Theory and analysis of partial diallel crosses. Parents and F_2 generations

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

*[1]Departamento de Biologia Geral, Universidade Federal de Viçosa, 36571-000, Viçosa, Minas Gerais, Brasil. [2]Departamento de Fitotecnia, Universidade Federal de Viçosa, 36571-000, Viçosa, Minas Gerais, Brasil. *Author for correspondence.*

ABSTRACT. This paper presents theory and analysis of partial diallels, using data from the parents and F_2 generations obtained from crossings among two groups of parents. Theory and analysis are based on Hayman's 1954 and 1958 suggestions. The diallel analysis allows the characterization of the polygenic systems under analysis through the estimation of genetic and non-genetic components of variation, in relation to each group of parents or to each parent, depending on the component and on the genetic parameter. The genetic parameters are the mean products of the allelic gene frequencies, mean degrees of dominance, mean proportions between dominant and recessive genes in the parents, direction of dominance and heritability. A diallel analysis for the common bean is included.

Key words: diallel analysis, partial diallel, biometrical genetics.

RESUMO. Teoria e análise de dialelos parciais. Pais e gerações F_2. Este trabalho apresenta teoria e análise de dialelos parciais, usando dados dos pais e das gerações F_2. A metodologia está fundamentada nos métodos propostos por Hayman, em 1954 e 1958. A análise dialélica possibilita a caracterização dos sistemas poligênicos em estudo, por meio da estimação de componentes genéticos e não genéticos de variação e de parâmetros genéticos, em relação a cada grupo de pais ou a cada pai, dependendo do componente de variação ou parâmetro de natureza genética. Os parâmetros genéticos são: médias dos produtos das freqüências de alelos, graus médios de dominância, proporções médias de genes dominantes e recessivos nos pais, direção de dominância e herdabilidade. É apresentada uma análise dialélica com feijoeiro.

Palavras-chaves: análise dialélica, dialelo parcial, genética biométrica.

The main aim of the diallel analysis is the study of the genetic control of quantitative traits, which is essential for planning and carrying out breeding programs. Several methods for analysis were proposed in which the parents may be pure lines or open pollinated varieties, among others (Jinks and Hayman, 1953; Hayman, 1954; Dickinson and Jinks, 1956; Griffing, 1956; Gardner and Eberhart, 1966). Information about general and specific combining ability, heterosis, reciprocal effects, maternal and paternal effects may also be obtained (Park and Davis, 1976; Cockerham and Weir, 1977; Foolad and Bassiri, 1983; Nienhuis and Singh, 1986; Yanchuk, 1996). Hayman (1958) presented the theory and analysis of diallel crosses based on the genotypic and phenotypic values of the parents, F_1 hybrids and F_2 generations. Although the diallel analysis using data from homozygous parents and their F_1 hybrids allows estimation of genetic parameters unbiased by linkage effects (Mather and Jinks, 1974) and the best assessment of dominance in the polygenic systems under study, the analysis using the information from the parents and the F_2 generations may be an excellent alternative when dealing with a species from which it is difficult to obtain a large number of F_1 hybrids.

The methodology developed by Hayman (1958) is not applicable to partial diallels which involve two groups of parents. This paper presents the theory and analysis of the partial diallel crosses, based on Hayman's proposals (1954, 1958), taking into consideration the data from the parents and their corresponding F_2 generations.

Theory

The theory presented here is a generalization of the methodology of Hayman (1954, 1958). If gene frequencies are the same for the two parent groups, results are equal to those presented by this author. Let **A** and **a** be the alleles of one of the k genes (locus a) that control a quantitative trait in a diploid species, in which **A** increases and **a** decreases the trait expression. Let N homozygous parents (N ≥ 6) be divided into two groups, one with n parents and the other with n' parents (n + n' = N, n and n' ≥ 3). The groups with N, n and n' parents define three polygenic systems to which the following assumptions are attached: a) Mendelian inheritance; b) no reciprocal effects; c) no interaction between non-allelic genes; d) only two allelic forms; and e) no correlation in the distribution of non-allelic genes in the parents. In the F_2 generation obtained from the cross between the rth (r = 1, ..., n) and sth (s = 1, ..., n') parents, the genotypic frequencies of AA, Aa and aa are:

$$P(AA)_{F_2 rs} = P(AA)_{rs} + \frac{1}{4}P(Aa)_{rs}$$

$$P(Aa)_{F_2 rs} = \frac{2}{4}P(Aa)_{rs}$$

$$P(aa)_{F_2 rs} = P(aa)_{rs} + \frac{1}{4}P(Aa)_{rs}$$

$P(AA)_{rs}$, $P(Aa)_{rs}$ and $P(aa)_{rs}$ are the probabilities of the hybrid from the rth and sth parents to be AA, Aa and aa, respectively, given by:

$$P(AA)_{rs} = \frac{1}{4}(1 + \theta_{ra} + \theta_{sa} + \theta_{ra}\theta_{sa})$$

$$P(Aa)_{rs} = \frac{2}{4}(1 - \theta_{ra}\theta_{sa})$$

$$P(aa)_{rs} = \frac{1}{4}(1 - \theta_{ra} - \theta_{sa} + \theta_{ra}\theta_{sa})$$

For all loci in the polygenic system under analysis the variable θ assumes the value −1 if the parent is homozygous for the gene that decreases trait expression, or 1 if it is homozygous in relation to the allele that increases trait expression. Thus, the genotypic mean of the F_2 population obtained from the cross between the rth and sth parents is:

$$g_{rs} = m + \frac{1}{2}\sum_{a=1}^{k}\left[d_a(\theta_{ra} + \theta_{sa}) + \frac{1}{2}h_a(1 - \theta_{ra}\theta_{sa})\right] = \frac{1}{2}(p_r + p_s) + \frac{1}{4}h_{rs}$$

where:

$$m = \sum_{a=1}^{k} m_a ;$$

$p_r = m + \sum_{a=1}^{k} d_a\theta_{ra}$ is the genotypic value of the rth parent;

$p_s = m + \sum_{a=1}^{k} d_a\theta_{sa}$ is the genotypic value of the sth parent, and

$h_{rs} = \sum_{a=1}^{k} h_a(1 - \theta_{ra}\theta_{sa})$ is their specific heterosis.

The parameter m_a is the mean of the genotypic values of the homozygotes, d_a is the difference between the genotypic value of the homozygote of largest expression and m_a, h_a is the difference between the genotypic value of the heterozygote and m_a. Means of the rth and sth arrays (a group of F_2 generations with a common initial parent) are:

$$g_r = \underset{r\ \text{fixed}}{E(g_{rs})} = \frac{1}{2}(p_r + m'_{L0}) + \frac{1}{4}h_r$$

$$g_s = \underset{s\ \text{fixed}}{E(g_{rs})} = \frac{1}{2}(p_s + m_{L0}) + \frac{1}{4}h_s$$

where:

$m_{L0} = E(p_r) = m + \sum_{a=1}^{k} d_a w_a$ is the genotypic mean of the n parents of a group;

$m'_{L0} = E(p_s) = m + \sum_{a=1}^{k} d_a w'_a$ is the genotypic mean of the n' parents of the other group;

$h_r = \underset{r\ \text{fixed}}{E(h_{rs})} = \sum_{a=1}^{k} h_a(1 - w'_a\theta_{ra})$ is the varietal heterosis of the rth parent, and

$h_s = \underset{s\ \text{fixed}}{E(h_{rs})} = \sum_{a=1}^{k} h_a(1 - w_a\theta_{sa})$ is the varietal heterosis of the sth parent.

The parameters $w_a = u_a - v_a$ and $w'_a = u'_a - v'_a$ are, respectively, the expected values of θ_{ra} and θ_{sa}. The parameters u_a and v_a and u'_a and v'_a are the frequencies of the alleles **A** and **a** in the groups with n and n' parents, respectively. The mean of the F_2 generations is:

$$E(g_{rs}) = m_{L2} = \frac{1}{2}(m_{L0} + m'_{L0}) + \frac{1}{4}h$$

where:

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

A discussion about the information provided by the specific, varietal and mean heteroses, and by the contrast $m_{L0} - m'_{L0}$, is presented by Viana *et al.* (1999).

Genetic components of variation and genetic parameters. The following second degree statistics can be used to estimate the genetic components of variation and, consequently, the genetic parameters: (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

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

where:

$$p_t = m + \sum_{a=1}^{k} d_a \theta_{ta} \quad (t = 1,\ldots,N),\ \overline{w}_a = pw_a + qw_a',\ p = n/N,\ q = n'/N$$

, and

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

(4) Covariance between genotypic mean of F_2 generation of the rth parent and the mean of the non-recurrent parent (covariance in the rth array)

$$W_{01(r)L02} = W_{r2} = \underset{r\ \text{fixed}}{\mathrm{Cov}(g_{rs}, p_s)} = \frac{1}{2}D(2) - \frac{1}{8}F_r,$$

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

(5) Variance of the genotypic means of the F_2 generations of the rth parent (variance in the rth array)

$$V_{1(r)L2} = V_{r2} = \underset{r\ \text{fixed}}{V(g_{rs})} = \frac{1}{4}D(2) - \frac{1}{8}F_r + \frac{1}{16}H_1(2),$$

$$H_1(2) = \sum_{a=1}^{k} h_a^2(1 - w_a'^2) = 4\sum_{a=1}^{k} u_a' v_a' h_a^2$$

(6) Covariance in the sth array

$$W_{01(s)L02} = W_{s2} = \underset{s\ \text{fixed}}{\mathrm{Cov}(g_{rs}, p_r)} = \frac{1}{2}D(1) - \frac{1}{8}F_s,$$

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

(7) Variance in the sth array

$$V_{1(s)L2} = V_{s2} = \underset{s\ \text{fixed}}{V(g_{rs})} = \frac{1}{4}D(1) - \frac{1}{8}F_s + \frac{1}{16}H_1(1),$$

$$H_1(1) = \sum_{a=1}^{k} h_a^2(1 - w_a^2) = 4\sum_{a=1}^{k} u_a v_a h_a^2$$

(8) Covariance between genotypic mean of F_2 generation of the rth parent and the mean of the non-recurrent parent array

$$W_{01(r)L2} = V(g_{rs}, g_s) = \frac{1}{4}D(2) - \frac{1}{16}F_r - \frac{1}{16}F(1) + \frac{1}{16}H_1(2) - \frac{1}{16}H_{2r},$$

$$F(1) = E(F_r) = 2\sum^{k} d_a h_a w_a(1 - w_a'^2) = 8\sum^{k} u_a' v_a'(u_a - v_a) d_a h_a$$

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

(9) Covariance between genotypic mean of F_2 generation of the sth parent and the mean of the non-recurrent parent array

$$W_{01(s)L2} = \underset{s\ \text{fixed}}{V(g_{rs}, g_r)} = \frac{1}{4}D(1) - \frac{1}{16}F_s - \frac{1}{16}F(2) + \frac{1}{16}H_1(1) - \frac{1}{16}H_{2s},$$

$$F(2) = E(F_s) = 2\sum_{a=1}^{k} d_a h_a w_a'(1 - w_a^2) = 8\sum_{a=1}^{k} u_a v_a (u_a' - v_a') d_a h_a$$

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

(10) Variance of the genotypic means of the F_2 generations

$$V(g_{rs}) = V_{F2} = \frac{1}{4}D(1) + \frac{1}{4}D(2) - \frac{1}{8}F(1) - \frac{1}{8}F(2) + \frac{1}{16}H_1(1) + \frac{1}{16}H_1(2) - \frac{1}{16}H_2,$$

$$E(H_2) = E(H_{2r}) = E(H_{2s}) = \sum_{a=1}^{k} h_a^2(1 - w_a^2)(1 - w_a'^2) = 16\sum_{a=1}^{k} u_a v_a u_a' v_a' h_a^2$$

(11) Variance of the genotypic values of individuals of the F_2 population obtained from the cross between the rth and sth parents

$$V_{2(rs)L2} = \sum_{a=1}^{k} \{P(AA)_{F_2 rs}(d_a)^2 + P(Aa)_{F_2 rs}(h_a)^2 + P(aa)_{F_2 rs}(-d_a)^2 -$$

$$[\frac{1}{2}d_a(\theta_{ra} + \theta_{sa}) + \frac{1}{4}h_a(1 - \theta_{ra}\theta_{sa})]^2\} = \frac{1}{4}D(rs) + \frac{1}{8}H_1(rs),$$

$$D(rs) = \sum_{a=1}^{k} d_a^2(1 - \theta_{ra}\theta_{sa})$$

$$H_1(rs) = \sum_{a=1}^{k} h_a^2(1 - \theta_{ra}\theta_{sa})$$

(12) Mean of the total genotypic variances of the F_2 generations of the rth parent

$$V_{2(r)L2} = \underset{r\ \text{fixed}}{E(V_{2(rs)L2})} = \frac{1}{4}D_r + \frac{1}{8}H_{1r},$$

$$D_r = \underset{r\ \text{fixed}}{E(D(rs))} = \sum_{a=1}^{k} d_a^2(1 - w_a' \theta_{ra})$$

$$H_{1r} = \underset{r\ \text{fixed}}{E(H_1(rs))} = \sum_{a=1}^{k} h_a^2(1 - w_a' \theta_{ra})$$

(13) Mean of the total genotypic variances of the F_2 generations of the sth parent

$$V_{2(s)L2} = \underset{s\ \text{fixed}}{E(V_{2(rs)L2})} = \frac{1}{4}D_s + \frac{1}{8}H_{1s},$$

$$D_s = \underset{s\ \text{fixed}}{E(D(rs))} = \sum_{a=1}^{k} d_a^2(1 - w_a \theta_{sa})$$

$$H_{1s} = \underset{s\ \text{fixed}}{E(H_1(rs))} = \sum_{a=1}^{k} h_a^2(1 - w_a \theta_{sa})$$

(14) Mean of the total genotypic variances of the F_2 generations of the parents of the diallel

$$V_{2L2} = E(V_{2(rs)L2}) = \frac{1}{4}D(3) + \frac{1}{8}H_1(3),$$

$$H_{1(3)} = E(H_{1(rs)}) = E(H_{1r}) = E(H_{1s}) = \sum_{a=1}^{k} h_a^2(1 - w_a w_a')$$

$$D_{(3)} = E(D_{(rs)}) = E(D_r) = E(D_s)$$

The sum $[(1/4)D_{(rs)} + (1/8)H_{1(rs)}]$ can be estimated, but the components $D_{(rs)}$ and $H_{1(rs)}$ cannot because there are not enough equations. Similarly, $[(1/4)D_r + (1/8)H_{1r}]$ and $[(1/4)D_s + (1/8)H_{1s}]$ can be estimated, but the components D_r, D_s, H_{1r} and H_{1s} cannot. In the absence of dominance the quadratic components $D_{(rs)}$, D_r and D_s are estimable and supply information which cannot be obtained from the linear components h_{rs}, h_r and h_s, because they are all zero. The additive component $D_{(rs)}$ is nil when the rth and sth parents have the same genotype ($\theta_{ra} = \theta_{sa}$, for every a). Its value is largest when the parents are carriers of distinct alleles for all the segregant genes in the polygenic system ($\theta_{ra}\theta_{sa} = -1$, for every a). The D component of a parent is minimum when it carries the most frequent genes in the group of parents to which it doesn't belong. Its value is largest when the parent is carrier of the less frequent genes in the group of parents to which it does not belong.

The mean of the products of the allele frequencies in the groups with n and n' parents are, respectively, $H_2/4H_{1(2)}$ and $H_2/4H_{1(1)}$. The average degree of dominance in the polygenic systems defined by the groups with n and n' parents are, respectively, $\sqrt{H_{1(1)}/D_{(1)}}$ and $\sqrt{H_{1(2)}/D_{(2)}}$. If the estimate of V_{2L2} is taken into account, the dominance component $H_{1(3)}$ may be estimated. This allows the estimation of a third average degree of dominance, which is valid for the polygenic system under study and not for a given group of parents, since, given that $h_a = h$ and $d_a = d$, $\sqrt{H_{1(3)}/D_{(3)}} = h/d$. The proportions between numbers of dominant and recessive genes in the groups with n and n' parents are, respectively:

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

The direction of dominance is $h^2 D_{(1)} D_{(2)} / H_2 D_{(3)}^2$. Detailed discussion about the genetic components of variation and genetic parameters are presented by Viana *et al.* (1999). When observations of the individuals in each F_2 generation are available, heritabilities in the broad (B) and narrow (R) senses can be defined. At the individual level the broad sense heritability in the F_2 generation obtained from the cross between the rth and sth parents is:

$$h_B^2 = \frac{V_{2(rs)L2}}{V_{2(rs)L2} + E_0}$$

The mean broad sense heritability, at the individual level, in the F_2 generations of the rth parent is:

$$h_B^2 = \frac{V_{2(r)L2}}{V_{2(r)L2} + E_0}$$

The mean broad sense heritability, at the individual level, in the F_2 generations of the sth parent is:

$$h_B^2 = \frac{V_{2(s)L2}}{V_{2(s)L2} + E_0}$$

The mean heritabilities, at the individual level, in the F_2 generations of the diallel are:

$$h_B^2 = \frac{V_{2L2}}{V_{2L2} + E_0}$$

$$h_R^2 = \frac{(1/4)D_{(3)}}{V_{2L2} + E_0}$$

E_0 is the non-genetic component of the variance of the phenotypic values of individuals in a single population. The broad sense heritability, at F_2 individual level, is the square of the correlation between the phenotypic and the genotypic values of same F_2 plant. The narrow sense heritability is the square of the correlation between the phenotypic and the additive genetic values of same F_2 plant (Falconer and MacKay, 1996). The later should be used to evaluate the efficiency of the mass selection.

The relationship between variance and covariance in the arrays. If the additive-dominant model is adequate, there is a functional relation between W_{r2} and V_{r2} and between W_{s2} and V_{s2} and the regression coefficients of W_{r2} on V_{r2} ($W_{r2} = \beta_0 + \beta_1 V_{r2}$) and of W_{s2} on V_{s2} ($W_{s2} = \beta_0 + \beta_1 V_{s2}$) are equal to unit, as the differences $W_{r2} - V_{r2} = (1/4)D_{(2)} - (1/16)H_{1(2)}$ and $W_{s2} - V_{s2} = (1/4)D_{(1)} - (1/16)H_{1(1)}$ are constants. The intercepts of the two regressions are, respectively,

$$\beta_0 = E(W_{r2} - V_{r2}) = \frac{1}{4}D_{(2)} - \frac{1}{16}H_{1(2)}$$

$$\beta_0 = E(W_{s2} - V_{s2}) = \frac{1}{4}D_{(1)} - \frac{1}{16}H_{1(1)},$$

which do not give an indication of the average degree of dominance.

The straight lines $W_{r2} = (1/4)[D_{(2)} - (1/4)H_{1(2)}] + V_{r2}$ and $W_{s2} = (1/4)[D_{(1)} - (1/4)H_{1(1)}] + V_{s2}$ are limited by the parabolas $W^2_{r2} = V_{0L0(2)}V_{r2}$ and $W^2_{s2} = V_{0L0(1)}V_{s2}$, respectively. The absence of a functional relation between variance and covariance in the parental arrays in each group, indicates that there is no dominance in the polygenic system under study (the coordinates of the points are $((1/2)D_{(2)}, (1/4)D_{(2)})$, for the arrays of the group with n parents, and $((1/2)D_{(1)}, (1/4)D_{(1)})$ for the arrays of the group with n' parents). If the regression coefficients are different from 1, the additive-dominant model is inadequate. If the regression coefficient of W_{t2} on V_{t2} ($W_{t2} = \beta_0 + \beta_1 V_{t2}$) is equal to unity, the genes are equally frequent in the two groups of parents. The regression has a coefficient equal to 1 because, given that $w_a = w'_a$, $D_{(1)} = D_{(2)} = D_{(3)} = D$, $F_{(1)} = F_{(2)} = F$ and $H_{1(1)} = H_{1(2)} = H_1$, where D, F and H_1 are parameters of the Hayman method. Then $W_{r2} - V_{r2} = W_{s2} - V_{s2} = W_{t2} - V_{t2} = (1/4)D - (1/16)H_1$. The straight line $W_{t2} = (1/4)[D - (1/4)H_1] + V_{t2}$ is limited by the parabola $W_{t2}^2 = DV_{t2}$.

The relationship between the sum of the variance and covariance in the arrays and the genotypic value of the common parent. The analysis of the regressions of p_r on $(W_{r2} + V_{r2})$ ($p_r = \beta_0 + \beta_1(W_{r2} + V_{r2})$) and of p_s on $(W_{s2} + V_{s2})$ ($p_s = \beta_0 + \beta_1(W_{s2} + V_{s2})$) supply information about the presence of dominance in the polygenic system under study and, also, whether dominance is predominantly uni or bi-directional. The regression coefficients are, respectively,

$$\beta_1 = \frac{-\frac{1}{2}\sum_{a=1}^{k} d_a^2 h_a (1 - w_a^2)(1 - w_a'^2)}{\frac{1}{4}\sum_{a=1}^{k} d_a^2 h_a^2 (1 - w_a^2)(1 - w_a'^2)^2}$$

$$\beta_1 = \frac{-\frac{1}{2}\sum_{a=1}^{k} d_a^2 h_a (1 - w_a^2)(1 - w_a'^2)}{\frac{1}{4}\sum_{a=1}^{k} d_a^2 h_a^2 (1 - w_a^2)^2 (1 - w_a'^2)}$$

When $\beta_1 > 0$ the deviations due to dominance are predominantly negative, contributing to a decrease in the trait expression. Positive unidirectional dominance makes $\beta_1 < 0$. If in each group of parents there is no functional relation between p and W + V ($\beta_1 = 0$) and the points on the graph of p versus (W + V) are randomly distributed, dominance is bi-directional. When there is no dominance in the polygenic systems, the points on the graphs of p_r over $(W_{r2} + V_{r2})$ and of p_s over $(W_{s2} + V_{s2})$ are $(p_r, (3/4)D_{(2)})$ and $(p_s, (3/4)D_{(1)})$, respectively.

Non-genetic components of variation. Let y_t and y_{rs} be, respectively, the mean phenotypic values of a parent and of the F$_2$ individuals obtained from the cross between the rth and sth parents. Thus,

$y_t = p_t + e_t$ ($e_t \sim N(0, E)$, independents)

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

p_t and g_{rs} are genotypic values and e_t and e_{rs} are non-genetic effects. Allowing e_t and e_{rs} be independent of the genotypic values and also between themselves, then:

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

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

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

$$\underset{r\ \text{fixed}}{\text{Cov}(y_{rs}, y_s)} = W_{r2}$$

$$\underset{s\ \text{fixed}}{\text{Cov}(y_{rs}, y_r)} = W_{s2}$$

$$\underset{r\ \text{fixed}}{V(y_{rs})} = V_{r2} + E''$$

$$\underset{s\ \text{fixed}}{V(y_{rs})} = V_{s2} + E''$$

$$\underset{r\ \text{fixed}}{\text{Cov}(y_{rs}, \bar{y}_{.s})} = W_{01(r)L2} + E''/n$$

$$\underset{s\ \text{fixed}}{\text{Cov}(y_{rs}, \bar{y}_{r.})} = W_{01(s)L2} + E''/n'$$

And,

$$E\{[\hat{m}_{L2} - \frac{1}{2}(\hat{m}_{L0} + \hat{m}'_{L0})]^2\} = [m_{L2} - \frac{1}{2}(m_{L0} + m'_{L0})]^2 + \frac{1}{nn'}[(N/4)E + E'']$$

$$= (1/16)h^2 + \frac{1}{nn'}[(N/4)E + E'']$$

The error mean square of the analysis of variance of the data of the F$_2$ generations, divided by the number of replications (b), is an estimator of E''. An estimator of E is the error mean square of the analysis of variance of the data of the parents, divided by b. If the means of the parents and F$_2$ generations have the same precision, then E and E'' can be estimated by the error mean square of the analysis of variance of the parents and F$_2$ generations data, divided by b. Let y_{rsij} be the phenotypic value of the ith individual (i = 1, ..., l) of the F$_2$ generation obtained from the cross between the rth and sth parents, in the jth replication (j = 1,..., b) ($y_{rsij} = g_{rsij} + e_{rsij}$ ($e_{rsij} \sim N(0, E_0)$, independents). Considering the jth replication:

$$E(\hat{V}_{2(rs)L2j}) = E\{\frac{1}{(l-1)}[\sum_{i=1}^{l} y_{rsij}^2 - \frac{(y_{rs.j})^2}{l}]\}$$

$$= V_{2(rs)L2j} + \frac{1}{(l-1)}E(\sum_{i=1}^{l} e_{rsij}^2 - l\bar{e}_{rs.j}^2)$$

$$= V_{2(rs)L2j} + E_0$$

Thus,

$$E\left(\frac{1}{b}\sum_{j=1}^{b}\hat{V}_{2(rs)L2j}\right) = V_{2(rs)L2} + E_0$$

An estimator of the environmental variance among individuals (E_0) is the mean variance of the phenotypic values of individuals in the same parent population (mean of N or bN variances, for the analyses considering one block and the b blocks, respectively).

Analysis

Estimation of the genetic and non-genetic components of variation. If individual measurements are not taken in each F_2 generation, the estimation of the genetic and non-genetic components may be carried out by fitting the general linear model $Y = X\beta + \varepsilon$ ($\varepsilon \sim N(\Phi, Cov(\varepsilon)))$, where:

$$Y' = [\hat{V}_{0L0(1)}\ \hat{V}_{0L0(2)}\ \hat{V}_{0L0}\quad \hat{W}_{01(r=1)L02} \cdots \hat{W}_{01(r=n)L02}$$
$$\hat{V}_{1(r=1)L2} \cdots \hat{V}_{1(r=n)L2}\quad \hat{W}_{01(s=1)L02} \cdots \hat{W}_{01(s=n')L02}$$
$$\hat{V}_{1(s=1)L2} \cdots \hat{V}_{1(s=n')L2}\quad \hat{W}_{01(r=1)L2} \cdots \hat{W}_{01(r=n)L2}$$
$$\hat{W}_{01(s=1)L2} \cdots \hat{W}_{01(s=n')L2}\quad [\hat{m}_{L2} - \frac{1}{2}(\hat{m}_{L0} + \hat{m}'_{L0})]^2\ \hat{E}\ \hat{E}'']$$

$$W_{01(r)L2} = (1/4)D_{(2)} - [(n+1)/16n]F_r - (1/16n)\sum_{\substack{r'=1 \\ r'\neq r}}^{n} F_{r'} +$$
$$(1/16)H_{1(2)} - (1/16)H_{2r} + E''/n$$

$$W_{01(s)L2} = (1/4)D_{(1)} - [(n'+1)/16n']F_s - (1/16n')\sum_{\substack{s'=1 \\ s'\neq s}}^{n'} F_{s'} +$$
$$(1/16)H_{1(1)} - (1/16)H_{2s} + E''/n'$$

$$\beta' = [D_{(1)}\ D_{(2)}\ D_{(3)}\quad F_{r=1} \cdots F_{r=n}\quad F_{s=1} \cdots F_{s=n'}$$
$$H_{1(1)}\ H_{1(2)}\quad H_{2r=1} \cdots H_{2r=n}\quad H_{2s=1} \cdots H_{2s=n'}$$
$$h^2\ \ E\ \ E'']$$

The estimation method utilized by Hayman (1954, 1958) was the weighted least squares ($Cov(\varepsilon) = \sigma^2 D$, where D is a diagonal matrix), since the means of the parents, F_1 hybrids and/or F_2 generations were associated to different number of observations. The elements of D were defined in these papers. If the diallel table means have same precision, the ordinary least squares method ($Cov(\varepsilon) = \sigma^2 I$) should be used. The maximum likelihood method ($Cov(\varepsilon) = V$, where V is a variance-covariance matrix) is an alternative. A simple iterative process for maximum likelihood estimation is presented by Hayman (1960). Under normality assumptions the maximum likelihood estimator reduces to the ordinary least squares estimator when $V = \sigma^2 I$ (Searle, 1971).

Application

Table 1 shows the grain yield per plant of nine lines of common bean (*Phaseolus vulgaris*), six as parents of group 1 (n = 6) and three as parents of group 2 (n' = 3), and 18 F_2 generations, obtained from the partial diallel among the two parent groups. Means refer to the observed values of one of four blocks of an experiment, carried out during the winter of 1993 at the Federal University of Viçosa, Viçosa, state of Minas Gerais, Brazil. In the case of the other blocks the additive-dominant model is not adequate to explain the observed results. Table 2 shows the results of the regression analyses of W_{r2} on V_{r2} and of W_{s2} on V_{s2}. Tests show absence of functional relation between variance and covariance in the arrays of the two groups. In this case there is no dominance in the polygenic systems under study. Result is confirmed by the analysis of variance of the diallel table (Viana *et al.*, 2000).

Table 1. Grain yield, in grams, of 9 lines of common bean and 18 derived F_2 generations

Parent		BAT-304 (1)	FT-84-835(2)	Batatinha (3)
		7.92	3.14	5.53
Ricopardo 896 (1)	9.71	5.56	5.73	6.19
Ouro Negro (2)	4.45	4.52	5.35	5.58
Antioquia 8 (3)	9.38	7.04	6.08	8.44
DOR. 241 (4)	4.12	4.82	3.88	4.86
RAB 94 (5)	7.26	8.41	9.46	6.64
Ouro (6)	5.52	5.10	5.32	8.44

Table 2. Summary of the regression analyses of W_{r2} on V_{r2} and of W_{s2} on V_{s2}, estimates of regression coefficients and level of significance of the test of the hypothesis H_0: $\beta_1 = 1$[1], in relation to grain yield of common bean plants, in grams

		Mean Square	
Source of Variation	Degrees of Freedom	Group 1	Group 2
Regression	1	0.19^{ns}	$1.65 \times 10^{-4\ ns}$
Error[2]		1.26	0.08
Regression Coefficient		–0.15⋆	-0.01^{+}

[1] Using t statistic, with four and one degrees of freedom for error in analyses considering groups 1 and 2, respectively; [2] With four and one degrees of freedom in the analyses for groups 1 and 2, respectively; ns: not significant (F less than one); ⋆: $0.01 < P < 0.05$; +: $P > 0.10$

The diallel analysis of the parents and F_1 hybrids (Viana *et al.*, 1999) showed that grain yield depended largely on dominance effects. The results of the analysis presented here show that one generation of selfing made these effects negligible. The estimation of the genetic and non-genetic components of variation was carried out by adjusting the additive model, using the ordinary least squares method. The estimates of $V_{2(rs)L2}$ were used in the calculations. The error mean squares from the analyses of variance considering only the parents and only the F_2 generations were 2.43 and 1.60, respectively. The former is an estimate of the E component and the later is an estimate of E". The estimate of E_0 was obtained as the weighted mean of the nine variances among plants in the same parental population (21.86).

Table 3 shows the estimates of the genetic and non-genetic components of variation. There is variability in group 1, but there is little or no genetic variability in group 2. Thus, the allelic frequencies of the non-fixed genes in the polygenic system defined by the parents of group 2 are near one and zero.

Table 3. Estimates of the additives and non-genetic (E, E" and E_0) components of variation, in relation to grain yield of common bean plants, in grams[1]

Parameter	Estimate	Standard Error	Probability[2]
$D_{(1)}$	3.31	1.13	0.0033
$D_{(2)}$	0.79	0.91	0.1937
$D_{(11)}$	-	6.39	-
$D_{(12)}$	-	6.39	-
$D_{(13)}$	-	6.39	-
$D_{(21)}$	-	6.39	-
$D_{(22)}$	6.41	6.39	0.1620
$D_{(23)}$	-	6.39	-
$D_{(31)}$	24.84	6.39	0.0003
$D_{(32)}$	21.42	6.39	0.0011
$D_{(33)}$	19.16	6.39	0.0028
$D_{(41)}$	-	6.39	-
$D_{(42)}$	-	6.39	-
$D_{(43)}$	-	6.39	-
$D_{(51)}$	48.86	6.39	0.0000
$D_{(52)}$	262.23	6.39	0.0000
$D_{(53)}$	-	6.39	-
$D_{(61)}$	-	6.39	-
$D_{(62)}$	-	6.39	-
$D_{(63)}$	130.91	6.39	0.0000
E	2.40	0.95	0.0088
E"	1.38	0.49	0.0041
E_0	26.71	0.82	0.0000
$D_{r=1}$	-	4.36	-
$D_{r=2}$	2.14	4.36	0.3140
$D_{r=3}$	21.80	4.36	0.0000
$D_{r=4}$	-	4.36	-
$D_{r=5}$	103.69	4.36	0.0000
$D_{r=6}$	43.64	4.36	0.0000
$D_{s=1}$	12.28	3.69	0.0012
$D_{s=2}$	48.34	3.69	0.0000
$D_{s=3}$	25.01	3.69	0.0000
$D_{(3)}$	28.55	3.16	0.0000

[1] Values not shown correspond to negative estimates, given as equal to zero; [2] Based on t test with 29 degrees of freedom

Analysis of the means of the two groups shows that the genes that increase grain yield have a greater frequency in group 1. In this group, Ricopardo 896, Ouro Negro and DOR. 241 have similar genotypes to those of parents in group 2 and carry the most frequent genes in this group, which, on the whole, decrease grain yield. Crosses between these parents or between them and the parents of group 2 should produce F_2 generations with little variability for selection purposes. This is also true for crosses between the parents of group 2. In group 2, BAT-304 has the highest mean and is, as expected, the carrier of the more frequent genes in group 1. In group 1, the parents Ricopardo 896 and Antioquia 8 produce the highest yield. Antioquia 8 should not have many of the more frequent genes in group 2, which, as already stated, decrease grain yield.

Therefore, these two parents may be crossed with one or more of the parents of group 2. During the breeding process the objective should be to fix in a line the desirable genes of each parent. As already mentioned, crosses involving Ricopardo 896 will produce F_2 generations with reduced variability, which may complicate the selection of lines with better performance than the best parent. In the F_2 generations obtained from these crosses the breeder can deal with a smaller number of plants and families. Crosses between Antioquia 8, RAB 94 and Ouro, or involving a parent from group 2, should produce segregant generations with larger variability. This may favor the selection of a genotype superior to the best parent, carrier of desirable genes from both parents. The breeder, however, will have to deal with many plants and families from these generations.

The narrow sense heritabilities values, at the individual level, in the F_2 generations of the diallel (Table 4), show that mass selection are inefficient to increase grain yield in most of the segregant generations obtained from the cross between any two lines. The exceptions are the F_2 generations derived from the crosses RAB 94 x FT-84-835 and Ouro x Batatinha. Even in these cases it is adequate that selection for grain yield be performed on family means, preferably in advanced generations, with high degree of homozygosis.

Table 4. Estimates of narrow sense heritabilities, at individual level, in the F_2 generations of the diallel, for grain yield of common bean plants, in grams[1]

Parent	BAT-304	FT-84-835	Batatinha	Mean
Ricopardo 896	-	-	-	-
Ouro Negro	-	0.06	-	0.02
Antioquia 8	0.19	0.17	0.15	0.14
DOR. 241	-	-	-	-
RAB 94	0.31	0.71	-	0.44
Ouro	-	-	0.55	0.25
Mean	0.09	0.27	0.16	0.43

[1] Values not shown correspond to negative estimates, given as equal to zero

References

COCKERHAM, C.C.; WEIR, B.S. Quadratic analyses of reciprocal crosses. *Biometrics*, Washington, v. 33, no. 1, p. 187-203, 1977.

DICKINSON, A.G.; JINKS, J.L. A Generalised analysis of diallel crosses. *Genetics*, Baltimore, v. 41, no. 1, p. 65-78, 1956.

FALCONER, D.S.; MACKAY, T.F.C. *Introduction to quantitative Genetics*. 4. ed. England: Longman, 1996.

FOOLAD, M.R.; BASSIRI, A. Estimates of combining ability, reciprocal effects and heterosis for yield and yield components in a common bean diallel cross. *J. Agricult. Sci.*, New York, v. 100, no. 1, 103-108, 1983.

GARDNER, C.O.; EBERHART, S.A. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics*, Washington, v. 22, no. 3, p. 439-452, 1966.

GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing system. *Austr. J. Biol. Sci.*, Melbourne, v. 9, no. 4, p. 463-493, 1956.

HAYMAN, B.I. The theory and analysis of diallel crosses. *Genetics*, Baltimore, v. 39, no. 6, p. 789-809, 1954.

HAYMAN, B.I. Maximum likelihood estimation of genetic components of variation. *Biometrics*, Washington, v. 16, no. 3, p. 349-381, 1960.

HAYMAN, B.I. The theory and analysis of diallel crosses. II. *Genetics*, Baltimore, v. 43, no. 1, p. 63-85, 1958.

JINKS, J.L.; HAYMAN, B.I. The analysis of diallel crosses. *Maize Genetics Cooperation Newsletter*, Columbia, v. 27, p. 48-54, 1953.

MATHER, K.; JINKS, J. L. *Biometrical Genetics*. 2.ed. Ithaca, New York: Cornell University Press, 1974.

NIENHUIS, J.; SINGH, S.P. Combining ability and relationships among yield, yield components, and architectural traits in dry bean. *Crop Science*, Madison, v. 26, no. 1, p. 21-27, 1986.

PARK, H.G.; DAVIS, D.W. Inheritance of interlocular cavitation in a six-parent diallel cross in snap beans (*Phaseolus vulgaris* L.). *J. Am. Soc. Horticult. Sci.*, Alexandria, v. 101, no. 2, p. 184-189, 1976.

SEARLE, S.R. *Linear Models*. New York: John Wiley & Sons, 1971.

VIANA, J.M.S. *et al.* Theory and analysis of partial diallel crosses. *Genetics and Molecular Biology*, Ribeirão Preto, v. 22, no. 4, p. 591-599, 1999.

VIANA, J.M.S. *et al.* Analysis of variance of partial diallel tables. *Genetics and Molecular Biology*, Ribeirão Preto, v. 23, no. 1, p. 229-234, 2000.

YANCHUK, A.D. General and specific combining ability from disconnected partial diallels of coastal douglas-fir. *Silvae Genetica*, Frankfurt, v. 45, no. 1, p. 37-45, 1996.

Received on August 25, 2000.

Accepted on March 02, 2001.