METHODOLOGY

THEORY AND ANALYSIS OF PARTIAL DIALLEL CROSSES

José Marcelo Soriano Viana[1], Cosme Damião Cruz[1] and Antonio Américo Cardoso[2]

ABSTRACT

This study presents theory and analysis of partial diallel crosses based on Hayman's methods. This genetic design consists of crosses among two parental groups. It should be used when there are two groups of parents, for example, dent and flint maize inbred lines, and the breeder is not interested in the assessment of crosses between parents of the same group. Analyses are carried out using data from the parents and their F_1 hybrids allowing a detailed characterization of the polygenic systems under study and the choice of parents for hybridization. Diallel analysis allows the estimation of genetic and non-genetic components of variation and genetic parameters and to assess the following: genetic variability in each group; genotypic differences between parents of distinct groups; if a parent has a common or a rare genotype in the group to which it does not belong; if there is dominance; if dominant genes increase or decrease trait expression (direction of dominance); average degree of dominance in each group; the relative importance of mean effects of genes and dominance in determining a trait; if, in each group, the allelic genes have the same frequency; if genes are equally frequent in the two groups; the group with the greatest frequency of favorable genes; the group in which dominant genes are most frequent; the relative number of dominant and recessive genes in each parent; if a parent has a common or a rare genotype in the group to which it belongs, and the genotypic differences between parents of the same group. An example with common bean varieties is considered.

INTRODUCTION

A diallel is a mating system that involves all possible crosses among a group of parents. This genetic design is used to study polygenic systems that determine quantitative traits. Depending on the nature of the parents, individuals from an open pollinated variety or an F_2 generation, or selected open pollinated varieties or line cultivars, inferences based on diallel results are useful in planning base populations for intra and/or interpopulation improvement or parents for hybridization. The partial diallel is one of many variations of the diallel (Griffing, 1956; Kempthorne and Curnow, 1961; Gardner and Eberhart, 1966) and consists of crosses among two parent groups. This mating system, also called experiment II, design II and factorial design, was proposed by Comstock and Robinson (1948, 1952). It is adequate when there are distinct groups of populations, for example one group of dent inbred lines and the other of flint inbred lines, derived from a reciprocal recurrent selection program, and the breeder is not interested in evaluating the crosses between parents of the same group. The two parent groups can involve adapted or commercial populations and exotic germplasm or plant introductions, for example, dwarf cashew tree and normal clones, small and large seeded cultivated common beans, among many others. As generally the parents are open-pollinated populations or pure lines defined by the breeder, there is no base population for inferences. The inferences must be established in relation to the set of parents.

Jinks and Hayman (1953) and Hayman (1954, 1958)

presented methods for diallel analysis, with a firm and refined theoretical basis, using data from parents and their F_1 progeny or parents, F_1 and F_2 generations. These methods have been used ever since (Jinks, 1954; Dickson, 1967; Chung and Stevenson, 1973; Kornegay and Temple, 1986; Nishimura and Hamamura, 1993), although the methodologies proposed by Griffing (1956) and Gardner and Eberhart (1966) are the most frequently employed. Limitations of these methods are discussed by Gilbert (1958), Nassar (1965), Coughtrey and Mather (1970), and Sokol and Baker (1977).

As Hayman's proposals (1954, 1958) are not valid for partial diallels, a separate theory is required. The theory and analysis of partial diallel crosses presented are a generalization of the methodology of Hayman (1954). If gene frequencies are the same for the two parent groups, results are equal to those presented by Hayman.

METHODOLOGY

General theoretical considerations

Consider a polygenic system with k genes controlling a quantitative trait in a diploid species and the following conditions: a) Mendelian inheritance; b) absence of reciprocal effects; c) absence of non-allelic interaction; d) no multiple allelism; e) no correlation in the distribution of non-allelic genes in the parents, and f) homozygous parents. If N parents (N $\geq$ 6) are divided into two groups, one with n and the other with n' parents (n + n' = N, n and n' $\geq$ 3), three polygenic systems are defined: one related to the group with n parents, one associated to n' parents and the last related to N parents. Let p_r be the genotypic value of the rth parent of the n

parental group and p_s the genotypic value of the sth parent of the n' parental group. Thus:

$$p_r = m + \sum_{a=1}^{k} d_a \theta_{ra} \, (r = 1, \ldots, n)$$

$$p_s = m + \sum_{a=1}^{k} d_a \theta_{sa} \, (s = 1, \ldots, n'), \, m = \sum_{a=1}^{k} m_a$$

If p_t is the genotypic value of the tth parental in the N parental group, then:

$$p_t = m + \sum_{a=1}^{k} d_a \theta_{ta} \, (t = 1, \ldots, N)$$

For all loci in the polygenic system under study, the parameter d_a is the difference between the genotypic value of the homozygote of largest expression and the genotypic mean of the homozygotes (m_a). The θ variable assumes the value -1 or 1 depending on if the parent is homozygous for a gene which diminishes or increases trait expression, respectively. If u_a and v_a are frequencies of alleles that increase and reduce trait expression, respectively, the following expectation and variance apply to the n parental group:

$$E(\theta_{ra}) = u_a - v_a = w_a$$

$$V(\theta_{ra}) = 1 - w_a^2$$

In relation to the n' parental group:

$$E(\theta_{sa}) = u'_a - v'_a = w'_a$$

$$V(\theta_{sa}) = 1 - w'^2_a$$

In the group formed by N parents:

$$E(\theta_{ta}) = \bar{u}_a - \bar{v}_a = \bar{w}_a = pw_a + qw'_a,$$

with $\bar{u}_a = pu_a + qu'_a$, $\bar{v}_a = pv_a + qv'_a$, $p = n/N$, and $q = n'/N$, and

$$V(\theta_{ta}) = 1 - \bar{w}_a^2 = 1 - p^2 w_a^2 - q^2 w'^2_a - 2pq w_a w'_a = 4\bar{u}_a \bar{v}_a$$

Genotypic mean of the n parental group is:

$$E(p_r) = m_{L0} = m + \sum_{a=1}^{k} d_a w_a$$

Genotypic mean of the n' parental group is:

$$E(p_s) = m'_{L0} = m + \sum_{a=1}^{k} d_a w'_a$$

Genotypic mean of N parents is:

$$E(p_t) = \bar{m}_{L0} = m + \sum_{a=1}^{k} d_a \bar{w}_a = pm_{L0} + qm'_{L0}$$

Diallel analysis of parents and their F$_1$ hybrids

The genotypic mean of the hybrid derived from the rth and sth parents is:

$$f_{rs} = m + \frac{1}{2} \sum_{a=1}^{k} [d_a (\theta_{ra} + \theta_{sa}) + h_a (1 - \theta_{ra}\theta_{sa})] =$$

$$= \frac{1}{2} (p_r + p_s) + \frac{1}{2} \sum_{a=1}^{k} h_a (1 - \theta_{ra}\theta_{sa}),$$

with $\sum_{a=1}^{k} h_a (1 - \theta_{ra}\theta_{sa}) = h_{rs}$ being the specific heterosis of the rth and sth parents. The parameter h_a is the difference between the genotypic value of the heterozygote and m_a.

The genotypic mean of the hybrids from the rth parent or mean of the rth array is:

$$\underset{r \text{ fixed}}{E(f_{rs})} = f_r = m + \frac{1}{2} \sum_{a=1}^{k} [d_a (\theta_{ra} + w'_a) + h_a (1 - w'_a\theta_{ra})] =$$

$$= \frac{1}{2} (p_r + m'_{L0}) + \frac{1}{2} \sum_{a=1}^{k} h_a (1 - w'_a\theta_{ra}),$$

where $\sum_{a=1}^{k} h_a (1 - w'_a\theta_{ra}) = h_r$ is the varietal heterosis of the rth parent.

The genotypic mean of the hybrids from the sth parent or mean of the sth array is:

$$\underset{s \text{ fixed}}{E(f_{rs})} = f_s = m + \frac{1}{2} \sum_{a=1}^{k} [d_a (\theta_{sa} + w_a) + h_a (1 - w_a\theta_{sa})] =$$

$$= \frac{1}{2} (p_s + m_{L0}) + \frac{1}{2} \sum_{a=1}^{k} h_a (1 - w_a\theta_{sa}),$$

where $\sum_{a=1}^{k} h_a (1 - w_a\theta_{sa}) = h_s$ is the varietal heterosis of the sth parent.

The genotypic mean of the F$_1$ hybrids is:

$$E(f_{rs}) = E(f_r) = E(f_s) = m_{L1} = m + \frac{1}{2} \sum_{a=1}^{k} [d_a (w_a + w'_a) +$$

$$+ h_a (1 - w_a w'_a)] = \frac{1}{2} (m_{L0} + m'_{L0}) + \frac{1}{2} \sum_{a=1}^{k} h_a (1 - w_a w'_a),$$

where $\sum_{a=1}^{k} h_a (1 - w_a w'_a) = h$ is the average heterosis.

Thus,

$$m_{L0} - m'_{L0} = \sum_{a=1}^{k} 2d_a (u_a - u'_a)$$

$$m_{L1} - \frac{1}{2} (m_{L0} + m'_{L0}) = \frac{1}{2} h$$

In the presence of dominance, specific heterosis of the rth and sth parents is nil if the parents have the same genotype ($\theta_{ra}\theta_{sa} = 1$, for every a) and greatest when the parents are carriers of distinct alleles at the k loci ($\theta_{ra}\theta_{sa} = -1$, for every a). Varietal heterosis of a parent is least when it carries the most frequent genes in the group of parents to which it does not belong and greatest if it is the carrier of the least frequent genes. If deviations due to dominance

are of the same magnitude and the varietal heterosis in a group is constant, the allelic genes in the other parental group have the same frequency (w_a or w'_a equal to zero, for every a). Since $(1 - w_a w'_a) \geq 0$, if $h > 0$, positive unidirectional dominance is established. In other words, the deviations due to dominance should be predominantly positive. If $h = 0$ and there is dominance, positive and negative deviations are present (bi-directional dominance). Dominance is negatively unidirectional when $h < 0$.

If $m_{L0} - m'_{L0}$ is greater than zero, there is evidence that genes which increase trait expression are more frequent in the n parental group. If this difference is negative, genes which increase trait expression should be more frequent in the n' parental group. When the difference is nil, genes are equally frequent in the two parental groups ($w_a = w'_a$, for every a) or, alternatively, some genes which increase trait expression are more frequent in one group of parents, while others are more frequent in the other group.

Genetic components of variation

Analysis of the diallel table allows the definition of the following second degree statistics:

(1) Variance of the genotypic means of the n parents

$$V_{0L0(1)} = V(p_r) = \sum_{a=1}^{k} d_a^2 (1 - w_a^2) = 4 \sum_{a=1}^{k} u_a v_a d_a^2 = D_{(1)}$$

(2) Variance of the genotypic means of the n' parents

$$V_{0L0(2)} = V(p_s) = \sum_{a=1}^{k} d_a^2 (1 - w_a'^2) = 4 \sum_{a=1}^{k} u'_a v'_a d_a^2 = D_{(2)}$$

(3) Variance of the genotypic means of the N parents

$$V_{0L0} = V(p_t) = \sum_{a=1}^{k} d_a^2 (1 - \overline{w}_a^2) = D = p^2 D_{(1)} + q^2 D_{(2)} + 2pq D_{(3)},$$

$$D_{(3)} = \sum_{a=1}^{k} d_a^2 (1 - w_a - w'_a)$$

(4) Covariance between genotypic mean of rth parent hybrid and the mean of the nonrecurrent parent (covariance in the rth array)

$$W_{01(r)L01} = W_r = \underset{r \text{ fixed}}{Cov(f_{rs}, p_s)} = \frac{1}{2} D_{(2)} - \frac{1}{4} F_r,$$

$$F_r = 2 \sum_{a=1}^{k} d_a h_a \theta_{ra} (1 - w_a'^2)$$

(5) Variance of the genotypic means of the rth parent hybrids (variance in the rth array)

$$V_{1(r)L1} = V_r = \underset{r \text{ fixed}}{V(f_{rs})} = \frac{1}{4} D_{(2)} - \frac{1}{4} F_r + \frac{1}{4} H_{1(2)},$$

$$H_{1(2)} = \sum_{a=1}^{k} h_a^2 (1 - w_a'^2) = 4 \sum_{a=1}^{k} u'_a v'_a h_a^2$$

(6) Covariance in the sth array

$$W_{01(s)L01} = W_s = \underset{s \text{ fixed}}{Cov(f_{rs}, p_r)} = \frac{1}{2} D_{(1)} - \frac{1}{4} F_s,$$

$$F_s = 2 \sum_{a=1}^{k} d_a h_a \theta_{sa} (1 - w_a^2)$$

(7) Variance in the sth array

$$V_{1(s)L1} = V_s = \underset{s \text{ fixed}}{V(f_{rs})} = \frac{1}{4} D_{(1)} - \frac{1}{4} F_s + \frac{1}{4} H_{1(1)},$$

$$H_{1(1)} = \sum_{a=1}^{k} h_a^2 (1 - w_a^2) = 4 \sum_{a=1}^{k} u_a v_a h_a^2$$

(8) Covariance between genotypic mean of rth parent hybrid and the mean of the nonrecurrent parent array

$$W_{01(r)L1} = Cov(f_{rs}, f_s) = \frac{1}{4} D_{(2)} - \frac{1}{8} F_r - \frac{1}{8} F_{(1)} + \frac{1}{4} H_{1(2)} - \frac{1}{4} H_{2r},$$
$$\underset{r \text{ fixed}}{}$$

$$F_{(1)} = E(F_r) = 2 \sum_{a=1}^{k} d_a h_a w_a (1 - w_a'^2) = 8 \sum_{a=1}^{k} u'_a v'_a (u_a - v_a) d_a h_a$$

$$H_{2r} = \sum_{a=1}^{k} h_a^2 (1 - w_a \theta_{ra}) (1 - w_a'^2)$$

(9) Covariance between genotypic mean of sth parent hybrid and the mean of the nonrecurrent parent array

$$W_{01(s)L1} = Cov(f_{rs}, f_r) = \frac{1}{4} D_{(1)} - \frac{1}{8} F_s - \frac{1}{8} F_{(2)} + \frac{1}{4} H_{1(1)} - \frac{1}{4} H_{2s},$$
$$\underset{s \text{ fixed}}{}$$

$$F_{(2)} = E(F_s) = 2 \sum_{a=1}^{k} d_a h_a w'_a (1 - w_a^2) = 8 \sum_{a=1}^{k} u_a v_a (u'_a - v'_a) d_a h_a$$

$$H_{2s} = \sum_{a=1}^{k} h_a^2 (1 - w'_a \theta_{sa}) (1 - w_a^2)$$

(10) Variance of the genotypic means of the F_1 hybrids

$$V(f_{rs}) = V_{F1} = \frac{1}{4} [D_{(1)} + D_{(2)} - F_{(1)} - F_{(2)} + H_{1(1)} + H_{1(2)} - H_2],$$

$$H_2 = E(H_{2r}) = E(H_{2s}) = \sum_{a=1}^{k} h_a^2 (1 - w_a^2) (1 - w_a'^2) =$$

$$= 16 \sum_{a=1}^{k} u_a v_a u'_a v'_a h_a^2$$

Analysis of the magnitude and significance of the additive components $D_{(1)}$ and $D_{(2)}$ allows establishment of several inferences about genetic variability of each parental group. If the additive component of a group is nil, the parents have the same genotype ($w_a^2 = 1$ or $w_a'^2 = 1$, for every a) and, therefore, there is no genetic variability. When nonfixed genes are involved, the parents have distinct genotypes and the additive component is different from zero. Its value is largest when alleles are equally frequent.

The following differences are also useful:

$$D_{(1)} - D_{(2)} = \sum_{a=1}^{k} d_a^2 (w_a'^2 - w_a^2) \text{ and } 2D_{(3)} - D_{(1)} - D_{(2)} =$$

$$= \sum_{a=1}^{k} d_a^2 (w_a - w'_a)^2$$

If the differences are nil, the genes are equally frequent in the two parental groups. If not, it can be concluded that the genes do not have the same frequency in the two groups. When $D_{(1)} - D_{(2)} < 0$, the n' parental group is more variable. If genetic variation is greater in the n parental group, then $D_{(1)} - D_{(2)} > 0$.

The F component of a parent can be negative, nil or positive. When negative, it indicates that the parent has more recessive than dominant genes ($h_a\theta_a < 0$). If positive, the parent has more dominant genes ($h_a\theta_a > 0$). If there is no dominance, F values of the parents are nil. When dominance exists, $F = 0$ indicates that the parent carries approximately the same number of dominant and recessive genes. For any parental group, the F value of one parent is directly proportional to the number of dominant genes it carries that are not fixed in the other group of parents. Therefore, the F value of a parent is largest when it carries all the dominant genes that are not fixed in the other group. The average F value of a parental group indicates the frequencies of dominant and recessive allelic genes. When it is positive, it indicates that dominant genes are more frequent than recessive alleles in that group ($h_a w_a > 0$ or $h_a w_a' > 0$). If recessive genes are more frequent in the group, its value is negative ($h_a w_a < 0$ or $h_a w_a' < 0$). If there is dominance and the average F value is nil, the alleles are equally frequent.

Given the presence of genetic variability among parents in a group, the H_1 component of the group will be nil in the absence of dominance and positive in the presence of dominance. The difference $H_{1(1)} - H_{1(2)}$ provides the same information as $D_{(1)} - D_{(2)}$. If there is genetic variability in the two parental groups, components H_{2r}, H_{2s} and H_2 will be positive in the presence of dominance and nil in its absence. Absence of variability in a parental group makes components F and H_2 of parents from the other group, and their respective average values, to be nil. Dominance components $H_{1(1)}$, $H_{1(2)}$ and H_2 have the same magnitude when allelic genes have the same frequency in the two parent groups ($u_a = u_a' = 1/2$, for every a).

In the presence of dominance, the relative magnitude of H_2 components of parents in a group supplies important information. The H_2 component of a parent is largest when it carries the least frequent genes of the group to which it belongs ($w_a\theta_{ra}$ or $w_a' \theta_{sa} < 0$, for every a), and smallest when it carries the most frequent genes ($w_a\theta_{ra}$ or $w_a'\theta_{sa} > 0$, for every a). Generally, the value of the H_2 component of a parent is inversely proportional to the concentration of the most frequent genes of its group, that are nonfixed in the other group. The difference between H_2 components of two parents from the same group can be nil, if the parents have the same genotype or if the most frequent genes in the parents are different. If all H_2 components of parents in a group have the same magnitude, allelic genes are equally frequent in the group.

Genetic parameters

Average values of uv and u'v'

If $u_a = u$, $H_2/4H_{1(2)} = uv$. If $u_a' = u'$, then $H_2/4H_{1(1)} = u'v'$. If $u_a = u_a' = u$, the two ratios are equal to uv. Thus $H_2/4H_{1(2)}$ and $H_2/4H_{1(1)}$ express the average value of the allelic frequency products in groups with n and n' parents, respectively.

Average degree of dominance (h/d)

If $h_a = h$ and $d_a = d$, then $\sqrt{H_{1(1)}/D_{(1)}}$ and $\sqrt{H_{1(2)}/D_{(2)}}$ are equal to h/d. Therefore, $\sqrt{H_{1(1)}/D_{(1)}}$ and $\sqrt{H_{1(2)}/D_{(2)}}$ express the average degree of dominance in the polygenic systems defined by the n and n' parental groups, respectively.

Proportion between dominant and recessive genes in parents of the same group

Following Hayman's approach, it can be demonstrated that the ratios

$$\frac{\sqrt{4D_{(2)}\,H_{1(2)}} + F_{(1)}}{\sqrt{4D_{(2)}\,H_{1(2)}} - F_{(1)}} \quad \text{and} \quad \frac{\sqrt{4D_{(1)}\,H_{1(1)}} + F_{(2)}}{\sqrt{4D_{(1)}\,H_{1(1)}} - F_{(2)}}$$

reflect the average value of the proportion between the number of dominant and recessive genes in the n and n' parental groups, respectively. A ratio close to zero indicates that parents in the group have few dominant genes and many recessive ones. A ratio near one indicates that parents in the group have the same number of dominant and recessive genes. A ratio larger than one indicates that parents in the group have many dominant genes and few recessive ones.

Direction of dominance

Defining k_+ and k_- as the number of dominant genes which increase and decrease trait expression, respectively, average heterosis can be expressed in the following way:

$$h = \sum_{a=1}^{k_+} h_a (1 - w_a w_a') - \sum_{a=1}^{k_-} h_a (1 - w_a w_a'), \quad k_+ + k_- \leq k,$$

If $h_a = h$, $d_a = d$, $w_a = w$ and $w_a' = w'$, then:

$$h^2 D_{(1)}D_{(2)} / H_2D_{(3)}^2 = \frac{(k_+ - k_-)^2}{k}$$

If $h_a = h$, $d_a = d$ and $w_a = w_a' = w$, then $h^2/H_2 = (k_+ - k_-)^2/k$. Therefore, if dominance exists in the polygenic system under study, but the above ratio is nil, the number of

dominant genes with a positive effect is the same as the number of dominant genes with a negative effect. In other words, dominance is bi-directional. If the ratio is positive, the numbers of dominant genes which increase and decrease trait expression are unequal, that is, dominance is predominantly unidirectional.

The relationship between variance and covariance in the arrays

Considering the additive-dominant model the differences $W_r - V_r = (1/4)(D_{(2)} - H_{1(2)})$ and $W_s - V_s = (1/4)(D_{(1)} - H_{1(1)})$ are constant for every r and s. Thus, the coefficients of the regressions of W_r as a function of V_r ($W_r = \beta_0 + \beta_1 V_r$) and W_s as a function of V_s ($W_s = \beta_0 + \beta_1 V_s$) are equal to 1. Intercepts of the regressions of W_r on V_r as well as W_s on V_s are, respectively:

$$\beta_0 = E(W_r - V_r) = \frac{1}{4}(D_{(2)} - H_{1(2)})$$

$$\beta_0 = E(W_s - V_s) = \frac{1}{4}(D_{(1)} - H_{1(1)})$$

Similarly to Hayman (1954), it can be shown that (W_r, V_r) and (W_s, V_s) values occur on the straight line delimited by the parabola $W_r^2 = V_{0L0(2)} V_r$ and $W_s^2 = V_{0L0(1)} V_s$, respectively. If analysis of variance of the differences $W_t - V_t$ indicates that $W_r - V_r$ and $W_s - V_s$ are constants and/or, if the regression analyses of W_r on V_r and W_s on V_s show that $\beta_1 = 1$, the additive-dominant model adequately describes the observed results. When, $\beta_1 \neq 1$, the additive-dominant model is inadequate. Hayman (1954, 1958) and Mather and Jinks (1974) presented detailed discussions about the causes of inadequacy of the additive-dominant model, including the presence of extranuclear gene effects, non-allelic interaction, multiple allelism, correlation in the distribution of non-allelic genes and heterozygosis in the parents. Hayman (1954, 1958) also discussed how to proceed with diallel analysis when one or more of the additive-dominant genetic model restrictions are not fulfilled.

Absence of a relationship between W_r and V_r and between W_s and V_s ($\beta_1 = 0$) indicates the absence of dominance in the polygenic system under study, since, when $h_a = 0$, $W_r = (1/2)D_{(2)}$, $V_r = (1/4)D_{(2)}$, $W_s = (1/2)D_{(1)}$ and $V_s = (1/4)D_{(1)}$. Thus, when W is plotted against V for each group, all points occur at the $((1/2)D, (1/4)D)$ position. When the additive-dominant model is adequate, the constant of the regression of W_r on V_r indicates the average degree of dominance in the polygenic system defined by the n' parental group. The intercept of the regression of W_s on V_s expresses the average degree of dominance in the polygenic system defined by the n parental group. If $\beta_0 < 0$, there is evidence of overdominance. When $\beta_0 = 0$, there is complete dominance. Partial dominance exists when $\beta_0 > 0$. If the genes are equally frequent in the two parental groups, then:

$$D_{(1)} = D_{(2)} = D_{(3)} = D$$

$$F_{(1)} = F_{(2)} = F$$

$$H_{1(1)} = H_{1(2)} = H_1$$

$$W_r - V_r = W_s - V_s = W_t - V_t = \frac{1}{4}(D - H_1)$$

Thus, the coefficient of the regression of W_t on V_t ($W_t = \beta_0 + \beta_1 V_t$) is equal to one, and the intercept is $(1/4)(D - H_1)$, if the additive-dominant genetic model is adequate. The straight line is delimited by the parabola $W_t^2 = DV_t$.

The relationship between the sum of the variance and covariance in the arrays and the genotypic value of the common parent

It has already been demonstrated that variance and covariance in an array are smallest when the common parent is a carrier of dominant genes and largest when it is a carrier of recessive genes. Thus, the relationship between the genotypic value of the common parent and the sum between the variance and covariance in the parent array ($p_r = \beta_0 + \beta_1(W_r + V_r)$ or $p_s = \beta_0 + \beta_1(W_s + V_s)$) defines the presence and direction of dominance. The regression coefficient of p_r on $W_r + V_r$ is:

$$\beta_1 = \frac{- \sum_{a=1}^{k} d_a^2 h_a (1 - w_a^2)(1 - w_a'^2)}{\sum_{a=1}^{k} d_a^2 h_a^2 (1 - w_a^2)(1 - w_a'^2)^2}$$

The coefficient of the regression of p_s as a function of $W_s + V_s$ is:

$$\beta_1 = \frac{- \sum_{a=1}^{k} d_a^2 h_a (1 - w_a^2)(1 - w_a'^2)}{\sum_{a=1}^{k} d_a^2 h_a^2 (1 - w_a^2)^2 (1 - w_a'^2)}$$

The existence of a relationship between the genotypic value of the common parent and the sum between variance and covariance in its array indicates the presence of unidirectional dominance. If the regression coefficient is positive, deviations due to dominance are predominantly negative, contributing to a decrease in trait expression. If the coefficient is negative, the deviations of dominance contribute, exclusively or mostly, to an increase in trait expression. If there is no dominance, the graphic points of p_r over $W_r + V_r$ and p_s over $W_s + V_s$ are $(p_r, (3/4)D_{(2)})$ and $(P_s, (3/4)D_{(1)})$, respectively. When there is dominance, but $\beta_1 = 0$, dominance is bi-directional. Points are randomly distributed on the p over W + V graph.

Nongenetic components of variation

Let y_t be the average phenotypic value of a parent and y_{rs} the average phenotypic value of the hybrid between the rth and sth parents. Then,

$$y_t = p_t + e_t \ (e_t \sim (0, E), \text{ independents})$$

$$y_{rs} = f_{rs} + e_{rs} \ (e_{rs} \sim (0, E'), \text{ independents})$$

p_t and f_{rs} are genotypic values, and e_t and e_{rs} are non-genetic effects. Letting e_t and e_{rs} be independent of the genotypic values as well as between themselves, then:

$$V(y_r) = V(p_r + e_r) = V_{0L0(1)} + E$$

$$V(y_s) = V_{0L0(2)} + E$$

$$V(y_t) = V_{0L0} + E$$

$$\underset{r \text{ fixed}}{\text{Cov}(y_{rs}, y_s)} = \text{Cov}(f_{rs} + e_{rs}, p_s + e_s) = W_r$$

$$\underset{s \text{ fixed}}{\text{Cov}(y_{rs}, y_r)} = W_s$$

$$\underset{r \text{ fixed}}{V(y_{rs})} = V_r + E'$$

$$\underset{s \text{ fixed}}{V(y_{rs})} = V_s + E'$$

$$\underset{r \text{ fixed}}{\text{Cov}(y_{rs}, \bar{y}_{.s})} = \text{Cov}[f_{rs} + e_{rs}, \frac{1}{n} \sum_{r=1}^{n} (F_{rs} + e_{rs})] = W_{01(r)L1} + E'/n$$

$$\underset{s \text{ fixed}}{\text{Cov}(y_{rs}, \bar{y}_{r.})} = W_{01(s)L1} + E'/n'$$

And also,

$$E\{[\bar{y}_{..} - \frac{1}{2} (\bar{y}_{.} + \bar{y}'_{.})]^2\} = [m_{L1} - \frac{1}{2} (m_{L0} + m'_{L0})]^2 +$$

$$+ E\{[\bar{e}_{..} - \frac{1}{2} (\bar{e}_{.} + \bar{e}'_{.})]^2\} = (1/4)h^2 + \frac{1}{nn'}[(N/4)E + E'],$$

with $\bar{y}_{..} = \sum_{r=1}^{n} \sum_{s=1}^{n'} y_{rs} / nn'$, $\bar{y}_{.} = \sum_{r=1}^{n} y_r / n$ and $\bar{y}'_{.} = \sum_{s=1}^{n'} y_s / n'$.

If the phenotypic values of the parents and their hybrids in the diallel table are averages of b replications, the error mean square of the analysis of variance considering N parents, divided by b, is an estimator of E. The error mean square of the analysis of variance of the nn' hybrids, divided by b, is an estimator of E'. If the parent and hybrid means have the same precision, then $\hat{E} = \hat{E}'$, and both are equal to the error mean square of the analysis of variance of parents and F_1 hybrids, divided by b.

Estimation of genetic and nongenetic components of variation

This is carried out based on fitting the linear model $Y = X\beta + \varepsilon$ ($\varepsilon \sim (\Phi, \text{Cov}(\varepsilon))$), where:

$$Y' = [\hat{V}_{0L0(1)} \ \hat{V}_{0L0(2)} \ \hat{V}_{0L0} \qquad \hat{W}_{01(r=1)L01} \ \dots \ \hat{W}_{01(r=n)L01}$$

$$\hat{V}_{1(r=1)L1} \dots \hat{V}_{1(r=n)L1} \qquad \hat{W}_{01(s=1)L01} \dots \hat{W}_{01(s=n')L01}$$

$$\hat{V}_{1(s=1)L1} \dots \hat{V}_{1(s=n')L1} \qquad \hat{W}_{01(r=1)L1} \dots \hat{W}_{01(r=n)L1}$$

$$\hat{W}_{01(s=1)L1} \dots \hat{W}_{01(s=n')L1} \quad [\hat{m}_{L1} - \frac{1}{2}(\hat{m}_{L0} + \hat{m}'_{L0})]^2 \ \hat{E} \ \hat{E}']$$

$$W_{01(r)L1} = (1/4)D_{(2)} - [(n+1)/8n]F_r -$$

$$- (1/8n) \sum_{\substack{r'=1 \\ r' \neq r}}^{n} F_{r'} + (1/4)H_{1(2)} - (1/4)H_{2r} + E'/n$$

$$W_{01(s)L1} = (1/4)D_{(1)} - [(n'+1)/8n']F_s -$$

$$- (1/8n') \sum_{\substack{s'=1 \\ s' \neq s}}^{n'} F_{s'} + (1/4)H_{1(1)} - (1/4)H_{2s} + E'/n'$$

$$\beta' = [D_{(1)} \ D_{(2)} \ D_{(3)} \quad F_{r=1} \dots F_{r=n} \quad F_{s=1} \dots F_{s=n'} \quad H_{1(1)} \ H_{1(2)}$$

$$H_{2r=1} \dots H_{2r=n} \quad H_{2s=1} \dots H_{2s=n'} \quad h^2 \ E \ E']$$

For estimation purposes the ordinary ($\text{Cov}(\varepsilon) = \sigma^2 I$), weighted or generalized least squares methods can be used (Hayman, 1954; Mather and Jinks, 1974). Alternatively, the maximum likelihood method (Hayman, 1960) can also be used.

APPLICATION

In Table I is presented the grain yield per plant, in grams, of nine common bean lines (*Phaseolus vulgaris* L.) and 18 F_1 hybrids. Six parent lines belong to one group (group 1) and three to the other (group 2). The former group includes cultivars with good performance in low temperature conditions. Parents and their hybrids were assessed during fall-winter of 1992 at the Federal University of Viçosa, in Viçosa, State of Minas Gerais, Brazil, using a completely randomized design, with four replications. The results of the regression analyses of W_r on V_r (r = 1, 2, ..., 6) and W_s on V_s (s = 1, 2 and 3) are in Table II. The existence of a relationship between covariance and variance in the arrays shows that the yield per plant depends on non-additive gene effects. As the regression coefficients are statistically equal to one, the additive-dominant model is suitable to describe the results. Tests of the hypothesis H_0: $\beta_0 = 0$ indicate that, in the polygenic systems defined by the two parent groups, dominance between nonfixed allelic genes is, on average, complete.

Table I - Grain yield per plant, in grams, of nine common bean lines and their hybrids.

Parent		Parent		
		BAT-304 (1)	FT-84-835 (2)	Batatinha (3)
		6.10	4.48	9.54
Ricopardo 896 (1)	9.26	10.11	7.46	16.57
Ouro Negro (2)	8.23	9.48	11.67	15.89
Antioquia 8 (3)	14.77	17.96	18.45	15.29
DOR. 241 (4)	13.55	7.79	9.58	15.62
RAB 94 (5)	8.41	12.16	9.85	9.46
Ouro (6)	5.87	7.82	11.23	16.39

Table II - Summary of the regression analyses of W_r on V_r and of W_s on V_s, regression coefficient estimates and significance level of the test of the hypothesis $H_0: \beta_1 = 1^a$, in relation to grain yield of common bean plants, in grams.

Source of variation	d.f.	Mean square	
		Group 1	Group 2
Intercept	1	25.26[+]	6.51[+]
Regression	1	189.51[**]	25.34[*]
Error[b]		4.42	0.06
Coefficient		0.74[+]	0.86[++]

[a]Using the *t* statistic, with four and one degrees of freedom (d.f.) for the error in the analyses considering groups 1 and 2, respectively. [b]With four and one degrees of freedom in the analyses for groups 1 and 2, respectively. [**]$P < 0.01$; [*]$0.01 < P < 0.05$; [+]$0.05 < P < 0.10$; [++]$P > 0.10$.

Estimation of genetic and nongenetic components of variation was carried out considering two nongenetic components and the ordinary least squares method. The error mean squares of the analyses of variance considering only parents and only F_1 hybrids were 8.79 and 19.20, respectively. The estimates are presented in Table III. Diallel analysis indicated the presence of variability in the two parent groups, with type I error probability of approximately 0.056 for group 2. Since $D_{(1)} - D_{(2)}$ is positive (P = 0.09815), the genes that determine yield per plant are not equally frequent in the two parent groups. Genetic variability is greater in group 1. Estimates of the mean values of the allelic gene frequency products, in the two groups, favor the previous inference. In group 1, there is evidence that allelic gene frequencies are close to 1/2 (uv = 0.24). In group 2, allelic genes have unequal frequencies (u'v' = 0.17). Considering that $\hat{m}_{L0} - \hat{m}'_{L0} = 3.32$ ($0.01 < P < 0.05$), it can be concluded that yield-increasing genes have a frequency less than 1/2 in the second parental group. In short, the frequency of genes which increase trait expression is approximately 1/2 in group 1 and less in group 2, which explains the greater genetic variability found in the former.

Parent F values from one group allow parents to be ranked according to the number of dominant genes they carry, not fixed in the other parental set. Ranking group 1

Table III - Estimates of the genetic and non-genetic components of variation and of genetic parameters, in relation to grain yield of common bean plants, in grams.

Parameter[a]	Estimate	Standard deviation	Probability[b]
$D_{(1)}$	9.53	2.47	0.00310
$D_{(2)}$	4.48	2.47	0.05630
$D_{(3)}$	10.82	4.66	0.02665
$F_{r=1}$	-45.06	7.27	0.00040
$F_{r=2}$	-11.39	7.27	0.16150
$F_{r=3}$	25.97	7.27	0.00910
$F_{r=4}$	-28.75	7.27	0.00550
$F_{r=5}$	21.37	7.27	0.02170
$F_{r=6}$	-30.87	7.27	0.00380
$F_{s=1}$	-10.56	7.54	0.20420
$F_{s=2}$	-8.11	7.54	0.31720
$F_{s=3}$	17.31	7.54	0.05550
$H_{1(1)}$	18.94	9.35	0.04125
$H_{1(2)}$	13.56	8.43	0.07615
$H_{2r=1}$	11.95	9.79	0.13085
$H_{2r=2}$	5.10	9.79	0.30970
$H_{2r=3}$	28.27	9.79	0.01170
$H_{2r=4}$	5.89	9.79	0.28370
$H_{2r=5}$	25.09	9.79	0.01870
$H_{2r=6}$	6.49	9.79	0.26420
$H_{2s=1}$	9.34	9.53	0.17990
$H_{2s=2}$	4.23	9.53	0.33410
$H_{2s=3}$	19.55	9.53	0.03965
h^2	62.41	7.05	0.00005
E	2.20	1.75	0.12455
E'	4.80	1.75	0.01425
$F_{(1)}$	-11.45	5.70	0.08470
$F_{(2)}$	-0.46	6.37	0.94560
H_2	12.88	6.35	0.04110
uv	0.24	-	-
u'v'	0.17	-	-
h/d(1)	1.41	-	-
h/d(2)	1.74	-	-
$k_D/k_R(1)$	0.15	-	-
$k_D/k_R(2)$	0.97	-	-
$(k_+ - k_-)^2/k$	1.77	-	-
R^2	0.99	-	-

[a]uv: Average value of the allelic frequency products in group 1. u'v': Average value of the allelic frequency products in group 2. h/d(1): Average degree of dominance in group 1. h/d(2): Average degree of dominance in group 2. k_D/k_R (1): Proportion between dominant and recessive genes in group 1. k_D/k_R (2): Proportion between dominant and recessive genes in group 2. $(k_+ - k_-)^2/k$: Indicator of the direction of dominance. R^2: Determination coefficient. [b]Based on the *t*-test with seven degrees of freedom.

parents, according to an increasing concentration of dominant genes not fixed in group 2, the following order is obtained: Ricopardo 896, Ouro, DOR. 241, Ouro Negro, RAB 94 and Antioquia 8. Ricopardo 896, Ouro and DOR. 241 should have more recessive than dominant genes. Ouro Negro should have, approximately, similar numbers of dominant and recessive genes. RAB 94 and Antioquia 8 have more dominant than recessive genes. In group 2, the variety Batatinha carries the largest number of dominant genes, not fixed in group 1. Both BAT-304 and FT-84-835 should have dominant and recessive genes in approximately equal numbers. Data indicate that recessive genes

generally act to decrease plant yield, while dominant ones increase yield. This fact, established in the analysis of variance (not presented), seems to be corroborated by the estimate of $h^2D_{(1)}D_{(2)}/H_2D_{(3)}^2$, which indicates unidirectional dominance, despite its small magnitude. The regression analyses of p_r on $(W_r + V_r)$ and p_s on $(W_s + V_s)$, however, do not confirm the above information. The results of these regression analyses, using the observed parent means as the genotypic values, are in Table IV.

Estimates of the regression coefficients show the presence of statistically nonsignificant positive unidirectional dominance. This seems to indicate that many nonfixed dominant genes act to decrease yield, although the majority of them increase it. Estimates of the average degree of dominance are larger than one, although previous results indicate complete dominance. It is reasonable, therefore, to admit that they are statistically equal to one. As there is evidence of complete dominance, estimates of the proportion among dominant and recessive genes suggest that the parents in group 1 have, on average, more recessive than dominant genes, while in group 2 the parents have, on average, the same number of recessive and dominant genes. These results agree with those from the analyses of the parent F values, as the majority of the parents in group 1 have more recessive than dominant genes, while in group 2 the majority have equal numbers of dominant and recessive genes.

The estimated F value for group 1 indicates that recessive genes, not fixed in group 2, are more frequent than dominant alleles. Taking the previous results into consideration, it can be concluded that this superiority is small. The estimate of $F_{(2)}$ suggests that dominant and recessive alleles, not fixed in group 1, are equally frequent in group 2, which is inconsistent with previous results. The presence of bi-directional dominance and fixed genes in the two parent groups may cause conflicting results in the genetic analysis. Many dominance component estimates show, as expected, the presence of dominance in the polygenic system under study. H_2 values of RAB 94, Antioquia 8 and Batatinha show that these genotypes are rare in their respective groups. They carry the least frequent genes in

their respective groups which are not fixed in the other group of parents. As already seen, they have more dominant genes than other parents in their group.

The difference between H_2 values of RAB 94 and Antioquia 8 is nil (P = 0.30045), indicating that they have similar genotypes, in regard to yield determining genes. The cross between them should not result in relevant heterosis, although it may be carried out to combine its nonfixed desirable genes in a single variety. As they have many dominant genes, which are rare in their group and not fixed in group 2, both can be used in crosses with the Batatinha variety, of group 2, which also has many dominant genes, rare in its group and not fixed in group 1. Estimates of specific heteroses show that RAB 94 and Antioquia 8 genotypes are not completely distinct from that of the variety Batatinha (not presented). The small magnitude of the two specific heteroses, nil from a statistical point of view, is not surprising, since these varieties have more dominant than recessive genes. Values of the varietal heteroses indicate that RAB 94 has more frequent genes of group 2 than Antioquia 8, while Batatinha has less frequent genes in group 1 than FT-84-835 and BAT-304 (not presented).

Use of RAB 94 and Batatinha and/or Antioquia 8 and Batatinha as parents is justifiable if the objective of the program is to obtain pure lines. This objective would combine, in a single line, the favorable dominant genes which are not fixed in the parents. The smaller variability expected in the segregant generations derived from crosses between Antioquia 8 and RAB 94 and, preferably, between Antioquia 8 and RAB 94 with Batatinha could favor selection of a line with superior gene combination, containing the favorable dominant and recessive nonfixed genes from both parents. Crosses of RAB 94, Antioquia 8 and Batatinha, with parents with many recessive genes, even in the same group, should show high levels of heterosis, due to their genotypic differences. If the objective is production of hybrids, several crosses, such as Antioquia 8 with FT-84-835 or BAT-304 and Batatinha with Ouro or Ricopardo 896 or Ouro Negro, may be considered. These crosses could also be considered in programs for pure lines, as there are various recessive genes that increase yield. Due to the accentuated genetic variability expected in the segregant generations of these crosses, it is necessary to use large number of plants and/or families to increase the probability of selecting a variety with a superior combination of genes.

Table IV - Summary, for each group of common bean lines, of the regression analyses of the parent genotypic value on the sum of the variance and covariance in its array, in relation to grain yield per plant, in grams.

		Mean square	
Source of variation	d.f.	Group 1	Group 2
Regression	1	6.73[++]	11.35[++]
Error[a]		12.97	2.00
Coefficient		-0.10	-0.31

[a]With four and one degrees of freedom (d.f.) in the analyses of group 1 and 2, respectively. [++]P > 0.10.

RESUMO

Neste trabalho são apresentadas teoria e análise de dialelos parciais, com base nos métodos propostos por Hayman em 1954 e 1958. A análise dialélica com os dados dos pais e de seus híbridos F_1 possibilita uma caracterização detalhada dos sistemas poligênicos sob estudo e a escolha de progenitores para hibridação. A partir da estimativa de componentes genéticos e não genéticos da variação e de parâmetros genéticos, a análise

dialélica permite avaliar: a variabilidade genética em cada grupo; as diferenças genotípicas entre pais de grupos distintos; se um genitor tem genótipo comum ou raro no grupo ao qual não pertence; se há dominância; a direção de dominância; o grau médio de dominância em cada grupo; a importância relativa dos efeitos médios dos genes e dos desvios de dominância em determinar a característica; as freqüências de genes alélicos em cada grupo; se os genes são igualmente freqüentes nos dois grupos; o grupo em que os genes favoráveis estão em maior freqüência; o grupo em que os genes dominantes são mais freqüentes; o número relativo de genes dominantes e recessivos em cada pai; se um genitor tem genótipo comum ou raro em seu grupo, e as diferenças genotípicas entre os pais de um mesmo grupo.

REFERENCES

Chung, J.H. and **Stevenson, E.** (1973). Diallel analysis of the genetic variation in some quantitative traits in dry beans. *New Zealand J. Agric. Res. 16*: 223-231.

Comstock, R.E. and **Robinson, H.F.** (1948). The components of genetic variance in populations. *Biometrics 4*: 254-266.

Comstock, R.E. and **Robinson, H.F.** (1952). Estimation of average dominance of genes. In: *Heterosis* (Gowen, J.W., ed.). Iowa State College Press, Ames, Iowa, pp. 494-516.

Coughtrey, A. and **Mather, K.** (1970). Interaction and gene association and dispersion in diallel crosses where gene frequencies are unequal. *Heredity 25*: 79-88.

Dickson, M.H. (1967). Diallel of seven economic characters in snap beans. *Crop Sci. 7*: 121-124.

Gardner, C.O. and **Eberhart, S.A.** (1966). Analysis and interpretation of the variety cross diallel and related populations. *Biometrics 22*: 439-452.

Gilbert, N.E.G. (1958). Diallel cross in plant breeding. *Heredity 12*: 477-492.

Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel crossing system. *Austr. J. Biol. Sci. 9*: 463-493.

Hayman, B.I. (1954). The theory and analysis of diallel crosses. *Genetics 39*: 789-809.

Hayman, B.I. (1958). The theory and analysis of diallel crosses. II. *Genetics 43*: 63-85.

Hayman, B.I. (1960). Maximum likelihood estimation of genetic components of variation. *Biometrics 16*: 349-381.

Jinks, J.L. (1954). The analysis of continuous variation in a diallel cross of *Nicotiana rustica* varieties. *Genetics 39*: 767-788.

Jinks, J.L. and **Hayman, B.I.** (1953). The analysis of diallel crosses. *Maize Genet. Coop. NewsL. 27*: 48-54.

Kempthorne, O. and **Curnow, R.N.** (1961). The partial diallel cross. *Biometrics 17*: 229-250.

Kornegay, J.L. and **Temple, S.R.** (1986). Inheritance and combining ability of leafhopper defense mechanism in common bean. *Crop Sci. 26*: 1153-1158.

Mather, K. and **Jinks, J.L.** (1974). *Biometrical Genetics*. 2nd edn. Cornell University Press, Ithaca, New York.

Nassar, R.F. (1965). Effect of correlated gene distribution due to sampling on the diallel analysis. *Genetics 52*: 9-20.

Nishimura, M. and **Hamamura, K.** (1993). Diallel analysis of cool tolerance at the booting stage in rice varieties from Hokkaido. *Jpn. J. Breed. 43*: 557-566.

Sokol, M.J. and **Baker, R.J.** (1977). Evaluation of the assumptions required for the genetic interpretation of diallel experiments in self-pollinating crops. *Can. J. Plant Sci. 57*: 1185-1191.