

# The parametric restrictions of the Griffing diallel analysis model: combining ability analysis

José Marcelo Soriano Viana

## Abstract

It was studied the parametric restrictions of the diallel analysis model of Griffing, method 2 (parents and  $F_1$  generations) and model 1 (fixed), in order to address the questions: i) does the statistical model need to be restricted? ii) do the restrictions satisfy the genetic parameter values? and iii) do they make the analysis and interpretation easier? Objectively, these questions can be answered as: i) yes, ii) not all of them, and iii) the analysis is easier, but the interpretation is the same as in the model with restrictions that satisfy the parameter values. The main conclusions were that: the statistical models for combining ability analysis are necessarily restricted; in the Griffing model (method 2, model 1), the restrictions relative to the specific combining ability (SCA) effects,  $\sum_{j=1}^N \sum_{j'=1}^N s_{jj'} = 0$  and  $s_{jj} + \sum_{j'=1}^N s_{jj'} = 0$ , for all  $j$ , do not satisfy the parametric values, and the same inferences should be established from the analyses using the model with restrictions that satisfy the parametric values of SCA effects and that suggested by Griffing. A consequence of the restrictions of the Griffing model is to allow the definition of formulas for estimating the effects, their variances and the variances of contrasts of effects, as well as for calculating orthogonal sums of squares.

## INTRODUCTION

The expressions general combining ability (GCA) and specific combining ability (SCA) have been used since the 1940s (Sprague and Tatum, 1942) to designate properties of endogamic families and inbred lines under selection in hybrid production programs. Since then, maize breeders have been aware of the need to evaluate endogamic families involved in the selfing process, normally  $S_3$  progenies, in order to reduce the number of inbred lines. For this, the families are normally crossed with a tester (a single- or double-cross hybrid or an open-pollinated population) and the progenies are then evaluated experimentally. Outstanding endogamic families, capable of generating superior progenies, i.e., with an elevated GCA, would continue to be selfed. The inbred lines produced would also be evaluated based on crossing between themselves, a system called diallel. The aim would be to identify pairs of inbred lines that produce the best hybrids. This second stage of evaluation is known as the SCA test.

Specific methods for estimating the effects of GCA and SCA or the variances of these effects have been described. These methodologies generally consist of an analysis of variance of data from progenies obtained by diallel, whether complete (Griffing, 1956a,b), partial (Gerald and Miranda Filho, 1988), circulant (Kempthorne and Curnow, 1961), or otherwise. The main characteristics of these methods are their generality, since they can be used for any species, and the easiness of analysis and interpretation.

Among several methods of combining ability analysis, that described by Griffing (1956b) is probably the most

used. Easy computer handling, guaranteed by the availability of appropriate formulas, and the care the author and others took to discuss in detail the value of the effect variance estimates or the effect estimates (Griffing, 1956a; Cruz and Vencovsky, 1989) in breeding programs, have contributed to the widespread use of this model. Nevertheless, not every aspect of this method has been evaluated in detail. If the diallel's parents are not a sample from a population, i.e., when the model is fixed, then the parametric restrictions associated with the statistical model must be addressed. Does the model really have to be restricted? Do the imposed restrictions satisfy the genetic parameters? Do the restrictions make analysis and interpretation easier? The objective of this study was to answer these and other questions.

## MATERIAL AND METHODS

The data used here are those reported by Gardner and Eberhart (1966) (Table I).

### Parametric values of the components of the Griffing (1956b) model

We consider a polygenic system with  $k$  genes, each with two allelic forms and no epistasis, that are responsible for determining a quantitative character in a diploid species with sexual reproduction. Independent of the reproduction system of the species (cross-pollination or self-pollination), the genotypic mean of a population can be expressed as

$$M_{jj} = \sum_{i=1}^k m_i + \sum_{i=1}^k (2p_{ij} - 1)a_i + 2 \sum_{i=1}^k (p_{ij} - p_{ij}^2)d_i$$

where  $m_i$  is the mean of the genotypic values of the homozygotes relative to locus  $i$ ,  $p_{ij}$  is the frequency in population  $j$  of the locus  $i$  gene that increases the trait expression,  $a_i$  is the difference between the genotypic value of the homozygote with highest expression and  $m_i$ , and  $d_i$  is the deviation due to dominance relative to locus  $i$ .

Note that  $p_{ij}$  is equal to 1 or to 0 if the species is autogamous. In the case of allogamous species, the population is in Hardy-Weinberg equilibrium and is not endogamic. If population  $j$  is one of the  $N$  parents of a diallel, the genotypic mean of the hybrid produced by crossing populations  $j$  and  $j'$  is

$$M_{jj'} = \sum_{i=1}^k m_i + \sum_{i=1}^k (p_{ij} + p_{ij'} - 1)a_i + \sum_{i=1}^k (p_{ij} + p_{ij'} - 2p_{ij}p_{ij'})d_i$$

The mean of the hybrids in which parent  $j$  participates, including population  $j$ , is

$$M_{j.} = \sum_{i=1}^k m_i + \sum_{i=1}^k (p_{ij} + \bar{p}_i - 1)a_i + \sum_{i=1}^k (p_{ij} + \bar{p}_i - 2p_{ij}\bar{p}_i)d_i$$

where  $\bar{p}_i$  is the average frequency in the parents of the diallel of the locus  $i$  gene that increases character expression.

The diallel mean is

$$M_{..} = \sum_{i=1}^k m_i + \sum_{i=1}^k (2\bar{p}_i - 1)a_i + 2 \sum_{i=1}^k \bar{p}_i (1 - \bar{p}_i)d_i$$

The effect of GCA of population  $j$  corresponds to

$$M_{j.} - M_{..} = \sum_{i=1}^k (p_{ij} - \bar{p}_i)[a_i + (1 - 2\bar{p}_i)d_i] = g_j$$

If the parents are open-pollinated populations, the greater the value of the GCA effect of a population, the greater the frequencies of the genes that increase the trait expression and the greater the differences between the gene frequencies of the population and the average frequencies in the diallel's parents. If the parents are inbred lines or pure lines, the greater the value of the GCA effect of a population, the greater the number of genes that increase the trait expression and, consequently, the greater the number of positive differences between the gene frequency of the

population and the average frequency in the diallel's parents. Therefore, the effect of GCA is an indicator of the superiority of the population and of its divergence relative to the diallel's parents, thus providing the same information as the parameters 'population effect' ( $v_j$ ) and 'variety heterosis' ( $H_j$ ) of the Gardner and Eberhart (1966) model. The correlation between the GCA effect and  $v_j$  approaches close to 1 as the degree of dominance approaches zero. For one gene and populations with  $p_{ij}$  values equal to 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1, the correlation values are 0.87 and 1 when  $|d/a| = 2$  and  $d/a = 0$ , respectively.

The genotypic means of population  $j$  and the hybrid of parents  $j$  and  $j'$  can be expressed as

$$M_{jj} = M_{..} + 2 \sum_{i=1}^k (p_{ij} - \bar{p}_i)[a_i + (1 - 2\bar{p}_i)d_i] + \left[ -2 \sum_{i=1}^k (p_{ij} - \bar{p}_i)^2 d_i \right] = M_{..} + 2g_j + s_{jj}$$

$$M_{jj'} = M_{..} + \sum_{i=1}^k (p_{ij} - \bar{p}_i)[a_i + (1 - 2\bar{p}_i)d_i] + \sum_{i=1}^k (p_{ij'} - \bar{p}_i)[a_i + (1 - 2\bar{p}_i)d_i] + 2 \sum_{i=1}^k \{ \bar{p}_i(p_{ij} - \bar{p}_i) + \bar{p}_i(p_{ij'} - \bar{p}_i) - [p_{ij}p_{ij'} - (\bar{p}_i)^2] \} d_i = M_{..} + g_j + g_{j'} + s_{jj'}$$

where  $s_{jj}$  is the effect of SCA of a population with itself, and  $s_{jj'}$  is the effect of SCA of populations  $j$  and  $j'$ .

When there is negative unidirectional dominance, the  $s_{jj}$  values are positive. If the deviations due to dominance are positive, the  $s_{jj}$  values are negative. When the SCA effect of a population with itself is null, the population has the same gene frequencies as the average frequencies in the group of the diallel's parents. Furthermore, the higher the absolute value of  $s_{jj}$ , the greater the differences between the gene frequencies in the population and the average frequencies in the diallel's parents. Therefore,  $s_{jj}$  is also an indicator of the population's divergence relative to the parental group. For one gene and populations with  $p_{ij}$  values equal to 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1, the correlation between the absolute values of  $g_j$  and  $s_{jj}$  is 0.965 for any degree of dominance ( $d/a \neq 0$ ). The  $s_{jj}$  values thus provide the same inferences as the estimates of the parameters 'average heterosis' and 'variety heterosis' of the Gardner and Eberhart (1966) model. Although  $s_{jj}$  is a measure of the population divergence relative to the diallel's parents and 'variety heterosis' is an indicator of the differences between the gene frequencies in the population and the average frequencies in the other genitors, the correlation between the absolute values of these parameters for one gene is 1, independent of the populations and of the degree of dominance ( $d/a \neq 0$ ).

**Table I** - Mean grain yield (bushels/acre) of six corn populations and their hybrids<sup>1</sup>.

	1 - M	2 - HG	3 - GR	4 - BR	5 - K	6 - K <sub>II</sub>
1 - M	91.0	98.8	91.1	95.3	93.5	100.7
2 - HG		91.7	92.7	97.1	94.1	105.4
3 - GR			87.9	101.3	91.6	103.3
4 - BR				96.6	95.4	102.7
5 - K					91.3	101.6
6 - K <sub>II</sub>						96.2

<sup>1</sup>From Gardner and Eberhart (1966).

When there is negative unidirectional dominance, the lowest values of  $s_{jj}$  identify the populations with the greatest differences in gene frequencies between themselves and in relation to the average frequencies in the diallel's parents. The highest values identify the populations with the smallest differences in gene frequencies, but with different gene frequencies relative to the average frequencies in the diallel's parents. When there is positive unidirectional dominance, the lowest  $s_{jj}$  values are associated with populations with the smallest differences in gene frequencies, but which have differences in gene frequencies relative to the average frequencies in the diallel's parents. The highest values indicate populations with the highest differences in gene frequencies between themselves and in relation to the average frequencies in the genitor group. Independent of the direction of the dominance effects,  $s_{jj}$  values close to the average value indicate populations with small differences in gene frequencies between themselves and relative to the average frequencies in the parental group. If the  $s_{jj}$  value equals zero, the gene frequencies in one of the populations are equal to the average frequencies in the diallel's parents. The average value of the SCA effects of different populations is

$$\frac{1}{N(N-1)/2} \sum_{j=1}^N \sum_{j'=1}^N s_{jj'} = - \frac{1}{N(N-1)} \sum_{j=1}^N s_{jj}$$

Therefore the SCA of populations  $j$  and  $j'$  is an indicator of the divergence between them and of their divergence from the diallel's parents, in a manner similar to the parameter 'specific heterosis' ( $S_{jj'}$ ) of the Gardner and Eberhart (1966) model. For one gene and populations with  $p_{ij}$  values equal to 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1, the correlation between the effect of SCA of two populations and specific heterosis is 0.998, independent of the degree of dominance ( $d/a \neq 0$ ). This and the other results given above show that the methods of Griffing (1956b) and of Gardner and Eberhart (1966) are not complementary but indeed absolutely equivalent in terms of inference.

The phenotypic means in a diallel table can be defined as

$$Y_{jj} = M_{..} + 2g_j + s_{jj} + \bar{e}_{jj}$$

$$Y_{jj'} = M_{..} + g_j + g_{j'} + s_{jj'} + \bar{e}_{jj'}$$

where  $\bar{e}_{jj}$  and  $\bar{e}_{jj'}$  are the average values of the residues associated with the observations for population  $j$  and the hybrid of parents  $j$  and  $j'$ , respectively.

The equations above define a statistical model that is necessarily restricted, since  $E(g_j) = 0$  and  $E(s_{jj'}) = 0$  ( $j$  fixed), in contrast to statistical models with fixed effects, where the researcher decides whether to impose parametric restrictions or not. Thus, the restrictions associated with the previously defined model for combining ability analysis are

$$(i) \sum_{j=1}^N g_j = 0 \text{ and } (ii) \sum_{j'=1}^N s_{jj'} = 0,$$

for all  $j$ , giving  $N + 1$  linearly independent restrictions.

This model is therefore not the same as that defined by Griffing (1956b) (method 2, model 1), whose restrictions are

$$(i) \sum_{j=1}^N g_j = 0, (ii) \sum_{j=1}^N \sum_{j'=1}^N s_{jj'} = 0 \text{ and } (iii) s_{jj} + \sum_{j'=1}^N s_{jj'} = 0,$$

for all  $j$ , giving  $N + 1$  linearly independent restrictions.

Note that restrictions (ii) and (iii) in the latter case do not satisfy the parametric values of the SCA effects, since

$$\bar{s}_{..} = \frac{1}{N(N+1)/2} \sum_{j=1}^N \sum_{j'=1}^N s_{jj'} = \frac{1}{N(N+1)} \sum_{j=1}^N s_{jj}$$

and, therefore,  $\sum_{j=1}^N \sum_{j'=1}^N s_{jj'} = \frac{1}{2} \sum_{j=1}^N s_{jj}$ . Why, then, were they

considered by Griffing (1956b)? The answer is simply because they allow one to obtain formulas for the estimation of the effects and the effect variances and for the calculation of the sums of squares of the GCA and SCA. The availability of formulas is an indispensable condition for the widespread use of methods in quantitative genetics, and allows the elaboration of software for data processing by professionals who are not specialized in this field and in the theory of linear models. The model described here is of full column rank and, therefore, it is possible to develop formulas for estimating the effects and their variances and for computing the sums of squares. As will be shown, when there is dominance, the sum of squares attributable to the null hypothesis for GCA effects is not orthogonal to the sum of squares due to the hypothesis of nullity of the SCA effects.

Despite the differences between the two models, the hypotheses that can be tested coincide and include:

1.  $H_{0(1)}$  - equality of the treatment means (parents and hybrids): to test this hypothesis is the same as testing that there are no gene frequency differences between the genitors ( $p_{ij} = p_i$  for all  $i$  and  $j$ ).
2.  $H_{0(2)}$  - nullity of the GCA effects ( $g_j = 0$  for all  $j$ ): in the case of rejection of the hypothesis  $H_{0(1)}$ , testing this hypothesis is the same as testing that the GCA effects of populations with different genetic structures are null. Of course, the rejection of  $H_{0(2)}$  means that there are differences in the gene frequencies between the parents; acceptance of the hypothesis does not imply the contrary.
3.  $H_{0(3)}$  - nullity of the SCA effects ( $s_{jj'} = 0$  for all  $j$  and  $j'$ ): if there are gene frequency differences between the parents, testing this hypothesis is the same as testing that there is no dominance ( $d_i$  for all  $i$ ).

## RESULTS AND DISCUSSION

Although there is a difference between the analyses of variance for grain yield of six corn populations and their

**Table II** - Analyses of variance of grain yield (bushels/acre) of six corn populations and their hybrids, based on the model with restrictions that satisfy the genetic parameter values and the Griffing (1956b) model (values in parentheses for general combining ability, GCA).

Source of variation	Degrees of freedom	Sum of squares	Mean square	F	Probability
Treatments	20	473.2	23.7	3.33	0.0002
GCA	5	226.2 (234.2)	45.2 (46.9)	6.37 (6.60)	0.0001 (0.0001)
SCA	15	238.9	15.9	2.24	0.0141
Error <sup>1</sup>	60	426.0	7.1		

<sup>1</sup>The degrees of freedom and the mean square are provided by Gardner and Eberhart (1966). SCA, Specific combining ability.

**Table III** - Estimates of the GCA effects ( $\hat{g}_j$ ), the SCA effects of a population with itself ( $\hat{s}_{jj}$ ), the variances of these effects and of the variances of contrasts between themselves for the model defined in this study and for the Griffing (1956b) model.

Population	Model defined in this study		Griffing model	
	$\hat{g}_j$	$\hat{s}_{jj}$	$\hat{g}_j$	$\hat{s}_{jj}$
1	-1.708	-2.358	-1.462	-2.232
2	-0.142	-4.792	-0.200	-4.057
3	-2.125	-4.625	-2.162	-3.932
4	1.292	-2.758	1.488	-2.532
5	-2.192	-1.092	-1.787	-1.282
6	4.875	-10.325	4.125	-8.207
	$\hat{V}(\hat{g}_j) = 0.8218$		$\hat{V}(\hat{g}_j) = 0.7396$	
	$\hat{V}(\hat{g}_j - \hat{g}_j) = 1.9722$		$\hat{V}(\hat{g}_j - \hat{g}_j) = 1.7750$	
	$\hat{V}(\hat{s}_{jj}) = 6.4097$		$\hat{V}(\hat{s}_{jj}) = 3.8036$	
	$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 12.6222$		$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 7.1000$	

**Table IV** - Estimates of the SCA effects of different populations ( $\hat{s}_{jj}$ ) for the diallel analyses using the model defined in this study (values above the diagonal) and the Griffing (1956b) model (values below the diagonal), as well as estimates of variances of the effects and of contrasts between them.

Parents	1	2	3	4	5	6
1		3.875	-1.842	-1.058	0.625	0.758
2	4.305		-1.808	-0.825	-0.342	3.892
3	-1.432	-1.095		5.358	-0.858	3.775
4	-0.882	-0.345	5.818		-0.475	-0.242
5	0.593	-0.070	-0.607	-0.457		2.142
6	1.880	5.318	5.180	0.930	3.105	
	Model defined in this study			Griffing model		
	$\hat{V}(\hat{s}_{jj}) = 4.832$			$\hat{V}(\hat{s}_{jj}) = 5.579$		
	$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 11.439$			$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 12.425$		
	$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 8.678$			$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 10.650$		
	$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 13.806$			$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 12.425$		
	$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 10.650$			$\hat{V}(\hat{s}_{jj} - \hat{s}_{jj}) = 8.875$		

hybrids using the model described here and that proposed by Griffing (1956b), in relation to the sum of squares attributable to the null hypothesis of the GCA effects (Table II), the inferences remain the same, namely, there are differences in the gene frequencies between the populations and there is dominance in the polygenic system under analysis.

The differences between the estimates of the GCA effects and the SCA effects of each population with itself in the two models did not change the inferences since the correlations between the estimated values of  $g_j$  and of  $s_{jj}$  are, respectively, 0.9955 and 1. Nevertheless, the variances of the estimates of the effects and the contrasts between the effects of the model defined in this study are greater (Table III). Population 6 has the genes that increase yield at the highest frequency, and is the most divergent relative to the parental group. This population ought, therefore, to be selected for an intrapopulation improvement program. Population 4 is the second best, but not the second most divergent, and can also be chosen for intrapopulation improvement. The negative  $s_{jj}$  values indicate unidirectional positive dominance.

Although the estimates of the SCA effects of different populations in the two models are different, the inferences that can be established are the same, since the correlation between the estimated values is 0.985. Differences between the variance values of the effects and the contrasts were also observed (Table IV). The variance estimates of  $\hat{s}_{jj}$  effects and of the contrasts between these effects are lowest for the model developed above. Populations 3 and 4, 2 and 6, 1 and 2, and 3 and 6 showed the greatest differences in gene frequencies between themselves and in relation to the average frequencies in the diallel's parents. Thus, for interpopulation improvement programs, the second pair ought to be selected, because of the superiority of population 6. There was little divergence between parents 1 and 6, and 1 and 5, and from them relative to the group of parents, since their SCA effects come close to the average value (0.865). There were only slight differences in gene frequency between populations 1 and 3, and 2 and 3, although they were divergent in relation to the diallel's parents. The frequencies in these populations ought to be lower than the average frequencies.

## CONCLUSIONS

The statistical models for combining ability analysis of a population group are obligatorily restricted. The restrictions  $\sum_{j=1}^N \sum_{j'=1}^N s_{jj'} = 0$  and  $s_{jj} + \sum_{j'=1}^N s_{jj'} = 0$ , for all  $j$ , of the

model proposed by Griffing (1956b) (method 2, model 1) do not satisfy the parametric values of SCA effects. Although there are differences between the analysis according to the model described here and that suggested by Griffing (1956b), the inferences should be the same. A consequence of the restrictions of the Griffing (1956b) model is to allow the definition of formulas for estimating the effects, their variances and the variances of contrasts of effects, as well as for the calculation of orthogonal sums of squares. In conclusion, it is generally quite safe to use the Griffing model.

## RESUMO

Com o propósito de estender os conhecimentos teóricos sobre o modelo de análise dialélica mais comumente empregado pelos melhoristas, o de Griffing, proposto em 1956, discute-se neste trabalho as restrições paramétricas associadas ao método 2 (pais e gerações  $F_1$ ), modelo 1 (fixo). As questões que conduziram à elaboração deste artigo foram: i) o modelo estatístico tem que ser restrito?; ii) as restrições satisfazem os valores dos parâmetros genéticos? e iii) elas tornam a análise e a interpretação mais fáceis? De forma objetiva, estas questões podem ser assim respondidas: i) sim; ii) nem todas; e iii) a análise sim, mas a interpretação é a

mesma do modelo restrito definido neste trabalho. As principais conclusões são: os modelos estatísticos de análise de capacidade de combinação de um grupo de populações são obrigatoriamente restritos; as restrições  $\sum_{j=1}^N \sum_{j'=1}^N s_{jj'} = 0$  e  $s_{jj} + \sum_{j'=1}^N s_{jj'} = 0$ , para todo  $j$ , do modelo proposto por Griffing, não satisfazem os valores paramétricos dos efeitos de capacidade específica de combinação; embora haja diferenças entre a análise segundo o modelo definido neste trabalho e a proposta por Griffing, as inferências a serem estabelecidas devem ser exatamente as mesmas; uma consequência das restrições do modelo de Griffing é possibilitar a definição de fórmulas para a estimação dos efeitos, das suas variâncias e das variâncias de contrastes entre efeitos, e para o cálculo das somas de quadrados, o que certamente foi fundamental para garantir o uso generalizado desta metodologia.

## REFERENCES

- Cruz, C.D.** and **Vencovsky, R.** (1989). Comparação de alguns métodos de análise dialélica. *Rev. Bras. Genet.* 12: 425-438.
- Gardner, C.O.** and **Eberhart, S.A.** (1966). Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* 22: 439-452.
- Geraldi, I.O.** and **Miranda Filho, J.B. de** (1988). Adapted models for the analysis of combining ability of varieties in partial diallel crosses. *Rev. Bras. Genet.* 11: 419-430.
- Griffing, B.** (1956a). A generalized treatment of the use of diallel crosses in quantitative inheritance. *Heredity* 10: 31-50.
- Griffing, B.** (1956b). Concept of general and specific combining ability in relation to diallel crossing system. *Aust. J. Biol. Sci.* 9: 463-493.
- Kemphorne, O.** and **Curnow, R.N.** (1961). The partial diallel cross. *Biometrics* 17: 229-250.
- Sprague, G.F.** and **Tatum, L.A.** (1942). General vs. specific combining ability in single crosses of corn. *J. Am. Soc. Agron.* 34: 923-932.

(Received February 21, 2000)

