THE DESCRIPTION OF GENE ACTION AND INTERACTION WITH MULTIPLE ALLELES IN CONTINUOUS VARIATION

P. NARAIN

Institute of Agricultural Research Statistics, New Delhi, India

Received January 8, 1965

THE description of gene control in characters showing continuous variation was first considered by FISHER (1918). He showed how to represent and estimate the average main and dominance effects of a large number of genes controlling the character under consideration, even when the genes are unequal in effect and show incomplete dominance. His theory was, later on, elaborated in the form of a "Biometrical" method by MATHER (1949) who showed how main and dominance variation can be estimated from various genetic experiments. HAYMAN (1955), however, gave an illuminating device of representing the gene action and interaction by making use of the mathematical representation of laws of genetics. His method gave a general formula for the variance of the family of a cross which leads to the formulae given by MATHER (1949) in particular cases. In all these studies, however, only two alleles at each locus were assumed. If one is dealing with a F_2 derived from two homozygous strains this assumption is quite justified, as only two alleles are then possible at each locus. But when several F_2 populations are considered or when one is dealing with a randomly breeding population, it is necessary to consider more than two alleles at each locus.

The main difficulty in considering more than two alleles at a locus lies in the fact that while with two alleles A_1, A_2 we can represent the three genotypes $A_1 A_1, A_1 A_2$ and $A_2 A_2$ by a stochastic variable θ taking values $-1, 0$ and 1 respectively so that the heterozygote is located at a point intermediate between the two homozygotes, we cannot do so with three alleles or more at a locus. STANTON (1960) suggested that with multiple alleles the homozygotes may be situated at the vertices of a regular simplex whereas the heterozygotes are at the mid-point of the edges. In particular, with three alleles A_1, A_2, A_3 involved at a locus there are six genotypes $A_1 A_1, A_1 A_2, A_2 A_2, A_2 A_3, A_3 A_3, A_1 A_3$ which can be arranged symmetrically by placing the homozygotes at the vertices of an equilateral triangle and the heterozygotes at the midpoint of the edges in a two-dimensional plane. Making use of this fact and representing the six genotypes by a two-dimensional stochastic vector $\theta = (x, y)$, NARAIN (1963) arrived at a mathematical formulation of a law of segregation analogous to that given by HAYMAN (1955) with two alleles at a locus. The genotypic value or the metric of an individual was then expressed as a polynomial function of the variables x and y . In this paper, this scheme is described and the effect of considering multiple alleles in continuous variation is discussed. While the results hold good

for an arbitrary number of alleles at a locus, only three alleles at a locus are assumed to illustrate the underlying principle. Linkage and nonallelic interactions are, however, not taken into account in this discussion.

Representation of genotype: Consider a panmictic population containing three alleles at a given locus. The six genotypes A_1A_1 , A_1A_2 , A_2A_2 , A_2A_3 , A_3A_3 , and A_1A_3 can be represented by the vector variable (X_1, X_2, X_3) where X_i is 0, 1, or 2, and denotes the number of A_i genes in the genotype. All these vectors lie on the plane $X_1 + X_2 + X_3 = 2$. Taking the origin at $(1, 1, 0)$ and making all the vectors lie on the plane $X_2 = 0$, gives two-dimensional vector variables $(1, 0)$, $(0, 0)$, $(-1, 0)$, $(-1, 1)$, $(-1, 2)$ and $(0, 1)$ which represent respectively the six genotypes A_1A_1 , A_1A_2 , A_2A_2 , A_2A_3 , A_3A_3 and A_1A_3 . An individual can then be represented by $\theta = (x, y)$ which takes anyone of the six pairs of values depending upon its genotype. It is obvious that x takes the value -1 when A_1 gene is absent, 0 when it is present in single dose and $+1$ when it is present in double dose. $\frac{1}{2}(1+x)$ will therefore take the values 0, $\frac{1}{2}$ and 1 respectively in the three cases, and these happen to be the corresponding probabilities for the production of an A_1 gamete by the individual. For if the A_1 gene is absent, the genotype will not produce a gamete carrying an A_1 gene, or it can be said to produce gamete A_1 with zero probability. If, however, the A_1 gene is present in single dose, the genotype (being a diploid) will produce a gamete carrying an A_1 gene with probability $\frac{1}{2}$, whereas when the gene is present in double dose, the genotype is certain to produce an A_1 gamete, or in the other words, the probability of its production is unity. The probability that an individual $\theta = (x, y)$ produces a gamete carrying the A_1 gene can, therefore, be represented by $\frac{1}{2}(1+x)$. Similar considerations show that the probabilities for the production of gametes carrying A_2 and A_3 genes can be represented respectively by $\frac{1}{2}(1-z)$ and $\frac{1}{2}y$ where $z = x+y$. When two individuals with genotypes $\theta_1 = (x_1, y_1)$ and $\theta_2 = (x_2, y_2)$ are crossed, the distribution of θ in the offspring can then be obtained as given in Table 1, where $z_1 = x_1 + y_1$, $z_2 = x_2 + y_2$.

In particular, if A_1A_2 and A_2A_3 are crossed, $x_1 = y_1 = 0$, $x_2 = -1$ and $y_2 = 1$ and we obtain from the Table 1 that the resulting generation will have offsprings with genotypes A_1A_2 , A_2A_2 , A_2A_3 and A_1A_3 each with probability $\frac{1}{4}$. This is what is expected otherwise from simple genetic principles. The result given in Table 1 can thus be regarded as a mathematical representation of the Mendel's law of

TABLE 1

Distribution of θ in the offspring of the cross $\theta_1 \times \theta_2$

Genotype	Value of θ	Probability
A_1A_1	(1,0)	$\frac{1}{4}[(1+x_1)(1+x_2)]$
A_1A_2	(0,0)	$\frac{1}{4}[(1+x_1)(1-z_2) + (1+x_2)(1-z_1)]$
A_2A_2	(-1,0)	$\frac{1}{4}[(1-z_1)(1-z_2)]$
A_2A_3	(-1,1)	$\frac{1}{4}[(1-z_1)y_2 + (1-z_2)y_1]$
A_3A_3	(-1,2)	$\frac{1}{4}[y_1y_2]$
A_1A_3	(0,1)	$\frac{1}{4}[(1+x_1)y_2 + (1+x_2)y_1]$

segregation with three alleles. It can easily be generalised to any number of alleles at a locus. For, in a panmictic population containing n alleles $A_1, A_2, \dots, A_n$ at a given locus, there will be $[n(n+1)]/2$ genotypes of form $A_i A_j$ which can be represented by the vector variable $(X_1, X_2, \dots, X_n)$ where X_i is 0, 1 or 2 and denotes the number of A_i genes in the genotype. All these vectors lie on the flat $X_1 + X_2 + \dots + X_n = 2$. Taking the origin at $(1, 1, 0, \dots, 0)$ and making all the vectors lie on the flat $X_2 = 0$ gives the $(n-1)$ -dimensional vector variables of the type $(u_1, u_2, \dots, u_{n-1})$. Here u_1 takes the values $-1, 0$ or 1 according as the gene A_1 is absent, present in single dose, or present in double dose whereas $(u_1 + u_2)$ takes the values $1, 0$ and -1 according as gene A_2 is absent, present in single dose or present in double dose. Also $(u_2 - u_3), (u_3 - u_4), \dots, (u_{n-2} - u_{n-1})$ take the values $0, 1$, or 2 in the three respective phases for the genes $A_3, A_4, \dots, A_{n-1}$ respectively, whereas u_{n-1} takes the values $0, 1$, or 2 in the three phases for the gene A_n . By the considerations given above, it can be seen that the individual $\theta = (u_1, u_2, \dots, u_{n-1})$ produces gametes $A_1, A_2, A_3, \dots, A_{n-1}$ and A_n with probabilities $\frac{1}{2}(1+u_1)$, $\frac{1}{2}(1-u_1-u_2)$, $\frac{1}{2}(u_2-u_3)$, $\dots$, $\frac{1}{2}(u_{n-2}-u_{n-1})$ and $\frac{1}{2}u_{n-1}$ respectively. These probabilities give the distribution of θ in the offspring of the cross $\theta_1 \times \theta_2$ which can be regarded as a mathematical representation of the law of segregation with an arbitrary number of alleles at a given locus.

When the two individuals with genotypes θ_1 and θ_2 happen to be members of a random mating population with gene frequencies p, q and r for the three alleles A_1, A_2 and A_3 respectively, it can be seen that the $E(\theta_1) = E(\theta_2)$ i.e. $E(x_1) = E(x_2) = p - q - r$ and $E(\gamma_1) = E(\gamma_2) = 2r$ so that in the next generation the six genotypes given in the Table 1 occur with frequencies respectively as

$$\begin{aligned} \frac{1}{4}E[(1+x_1)(1+x_2)] &= p^2 \\ \frac{1}{4}E[(1+x_1)(1-z_2) + (1+x_2)(1-z_1)] &= 2pq \\ \frac{1}{4}E[(1-z_1)(1-z_2)] &= q^2 \\ \frac{1}{4}E[(1-z_1)\gamma_2 + (1-z_2)\gamma_1] &= 2qr \\ \frac{1}{4}E[\gamma_1\gamma_2] &= r^2 \\ \text{and } \frac{1}{4}E[(1+x_1)\gamma_2 + (1+x_2)\gamma_1] &= 2pr. \end{aligned}$$

The gene frequencies in this generation thus turns out to be the same as in the previous generation as is otherwise expected in view of the HARDY-WEINBERG law of equilibrium in random mating populations. This further corroborates the soundness of the proposed representation.

From the distribution given above it is possible to work out the expectations of the stochastic variables x and γ and their various powers, as well as of various products between them. These expectations can be used to obtain the variances of $x, \gamma, x^2, \gamma^2, x\gamma$, and covariances between x and γ, x and x^2, x and γ^2, x and $x\gamma, \gamma$ and x^2, γ and γ^2, γ and $x\gamma, x^2$ and γ^2, x^2 and $x\gamma$ and γ^2 and $x\gamma$. The results obtained for various expectations, variances and covariances are given by

Expectations

$$\begin{aligned} E(x) &= E(x^3) = \frac{1}{2}(x_1 + x_2) \\ E(x^2) &= E(x^4) = \frac{1}{2}(1 + x_1 x_2) \end{aligned}$$

$$\begin{aligned}
E(\gamma) &= \frac{1}{2}(\gamma_1 + \gamma_2) \\
E(\gamma^2) &= \frac{1}{2}(\gamma_1 + \gamma_2 + \gamma_1\gamma_2) \\
E(\gamma^3) &= \frac{1}{2}(\gamma_1 + \gamma_2 + 3\gamma_1\gamma_2) \\
E(\gamma^4) &= \frac{1}{2}(\gamma_1 + \gamma_2 + 7\gamma_1\gamma_2) \\
-E(xy) &= -E(x^3y) = E(x^2y) = \frac{1}{4}\{\gamma_1(1-x_2) + \gamma_2(1-x_1)\} \\
-E(x\gamma^2) &= E(x^2\gamma^2) = \frac{1}{4}\{\gamma_1(1-x_2) + \gamma_2(1-x_1) + 2\gamma_1\gamma_2\}
\end{aligned} \tag{1}$$

Variances

$$\begin{aligned}
V(x) &= \frac{1}{4}(2-x_1^2-x_2^2) \\
V(\gamma) &= \frac{1}{4}\{2(\gamma_1+\gamma_2) - \gamma_1^2 - \gamma_2^2\} \\
V(x^2) &= \frac{1}{4}(1-x_1^2x_2^2) \\
V(\gamma^2) &= \frac{1}{4}\{1 - (\gamma_1+\gamma_2+\gamma_1\gamma_2-1)^2 + 12\gamma_1\gamma_2\} \\
V(xy) &= \frac{1}{16}\{\gamma_1(1-x_2)(4-\gamma_1+\gamma_1x_2) + \gamma_2(1-x_1)(4-\gamma_2+\gamma_2x_1) \\
&\quad + 2\gamma_1\gamma_2(3-x_1x_2+x_1+x_2)\}
\end{aligned} \tag{2}$$

Covariances

$$\begin{aligned}
Cov(x, \gamma) &= -\frac{1}{4}\{\gamma_1(1+x_1) + \gamma_2(1+x_2)\} \\
Cov(x, x^2) &= \frac{1}{4}(x_1+x_2)(1-x_1x_2) \\
Cov(x, \gamma^2) &= -\frac{1}{4}\{\gamma_1(1+x_1) + \gamma_2(1+x_2) + \gamma_1\gamma_2(x_1+x_2+2)\} \\
Cov(x, xy) &= \frac{1}{8}\{\gamma_1(1-x_2) + \gamma_2(1-x_1)\}(2+x_1+x_2) \\
Cov(x^2, \gamma) &= -\frac{1}{4}\{x_1(\gamma_2-x_2) + x_2(\gamma_1-x_1) + x_1x_2(\gamma_1+\gamma_2+2)\} \\
Cov(\gamma, \gamma^2) &= \frac{1}{4}\{1 - (\gamma_1+\gamma_2+\gamma_1\gamma_2-1)(\gamma_1+\gamma_2-1) + 5\gamma_1\gamma_2\} \\
Cov(\gamma, xy) &= \frac{1}{8}\{(\gamma_1-\gamma_2)^2 - 2(\gamma_1+\gamma_2) - (\gamma_1+\gamma_2-2)(x_2\gamma_1+x_1\gamma_2)\} \\
Cov(x^2, \gamma^2) &= \frac{1}{4}\{\gamma_1\gamma_2(1-x_1x_2) - \gamma_1x_2(1+x_1) - \gamma_2x_1(1+x_2)\} \\
Cov(x^2, xy) &= \frac{1}{8}\{\gamma_1(1-x_2) + \gamma_2(1-x_1)\}(x_1x_2-1) \\
Cov(\gamma^2, xy) &= \frac{1}{8}\{(\gamma_1+\gamma_2+\gamma_1\gamma_2-2)(\gamma_1+\gamma_2-\gamma_1x_2-\gamma_2x_1) - 12\gamma_1\gamma_2\}
\end{aligned} \tag{3}$$

Now consider k loci, $A, B, C, \dots K$ each with three alleles. The whole genotype can be represented by a set of k two-dimensional stochastic variables $G = (\theta_a, \theta_b, \dots \theta_k)$ where $\theta_a = (x_a, \gamma_a)$, $\theta_b = (x_b, \gamma_b)$, $\dots$, $\theta_k = (x_k, \gamma_k)$. Since we are not taking into account linkage between loci, x_a, γ_a can be assumed to be uncorrelated with x_b, γ_b etc. so that $Cov(x_a, x_b) = Cov(x_a, \gamma_b) = Cov(x_b, \gamma_a) = Cov(\gamma_b, \gamma_a) = 0$ etc.

Relationship between the metric and the genotype: Consider first a single locus involving three alleles A_1, A_2, A_3 determining the genotype $\theta_a = (x_a, \gamma_a)$. Since the genotype is two-dimensional, the metric $M(\theta_a)$, in accordance with HAYMAN's notations is described as a suitable polynomial function in x_a and γ_a given by

$$M(x_a, \gamma_a) = \lambda_{1a}x_a + \lambda_{2a}\gamma_a + \lambda_{3a}x_a^2 + \lambda_{4a}\gamma_a^2 + \lambda_{5a}x_a\gamma_a \tag{4}$$

In this case five parameters are needed to describe the metric completely. There are two main effects which may be represented by d_{1a} and d_{2a} . While d_{1a} measures the effect of A_1 relative to A_2 , d_{2a} measures that of A_2 relative to A_3 . The main effect of A_3 relative to A_1 , would obviously be $(d_{1a}-d_{2a})$. As regards dominance deviations, we may represent them by h_{1a}, h_{2a} and h_{3a} respectively measuring the

dominance deviations for the pairs A_1 - A_2 , A_2 - A_3 and A_3 - A_1 . The relationship between the metrics of the various genotypes and the parameters should, therefore, satisfy the following relations.

$$\begin{aligned}
 \frac{1}{2}\{M(1, 0) - M(-1, 0)\} &= d_{1a} \\
 \frac{1}{2}\{M(-1, 2) - M(-1, 0)\} &= d_{2a} \\
 \frac{1}{2}\{M(1, 0) - M(-1, 2)\} &= (d_{1a} - d_{2a}) \\
 M(0, 0) - \frac{1}{2}\{M(1, 0) + M(-1, 0)\} &= h_{1a} \\
 M(-1, 1) - \frac{1}{2}\{M(-1, 0) + M(-1, 2)\} &= h_{2a} \\
 M(0, 1) - \frac{1}{2}\{M(-1, 2) + M(1, 0)\} &= h_{3a}
 \end{aligned} \tag{5}$$

Consistent with (5) the metric (4) can be shown to take the six values corresponding to the six genotypes as given in Table 2.

The coefficients in the polynomial (4) when expressed in terms of the five parameters are given by

$$\begin{aligned}
 \lambda_{1a} &= d_{1a} \\
 \lambda_{2a} &= d_{2a} - h_{1a} + h_{2a} + h_{3a} \\
 \lambda_{3a} &= -h_{1a} \\
 \lambda_{4a} &= -h_{2a} \\
 \lambda_{5a} &= -h_{1a} - h_{2a} + h_{3a}
 \end{aligned} \tag{6}$$

When k loci are considered simultaneously the metric is given by

$$\sum_a [\lambda_{1a}x_a + \lambda_{2a}y_a + \lambda_{3a}x_a^2 + \lambda_{4a}y_a^2 + \lambda_{5a}x_a y_a] \tag{7}$$

where nonallelic interactions are assumed to be absent.

The relationship between the metric and the genotype given above for the case of three alleles at each locus can be generalised to any number of alleles at each of the loci. Consider first a single locus involving n alleles $A_1, A_2, \dots, A_n$ determining the genotype $\theta_a = (u_{1a}, u_{2a}, \dots, u_{(n-1)a})$. The metric $M(\theta_a)$ is now described as a suitable polynomial function in $(n-1)$ variables $u_{1a}, u_{2a}, \dots, u_{(n-1)a}$ given by

$$M(u_{1a}, u_{2a}, \dots, u_{(n-1)a}) = \sum_{j=1}^{(n-1)} \lambda_{1ja} u_{ja} + \sum_{j=1}^{(n-1)} \lambda_{2ja} u_{ja}^2 + \sum_{i=1}^{(n-2)} \sum_{j>i}^{(n-1)} \lambda_{(i+2), (j-i)a} u_{ia} u_{ja} \tag{4a}$$

In this case $(n-1)(n+2)/2$ parameters are needed to describe the metric com-

TABLE 2
Value of the metric corresponding to different genotypes

Genotype	(x_a, y_a)	Metric $M(x_a, y_a)$ or the genotypic value
A_1A_1	(1,0)	$d_{1a} - h_{1a}$
A_1A_2	(0,0)	0
A_2A_2	(-1,0)	$-d_{1a} - h_{1a}$
A_2A_3	(-1,1)	$-d_{1a} - h_{1a} + d_{2a} + h_{2a}$
A_3A_3	(-1,2)	$-d_{1a} - h_{1a} + 2d_{2a}$
A_1A_3	(0,1)	$-h_{1a} + d_{2a} + h_{3a}$

pletely. There are now $(n-1)$ main effects and $n(n-1)/2$ dominance deviations which make up the total of $(n-1)(n+2)/2$. The coefficients λ 's in (4a) can be expressed as linear functions of these main effects and dominance deviations. When k loci are considered simultaneously the metric is given by

$$\sum_a \sum_j \lambda_{1ja} u_{ja} + \sum_a \sum_j \lambda_{2ja} u_{ja}^2 + \sum_a \sum_i \sum_{j>i} \lambda_{(i+2), (j-i), a} u_{ja} u_{ia} \quad (7a)$$

where nonallelic interactions are assumed to be absent.

Mean and variance of the offspring of the cross $\theta_1 \times \theta_2$: The mean of the offspring generation resulting from a cross $\theta_1 \times \theta_2$ can be obtained by finding the expectation of the metric $M(x, y)$ with the help of (1). Thus

$$\begin{aligned} E\{M(x, y)\} &= \frac{1}{2} \lambda_1 (x_1 + x_2) + \frac{1}{2} \lambda_3 (1 + x_1 x_2) \\ &+ \frac{1}{4} \{2(\lambda_2 + \lambda_4) - \lambda_5\} (y_1 + y_2) \\ &+ \frac{1}{2} \lambda_4 y_1 y_2 + \frac{1}{4} \lambda_5 (y_1 x_2 + y_2 x_1) \end{aligned} \quad (8)$$

The suffix "a" is dropped for simplicity, it being understood that we are considering only a single locus. It is interesting to note the difference between the mean given by (8) and the parental mean given by

$$\begin{aligned} \frac{1}{2} \{M(x_1, y_1) + M(x_2, y_2)\} &= \frac{1}{2} \lambda_1 (x_1 + x_2) + \frac{1}{2} \lambda_2 (y_1 + y_2) \\ &+ \frac{1}{2} \lambda_3 (x_1^2 + x_2^2) + \frac{1}{2} \lambda_4 (y_1^2 + y_2^2) \\ &+ \frac{1}{2} \lambda_5 (x_1 y_1 + x_2 y_2) \end{aligned} \quad (9)$$

This difference is given by

$$\begin{aligned} &+ \frac{1}{4} h_1 [2\{(1 + x_1 x_2) - x_1^2(1 - y_1) - x_2^2(1 - y_2)\} - \{y_1(1 - x_2) + y_2(1 - x_1)\}] \\ &+ \frac{1}{4} h_2 [2 y_1 y_2 + y_1(1 + 2x_1^2) + y_2(1 + 2x_2^2) + y_1(x_2 - 2y_1) + y_2(x_1 - 2y_2)] \\ &+ \frac{1}{4} h_3 [y_1(1 - 2x_1^2) + y_2(1 - 2x_2^2) - (y_1 x_2 + y_2 x_1)] \end{aligned} \quad (10)$$

It can be seen that the difference given by (10) clearly depends on the three dominance deviations and the component variables of the genotype of the parents. That this expression would reduce to the ones expected in particular cases is demonstrated in Table 3.

The variance of the offspring generation from the cross $\theta_1 \times \theta_2$ is obtained by finding the variance of the metric $M(x, y)$ with the help of (2), (3) and (6). The resulting expression is given by

TABLE 3
Differences in the mid-parent and offspring means in various crosses

Type of cross	Value of				Difference
	x_1	y_1	x_2	y_2	
Selfing or intercrossing					
$A_1A_2 \times A_1A_2$	0	0	0	0	$\frac{1}{2} h_1$
$A_2A_3 \times A_2A_3$	-1	1	-1	1	$\frac{1}{2} h_2$
$A_1A_3 \times A_1A_3$	0	1	0	1	$\frac{1}{2} h_3$
Backcrossing					
$A_1A_2 \times A_1A_1$	0	0	1	0	0
$A_1A_2 \times A_2A_2$	0	0	-1	0	0
$A_2A_3 \times A_2A_2$	-1	1	-1	0	0
$A_2A_3 \times A_3A_3$	-1	1	-1	2	0
$A_1A_3 \times A_3A_3$	0	1	-1	2	0
$A_1A_3 \times A_1A_1$	0	1	1	0	0

$$\begin{aligned}
V\{M(x, y)\} = & \frac{1}{4} d_1^2 \{2(1+x_1x_2) - (x_1+x_2)^2\} \\
& + \frac{1}{4} d_2^2 \{(1+2y_1y_2) - (y_1+y_2-1)^2\} \\
& + \frac{1}{16} h_1^2 \{4 - [(\gamma_1+\gamma_2) + (z_1x_2+z_2x_1)]^2\} \\
& + \frac{1}{16} h_2^2 \{(\gamma_1+\gamma_2) - (z_1y_2+z_2y_1)\} \{4 - (\gamma_1+\gamma_2) + (z_1y_2+z_2y_1)\} \\
& + \frac{1}{16} h_3^2 \{\gamma_1(1+x_2) + \gamma_2(1+x_1)\} \{4 - \gamma_1(1+x_2) - \gamma_2(1+x_1)\} \\
& - \frac{1}{2} d_1d_2 \{\gamma_1(1+x_1) + \gamma_2(1+x_2)\} \\
& + \frac{1}{4} d_1h_1 \{x_1+x_2\} \{(z_1x_2+z_2x_1) + (\gamma_1+\gamma_2-2)\} \\
& + \frac{1}{4} d_1h_2 \{x_1+x_2+2\} \{(z_1y_2+z_2y_1) - (\gamma_1+\gamma_2)\} \\
& - \frac{1}{4} d_1h_3 \{x_1+x_2\} \{\gamma_1(1+x_2) + \gamma_2(1+x_1)\} \\
& + \frac{1}{4} d_2h_1 \{\gamma_1+\gamma_2\} \{(z_1x_2+z_2x_1) + (\gamma_1+\gamma_2-2)\} \\
& + \frac{1}{4} d_2h_2 \{\gamma_1+\gamma_2-2\} \{(z_1y_2+z_2y_1) - (\gamma_1+\gamma_2)\} \\
& - \frac{1}{4} d_2h_3 \{\gamma_1+\gamma_2-2\} \{\gamma_1(1+x_2) + \gamma_2(1+x_1)\} \\
& + \frac{1}{8} h_1h_2 \{(z_1y_2+z_2y_1) - (\gamma_1+\gamma_2)\} \{(z_1x_2+z_2x_1) + (\gamma_1+\gamma_2-2)\} \\
& + \frac{1}{8} h_1h_3 \{\gamma_1(1+x_2) + \gamma_2(1+x_1)\} \{(z_1x_2+z_2x_1) + (\gamma_1+\gamma_2-2)\} \\
& + \frac{1}{8} h_2h_3 \{\gamma_1(1+x_2) + \gamma_2(1+x_1)\} \{(z_1y_2+z_2y_1) - (\gamma_1+\gamma_2)\} \quad (11)
\end{aligned}$$

In particular, the variances of the families obtained by selfing the heterozygotes can be obtained by substituting the particular set of values of x_1, y_1, x_2, y_2 in (11). The expressions obtained are in perfect agreement with those which can be obtained from first principles. This is demonstrated in Table 4.

When k loci are considered simultaneously and linkage and epistasis are assumed to be absent, the mean and variance of the offspring generation from the cross $G_1 \times G_2$ are given by $\sum_a E\{M(x_{1a}, y_{1a})\}$ and $\sum_a V\{M(x_{1a}, y_{1a})\}$ respectively.

Randomly breeding populations: Consider a random mating population with gene frequencies p_a, q_a, r_a for the alleles A_1, A_2, A_3 respectively at a given locus. With the help of metric values given in Table 2 and the genotypic frequencies ($p_a^2, 2p_aq_a, q_a^2, 2q_ar_a, r_a^2, 2p_ar_a$) the contribution of this locus to the average of the random mating population is given by

$$d_a = (p_a - q_a - r_a)d_{1a} + 2r_ad_{2a} - (1 - 2p_aq_a)h_{1a} + 2q_ar ah_{2a} + 2p_ar ah_{3a} \quad (12)$$

The metric values can then be represented, alternatively, in terms of d_a and other quantities leading to symmetric expressions as given in Table 5.

The average of the random mating population is $\sum_a d_a$ where d_a is given by (12).

The contribution of this locus to the genetic variance of the random mating population is given by

$$\begin{aligned}
& p_a^2(d_{1a} - h_{1a})^2 + q_a^2(d_{1a} + h_{1a})^2 + 2q_ar_a(d_{1a} - d_{2a} + h_{1a} - h_{2a})^2 \\
& + r_a^2(d_{1a} - 2d_{2a} + h_{1a})^2 + 2p_ar_a(d_{2a} - h_{1a} + h_{3a})^2 - d_a^2 \quad (13)
\end{aligned}$$

TABLE 4

Variances in various types of selfed families

Selfing of	Value of				$V\{M(x, y)\}$
	x_1	y_1	x_2	y_2	
A_1A_2	0	0	0	0	$\frac{1}{2} d_1^2 + \frac{1}{4} h_1^2$
A_2A_3	-1	1	-1	1	$\frac{1}{2} d_2^2 + \frac{1}{4} h_2^2$
A_1A_3	0	1	0	1	$\frac{1}{2} (d_1 - d_2)^2 + \frac{1}{4} h_3^2$

TABLE 5

Metric values of various genotypes in a random mating population

Genotype	(x, y)	M(x, y)
A_1A_1	(1,0)	$d_a + 2(1-p_a)d_{1a} - 2r_ad_{2a} - 2p_aq_ah_{1a} - 2q_ar_ah_{2a} - 2p_ar_ah_{3a}$
A_1A_2	(0,0)	$d_a + (1-2p_a)d_{1a} - 2r_ad_{2a} + (1-2p_aq_a)h_{1a} - 2q_ar_ah_{2a} - 2p_ar_ah_{3a}$
A_2A_2	(-1,0)	$d_a - 2p_ad_{1a} - 2r_ad_{2a} - 2p_aq_ah_{1a} - 2q_ar_ah_{2a} - 2p_ar_ah_{3a}$
A_2A_3	(-1,1)	$d_a - 2p_ad_{1a} + (1-2r_a)d_{2a} - 2p_aq_ah_{1a} + (1-2q_ar_a)h_{2a} - 2p_ar_ah_{3a}$
A_3A_3	(-1,2)	$d_a - 2p_ad_{1a} + 2(1-r)d_{2a} - 2p_aq_ah_{1a} - 2q_ar_ah_{2a} - 2p_ar_ah_{3a}$
A_1A_3	(0,1)	$d_a + (1-2p_a)d_{1a} + (1-2r_a)d_{2a} - 2p_aq_ah_{1a} - 2q_ar_ah_{2a} + (1-2p_ar_a)h_{3a}$

When the genes are independent in action and uncorrelated in distribution, the total genetic variation can be obtained by summing contributions given by (13) over all loci. This can finally be given by

$${}_H V_R = \frac{1}{2} D + \frac{1}{4} H \quad (14)$$

where

$$D = \sum_a [4p_aq_a\{d_{1a} + (2q_a - 1)h_{1a} + r_a(h_{1a} - h_{2a} + h_{3a})\}^2 + 4q_ar_a\{d_{2a} + (2q_a - 1)h_{2a} + p_a(h_{2a} - h_{1a} + h_{3a})\}^2 + 4r_ap_a\{(d_{1a} - d_{2a}) + (2r_a - 1)h_{3a} + q_a(h_{1a} - h_{2a} + h_{3a})\}^2]$$

and

$$H = \sum_a [16\{p_aq_ah_{1a} + q_ar_ah_{2a} + r_ap_ah_{3a}\}^2 + 8p_aq_ar_a\{(h_{1a} - h_{2a})^2 + h_{3a}(h_{3a} - 2h_{1a} - 2h_{2a})\}]$$

It is interesting to compare the definitions of D and H given under (14) with those given by MATHER (1949, p. 75) for randomly breeding populations. The definitions are the same as given by MATHER (1949) when either $p_a=0$, or $q_a=0$ or $r_a=0$ i.e. when the third allele at each locus is assumed to be nonexistent. When the two main effects d_{1a} and d_{2a} are identical, and equal to m_a say, the third main effect is zero. Also let $h_{1a}=h_{2a}=h_a$ and h_{3a} be zero. With these conditions, the definitions would become

$$D = \sum_a 4q_a(p_a + r_a)\{m_a + (2q_a - 1)h_a\}^2 \quad (15)$$

$$\text{and } H = \sum_a 16q_a^2(p_a + r_a)^2 h_a^2$$

In such a case, therefore, the three alleles at each locus can be regarded as a two-allele system with allelic frequencies as q_a and $(p_a + r_a)$.

The offsprings resulting from a particular cross $\theta_1 \times \theta_2$ in the random mating population form a family of full-sibs. The mean and variance of such a family are given by (8) and (11) respectively. If we take expectation of (8) and substitute $E(x_{1a}) = E(x_{2a}) = (p_a - q_a - r_a)$ and $E(y_{1a}) = E(y_{2a}) = 2r_a$ we get the contribution of the locus to the mean of the offspring generation which is found to be the same as the contribution to the mean of the parental generation given by (12). If we take the expectation of (11) and substitute $E(x_{1a}) = E(x_{2a}) = (p_a - q_a - r_a)$, $E(y_{1a}) = E(y_{2a}) = 2r_a$, $E(x_{1a}^2) = E(x_{2a}^2) = p_a^2 + (q_a + r_a)^2$, $E(y_{1a}^2) = E(y_{2a}^2) = 2r_a(1 + r_a)$ and $E(x_{1a}y_{1a}) = E(x_{2a}y_{2a}) = -2r_a(1 - p_a)$ we would get the contri-

bution of the locus to the mean variance of full-sib families. The variance of (8) after necessary substitutions would give the contribution to covariance between full-sibs. If expectation of (8) is taken over either $\theta_1 = (x_1, y_1)$ or $\theta_2 = (x_2, y_2)$ the contribution to the mean of a particular family of half-sibs is obtained. For instance, the contribution of the locus to the mean of half-sib family with parent as θ_2 is given by

$$\begin{aligned} & \{ \frac{1}{2} \lambda_{1a}(2p_a - 1) + (\lambda_{2a} + \lambda_{4a} - \frac{1}{2} \lambda_{5a})r_a + \frac{1}{2} \lambda_{3a} \} + \frac{1}{2} \{ \lambda_{1a} + \lambda_{3a}(2p_a - 1) + \lambda_{5a}r_a \} x_{2a} \\ & + \frac{1}{2} \{ \lambda_{2a} + \lambda_{4a}(1 + 2r_a) - \lambda_{5a}(1 - p_a) \} y_{2a} \end{aligned} \quad (16)$$

Variance of (16) would give the contribution to the covariance between half-sibs.

Effect of multiple allelism in continuous variation: It is well known that the continuous variation observed in the phenotype arises from the discontinuous variation of the genotype when the genes have effects similar to one another, supplementing each other and small in relation to environmental variation. If there are two alleles at each locus with frequencies p and q and there are k loci, the frequency distribution of the phenotypes ignoring environmental effects and assuming no dominance would be described by the various terms in the binomial $(q+p)^{2k}$ which tends to a continuous distribution when k is quite large. When, however, there are more than two alleles at each locus, say for instance three alleles with frequencies p , q , and r , the corresponding distribution is given by the multinomial $(q+p+r)^{2k}$. Now when k tends to become very large, a continuous distribution in two dimensions is approximated. It is, therefore, expected that with multiple allelism, contribution of continuous variation from each dimension is to be added over for the realization of the total variation. When such a contribution is of the same order, whatever may be the dimension under consideration, the existing partition of the total variation is not going to be affected. The case of three alleles, or in fact more than three, can be looked upon as a case of two alleles by dividing the number of alleles into two groups, one consisting of one allele and the other group containing the rest of them and treating the two groups as a two-allele system. Such a situation gives rise to the definitions of D and H as under (15). It, therefore, appears that the nonequality of d_{1a} and d_{2a} , h_{1a} and h_{2a} and the nonvanishing of h_{3a} may reveal some component of variation contributing towards the total variation. Such a component, of course, expected to be existing only in a population where several alleles are operating at each locus, may be described as due to multiple allelism. It may also be looked upon in a manner similar to that of dominance. Dominance is regarded as due to the differences in the effect of gene substitution when the other allele is changed. In a similar way, in a population where several alleles are operating at each locus the effects of gene substitution with respect to any fixed particular allele can be considered. The differences in the effects of gene substitution can be regarded as effects due to multiple allelism. When the effects of gene substitution with respect to the fixed allele are the same, we can say that there is no effect due to the multiple alleles, though there may be several of them operating.

In the first instance, we may regard dominance effects as absent so that h_{1a} , h_{2a} and h_{3a} are zero and the D in (14) reduces to

$$D = \sum_a [4p_a q_a d_{1a}^2 + 4q_a r_a d_{2a}^2 + 4r_a p_a (d_{1a} - d_{2a})^2] \quad (17)$$

Let $d_{1a}=m_a+\alpha_{1a}$ and $d_{2a}=m_a+\alpha_{2a}$, when m_a is the mean of d_{1a} and d_{2a} given by

$$m_a = \frac{p_a d_{1a} + r_a d_{2a}}{p_a + r_a} \quad (18)$$

Here allele A_2 is taken as fixed and A_1 and A_3 with frequencies p_a and r_a are considered with respect to this allele. Since in view of (18), $p_a \alpha_{1a} + r_a \alpha_{2a} = 0$, D can be regarded as consisting of two components

$$D_A = \sum_a 4q_a(1-q_a)m_a^2$$

$$\text{and } D_B = \sum_a 4(p_a \alpha_{1a}^2 + r_a \alpha_{2a}^2) \quad (19)$$

When there are more than three alleles, say four alleles it is expected that in (19) D_A would still have the same definition but D_B would be given by

$$D_B = \sum_a 4\{p_a \alpha_{1a}^2 + r_a \alpha_{2a}^2 + s_a \alpha_{3a}^2\} \quad (20)$$

where $q_a + p_a + r_a + s_a = 1$, and α_{3a} corresponds to the third independent main effect. Thus it appears that D_B would reflect the contribution of multiple allelic effect to the additive genetic variance.

When there is dominance also present, the nonequality of the dominance effects may be taken into account in the same manner by letting $h_{1a} = h_a + \beta_{1a}$ and $h_{2a} = h_a + \beta_{2a}$ where h_a is given by

$$h_a = \frac{p_a h_{1a} + r_a h_{2a}}{p_a + r_a} \quad (21)$$

Again, since in view of (21), $p_a \beta_{1a} + r_a \beta_{2a} = 0$, it can be shown that D and H of (14) take the form

$$D = \sum_a [4q_a(1-q_a)\{m_a + (2q_a - 1)h_a + \frac{2r_a p_a}{1-q_a} h_{3a}\}^2]$$

$$+ \sum_a [4p_a\{\alpha_{1a} + q_a \beta_{1a} + r_a h_{3a}\}^2 + 4r_a\{\alpha_{2a} + q_a \beta_{2a} + p_a h_{3a}\}^2]$$

$$- \sum_a \left[\frac{4r_a^2 p_a^2}{1-q_a} h_{3a}^2 \right] \quad (22)$$

$$H = \sum_a [16q_a^2(1-q_a)^2\{h_a - \frac{r_a p_a}{(1-q_a)^2} h_{3a}\}^2]$$

$$+ \sum_a [8q_a(1-q_a)\{p_a(\beta_{1a} - \frac{r_a}{1-q_a} h_{3a})^2 + r_a(\beta_{2a} - \frac{p_a}{1-q_a} h_{3a})^2\}]$$

$$+ \sum_a \left[\frac{16r_a^2 p_a^2(1-2q_a)}{(1-q_a)^2} h_{3a}^2 \right]$$

Thus the genetic variance of a randomly breeding population can be expressed as

$${}^H V_R = \frac{1}{2} D_A + \frac{1}{4} H_A + \frac{1}{2} (D_B + H_B + H_C) \quad (23)$$

where

$$D_A = \sum_a [4q_a(1-q_a)\{m_a + (2q_a - 1)h_a + \frac{2r_a p_a}{1-q_a} h_{3a}\}^2]$$

$$D_B = \sum_a [4p_a\{\alpha_{1a} + q_a \beta_{1a} + r_a h_{3a}\}^2 + 4r_a\{\alpha_{2a} + q_a \beta_{2a} + p_a h_{3a}\}^2]$$

$$H_A = \sum_a [16q_a^2(1-q_a)^2\{h_a - \frac{r_a p_a}{(1-q_a)^2} h_{3a}\}^2]$$

$$H_B = \sum_a \left[4q_a(1-q_a) \left\{ p_a(\beta_{1a} - \frac{r_a}{1-q_a} h_{3a})^2 + r_a(\beta_{2a} - \frac{p_a}{1-q_a} h_{3a})^2 \right\} \right]$$

and

$$H_C = \sum_a \left[\frac{4r_a^2 p_a^2 (1-3q_a)}{(1-q_a)^2} h_{3a}^2 \right]$$

In (23), D_A and H_A correspond to the components due to additive genetic and dominance effects. D_A , however, includes some effect of the mean dominance h_a taken over the A_1 and A_3 and also of the dominance h_{3a} operating between A_1 and A_3 . Similarly H_A includes some effect of the h_{3a} . The remaining three components D_B , H_B and H_C can be regarded as components corresponding to the multiple allelic effects. While D_B includes some effect of the nonequality of the two dominance effects and also of the third dominance effect h_{3a} , H_B includes some effect of the h_{3a} only. The multiple allelic effects would not contribute towards the total variation whenever $d_{1a} = d_{2a}$, $h_{1a} = h_{2a}$ and $h_{3a} = 0$ which correspond to the simultaneously vanishing of D_B , H_B and H_C .

The above investigation shows that in addition to the hitherto described two components of continuous variation in the literature *viz.* additive and dominance, there may be a necessity of describing a third multiple allelic component in genetic populations where several alleles are required to be considered at each of the loci. How this new component having three subcomponents D_B , H_B and H_C would behave in the constitution of various second degree statistics which are utilized for the estimation purposes, remains to be seen in future investigations.

SUMMARY

The effect of three alleles at each locus in continuous variation has been considered. HAYMAN's (1955) device of representing the gene action and interaction by making use of the mathematical representation of laws of genetics can be extended to multiple allelic systems. In a randomly breeding population where several alleles are supposed to be operating at each loci, a new component of genetic variation due to multiple allelic effects is described. This component consists of three subcomponents, and the multiple allelism can be regarded as contributing nothing to the total genetic variation when the three subcomponents vanish simultaneously.

LITERATURE CITED

- FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.* **52**: 399-433.
- HAYMAN, B. I., 1955 Description and analysis of gene action and interaction. Cold Spring Harbor Symp. Quant. Biol. **20**: 79-86.
- MATHER, K., 1949 *Biometrical Genetics*. Methuen, London.
- NARAIN, P., 1963 On mathematical representation of gene action and interaction. *J. Indian Soc. Agri. Stat.* **15**: 270.
- STANTON, R. G., 1960 Genetic correlations with multiple alleles. *Biometrics* **16**: 235-244.