j_{b/a})+\frac{1}{4}l_{/ab}]$ $\pm[e_1+g_{d_{a_1}}+g_{d_{b_1}}+g_{iab/_1}-\frac{1}{2}(g_{ja/b_1}+g_{jb/a_1})+\frac{1}{4}g_{l/ab_1}]$	$[d_a+d_b+i_{ab/}-\frac{1}{2}(j_{a/b}+j_{b/a})+\frac{1}{4}l_{/ab}]$
AABb	..	$[d_a+\frac{1}{2}h_b]\pm[e_1+g_{d_{a1}}+\frac{1}{2}g_{hb_1}]$	$[d_a+\frac{1}{2}h_b]$
AAbb	...	$[d_a-d_b-i_{ab/}-\frac{1}{2}(j_{a/b}-j_{b/a})+\frac{1}{4}l_{/ab}]$ $\pm[e_1+g_{d_{a_1}}-g_{d_b}-g_{iab/_1}-\frac{1}{2}(g_{ja/b_1}-g_{jb/a_1})+\frac{1}{4}g_{l/ab_1}]$	$[d_a-d_b-i_{ab/}-\frac{1}{2}(j_{a/b}-j_{b/a})+\frac{1}{4}l_{/ab}]$
AaBB	..	$[\frac{1}{2}h_a+d_b]\pm[e_1+\frac{1}{2}g_{h_{a_1}}+g_{d_{b_1}}]$	$[\frac{1}{2}h_a+d_b]$
AaBb	...	$[\frac{1}{2}h_a+\frac{1}{2}h_b]\pm[e_1+\frac{1}{2}g_{h_{a_1}}+\frac{1}{2}g_{hb_1}]$	$[\frac{1}{2}h_a+\frac{1}{2}h_b]$
Aabb	..	$[\frac{1}{2}h_a-d_b]\pm[e_1+\frac{1}{2}g_{h_{a_1}}-g_{db_1}]$	$[\frac{1}{2}h_a-d_b]$
aaBB	..	$[-d_a+d_b-i_{ab/}+\frac{1}{2}(j_{a/b}-j_{b/a})+\frac{1}{4}l_{/ab}]$ $\pm[e_1-g_{d_{a_1}}+g_{d_{b_1}}-g_{iab/_1}+\frac{1}{2}(g_{ja/b_1}-g_{jb/a_1})+\frac{1}{4}g_{l/ab_1}]$	$[-d_a+d_b-i_{ab/}+\frac{1}{2}(j_{a/b}-j_{b/a})+\frac{1}{4}l_{/ab}]$
aaBb	...	$[\frac{1}{2}h_b-d_a]\pm[e_1+\frac{1}{2}g_{hb_1}-g_{d_{a_1}}]$	$[\frac{1}{2}h_b-d_a]$
aabb	..	$[-d_a-d_b+i_{ab/}+\frac{1}{2}(j_{a/b}+j_{b/a})+\frac{1}{4}l_{/ab}]$ $\pm[e_1-g_{d_{y_1}}-g_{d_{b_1}}+g_{iab/_1}+\frac{1}{2}(g_{ja/b_1}+g_{jb/a_1})+\frac{1}{4}g_{l/ab_1}]$	$[-d_a-d_b+i_{ab/}+\frac{1}{2}(j_{a/b}+j_{b/a})+\frac{1}{4}l_{/ab}]$
Mean of F$_3$ means		$[\frac{1}{4}h_a+\frac{1}{4}h_b+\frac{1}{16}l_{/ab}]\pm[e_1+\frac{1}{4}g_{h_{a_1}}+\frac{1}{4}g_{hb_1}+\frac{1}{16}g_{l/ab_1}]$	$[\frac{1}{4}h_a+\frac{1}{4}h_b+\frac{1}{16}l_{/ab}]$

Note.—Environment P will be having $+$ve sign whereas Q will be having $-$ve sign.

that so far as the heritable portion of V_{F_2} is concerned it is capable of separation into

$$\tfrac{1}{2} D' + \tfrac{1}{16} H' + \tfrac{1}{4} I + \tfrac{1}{32} J + \tfrac{1}{256} L \quad \text{for} \quad V_{1F_3} \quad \text{and}$$

$$\tfrac{1}{4} D' + \tfrac{1}{8} H' + \tfrac{9}{16} I + \tfrac{7}{64} J + \tfrac{1}{32} L \quad \text{for} \quad V_{2F_3}$$

whatever may be the sizes of the families and howsoever the families are distributed over the environments. The separation of the associated interactive variation into between families and within families would, however, depend on the structure of the population and the design of the experiment in which the variation is measured.

(e) *Variance of Bi-parental generation :* (V_{S_3}):—A bi-parental generation raised from F_2 population would contain the nine genotypes in the same overall proportion as in case of F_2. The total variance V_{S_3} would, therefore, be same as V_{F_2}. It is, however, capable of breaking into two parts: V_{1S_3} and V_{2S_3} representing respectively the variance of means of bi-parental progenies and mean variance of biparental progenies, provided the families are quite large in size to make sampling variations negligible and the families are distributed randomly over the environments. Under these conditions, which can be met by appropriately designing the experiment, the whole of the interactive variation would tend to appear in V_{2S_3}.

With two gene differences, there would be eighty-one bi-parental families. With random distribution, the mean of the families would contain no environmental and associated interaction effects, being averaged out. The variance of means of the families would then be the same as it would have been, if genotype $\times$ environment interactions had been assumed absent. The expression for this variance for several gene differences would, then, turn out to be the same as given by Hayman and Mather (1958), *i.e.*,

$$V_{1S_3} = \tfrac{1}{4} D + \tfrac{1}{16} H + \tfrac{1}{16} I + \tfrac{1}{64} J + \tfrac{1}{256} L \tag{7}$$

Since $V_{S_3} = V_{F_2}$, we get the expression for V_{2S_3} as

$$V_{2S_3} = \tfrac{1}{4} D + \tfrac{3}{16} H + \tfrac{3}{16} I + \tfrac{7}{64} J + \tfrac{15}{256} L + \tfrac{1}{2} G_D + \tfrac{1}{4} G_H$$
$$+ \tfrac{1}{4} G_I + \tfrac{1}{8} G_J + \tfrac{1}{16} G_L + E_{F_2} \tag{8}$$

Alternately, we can also obtain the variance of each of the eighty-one families considered over two environments and combine these variances in the proportion in which these families occur. The eighty-one groups, however, reduce to sixteen groups in view of their similarity. The algebraic exercise is too elaborate to be reproduced here. The resulting expression is given by

$$\tfrac{1}{4}d_a{}^2 + \tfrac{1}{4}d_b{}^2 + \tfrac{3}{16}h_a{}^2 + \tfrac{3}{16}h_b{}^2 + \tfrac{3}{16}i^2{}_{ab/} + \tfrac{7}{64}(j^2{}_{a/b} + j^2{}_{b/a})$$
$$+ \tfrac{15}{256}l^2{}_{/ab} + \tfrac{1}{2}(g^2{}_{d_{a_1}} + g^2{}_{d_{b_1}}) + \tfrac{1}{4}(g^2{}_{h_{a_1}} + g^2{}_{h_{b_1}})$$
$$+ \tfrac{1}{4}g^2{}_{i_{ab/_1}} + \tfrac{1}{8}(g^2{}_{j_{a/b_1}} + \tfrac{1}{8}g^2{}_{j_{b/a_1}}) + \tfrac{1}{16}g^2{}_{l/_{ab_1}}$$
$$+ (e_1 + \tfrac{1}{2}g_{h_{a_1}} + \tfrac{1}{2}g_{h_{b_1}})^2$$

which can be generalised to (8) by considering several gene differences and several environmental comparisons.

(f) *Variance of maternal progenies:* (V_{M_3}):—Maternal progenies derived from F_2 contain the nine genotypes in the same overall proportion as in case of F_2. As such $V_{M_3} = V_{F_2}$. As in the case of bi-parental generation V_{M_3} can, however, be divided into V_{1M_3} and V_{2M_3} respectively representing the variance of means of maternal progenies and mean variance of maternal progenies, if the maternal families are quite large and the families are distributed randomly over the environments. The whole of interactive variation would, then, tend to appear in V_{2M_3}.

With two gene differences, there would be nine maternal families in the proportion $1:2:1:2:4:2:1:2:1$. The means of these families, with a random distribution, would not contain any environmental or associated interactive effects and would have values as shown in Table IV illustrated with two environments.

The contribution of two gene differences to V_{1M_3} is then obtained as

$$\tfrac{1}{8}(d_a{}^2 + d_b{}^2) + \tfrac{1}{64}i^2{}_{ab/}$$

V_{1M_3} is then given by

$$V_{1M_3} = \tfrac{1}{8}D + \tfrac{1}{64}I \qquad (9)$$

Then

$$V_{2M_3} = V_{M_3} - V_{1M_3}$$
$$= V_{F_2} - (\tfrac{1}{8}D + \tfrac{1}{64}I)$$
$$= \tfrac{3}{8}D + \tfrac{1}{4}H + \tfrac{15}{64}I + \tfrac{1}{8}J + \tfrac{1}{16}L$$
$$+ \tfrac{1}{2}G_D + \tfrac{1}{4}G_H + \tfrac{1}{4}G_I + \tfrac{1}{8}G_J + \tfrac{1}{16}G_L + E_{F_2} \qquad (10)$$

Alternatively, the same expression is obtained for V_{2M_3} when the variances of various maternal families are pooled.

4. DISCUSSION

From the derivations made in Section 3, it is apparent that the interactive effects get confounded with the environmental comparisons, in separating the environmental component of continuous variation, from the genetic components and components due to associated interactions. In case of distinguishable environments, for instance, where each environment is a block in a replicated experiment, the mean phenotype of each environment or block is given by $\frac{1}{2}h_a + \frac{1}{2}h_b + \frac{1}{2}g_{h_{a_1}} + \frac{1}{2}g_{h_{b_1}} + e_1$ (considering two gene differences and two environments). The variance of the block means round the grand mean is, then given by:

$$\sum_k (e_k + \tfrac{1}{2}g_{h_{a_k}} + \tfrac{1}{2}g_{d_{b_k}})^2 = V_{(e+\frac{1}{2}g_{h_a}+\frac{1}{2}g_{h_b})}.$$

With several gene differences, this is $V_{\left(e + \frac{1}{2} \sum_a g_{h_a}\right)}$ which is the last term in V_{F_2} given by (3). Thus the effect of the heterozygotes interaction is expressed partly by inflation of the estimate of environmental variation. It is interesting to note, that when epistasis is assumed to be non-existing, this inflation will be still there as observed by Mather and Jones (1958). In F_3 generation, however, the environmental component is given by $V_{\left(e + \frac{1}{4} \sum_a g_{h_a} + \frac{1}{16} \sum_{a<b} g_{l/ab}\right)}$ which shows that the proportion of g_h confounded with e is halved whereas there is further confounding of $g_{l/ab}$ with e, the confounded proportion being $\frac{1}{16}$.

The effect of interactions on the environmental component under varying conditions of dominance and epistasis is summarised in Table V.

It is apparent from the above table that so far as mean variance of true breeding parents (which is composed of only environmental component) is concerned the presence and absence of epistasis affects the confounding of environmental with interaction effects and dominance does not play any part in it. Contrary to this the confounding in the environmental component of V_{F_2}, V_{S_3} or V_{M_3} is, however, affected by the presence and absence of dominance, and here epistasis does not have any effect. In V_{F_1} (which is composed of only environmental component) and V_{F_3}, however, both the dominance and epistasis play their roles in producing confounded effects.

TABLE V

Effect of interactions on the environmental component

Conditions of Dominance and Epistasis	Environmental component in			
	$\frac{1}{2}(V_{P_1} + V_{P_2})$	V_{F_1}	$V_{F_2}, V_{S_3}, V_{M_3}$	V_{F_3}
1. No dominance and no epistasis	$V_e + V_{gd}$	V_e	V_e	V_e
2. Dominance and no epistasis	$V_e + V_{gd}$	$V_{(e+gh)}$	$V_{(e+\frac{1}{2}gh)}$	$V_{(e+\frac{1}{4}gh)}$
3. No dominance and epistasis	$V_{(e+gi+\frac{1}{4}gl)} + V_{(gd-gj)}$	$V_{(e+\frac{1}{4}gl)}$	V_e	$V_{(e+\frac{1}{16}gl)}$
4. Dominance and epistasis	$V_{(e+gi+\frac{1}{4}gl)} + V_{(gd-gj)}$	$V_{(e+gh+\frac{1}{4}gl)}$	$V_{(e+\frac{1}{2}gh)}$	$V_{(e+\frac{1}{4}gh\frac{1}{16}gl)}$

It has been shown in Section 3, how the total phenotypic variation can be expressed in terms of quantities denoted by D, H, I, J, L, G_D, G_H, G_I, G_J, G_L, D', H', $G_{D'}$, $G_{H'}$ and E. In the presence of epistasis and genotype $\times$ environment interactions, therefore, additional components of variation like D', H', I, J, L, G_D, G_H, $G_{D'}$, $G_{H'}$, G_I, G_J, G_L come into picture. The biometrical analysis for the separation of the various components of variation and their estimation would, however, depend on the design of experiment and the multiplicity of the statistics available for this purpose. There is, however, one complication when epistatic interactions are taken into account. The definitions of D, H, G_D and G_H as in F_2 do not remain the same in the subsequently selfed generation, i.e., F_3, F_4, etc., but gets changed to D', H', $G_{D'}$ and $G_{H'}$. As such when D, H, G_D, G_H are to be estimated from selfed generations F_3 onwards, they do not strictly measure the variance components due to additive effects of genes, dominance deviations and associated interaction effects. The additive effects get confounded with additive $\times$ dominance type epistatic interactions and the dominance deviation effects with dominance $\times$ dominance type epistatic interactions. Similar is the case with associated interaction effects. But if F_2 generation and their subsequent bi-parental and maternal generations are to be utilised for estimation purposes, we get D, H, G_D, G_H clear of any confounding.

As has been found in Section 3, the separation of the total bi-parental or maternal variance into between-family and within-family components would however depend on whether the families are randomly distributed over the environments or not and whether the families are large enough to make the sampling variation negligible. Even when these conditions are met, a successful estimation of various components of variation would require at least as many statistics as the number of components, *viz.*, eleven in this case. The method of least square estimation given by Mather (1949) would, however, require more than eleven statistics. It is a matter for future investigation to see what proper design of experiment should be followed in raising bi-parental and maternal generation from a F_2 so as to produce sufficient number of statistics for the estimation of various components of variation.

5. SUMMARY

The components of phenotypic variation in presence of epistasis and genotype $\times$ environment interactions have been described. The effect of interactions on the environmental component of variation has been studied.

With two environments and two gene differences, the differences among the eighteen phenotypes can be described in terms of seventeen parameters. Eight of these parameters are d_a, d_b, h_a, h_b, $i_{ab/}$, $j_{a/b}$, $j_{b/a}$, $l_{/ab}$, measuring the average effects of the genic differences, the allelic interactions and non-allelic digenic type of interactions. One is e_1 measuring the average effect of difference in environments and the rest eight parameters are $g_{d_{a_1}}$, $g_{d_{b_1}}$, $g_{h_{a_1}}$, $g_{h_{b_1}}$, $g_{iab/_1}$, g_{ja/b_1}, g_{jb/a_1} and g_{l/ab_1} measuring the interactions of the eight former parameters with e_1. More than two environments can be accommodated by partitioning the environmental differences into orthogonal e components. Then the interaction terms can be described in terms of corresponding sets of orthogonal g components.

The variance of F_2 derived from a cross between two true breeding strains and of segregating S_3 (Bi-parental) and M_3 (Maternal) families derived from F_2 are shown to be expressible in terms of eleven components of variation, *viz.*, D, H, I, J, L, G_D, G_H, G_I, G_J, G_L and E_{F_2}. In case of variances for F_3, derived from F_2 the terms D, H, G_D, G_H and E_{F_2} change to D', H', $G_{D'}$, $G_{H'}$ and E_{F_3} respectively. The first four quantities change their definitions in the sense that j's and

g_j's get confounded with d's and g_d's, l's and g_l's get confounded with h's and g_h's respectively.

It is shown that e comparisons occur with g_h terms in case of variances of F_2, S_3 and M_3 so that E_{F_2} measuring environmental variance gets inflated by part of g_h interaction. It is interesting to note that this confounding in the environmental component is independent of the presence or absence of epistasis. In E_{F_3}, however, apart from the confounding of g_h with e, we also get confounding of g_l with e, the proportion of g_h confounded being $\frac{1}{4}$ and that of g_l being $\frac{1}{16}$.

6. ACKNOWLEDGMENT

7. REFERENCES

1. Hayman, B. I. and Mather, K.   "The description of genic interactions in continuous variation," *Biometrics*, 1955, **11**, 69–82.

2. Mather, K. and Jones, R. Morley   "Interaction of genotype and environment in continuous variation, I. Description," *Ibid.*, 1958, **14**, 343–59.

3. Mather, K.   ..   *Biometrical Genetics*, Mathuen & Co., London, 1949.