

GENETIC DISTRIBUTIONS AND ESTIMATION OF NUMBER OF GENES IN DIALLEL CROSS ANALYSIS

by K. R. KHANNA, *National Botanic Gardens, Lucknow*

(Received 22 September 1975; after revision 22 February 1976)

The distribution of additive and dominance components has been studied in diallel cross analysis for array totals, V_r and W_r under complete binomial and under certain selected genotype distributions. The deviations as a result of omitting certain arrays, from the usual pattern of W_r , V_r graphs under independent gene distribution, and the regular shape of these graphs under parental selection have been brought out. A method for the estimation of the number of genes from the length of W_r , V_r line under various distributions is presented and this has been illustrated from data on fruit size in tomato.

INTRODUCTION

The graphic analysis of the diallel cross, developed by Jinks (1954) and Hayman (1954), by virtue of consistent unit correlation between W_r and V_r (except in case of epistasis which, when present, is removed by omitting the appropriate arrays) gives a visual depiction of the degree of dominance determined by the point where the W_r , V_r regression line intersects the W_r axis. Besides the degree of dominance, it shows the relative number of dominant and recessive genes possessed by the parents involved in the diallel cross, and the number of effective factors that can be estimated. Because of early generation evaluation of the hybrids, it has found frequent application in breeding and evolutionary problems. Both Jinks and Hayman develop their analysis on single gene model and then generalize the treatment for polygenes with the assumption that the genes are independently distributed. Gilbert (1958) has pointed out the shortcomings of such an approach. The present study takes into account the situations where the genes are taken as concrete units of segregation, and their additive and dominance components have been followed under regular and incomplete binomial distributions. The latter introduces a degree of association/dispersion amongst the genes and Coughtrey and Mather (1970) have worked out their effects in diallel cross analysis. They have taken a two gene case and have generalized for polygenes. In the present study as it was felt that the number of genes determining a quantitative character is likely to be large and a perfectly independent distribution of genes within the limited number of parents that are involved in the diallel cross is not possible, certain selected distributions within the totality of different possible genotypes and the graphs that develop from them have been worked out. A method for estimating the number of genes from W_r , V_r graph is presented. The applicability of these methods has been illustrated.

The meanings of the terms d and h , the additive and dominance components of variance respectively, are the same as given initially by Mather (1949). The assumptions of the analysis are the same as given by Jinks (1954) and Hayman (1954). Further, d is assumed to be the same for all genes and likewise h is also assumed to be the same for all genes determining a particular character.

INDEPENDENT DISTRIBUTION OF GENES AND Vr, Wr VALUES

The possible number of parental arrays depends upon the number of genes that govern the character. Keeping in view the assumption that the parents are homozygous, with a single gene there are only two non-segregating arrays AA and aa, while with two genes there would be four (AA BB, AA bb, aa BB, aa bb) and with three genes there would be eight (AA BB CC, AA BB cc, AA bb CC, AA bb cc, aa BB CC, aa BB cc, aa bb CC, aa bb cc) and with n genes, there would be $(2)^n$ such arrays. The array totals as well as Vr and Wr follow a definite distribution with regard to d and h components. Their values of d and h components for 1-4 genes are given in Table I by taking the initial genotype frequencies to be equal ($u=v=0.5$) for all cases, and such values for any number of genes can be worked out. In this table the number at upper left outside the brackets shows the number of arrays having this value. These numbers follow the usual pattern of binomial distribution. In the array totals, h is distributed

TABLE I
The complete binomial distribution of d and h components

No. of genes	Array totals				
1	$1(h+d)$	$1(h-d)$			
2	$1(4h+4d)$	$2(4h)$	$1(4h-4d)$		
3	$1(12h+12d)$	$3(12h+4d)$	$3(12h-4d)$	$1(12h-12d)$	
4	$1(32h+32d)$	$4(32h+16d)$	$6(32h)$	$4(32h-16d)$	$1(32h-32d)$
			Vr		
1	$1(0.25d^2+0.25h^2-0.5dh)$	$1(0.25d^2+0.25h^2+0.5dh)$			
2	$1(0.5d^2+0.5h^2-dh)$	$2(0.5d^2+0.5h^2)$	$1(0.5d^2+0.5h^2+dh)$		
3	$1(0.75d^2+0.75h^2-1.5dh)$	$3(0.75d^2+0.75h^2-0.5dh)$	$3(0.75d^2+0.75h^2+0.5dh)$	$1(0.75d^2+0.75h^2+1.5dh)$	
4	$1(d^2+h^2-2dh)$	$4(d^2+h^2-dh)$	$6(d^2+h^2)$	$4(d^2+h^2+dh)$	$1(d^2+h^2+2dh)$
			Wr		
1	$1(0.5d^2-0.5dh)$	$1(0.5d^2+0.5dh)$			
2	$1(d^2-dh)$	$2(d^2)$	$1(d^2+dh)$		
3	$1(1.5d^2-1.5dh)$	$3(1.5d^2-0.5dh)$	$3(1.5d^2+0.5dh)$	$1(1.5d^2+1.5dh)$	
4	$1(2d^2-2dh)$	$4(2d^2-dh)$	$6(2d^2)$	$4(2d^2+dh)$	$1(2d^2+2dh)$

throughout according to the formula $(2)^n \times n/2$, while d observes this distribution in the marginal classes with its value being positive in the first and negative in the last class. In the central classes, d is distributed according to the rule of the Pascal triangle, i.e., its value is the sum of the values of the two d elements immediately above it. In V_r , h^2 and d^2 follow the distribution formula, $n/4$. The marginal dh follow the formula, $n/2$ with the first value positive and the last value negative, and the central dh follow the modified Pascal triangle rule, i.e., each element being composed of two elements above and divided by 2. Similarly for W_r the distribution of the different components can be explained.

INCOMPLETE BINOMIAL DISTRIBUTION AND W_r , V_r GRAPH

The representation of genotypes in the parental arrays affects the shape of the W_r , V_r graph. This would, of course, depend basically upon the number of the genes determining the character as they determine the possible number of different genotypes and hence of the parental arrays. The combinations for each value of n can be worked

out by the expression $\frac{n!}{(n-r)! r!}$ and would depend upon r , where r is the

number of the parental arrays taken at any particular time. Totalling the number of combinations for different values of r from 2 to n , all the possible types of combinations for a given gene number (and its possible parental arrays) can be worked out. With 3 genes (having 8 parental genotypes) the total number of possible combinations is 247. With a large number of genes determining a character, the number of such combinations can be very high. Obviously, not all the possible combinations can be taken into account. The effect of omitting certain arrays is shown in Figs. 1 and 2 for 2 and 3 genes respectively, with $h=0.5$, 1.0 and 1.5.

In the two gene model, the omission of any one of the arrays leads to deviation of W_r , V_r line from the unit slope each leaving its characteristic effect. The omission of $++$ or $--$ array reduces the length of the W_r , V_r line and the two arrays $-+$ and $+ -$ fall at the same point which is the lower point when $++$ array is omitted (dotted line) and the same is the upper point when $--$ array is omitted (dashed line). As in Fig. 1, the W_r , V_r lines after omitting the extreme genotypes individually have been drawn together, the central point where the two lines (i.e. dotted and dashed line) meet actually represents four arrays whose values fall at the same place. The angle formed at this point is convex downwards. The position of these lines is lower in relation to the line of unit slope and they thus tend to inflate the degree of dominance. The omission of either of the genotypes $+ -$ and $- +$ has the same effect and produces a line which is convex upwards and is on the upper side of the line of unit slope thus tending to lower the degree of dominance.

The three gene graphs also develop in a similar fashion. In line of unit shape, as expected, $---$ array occupies the top position, $+++$ the bottom position and three $--+$ and three $++-$ arrays occupy two central positions. With the omission of the extreme genotype $+++$ (dotted line) the most recessive array is at the top and the two intermediate genotypes $--+$ and $++-$ representing three arrays each occupy the two lower positions. With the omission of $---$ genotype,

(dashed line) the + + + array takes the lowest position, the position of the intermediate genotypes remaining unchanged. In both of these situations, W_r , V_r lines have downward convexity. The omission of any of — + or + + — arrays produces an upward convexity though at different locations. With more than three genes determining the character and the omission of genotypes as already described, the shape of W_r , V_r graph can be extrapolated.

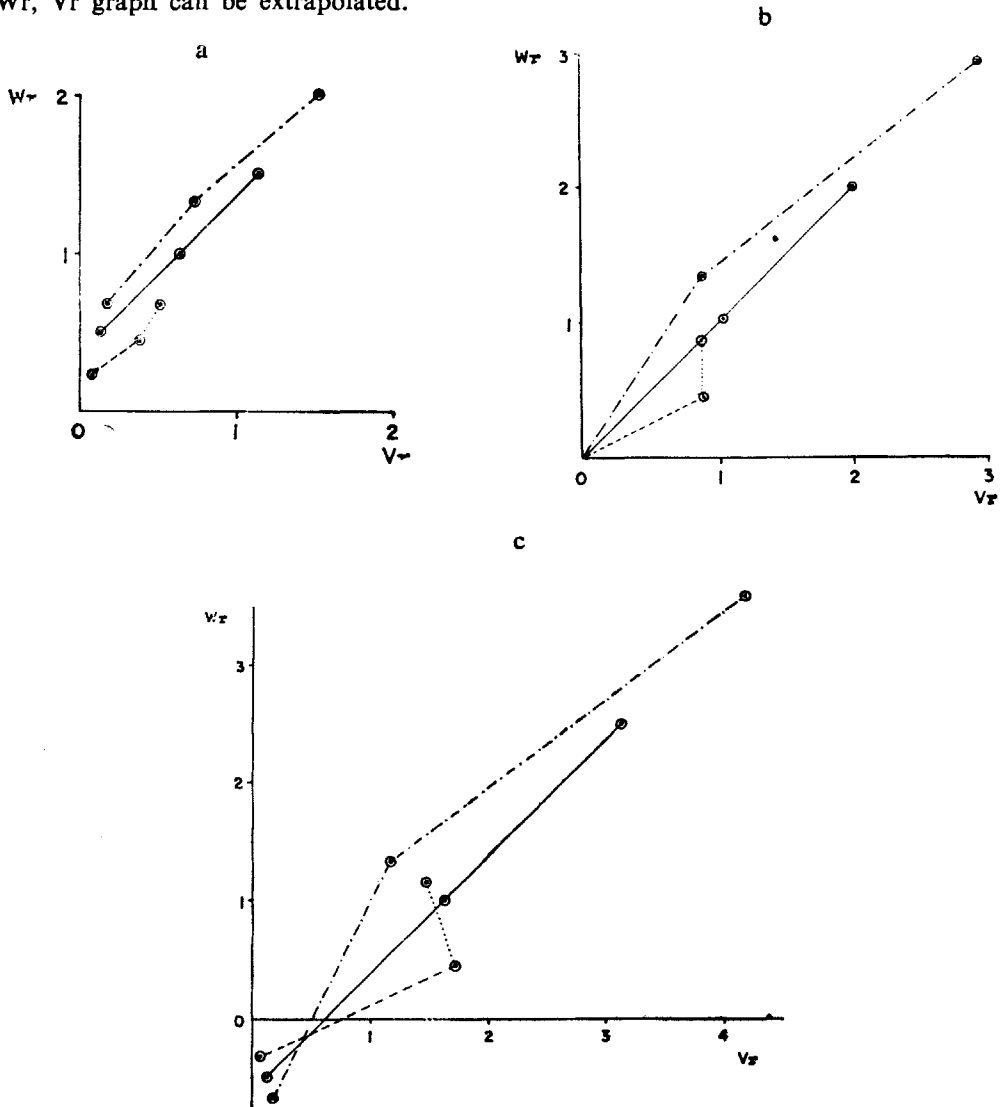


FIG. 1a—c : W_r , V_r graph for 2 genes showing partial dominance, $h=0.5$ (a); complete dominance, $h=1.0$ (b); and over dominance, $h=1.5$ (c). The line of unit slope when all the arrays are present is shown in solid. The central point represents arrays +— and —+ which fall at the same place. The dot line shows the absence of + + array, the dash line shows the absence of — — array and the dot and dash line shows the absence of +— or —+ array.

DISTRIBUTIONS UNDER PARENTAL SELECTIONS

Selection of first order—If the parents are selected in a way that each class of genotype is represented only once (as for instance, with four genes the 16 possible genotypes fall into 5 classes), still a regular and characteristic distribution of array totals, W_r , and V_r is obtained. In Tables II–IV, the values upto 5 genes have been worked out.

TABLE II
Distribution of array totals in selection of first order

No. of genes						
*1	$d+h$	$h-d$				
2	$3d+3h$	$2h$	$3h-3d$			
3	$6d+6h$	$2d+4h$	$4h-2d$	$6h-6d$		
4	$10d+10h$	$5d+7h$	$6h$	$7h-5d$	$10h-10d$	
5	$15d+15h$	$9d+11h$	$3d+9h$	$9h-3d$	$11h-9d$	$15h-15d$

*Presented for comparison, no selection.

The values for array totals are given in Table II. The marginal d and h coefficients are the same and can be calculated by the formula: $n(n/2+0.5)$. The values for central classes can be calculated from the distribution of the previous n , i.e. $n-1$, by taking its left marginal value and successively adding $1h, 2h, 3h, \dots$ and subtracting $1d, 2d, 3d, \dots$. The V_r values (Table III) are actually the sum of squares and the value in each cell must be divided by the number of arrays for a particular n (i.e., $n+1$) to get V_r . Obviously, this would apply also to formulae now discussed for working out V_r . The coefficients d^2, h^2 and dh follow different patterns of distribution. The value of coefficients of d^2 is the same for all classes of any genotypic array and it is also the same as the value of coefficients of h^2 in the marginal cells of that array and can, therefore, be worked out from the distribution of h^2 given below:

No. of genes	Coefficients of h^2	
1	$X_1^2 - \frac{(X)^2}{2}$	$X_1^2 - \frac{(X)^2}{2}$
2	$X_2^2 - \frac{(X)^2}{3}$	$X_1^2 + (1)^2 - \frac{(X+1)^2}{3}$
3	$X_3^2 - \frac{(X)^2}{4}$	$X_2^2 + (1)^2 - \frac{(X+1)^2}{4}$
		$X_1^2 + (1)^2 + (2)^2 - \frac{(X_1+1+2)^2}{4}$
n	$X_n^2 - \frac{(X)^2}{n+1}$	$X_{n-1}^2 + (1)^2 - \frac{(X_{n-1}+1)^2}{n+1}$
		\dots
		$X_1^2 + (1)^2 + (2)^2 \dots + (n)^2 - \frac{(X_1+1+2 \dots +n)^2}{n+1}$

Where $X_n = 1 + 2 \dots + n, X_n^2 = 1^2 + 2^2 + \dots + n^2, (X_n)^2 = (1 + 2 \dots + n)^2$

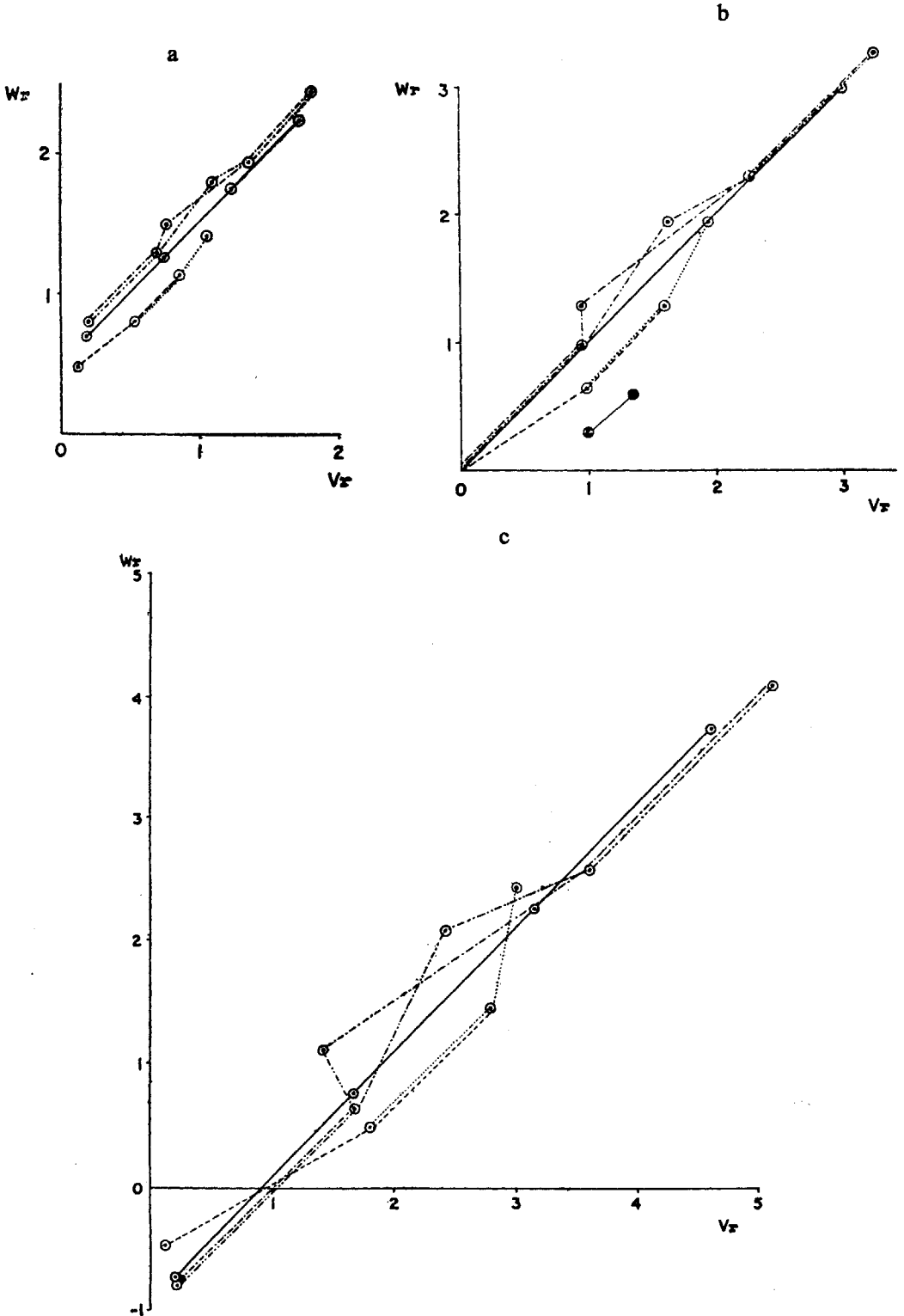


FIG. 2

It can be noted that while in the above it is the cell to the upper left which is important for calculating the value for any cell, in dh it is the difference between the two left upper cells which is important from this point of view and it is these differences (underlined) that follows a simple linear distribution with increasing number of genes. This is illustrated below:

No. of genes		<u>2</u>								
1	-1	<u>4</u>	1	<u>4</u>						
2	-4	<u>6</u>	0	<u>8</u>	4	<u>6</u>				
3	-10	<u>8</u>	-4	<u>12</u>	4	<u>12</u>	10	<u>8</u>		
4	-20	<u>10</u>	-12	<u>16</u>	0	<u>18</u>	12	<u>16</u>	20	<u>10</u>
5	-35		-25		-9		9		25	

$$\text{If } k = -2 \left\{ X_n^2 - \frac{(X_n)^2}{n+1} \right\}$$

then the various elements for the distribution for n number of genes can be calculated as:

- k
- $k + 2n$
- $k + 2n + 2(n-1)2$
- $k + 2n + 2(n-1)2 + 2(n-2)3$
- \vdots
- $k + 2n + 2(n-1)2 + 2(n-2)3 \dots 2(n-n+1)n$

i. e. by adding each time the following in succession to the previous sum:

$$K, 2n, 2(n-1)^2, 2(n-2)^3, 2(n-3)^4 \dots 2(n-n+1)^n, \text{ or } 2n$$

The values of Wr given in Table IV, too are cross products and every value must be divided by appropriate value for n to get Vr . They are simply related to Vr as the values of d^2 here are double to those in Vr , and dh follows the same distribution as in Vr .

The graphs that develop from such selections are shown in Figs. 3a & b. The degree of dominance may be measured by the distance from the point of no dominance

FIG. 2a-c. Wr, Vr graph for 3 genes showing partial dominance, $h=0.5$ (a); complete dominance, $h=1.0$ (b); and over dominance, $h=1.5$ (c). The line of unit slope when all the arrays are present is shown in solid. The dot line shows the absence of + + + array, the dash line shows the absence of - - - array, one dot and dash line shows the absence of one of the - - + arrays, and two dots and dash line shows the absence of one of the + + - arrays. In (b) the short line between the two solid points shown in lower portion is obtained when both + + + and - - - arrays are absent.

TABLE III
Distribution of Vr sum of squares in selection of first order

No. of genes								
*1	$0.5d^2 + 0.5h^2 - dh$	$0.5d^2 + 0.5h^2 + dh$						
2	$2d - 4dh + 2h^2$	$2d^2 + 4dh + 2h^2$						
3	$5d^2 - 10dh + 5h^2$	$5d^2 + 2h^2 - 4dh$	$5d^2 + 5h^2 + 10dh$					
4	$10d^2 + 10h^2 - 20dh$	$10d^2 + 2h^2 + 4dh$	$10d^2 + 5.2h^2 + 12dh$	$10d^2 + 10h^2 + 20dh$				
5	$17.5d^2 + 17.5h^2 - 35dh$	$10d^2 + 2.8h^2$	$17.5d^2 - 5.5h^2 - 9.0dh$	$17.5d^2 + 10.9h^2 + 25.0dh$	$17.5d^2 + 17.5h^2 + 35dh$			

*Presented for comparison, no selection.

TABLE IV
The distribution of Wr cross products in selection of first order

No. of genes							
*1	$d^2 - dh$	$d^2 + dh$					
2	$4d^2 - 4dh$	$4d^2$	$4d^2 + 4dh$				
3	$10d^2 - 10dh$	$10d^2 - 4dh$	$10d^2 + 10dh$				
4	$20d^2 - 20dh$	$20d^2 - 12dh$	$20d^2 + 12dh$	$20d^2 + 20dh$			
5	$35d^2 - 35dh$	$35d^2 - 25dh$	$35d^2 - 9dh$	$35d^2 + 25dh$	$35d^2 + 35dh$		

*Presented for comparison, no selection.

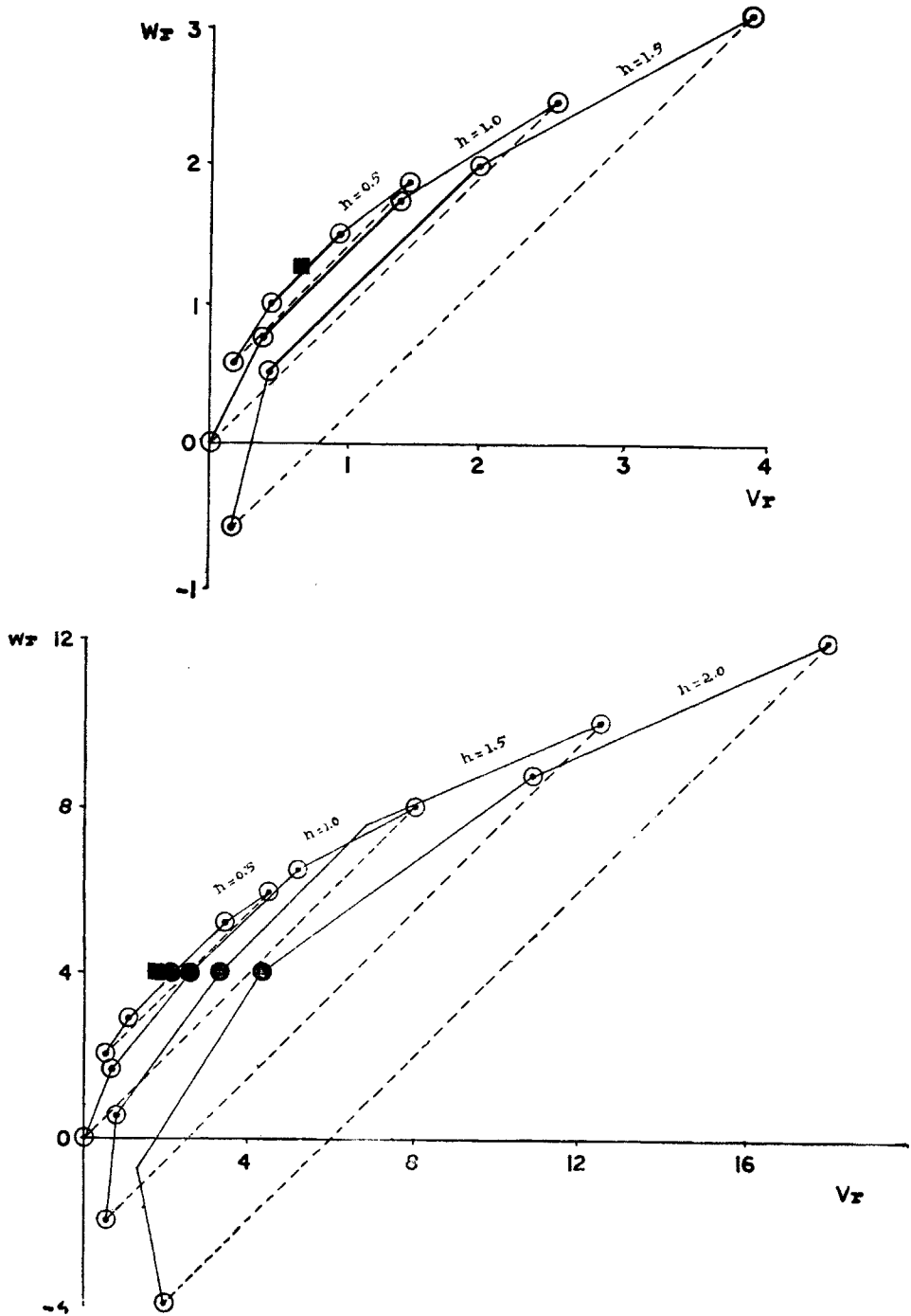


FIG. 3a & b. W_r , V_r graph when one array from each class of genotype expected under binomial distribution is represented, (a) 3 genes, (b) 4 genes. The point of no dominance $W_r/V_r=2$ is represented in solid square. The curve for each value of dominance based on four arrays (a), and on 5 arrays (b), is given by the solid line. In (a) the portion connecting the two middle points in each case is shown by a thicker line and its central point is in line horizontally with the point of no dominance. In (b), the central point (shown solid), representing the central array, is in line with the point of no dominance. The lines connecting (dashed) the end points of the above lines (curves) are lines of unit slope and are parallel to one another.

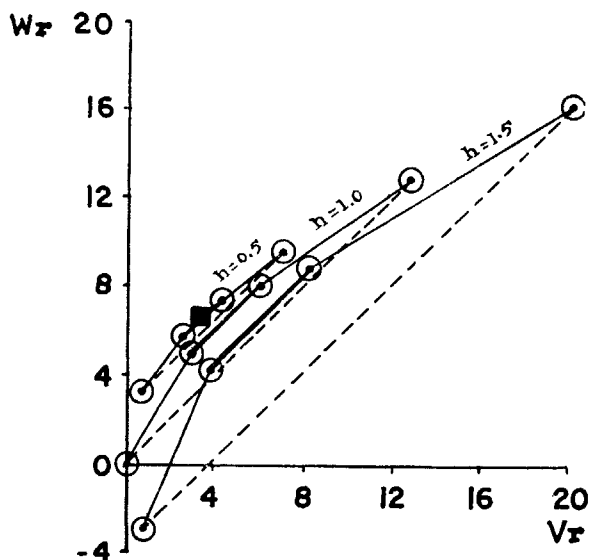
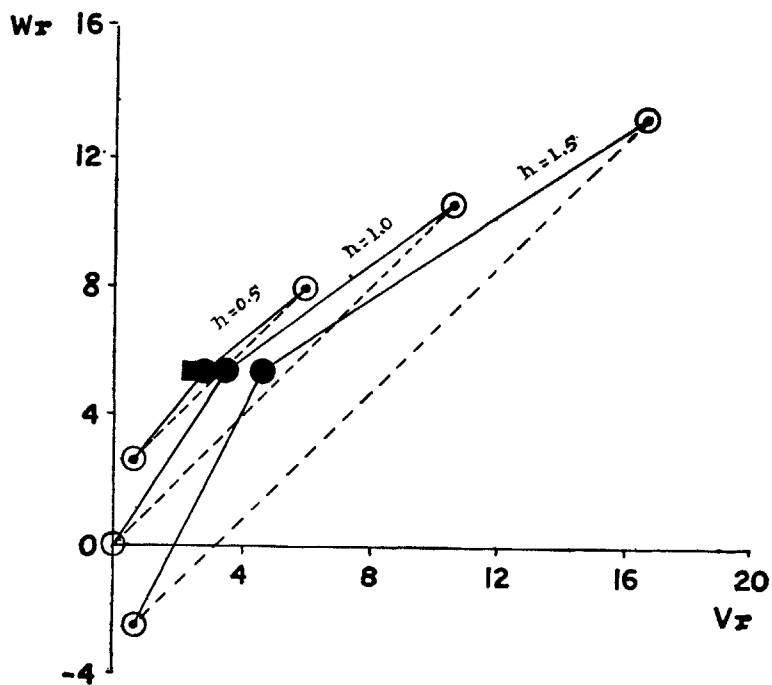


FIG. 4a & b. W_r , V_r graph, when 4 genes with the two genotypes at the extreme margins and the central genotype (a) or five genes with the two genotypes at the extreme margins and the two central genotypes (b) expected under the binomial distribution are taken in the diallel cross. The general shape of the graphs is similar to those in Fig. 3.

(solid square) to the middle point of the thick line in case of three genes and to the central point (solid circle) in the case of four genes. The degree of dominance could also be measured from the point of intersection of dashed lines at the W_r axis as in the normal W_r, V_r graph.

Selection of second order—The parental selections in which the two extreme genotypes and the central genotype or two central genotypes depending whether the number of genes determining the character is even or odd, are taken in account, the diallel cross statistics still give a regular pattern of distribution. This would greatly reduce the size of the experiment and the same number of genotypes, i.e., three or four has to be considered no matter whatever be the possible number of the genotypes. The distributions for array totals, V_r and W_r are given in Tables V–VII along with the formulae for working out any particular distribution.

The W_r, V_r graphs for four and five genes under second order of selection are given in Figs. 4a and b respectively. It is observed that as in the selection of first order, the central array (or mid point of two central arrays) falls in line with the point of no dominance (dark square) and the degree of dominance can be measured as described earlier.

ESTIMATION OF THE NUMBER OF GENES

The variation in the length of the W_r, V_r line in relation to the degree of dominance and number of genes is given in Fig. 5, but even for all types of regular distributions presented above, the number of genes can be estimated from the length of the W_r, V_r line. Obviously, the length of the W_r, V_r line is determined by the position of the top recessive and bottom dominant array. In the independent distribution given in Table I, the required values are:

V_r		W_r	
Top array (a)	Bottom array (b)	Top array (a')	Bottom array (b')
$\frac{n}{4}(d^2 + h^2) + \frac{n}{2}dh$	$\frac{n}{4}(d^2 + h^2) - \frac{n}{2}dh$	$\frac{n}{2}(d^2 + dh)$	$\frac{n}{2}(d^2 - dh)$
Length of V_r line		Length of W_r line	
$a - b = ndh$		$a' - b' = ndh$	

$$\begin{aligned} \text{Length of } W_r, V_r \text{ line} &= \{(ndh)^2 + (ndh)^2\}^{\frac{1}{2}} \\ &= (2)^{\frac{1}{2}} (ndh) = 1.4142 (ndh) \end{aligned}$$

By taking $d = 1$ and expressing h relative to d ,

$$n = \frac{\text{length of } W_r, V_r \text{ line}}{h \times 1.4142}$$

This is borne out by the graphs in Fig. 5.

Similarly for selections of 1st order it is seen that the length of W_r, V_r line is:

$$\{(2kdh)^2 + (2kdh)^2\}^{\frac{1}{2}}$$

where k is the coefficient of dh and is the same for the first as well as the last arrays of both V_r and W_r . Once the values of d and h are known, by substituting

$$k = -2 \left\{ X^2_n - \frac{(X_n)^2}{n+1} \right\}$$

as given already for different values of n , the length of W_r , V_r line that comes closest to a particular n , would give the number of genes involved.

TABLE V
The distribution of array totals in selection of second order

No. of genes				
3	$6d+6h$	$2d+4h$	$4h-2d$	$6h-6d$
4	$6d+6h$	$4h$	$6h-4d$	
5	$10d+10h$	$2d+6h$	$6h-2d$	$10h-10d$
6	$9d+9h$	$6h$	$9h-9d$	
7	$14d+14h$	$2d+8h$	$8h-2d$	$14h-14d$
8	$12d+12h$	$8h$	$12h-12d$	

Marginal d and h coefficients:

- (i) Odd number of genes : $n \times 2$
- (ii) Even number of genes : $n + n/2$

Central d and h coefficients:

- (i) Odd arrays : $d=2, h=n+1$
- (ii) Even arrays : $d=0, h=n$

TABLE VI
The distribution of V_r sum of squares in selection of second order

No. of genes				
3	$5d^2 + 5h^2 - 10dh$	$5d^2 + 2h^2 - 4dh$	$5d^2 + 2h^2 - 4dh$	$5d^2 + 5h^2 + 10dh$
4	$8d^2 + 8h^2 - 16dh$	$8d^2 + 2.67h^2$	$8d^2 + 8h^2 + 16dh$	
5	$13d^2 + 13h^2 - 26dh$	$13d^2 + 5h^2 - 6dh$	$13d^2 + 5h^2 + 6dh$	$13d^2 + 13h^2 + 26dh$
6	$18d^2 + 18h^2 - 36dh$	$18d^2 + 6h^2$	$18d^2 + 18h^2 + 36dh$	
7	$25d^2 + 25h^2 - 50dh$	$25d^2 + 10h^2 - 8dh$	$25d^2 + 10h^2 + 8dh$	$25d^2 + 25h^2 + 50dh$
8	$32d^2 + 32h^2 - 64dh$	$32d^2 + 10.7h^2$	$32d^2 + 32h^2 + 64dh$	

Marginal d^2 and h^2 coefficients:

- (1) Arrays with odd number of genes : $(n^2+1)/2$
- (2) Arrays with even number of genes : $(n^2)/2$

Central h^2 coefficients:

$$= 2 + (x_1 + x_2) + (x_2 + x_3) \dots + (x_{i-1} + x_i) - \frac{(n_i + 1)^2}{4} \text{ (for odd } n_i \text{ arrays)}$$

O

$$= 2 + (x_1 + x_2) + (x_2 + x_3) \dots + (x_{i-1} + x_i) - \frac{n_i^2}{3} \text{ (for even } n_i \text{ arrays)}$$

Where for the odd n arrays:

$$x_1=(n=1), x_2=(n=3), x_3=(n=5)$$

and for the even n arrays:

$$x_1=(n=2), x_2=(n=4), x_3=(n=6)$$

Marginal dh coefficients:

They are the sum of marginal d^2 and h^2 coefficients.

Central dh coefficients:

1. Arrays with odd number of genes: $n_1=2, n_3=4, n_5=6, n_7=8\dots$
2. Arrays with even number of genes: 0.

TABLE VII

The distribution of Wr cross products in selection of second order

No. of genes				
3	$10d^2-10dh$	$10d^2-4dh$	$10d^2+4dh$	$10d^2+10dh$
4	$16d^2-16dh$	$16d^2$	$16d^2+16dh$	
5	$26d^2-26dh$	$26d^2-6dh$	$26d^2+6dh$	$26d^2+26dh$
6	$36d^2-36dh$	$36d^2$	$36d^2+36dh$	
7	$50d^2-50dh$	$50d^2-8dh$	$50d^2+8dh$	$50d^2+50dh$
8	$64d^2-64dh$	$64d^2$	$64d^2+64dh$	

Marginal d^2 and dh coefficients:

Arrays with odd number of genes: n^2+1

Arrays with even number of genes: n^2

Central d^2 and dh^2 coefficients:

Coefficient of d^2 is the same for whole of the genotypic array dh :

Odd n array : $n+1$

Even n array: 0

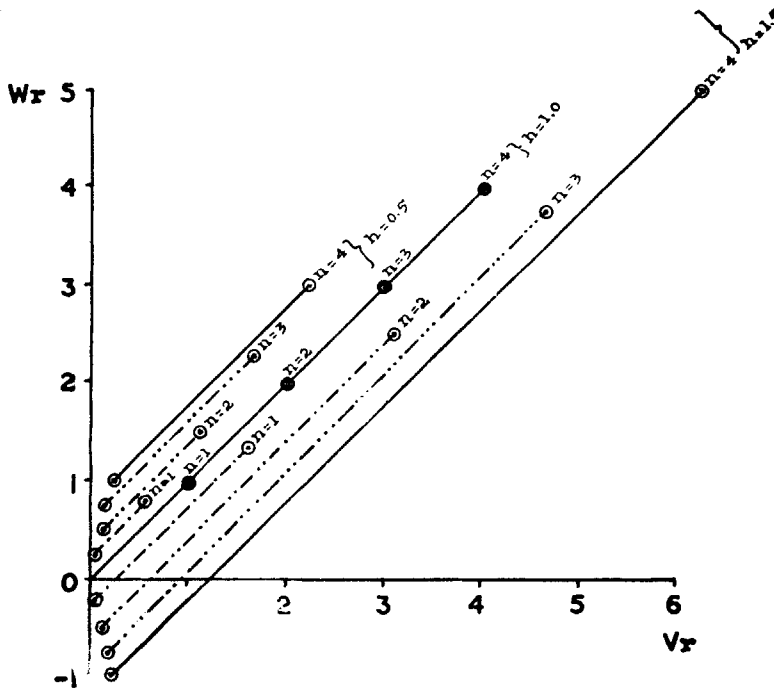


FIG. 5. The length of Wr, Vr line in relation to the number of genes and degree of dominance.

The selection of second order is more important as it gives the desired information by taking only the smallest number of crosses. In this case the results are dependent upon whether the number of genes determining the character is odd or even. From Tables VI and VII, taking out d^2 and h^2 which are common for the marginal arrays, the length of Vr line is:

Top array a	Bottom array b	Length of Vr line (a—b)
Odd (n): $(n^2 + 1) dh$	— $(n^2 + 1) dh$	$2 (n^2 + 1) dh$
Even (n): $n^2 dh$	— $n^2 dh$	$2n^2 dh$

The values for Wr are the same as for Vr.

The length of Wr, Vr line is, therefore,

$$\begin{aligned} \text{Odd (n): } & \{ (2dh (n^2 + 1))^2 + (2dh (n^2 + 1))^2 \}^{\frac{1}{2}} \\ & = (8)^{\frac{1}{2}} dh (n^2 + 1) \\ & = 2.8284 dh (n^2 + 1) \end{aligned}$$

taking the value of d as 1 and expressing h in relation to d , as usual, and

$$n = \left\{ \frac{\text{Length of Wr, Vr line} - 2.8284 h}{2.8284 h} \right\}^{\frac{1}{2}}$$

$$\begin{aligned} \text{Even (n): } & \{ (2n^2 dh)^2 + (2n^2 dh)^2 \}^{\frac{1}{2}} \\ & = (2)^{\frac{1}{2}} (2n^2 dh) \\ & = 2.8284 hdn^2, \text{ and similarly,} \end{aligned}$$

$$n = \left\{ \frac{\text{length of Wr, Vr line}}{2.8284h} \right\}^{\frac{1}{2}}$$

It must be noted that value of d is taken to be 1 for all the models presented above and value of h has been expressed relative to d . However, for application to practical data the Wr, Vr line must be scaled in this way and this can be done by dividing the value of Wr, Vr line by the value obtained for d . Secondly, as mentioned earlier, the various distributions under parental selections are presented as sum of squares and to calculate Vr and Wr, they must be divided by the number of arrays taken in account. These points become clear in the illustration given below.

ILLUSTRATION

A diallel cross of eight varieties of tomato was analysed for fruit size (Khanna and Chaudhary 1974) according to the original model of Jinks (1954) and Hayman (1954). A reduced diallel from this after selection of the second order is given in Table VIII. The half above the main diagonal represents the selection of the two central arrays and the half below represents the selection of one central array. The Wr, Vr graphs that develop from these selections are given in Figs. 6a & b. To work out the number of genes from the length of the Wr, Vr line, the values of d and h must be known. The parental

TABLE VIII
Reduced diallel for fruit size in tomato after selection of second order

	RJ	CP	HN	PP
RJ	110.22	85.20	74.26	32.31
CP	85.20	66.60	66.35	32.11
HN	—	—	56.48	29.98
PP	32.31	32.11	—	13.18

RJ, Red Jacket; CP, Crack Proof; HN, Harkness; PP, Pusa Red Plum.

variance or D equals $4 \sum uv d^2$ and H_1 equals $4 \sum uv h^2$ (Jinks 1954). Ideally, the top-most array would have all the recessive genes and the bottommost array would have all the dominant genes, thus giving $uv=0.25$ when these arrays are taken in account. In the present case by taking only the parents RJ and PP the values obtained are: $D=2352.75$, $H_1=969.28$ and $\bar{uv}=0.25$, $H_2/H_1=0.2493$ which is very close to the expectations. From the above, the workable values of d and h may be taken as 48.50 and 31.10 respectively. The degree of dominance is $\left(\frac{H_1}{D}\right)^{\frac{1}{2}}=0.64$, a value quite similar to the ones obtained for the full diallel where $\left(\frac{H_1}{D}\right)^{\frac{1}{2}}=0.57$ and $VILI/WOLOI = 0.69$ (Khanna & Chaudhary 1974). The length of the W_r, V_r line (Fig. 6a & b) based on possible odd n (four arrays) is 990 and that based on even n (three arrays) is 1293. The value of n in the two cases is, therefore:

$$(1) \quad n = \left\{ \frac{d (\text{Length of } W_r, V_r \text{ line} \times 4 - 2.8 h)}{d (2.8 h)} \right\}^{\frac{1}{2}}$$

$$= \left\{ \frac{3960 - 87.08}{87.08} \right\}^{\frac{1}{2}} = 6.6$$

$$(2) \quad n = \left\{ \frac{d (\text{Length of } W_r, V_r \text{ line} \times 3)}{d (2.8 h)} \right\}^{\frac{1}{2}}$$

$$= \left\{ \frac{3879}{87.08} \right\}^{\frac{1}{2}} = 6.7$$

It is thus seen that the number of genes determining fruit size in tomato is about 6 or 7.

The term 'genes' is, of course, applied within the limitations of the assumptions given earlier.

The scaling procedures used above are being investigated further.

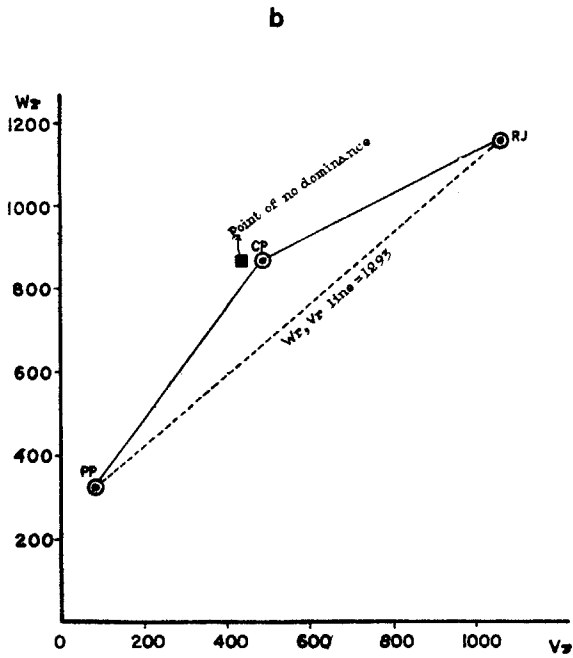
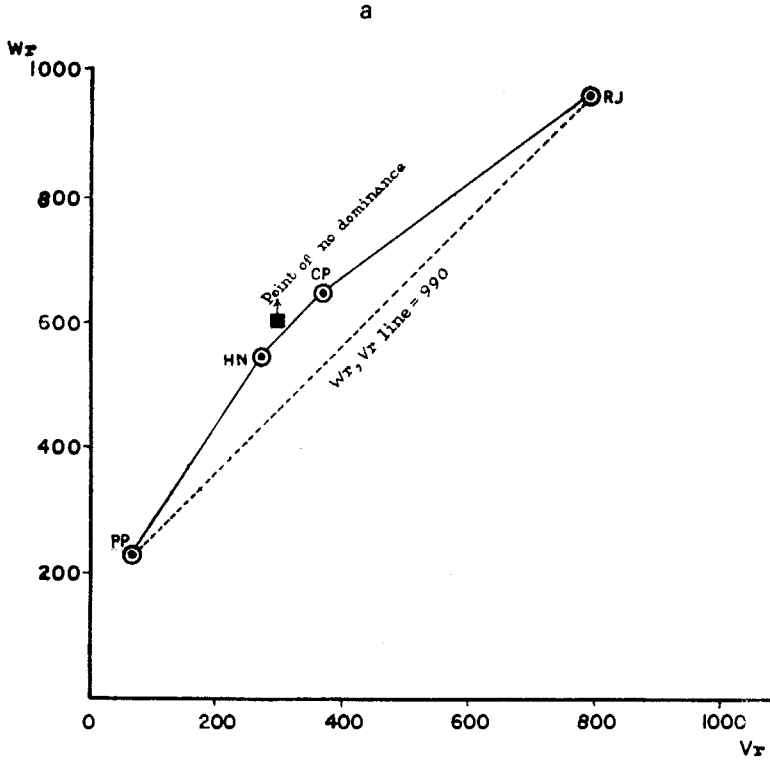


FIG. 6a & b. W_r, V_r graph for four varieties (a), and three varieties (b) of tomato.

REFERENCES

- Coughtrey, A., & Mather, K. (1970). Interaction and gene association and dispersion in diallel crosses where gene frequencies are unequal. *Heredity*, **25**, 79-88.
- Gilbert, N. (1958). Diallel cross in plant breeding. *Heredity*, **12**, 477-492.
- Hayman, B. I. (1954). The theory and analysis of diallel crosses. *Genetics*, **39**, 789-809.
- Jinks, J. L. (1954). The analysis of continuous variation in diallel cross of *Nicotiana rustica* varieties. *Genetics*, **39**, 767-788.
- Khanna, K. R. & Chaudhary, R. C. (1974). The nature of gene action for economic characters in tomato and its bearing on yield improvement. In: *Advancing Frontiers of Cytogenetics*, pp. 150-161, New Delhi.
- Mather, K. (1949). *Biometrical Genetics*, London.