Griffing’s methods: review of their importance and application in conventional plant breeding

Claudia Saavedra Guevara¹
Delfina de Jesús Pérez López²§
Andrés González Huerta²
J. Ramón Pascual Franco Martínez²
Martín Rubí Arriaga²
José Francisco Ramírez Dávila²

¹Doctoral Program in Agricultural Sciences and Natural Resources-Autonomous University of the State of Mexico-Institute of Agricultural and Rural Sciences-‘El Cerrillo’ University Campus. Toluca, State of Mexico, Mexico. Tel. 722 2965552, ext. 117. (pcarn@uaemex.mx). ²Center for Research and Advanced Studies in Plant Breeding-Faculty of Agricultural Sciences. Toluca, State of Mexico, Mexico. AP. 435. Tel. 722 2965518, ext. 148. (csaavedrag001@alumno.uaemex.mx; agonzalez@uaemex.mx; jrfancom@uaemex.mx; mrubia@uaemex.mx; jframirezd@uaemex.mx).

§Corresponding author: djperezl@uaemex.mx.

Abstract

The estimation of genetic parameters in plants and animals is of great relevance in the agro-livestock and biological sciences. In this context, the effects and variances of combining ability, heritability, heterosis, response to selection, identification of parents and outstanding simple crosses, as well as the prediction of hybrids of higher production and quality, depend on the type of mating and experimental design selected. This study analyzes Griffing’s four methodologies in relation to those of Hayman and Jinks and Gardner and Eberhart; the assumptions considered in these are underlined and the similarities that exist between them are commented. It includes the eight mathematical models that Griffing discussed and that are frequently used to apply some statistical package, cites some studies carried out in the last decade, and suggests some software for its genetic-statistical analysis.

Keywords: diallel crosses, genetic parameters, genetic-statistical models and packages.

Reception date: March 2021
Acceptance date: June 2021
Introduction

Diallelic experiments are related to genetic studies, as well as to plant breeding programs, seed production, generation, validation, application and transfer of technology (Martínez, 1988; González et al., 2007a, b; Pérez et al., 2019; González et al., 2019). The genetic designs used to form families of self-siblings, half-siblings and full siblings, within and between populations of the same variety, race, species and genus, are crosses between two parents, polycross, test cross, North Carolina designs I, II and III, Griffing’s methods 1, 2, 3 and 4, line x tester (Shattuck et al., 1993; Harriman and Nwammadu, 2016; Nduwumuremyi et al., 2013; Awata et al., 2018), triple test cross, triallel and quadriallel analysis, backcross and incomplete diallelics (Kempthorne and Curnow, 1961; Soriano, 2000; Muntau et al., 2015).

Another analysis proposed in autogamous species or in which pure lines can be easily derived is attributed to Jinks and Hayman (1953); Hayman (1954a, b); Jinks (1954), based on the estimation of six types of variances and covariances between relatives. Gardner and Eberhart (1966) designed another methodology in which each parent is a population with random mating, is in Hardy-Weinber equilibrium, and there are two alleles per locus. In the evaluation of genetic material in homogeneous areas, the completely randomized design has been applied. With an undesirable but predictable variation gradient, randomized complete blocks have been selected.

With two gradients of environmental heterogeneity, one perpendicular to the other, a Latin square or some lattice has been used (Gomez and Gomez, 1984; Martínez, 1988; Borojevic, 1990), but the series of experiments in time and space in randomized complete blocks or in lattices have been more used (González et al., 2007a, b; Martínez, 1988; Moore and Dixon, 2015; González et al., 2019). Mating and experimental designs are used to estimate genetic and environmental effects and variances, heritability, heterosis, response to selection, prediction of hybrids, as material for new plant breeding programs or as a suggestion for commercial planting (Rodríguez et al., 2016; Vesali et al., 2020; Aditika et al., 2020; Vasconcelos et al., 2020). In the above context, emphasis will be placed on Griffing’s (1956a, b) four methods, which are the most widely used since the 1950s. Some statistical packages available on their websites, such as academic or commercial versions, are also cited.

Materials and methods

Assumptions used in a diallel cross

Christie and Shattuck (1992) discussed four basic designs considering GCA and SCA (Griffing, 1956b), the arrangements of variances and covariances (Jinks and Hayman, 1953; Hayman, 1954b; Jinks, 1954; Jinks, 1956), additive and dominance effects (Gardner and Eberhart, 1966; Eberhart and Gardner, 1966) and incomplete diallelics (Kempthorne and Curnow, 1961). In these, as in other diallelic techniques, there is some assumption or restriction for the estimation of genetic-statistical parameters (Martínez, 1988; Sahagún, 1998; Awata et al., 2018; Soriano, 2000).

Griffing’s four methods are not conditioned by genetic hypotheses or assumptions (Wright, 1985; Christie and Shattuck, 1992; Shattuck et al., 1993; Hallauer et al., 2010). In Hayman and Jinks’s technique, diploid segregation, homozygous parents, no differences between direct and reciprocal
crosses, absence of epistasis and non-linkage between genes, absence of multiple alleles, and independent distribution of genes in the parents are considered. In Gardner and Eberhart’s methodologies, it is only assumed that each parent is in Hardy-Weimberg equilibrium and that in each locus, there are only two alleles.

**Griffing’s methods**

Griffing (1956b) proposed a diallelic technique for estimating GCA and SCA, as well as the nature and type of gene action involved in the inheritance of quantitative characters that manifest in plants and animals; this may vary in its design and biometric analysis if parents, direct simple crosses and/or reciprocal simple crosses (p, DC and RC, respectively) are included. Based on the combinations that can be formed when considering p parents, in different pairs \( (C_p^2) \), one will have:

- Method 1: the \( p^2 \) families are called complete diallelic and include \( p, \) DC and RC \( (p+C_p^2+C_p^2) \).
- Method 2: in this half diallelic, there are \( p + C_p^2 = p(p + 1)/2 \) families \( (p \) and DC). Method 3: This modified diallelic analyzes DC and RC; there are \( C_p^2 + C_p^2 = p^2 - p \) families. Method 4: \( C_p^2 = p (p-1)/2 \) families (DC only) are obtained.

**Genetic-statistical models**

In the eight analyses that originate by combining fixed-effect (I) and random-effect (II) models with their four methods, Griffing (1956b) focuses his discussion on a randomized complete block design, for one environment, there are b repetitions, a genotypes and c observations in each experimental plot.

**Method 1**

Model I: \( x_{ij} = \mu + g_i + g_j + s_{ij} + r_{ij} \) \( + \frac{1}{bc} \sum_k \sum_l e_{ijkl} \left\{ \begin{array}{l} \,i,j =1,\ldots,p \\ \,k=1,\ldots,b \\ \,l=1,\ldots,c \end{array} \right\} \)

Where: \( \mu \): is the population mean; \( g_i, g_j \) are the effects of GCA for parents, are \( i, j \), \( s_{ij} \) are the effects of SCA for that pair of parents; \( r_{ij} \) is the effect caused by their reciprocal cross; \( e_{ijkl} \) is the environmental effect that affects each \( ijkl \)-th observation, the latter is also known as the residual of the model. The components of models I shown below, are defined in the same terms as for method 1.

Model II: \( x_{ij} = \mu + g_i + g_j + s_{ij} + r_{ij} + \frac{1}{b} \sum_k b_k + \frac{1}{b} \sum_k (bv)_{ijk} + \frac{1}{bc} \sum_k \sum_l e_{ijkl} \).

In all four methods, except for \( \mu \), the other components are random variables.

**Method 2**

Modelo I: \( x_{ij} = \mu + g_i + g_j + s_{ij} + \frac{1}{bc} \sum_k \left\{ \begin{array}{l} \,i,j =1,\ldots,p \\ \,k=1,\ldots,b \\ \,l=1,\ldots,c \end{array} \right\} \)

Modelo II: \( x_{ij} = \mu + g_i + g_j + s_{ij} + \frac{1}{b} \sum_k b_k + \frac{1}{b} \sum_k (bv)_{ijk} + \frac{1}{bc} \sum_k \sum_l e_{ijkl} \)
Method 3

Modelo I: \( x_{ij} = \mu + g_i + g_j + s_{ij} + r_{ij} + \frac{1}{bc} \sum_{k} \sum_{l} e_{ijkl} \) for \( i,j =1,\ldots,p \), \( k=1,\ldots,b \), \( l=1,\ldots,c \)

Modelo II: \( x_{ij} = \mu + g_i + g_j + s_{ij} + r_{ij} + \frac{1}{b} \sum_{k} b_k + \frac{1}{b} \sum_{k} (bv)_{ijk} + \frac{1}{bc} \sum_{k} \sum_{l} e_{ijkl} \)

Method 4

Modelo I: \( x_{ij} = \mu + g_i + g_j + s_{ij} + \frac{1}{bc} \sum_{k} \sum_{l} e_{ijkl} \) for \( i,j =1,\ldots,p \), \( k=1,\ldots,b \), \( l=1,\ldots,c \)

Modelo II: \( x_{ij} = \mu + g_i + g_j + s_{ij} + \frac{1}{b} \sum_{k} b_k + \frac{1}{b} \sum_{k} (bv)_{ijk} + \frac{1}{bc} \sum_{k} \sum_{l} e_{ijkl} \)

Applications in the four methodologies

Method 1 has been used more frequently in plant breeding to estimate effects and variances of combining ability, reciprocal and maternal effects, gene action, heterosis, heritability, prediction of outstanding hybrids and response to selection, as can be seen in the studies carried out by González et al. (2007b); Amissah et al. (2019); Lima et al. (2019); Yaw et al. (2020), among others. Estimation of genetic parameters with method 2 in potato (Solanum tuberosum L.); corn (Zea mays L.); gerbera (Gerbera x hybrida), among other species, have been made by Muhiyyuza et al. (2016); Oliveira et al. (2016); Mbusa et al. (2017); Rivera-Colín et al. (2019); Muhumuza et al. (2020). This has been used to estimate combining ability, heterosis, heritability, gene action, inbreeding depression, and prediction of superior hybrids.

With method 3, Rodríguez-Pérez et al. (2016); Ngali et al. (2019); Barreta et al. (2019); Vesali et al. (2020) have estimated genetic parameters in corn, sweet potato (Ipomoea batatas L.) and potato, but in these, heterosis was not estimated because the analysis does not include parents. In the studies of Muhinyuza et al. (2016); Anyanga et al. (2016); Barroso et al. (2019); Aditika et al. (2020); Vasconcelos et al. (2020), the relevance of method 4 in the estimation of genetic parameters has been highlighted, in these, priority has been given to the analysis and discussion of the effects of general and specific combining ability, heritability in a broad and narrow sense, and in the prediction of hybrids of higher production.

Available software

Some statistical packages that are available on the Internet, for the analysis of experiments of diallel crosses, are: statistical analysis system (SAS), Agrobase, Indostat, AGD-R (analysis of genetic design in R), PBTools (plant breeding tools in R), TNAUSTAT (plant breeding-heterosis), OPStat, and Genes, among others.
Results and discussion

Plant breeding can be divided into three stages: the assembly or creation of a variable germplasm pool, the selection of superior individuals in the pool and their use to create a new and better variety. In these, the estimation of genetic parameters, such as genetic variance and heritability, can be of great value (Dudley and Moll, 1969). In genetic and plant breeding studies, the mating and environmental (experimental) design, as well as the sowing arrangement selected are very important to generate the type of family required, to reduce the residual of the mathematical model, to optimize the time allocated to the trial, and to estimate with greater precision the population parameters that allow obtaining correct conclusions (Sahagún, 1998; Mumtaz et al., 2015; Harriman and Nwammadu, 2016; Awata et al., 2018).

Nduwumuremyi et al. (2013) pointed out that the selection of an appropriate group of parents and genetic design are key elements in developing successful plant breeding programs and that these are conditioned by the objectives of the study, time, space, costs and other biological constraints. Regarding the mating design, the predominant type of pollination (by wind or insects), the form of dissemination of pollen (by wind or insects), whether there is genetic or cytoplasmic sterility, the objectives of the study, and the size of the required population, among others, should be considered (Nduwumuremyi et al., 2013; Mumtaz et al., 2013; Fasahat et al., 2016).

Since the 1950s, various diallelic techniques have been designed but the most used are those of Griffing, Hayman and Jinks, and Gardner and Eberhart (Christie and Shattuck, 1992; Shattuck et al., 1993; Hallauer et al., 2010; Nduwumuremyi et al., 2013). Cockerham (1963) classified mating designs into categories of one, two, three, and four factors, depending on the number of ancestors by progeny over which control is exercised. Polycrossing (a factor) is sufficient to detect genetic variability; with the diallel cross, with North Carolina designs I, II and III and with the partial diallelic (two factors), additive and dominance variance can be estimated, epistasis or the linkage between genes can be calculated using triallel and quadrallel crosses (three and four factors), or with the combination of any of these designs.

Fasahat et al. (2016) commented that the genetic designs that have been most frequently used in recent decades are diallel cross, the North Carolina factorial design (method II) and the line x tester cross (mestizos). In these, the results have been discussed based on the estimation of the effects and variances of general and specific combining ability in crops such as wheat, safflower (Carthamus tinctorius L.), rice (Oryza sativa L.), sorghum (Sorghum bicolor (L) Moench), alfalfa (Medicago sativa L.), chili (Capsicum annum L.), cowpea (Vigna unguiculata L.), tomato (Solanum lycopersicum L.), sesame (Sesamum indicum L.) and flaxseed (Linum usitatissimum L.), among others.

The evaluation of the progenies of a diallel cross in several years and/or localities using an appropriate design and genetic-statistical analysis could be complex, but essential, to determine their genetic, environmental and genotype-environment interaction structure, their contribution in estimating the effects and variances of combining ability, heterosis, heritability, response to the selection and prediction of outstanding hybrids (Sahagún, 1990; González et al., 2007a, b; Harriman and Nwammadu, 2016; Fasahat et al., 2016; Awata et al., 2018).
The components of genetic variance estimated by mating designs can be equated with covariances between relatives. The only necessary assumptions are that the parents are random members of the genetic population and that the experimental errors are independent. Environmental correlations are avoided by randomizing progenies; a negative variance can be attributed to a deviation caused by the type of sampling and by the sample size (Dudley and Moll, 1969).

The analysis of variance (ANOVA) technique is used to divide the variability estimated in one or more experiments into different components according to the genetic-statistical model applied to quantitative variables. It is also used in qualitative traits after transforming the original data so that they meet the assumptions that the genotypic and environmental effects are additive, and that the experimental or residual errors of the models are distributed normally and independently with a common variance (Sahagún, 1990; Sahagún, 1998). If the differences between genotypes are significant \( p \leq 0.05 \), the diallelic analysis is performed with any of the methodologies considered in this or in other studies (Griffing, 1956b; Shattuck et al., 1993; González et al., 2007a, b; Pérez et al., 2020).

González et al. (2007b); Pérez et al. (2020) divided the variability of a complete diallelic formed with inbred corn lines into partial sums of squares and mutually orthogonal contrasts to test hypotheses about parents (P), direct crosses (DC), reciprocal crosses (RC), P versus CD and DC versus RC, when their effects were significant, Griffing’s (1956b) method 1 was applied, as suggested by Shattuck et al. (1993). The first contrast calculates mean heterosis, also estimable in ANOVA with Gardner and Eberhart’s (1966) method, but with none of Griffing (1956b) and the second contrast indirectly calculates maternal and/or reciprocal effects, also estimable with ANOVA with Griffing’s (1956b) methods 1 and 3.

Singh (1973a, b) developed the statistical procedures for the analysis of data from a diallel cross in various environments, and Zhang and Kang (1997) and Zhang et al. (2005) showed how to estimate reciprocal effects using SAS with emphasis on its partition in maternal and non-maternal effects in series of experiments, as suggested by Cockerham (1963). They commented that this type of partition allows knowing if maternal effects or extranuclear factors are involved in the expression of a phenotypic trait.

Kearsey (1965) derived families from the same population using crosses between two parents, North Carolina designs I and II, Hayman’s (1954) half diallelic and Kempthorne and Curnow’s (1961) partial diallelic. The only restriction was that for the estimation of the components of variance, the number of crosses in each methodology, including self-fertilizations, was equal to or close to 45. It was concluded that Hayman and Jinks’s (1954) method provides more information than any other design but requires that a greater number of genetic assumptions be met. Baker (1978) grouped Gardner and Eberhart’s (1966) methods II III, Griffing’s (1956b) methods 2 and 4 and Hayman’s (1954) to discuss the problems that arise in a diallelic analysis, from the statistical point of view, there are critical controversies about the selection of a fixed and random model of genotypic effects, and from the genetic point of view, assumptions about the independent distribution of genes in parents are difficult to meet and the hypothesis of the absence of epistasis and linkage of genes may also be incorrect.

Griffing’s (1956b) methods have been preferred to those of Hayman and Jinks and those of Gardner and Eberhart (1966) because they are not conditioned by genetic assumptions. In addition, the analysis, if specialized software is not available, is more complex in the last two. Griffing’s (1956b)
methods and Gardner and Eberhart’s (1966) methods can be applied in autogamous and allogamous species, including polyploids such as potato (Solanum tuberosum L.) and those of Hayman and Jinks (1954) are restricted to homozygous or inbred parents (Baker, 1978; Christie and Shattuck, 1992; Awata et al., 2018). Christie and Shattuck (1992) commented that Gardner and Eberhart’s (1966) analyses are like those of Hayman (1954a, b) and one of Griffing (1956b). Zhang and Kang (1997) analyze other similarities and differences that exist between Griffing’s four methods.

The main limitation shown by Griffing’s (1956a, b) methodologies has to do with the number of parents selected: with four parents there are three degrees of freedom that do not allow the hypothesis for GCA to be reliably estimated, and with 20 of them there will be greater precision in its estimation, but an unmanageable trial (400 crosses) will be generated if method 1 is chosen, regardless of the application of model I or II (González et al., 2007a, b; Awata et al., 2018; González et al., 2020; Pérez et al., 2020). Christie and Shattuck (1992) commented that most of the diallelic experiments they reviewed included between 6 and 10 parents.

Using a larger number of parents will allow more reliable information to be obtained, but their optimal number should be such as to justify the additional information obtained and it is necessary to define the reference population in advance. When parents were considered as a sample from an ancestral population, between 17 and 19 parents or between three and five trials with eight of them each were suggested (Hayward, 1979). In the investigations that were considered in the present study, it was observed that the number of parents for Griffing’s (1956b) methods 1, 2, 3 and 4 varied from 4 to 12, 5 to 12, 7 to 13, and 4 to 16, respectively. Fasahat et al. (2016) reported that for the diallelic, the number of parents tested varied from 4 to 15 and for the half diallelic, it ranged from 5 to 12.

Griffing (1956b) discussed eight analyses that originate from combining his four methods with the two assumptions made regarding the sampling nature of experimental material (models I and II). Shattuck et al. (1993) highlighted that there are 16 possible analyses, if in addition to the above, mixed models A and B are included. There would be more combinations if one considers the series of experiments in time and space or the arrangements of split plots tested under the completely randomized experimental designs, randomized complete blocks or some lattice (Gomez and Gomez, 1984; Sahagún, 1998; Moore and Dixon, 2014; Pérez et al., 2019).

The selection of any of Griffing’s (1956b) methodologies depends on the type of experimental material considered and the objectives of the research. When only F1 crosses are of interest, the application of methods 3 and 4 is more appropriate to estimate the effects (model I) or variances (model II) of GCA and SCA. If there is doubt regarding the presence of genes linked to sex or maternal effects, method 3 is the most appropriate. If the main objective is to identify outstanding lines that will be used in the formation of a synthetic variety, the parents should be included in the experiment and methods 1 or 2 will be used for their genetic-statistical analysis (Griffing, 1956 b; Baker, 1978; Christie and Shattuck, 1992; Shattuck et al., 1993; González et al., 2007a, b).

GCA is mainly related to additive gene effects and SCA to dominance or epistasis (Sprague and Tatum, 1942; Dudley and Moll, 1969; Baker, 1978; Christie and Shattuck, 1992; Awata et al., 2018). In the different selection techniques, the effects of additivity are more important, and in the
hybridization programs, the non-additive effects are essential (Borojevic, 1990; Mumtaz et al., 2015; Harriman and Nwammadu, 2016; Awata et al., 2018). Regarding the prediction of F₁ hybrids, Dudley and Moll (1969) pointed out that this is only justified when the proportion of the dominance variance to the additive variance is in the overdominance range or when estimates of the epistatic variance including dominance variance are obtained. They do not recommend the use of some type of design in lattice because in these, genetic and environmental differences are confused in the means, so these are not satisfactory.

In open-pollinated species (allogamous plants), when self-fertilization causes inbreeding depression that causes high plant mortality, as often happens in alfalfa (Medicago sativa L.), studies of general and specific combining ability could be restricted to the use of non-inbred or partially inbred parents, with a single generation of self-fertilization; if F, the inbreeding coefficient, is equal to or less than 0.75, the diallel cross is not recommended. In the latter situation and when the main objective is to estimate the effects of GCA, the methods of polycrossing, line x tester and partial diallelic could be the most appropriate (Griffing, 1956b; Christie and Shattuck, 1992; Shattuck et al., 1993; González et al., 2007a).

In husk tomato (Physalis ixocarpa Brot. ex Horm.), there is gametophytic incompatibility caused by two independent genes with multiple alleles that prevent the formation of highly inbred lines and hybrids using conventional plant breeding techniques (Mulato et al., 2020) so, with this biological constrain, it would be more advisable to apply some selection technique (Peña and Márquez, 1990), practice intervarietal hybridization with families of maternal half-siblings (Peña et al., 1999), or plant-to-plant crosses (Santiaguillo et al., 2004).

None of Griffing’s (1956b) methods considers epistasis and linkage; for its estimation, one of these must be used together with Hayman’s (1954) methodology or if the objective was also to estimate heterosis, both techniques together with that of Gardner and Eberhart (1966), even if they would increase the costs and time allocated to the evaluation of a group of diallel crosses, they would be the most desirable to carry out a comprehensive evaluation (Hallauer et al., 2010; Muntaz et al., 2015; Awata et al., 2018; Pérez et al., 2020).

For SAS, the codes and programs for the methodologies of Griffing (1956b), Hayman (1954) and Gardner and Eberhart (1966), among others, have been designed to analyze data with completely randomized experimental designs, randomized complete blocks and some lattice in one or more factors (Martínez, 1988; Kang and Zhang, 1977; Zhang et al., 2005; Makumbi et al. (2018), if one does not have the resources to purchase a commercial license, a trial academic version may be downloaded, valid for up to one year.

AGD-R, PBTTools and Genes are also good choices; these can be downloaded free of charge from the websites of the International Maize and Wheat Improvement Center (CIMMYT, for its acronym in Spanish; Rodríguez et al., 2018), from the International Rice Research Institute (IRRI, 2014) and the Federal University of Viçosa (Brazil; Cruz, 2013), respectively. Agrobase and Indostat can only be used with a commercial license, but both are friendlier than previous software. TnauStat and OPStat, available free of charge on their websites, only analyze experiments for one environment (year or locality). In Mastache et al. (1999a, b); Mastache and Martínez (2003); Sestras et al. (2018); González et al. (2020) are considered codes, programs and algorithms for Griffing’s analysis with SAS or other software.
Conclusions

The selection of suitable parents and a good mating and experimental design are the keys to achieving a successful breeding scheme. From a statistical point of view, the critical part of the use of diallelic designs would be the selection of a fixed or random effect model, from the genetic point of view, to interpret the results, the assumption of the independent distribution of genes in the parents is not very acceptable in practice, the second assumption, the absence of epistasis and linkage between genes, is sometimes also incorrect. Epistasis affects the estimation of GCA and SCA variances. As an alternative to the genetic interpretation of the results, the statistical description of the diallelic analysis can be used to answer questions concerning the importance of SCA and the predictability of hybrid behavior using GCA or the behavior of the parents.

Cited literature


