Theor Appl Genet (1994) 89: 693–697

Y. Naciri · A. Gallais

Predicting mean and variance of all possible lines and hybrids from designs with partially inbred progenies

Received: 18 January 1994 / Accepted: 29 April 1994

Abstract Two-factor mating designs at consecutive S_n and $S_{(n+1)}$ levels (S_0 and S_1, S_1 and S_2, or F_2 and F_3) allow estimation of all components of the variation among homozygous lines and F_1 hybrids that can be derived from a given population. They also allow for the prediction of the mean of these lines and single-cross hybrids. Some tests for the presence of epistasis are possible at the levels of means and of variances. Such mating designs can be very useful for predicting the value of the best possible lines or the best possible F_1 hybrids when it is difficult to produce, at an experimental level for exploratory purposes, either lines or hybrids.

Key words Two-factor mating design
Variance components · Inbred relatives · Epistasis

Introduction

Since Cockerham's (1954) and Kempthorne's (1957) major contributions to quantitative genetics, progenies derived from mating designs have been widely used to estimate genetic variance components in random mating populations. In the absence of epistasis, these designs provide for an estimation of additive and dominance variances. For breeders who improve *per se* or combining ability values of their material, such components have implications in choosing the more efficient breeding method and in predicting the expected genetic advance in recurrent selection. However, for the development of completely inbred lines, additive and dominance variances are not sufficient because they do not describe all of the variation involved at the homozygous level. At is this level, besides the additive variance, two additional parameters must be considered: the covariance between additive effect and interaction effect between identical alleles, and the variance of interaction between identical alleles (Gillois 1964; Harris 1964; Gallais 1970, 1974; Cornelius 1975). According to Gallais (1988), it is possible to estimate such parameters from two-factor mating designs among S_0 plants including $S_0 \times S_0$ crosses and the self (S_1) of the parents. The simultaneous study of $S_0 \times S_0$ crosses and of S_1 progenies does provide estimation of the additive and dominance variances and of the two additional components for variance among lines. The reconstruction of total genetic variance among lines and among hybrids is then possible. Such designs are well adapted for species where it is easy to test $S_0 \times S_0$ crosses. This will not be the case, in particular for self-pollinating species, when $S_0 \times S_0$ crosses give reduced seed numbers that prevent any direct cross performance study.

In this paper, a two-factor (factorial, diallelic or hierarchical) mating design among S_0 plants is proposed, for which each cross among S_0 plants is used to produce S_1 and S_2 generations in bulk. In the absence of epistasis, combined analysis of the two-factor mating designs at the S_1 and S_2 levels allows estimation of the three variance components among homozygous lines and the homozygous line mean. In the same way, this design provides an estimate of dominance variance and of the F_1 hybrid mean . In the presence of epistasis restricted to the additive×additive type, it is also possible to estimate the associated component of variance. The mean and variance of all possible F_1 hybrids can then be compared to those of homozygous lines. This approach can be useful for choosing between line or hybrid developments.

Variance component estimation

The aim is to estimate the four variance components involved in the variances among homozygous lines and F_1

Communicated by A. R. Hallauer

Y. Naciri[1]
IUT Louis Pasteur, Département de Biologie Appliquée, 3, rue de l'Argonne 67 000 Strasbourg. France

A. Gallais (✉)
Ferme du Moulon, INA PG-INRA-UPS, 91190 Gif sur Yvette, France

Present address:
[1] IFREMER, Unité de Recherche en Génétique et Ecloserie, BP 133, Ronce les Bains 17 390 La Tremblade, France

hybrids. Using the notation introduced by Kempthorne (1957) and already used by Gallais (1974), in the absence of epistasis we can write $\sigma^2_L=2\sigma^2_A+2\sigma_{ADo}+\sigma^2_{Do}$ for the variance among all lines that can be derived from a random mating population and $\sigma^2_H=\sigma^2_A+\sigma^2_D$ for the variance among all single-cross hybrids.

σ^2_A is the additive variance, σ^2_D the dominance variance, σ^2_{Do} the homozygous dominance variance, and σ^2_{ADo} the covariance between additive effect and homozygous dominance effects.

Definitions and equivalents for the notation of parameters in terms of the Q-model (Cornelius and Van Sanford 1988), C-model (Cornelius 1975; Cornelius and Dudley 1976), and D-model (Cockerham 1983) are given in Appendix 1. The parameter $\Sigma_l \mathrm{Do}^2=\Sigma_l [\mathrm{E}\ (\beta_{ii})_l]^2$, as already noticed by Wright and Cockerham (1986), is not involved in the variance among bulked partially inbred families derived from S_0. Therefore, there are only four parameters and not five, as in the general case of covariances among inbred relatives (Gillois 1964; Harris 1964). From the point of view of variance component estimation, the situation considered in this paper is simpler than the general one considered by Gallais (1976), Cornelius and Dudley (1976), Cornelius (1988), and Cornelius and Van Sanford (1988).

Consider a two-way mating design at the S_1 and S_2 levels. The cross of two S_0 plants gives a full-sib family that is evaluated for its value at the S_1 and S_2 levels. This evaluation is equivalent to the evaluation of the full-sib family for two specific characters, the S_1- and S_2-value (Gallais 1990b). It is then possible to estimate the covariances among full-sibs and among half-sibs as with the test of $S_0\times S_0$ crosses. According to Gallais (1988, 1990a), in the absence of epistasis it is possible to write directly:

$$\begin{aligned}
&\mathrm{Cov}\,(FS)_{S1}=1/2\ \sigma^2_{AS1}+1/4\ \sigma^2_{DS1}\\
&\mathrm{Cov}\,(HS)_{S1}=1/4\ \sigma^2_{AS1} \qquad (1)\\
&\mathrm{Cov}\,(FS)_{S2}=1/2\ \sigma^2_{AS2}+1/4\ \sigma^2_{DS2}\\
&\mathrm{Cov}\,(HS)_{S2}=1/4\ \sigma^2_{AS2}
\end{aligned}$$

with σ^2_{ASn} (n=1 or 2) and σ^2_{DSn} being the additive and dominance variances for genotypic values at the S_n level.

This approach allows condensed expressions of $\mathrm{cov}(FS)_{Sn}$ and $\mathrm{cov}(HS)_{Sn}$. Obviously, to go further it is necessary to express each component of the variance in terms of elementary components defined for covariances between inbred relatives. With α_i the additive effect of allele A_i and β_{ij} the dominance effect for alleles A_i and A_j, the additive effect defined for bulked inbred progenies at the S_n level is given by: ${}_{Sn}\alpha_i=\alpha_i+1/2(1-1/2^n)\ \beta'_{ii}$, where $\beta'_{ii}=\beta_{ii}-\mathrm{E}(\beta_{ii})$, with α_i being the additive effect defined in random mating population, and β_{ii} the interaction between identical alleles ii, i.e., homozygous dominance. E( ) means expectation. Concerning the dominance effects, ${}_{Sn}\beta_{ij}$, it is straightforward that they are halved at each generation of selfing, then ${}_{Sn}\beta_{ij}=1/2^n\ \beta_{ij}$. Finally, it results:

$$\sigma^2_{ASn}=\sigma^2_A+1/2\ (1-1/2^n)^2\sigma^2_{Do}+(1-1/2^2)\ \sigma_{ADo} \qquad (2)$$

and,

$$\sigma^2_{DSn}=(1/2)^{2n}\ \sigma^2_D \qquad (3)$$

Covariance between the S_1 and S_2 generations can also be computed to give a new set of two equations:

$$\begin{aligned}
&\mathrm{Cov}[(FS)_{S1},\ (FS)_{S2}]=1/2\ \sigma_{AS1\,AS2}+1/4\ \sigma_{DS1\,DS2} \qquad (4)\\
&\mathrm{Cov}[(HS)_{S1},\ (HS)_{S2}]=1/4\ \sigma_{AS1\,AS2}
\end{aligned}$$

with:

$$\sigma_{DS1\,DS2}=1/8\ \sigma^2_D \qquad (5)$$

$$\begin{aligned}
\sigma_{AS1\,AS2}&=2\Sigma\,\mathrm{E}(\alpha_i+1/4\ \beta'_{ii})\ (\alpha_i+3/8\ \beta'_{ii}) \qquad (6)\\
&=\sigma^2_A+5/8\ \sigma_{ADo}+3/16\ \sigma^2_{Do}\ .
\end{aligned}$$

General expressions at the n level can be given for $\sigma_{ASnAS(n+1)}$ and $\sigma_{DSnDS(n+1)}$:

$$\sigma_{DSnDS(n+1)}=(1/2)^{(2n+1)}\ \sigma^2_D \qquad (7)$$

$$\begin{aligned}
\sigma_{ASnAS(n+1)}=\sigma^2_A&+[1/2+3/2^{(n+2)} \qquad (8)\\
&+1/2^{(2n+2)}]\ \sigma_{ADo}+[1-3/2^{(n+2)}]\ \sigma^2_{Do}
\end{aligned}$$

Table 1 summarizes the coefficients for variance components, including those for additive×additive epistasis, involved in the six covariances: $\mathrm{cov}(FS)_{S1}$, $\mathrm{cov}(HS)_{S1}$, $\mathrm{cov}(FS)_{S2}$, $\mathrm{cov}(HS)_{S2}$, $\mathrm{cov}(FS)_{S1,S2}$, $\mathrm{cov}(HS)_{S1,S2}$. Such coefficients could have been directly derived by computing the kinship coefficients that are involved in covariance among inbred relatives (Gillois 1964; Harris 1964; Gallais 1970, 1974). In particular, the σ^2_A coefficient is twice the Malécot coefficient of kinship. Note that if linkage is supposed to be absent, coefficients of covariance between relatives for σ^2_{AA} correspond to the square of σ^2_A coefficients.

Assuming the absence of epistasis, the four parameters (σ^2_A , σ^2_{ADo}, σ^2_{Do} and σ^2_D) can be estimated using iterative weighted least square or maximum likelihood procedures, as there are six equations for four parameters. In addition, we notice that:

$$\mathrm{Cov}\,(HS)_{Sn}=\sigma^2_{gSn} \qquad (9)$$

$$\mathrm{Cov}\,(FS)_{Sn}=2\ \sigma^2_{gSn}+\sigma^2_{sSn} \qquad (10)$$

with σ^2_{gSn} and σ^2_{sSn} being the general combining ability (GCA) and the specific combining ability (SCA) variances, respectively at the S_n level. The Table 1 matrix can be transformed into the matrix given in Table 2. From Table 2, it is now clear that, in the absence of epistasis, the last three rows are only relative to the σ^2_D estimation. Thus three independent estimates of σ^2_D are possible, and one estimate of σ^2_D is a weighted mean of these three estimates. However, the best estimates will be by a simultaneous estimation of the four parameters.

In the presence of epistasis, five or even six parameters could be estimated. However, it can be predicted that the accuracy of estimates will be generally poorer than by neglecting epistasis.

Two particular designs

Two particular designs are interesting to consider: the first is the one developed by the simultaneous study of $(S_0\times S_0)$

crosses and their S_1 bulks, and the second is that developed by the study of F_2s and F_3s derived from F_1 crosses between completely inbred lines. Coefficients of variance and covariance components are given in Table 3. In both cases the four parameters (σ^2_A, σ^2_{ADo}, σ^2_{Do} and σ^2_D) can always be estimated. The second design is useful to estimate $\sigma^2_H=\sigma^2_A+\sigma^2_D$ (+ σ^2_{AA}) when it is not possible to produce enough seeds for F_1 tests.

Table 1 Variance component coefficients for covariances that can be estimated with a two-factor mating design at S_1 and S_2 levels

	σ^2_A	σ^2_{ADo}	σ^2_{Do}	σ^2_D	σ^2_{AA}
$Cov(FS)_{S1}$	1/2	1/4	1/16	1/16	1/4
$Cov(HS)_{S1}$	1/4	1/8	1/32	0	1/16
$Cov(FS)_{S2}$	1/2	3/8	9/64	1/64	1/4
$Cov(HS)_{S2}$	1/4	3/16	9/128	0	1/16
$Cov(FS)_{S1,S2}$	1/2	5/16	3/32	1/32	1/4
$Cov(HS)_{S1,S2}$	1/4	5/32	3/64	0	1/16

Table 2 Variance component coefficients for σ^2_{gSn}, σ^2_{sSn}, [with n=(1, 2)], $\sigma_{gS1,gS2}$ and $\sigma_{sS1,sS2}$

	σ^2_A	σ_{ADo}	σ^2_{Do}	σ^2_D	σ^2_{AA}
σ^2_{gS1}	1/4	1/8	1/32	0	1/16
$\sigma_{gS1,gS2}$	1/4	5/32	3/64	0	1/16
σ^2_{gS2}	1/4	3/16	9/128	0	1/16
σ^2_{sS1}	0	0	0	1/16	1/8
$\sigma_{sS1,sS2}$	0	0	0	1/32	1/8
σ^2_{sS2}	0	0	0	1/64	1/8

Table 3 Variance component coefficients for σ^2_{gSn}, σ^2_{sSn} [for n=(0, 1)], $\sigma_{gS0,gS1}$ and $\sigma_{sS0,sS1}$ expressions and for σ^2_{gFn}, σ^2_{sFn} [for n=(2, 3)], $\sigma_{gF2,gF3}$ and $\sigma_{sF2,sF3}$ expressions

	σ^2_A	σ_{ADo}	σ^2_{Do}	σ^2_D	σ^2_{AA}
σ^2_{gS0}	1/4	0	0	0	1/16
σ^2_{gF2}	1/2	1/4	1/16	0	1/4
$\sigma_{gS0,gS1}$	1/4	1/16	0	0	1/16
$\sigma_{gF2,gF3}$	1/2	5/16	3/32	0	1/4
σ^2_{gS0}	1/4	1/8	1/32	0	1/16
σ^2_{gF3}	1/2	1/4	1/16	0	1/4
σ^2_{sS0}	0	0	0	1/4	1/8
σ^2_{sF2}	0	0	0	1/4	1/2
$\sigma_{sS0,sS1}$	0	0	0	1/16	1/8
$\sigma_{sF2,sF3}$	0	0	0	1/8	1/2
σ^2_{sS1}	0	0	0	1/16	1/8
σ^2_{sF3}	0	0	0	1/16	1/2

Prediction of the best lines and the best hybrids

Estimation of variances among lines and among single-cross hybrids

Estimates for the variance among lines (σ^2_L) and for the variance among single-cross hybrids (σ^2_H) can be deduced from the previous estimates of variance components:

$$\sigma^2_L=2\ \sigma^2_A+2\ \sigma_{ADo}+\sigma^2_{Do} \qquad (11)$$

$$\sigma^2_H=\sigma^2_A+\sigma^2_D\ . \qquad (12)$$

Estimation of line and hybrid means

S_1 and S_2 family means can be written:

$$\mu_{S1}=1/2\ \mu_H+1/2\ \mu_L \qquad (13)$$

$$\mu_{S2}=1/4\ \mu_H+3/4\ \mu_L$$

where μ_H is the mean of the random mating population (i.e. the mean of all possible F_1 hybrids among lines), and μ_L the mean of all lines. Solving Eq. 13 for μ_H and μ_L gives:

$$\mu_L=2\ \mu_{S2}-\mu_{S1} \qquad (14)$$

$$\mu_H=3\ \mu_{S1}-2\ \mu_{S2} \qquad (15)$$

For a given cross among two S_0 plants i and j, the value of all lines [$L_{(ij)}$] that can be derived from i is:

$$L_{(ij)}=2\ S2_{(ij)}-S1_{(ij)} \qquad (16)$$

where $S1_{(ij)}$ ans $S2_{(ij)}$ are the means of bulked S1 and S2 from the cross $S_{0(ij)}$; the value of a cross $C_{(ij)}$ between S_0 plants is:

$$C_{(ij)}=3\ S1_{(ij)}-2\ S2_{(ij)} \qquad (17)$$

The last formula is not very useful for S_0 plants but becomes so when inbred lines are used as parents. It then allows for the prediction of the F_1 hybrids' values. Such predictions are very useful when it is not possible to produce enough seeds for F_1 tests. It must be noted that Eqs. 14, 15, 16 and 17 remain valid even when additive×additive epistasis is present.

From the prediction formulae for the means and variances of distribution of all possible lines and F_1 hybrids, it is then possible to predict the expected mean of the p% selected lines, L_{max}, and of the p% selected F_1 hybrids H_{max} by:

$$L_{max}=\mu_L+i\ h_L\ \sigma_L \qquad (18)$$

$$H_{max}=\mu_H+i\ h_H\ \sigma_H \qquad (19)$$

where i is the selection intensity corresponding to the rate of selection p, and h_L and h_H are the square root of broad-sense heritabilities associated to the lines and to the single-cross hybrids, respectively.

Investigating the presence of epistasis

At the level of variances

At the level of variances, testing for additive×additive epistasis is possible by comparing σ^2_{sS1} and 4 σ^2_{sS2}. If

(σ^2_{sS1}–4 σ^2_{sS2}) is significantly different from 0, it can be concluded that epistasis is present. Another contrast could be (2 σ^2_{sS1}–σ_{sS1S2}–6 σ^2_{sS2}). The test of epistasis could be also obtained by fitting the model by iterative weighted least squares or by maximum likelihood, without and with the additive×additive component: in large samples the ratio of the weighted residual sums of squares from the two models is approximately distributed as chi-square with one degree of freedom.

Another method for investigating the presence of epistasis can be proposed. It is interesting to note that when mating designs are considered at a given S_n level, the σ^2_D coefficient tends toward 0 with increasing n values, and so does σ^2_s in the absence of epistasis. In the particular case of doubled haploid lines (S_∞) when epistasis is absent, specific combining ability (SCA) variance will be 0 (Gallais 1990 c). An estimate significantly different from 0 would then indicate the occurrence of epistatic effects.

At the level of means

The design described in this paper provides a test of epistasis at the level of general means or at the level of each cross. With the design involving S_1 and S_2 families, and when the S_0 plants (or plants from $S_0 \times S_0$ crosses) can be studied simultaneously, a test of epistasis at the level of general means is given by (according to Eq. 13):

$$3\ \mu_{S1}-2\ \mu_{S2}-\mu_H=0 \qquad (20)$$

or

$$3\ \mu_{S1}-2\ \mu_{S2}-\mu=0 \qquad (21)$$

with μ being the mean of the random mating population. Note that the test according to Eq. 21 will detect only the epistasis affecting the mean, i. e., the average effect of dominance×dominance epistasis at the homozygous level. Therefore, if Eq. 21 is confirmed, this does not mean the absence of epistasis.

It is also possible to test the presence of epistasis at the level of each cross by using a similar contrast to Eq. 21:

$$3\ S_{1(ij)}-2\ S_{2(ij)}-C_{ij}=0 \qquad (22)$$

When parents are lines, it is easier to study simultaneously parents, F_2, and F_3 families, or parents and F_1 and F_2 families. The test for epistasis is then the one given by Hayman (1958):

$$2\ \bar{F}_3-\bar{F}_2-\bar{P}=0 \qquad (23)$$

$$2\ \bar{F}_2-\bar{F}_1-\bar{P}=0 \qquad (24)$$

or:

$$2\ \bar{F}_3-3\ \bar{F}_2+\bar{F}_1=0\ . \qquad (25)$$

Conclusion

The mating design presented in this paper can be very useful for the simultaneous estimation of components for variance among lines and variance among single-cross hybrids. Such a mating design is interesting for cross-pollinating species when it is both difficult to produce crosses with enough seeds and time consuming and expensive to develop completely inbred lines (i.e., by SSD or HD). It is also interesting for self-pollinating species where lines exist or are easy to produce, but where single-cross hybrids can also be difficult to produce with enough seeds. The main problem will be the accuracy of the estimates. Such a problem has not been considered . However, it is clear that the accuracy will be determined by the accuracy for the estimates of the covariances between full-sibs and between half-sibs evaluated at the S_1 and S_2 levels. From this point of view a diallel mating design, with only one set, will be a poor design; however with several small disconnected sets of parents it could be a good mating design. Similarly, a factorial design (or design II NC) with several disconnected sets could be an efficient mating design, and is in particular more efficient than the nested design (design I NC) for the same number of developed and tested crosses.

Acknowledgments The authors are especially grateful to one of the reviewers for his helpful criticism and suggestions.

Appendix 1

Author's notation	C-model[a]	D-model[b]	Q-model[c]
$\sigma^2_A=2\ \Sigma_l\,[E(\alpha_i)^2_l]$	σ^2_A	σ^2_A	σ^2_A
$\sigma_{ADo}=2\ \Sigma_l\,[E(\alpha_i\,\beta_{ii})_l]$	$C-2\,\sigma^2_A$	$2D_1$	$(Q_{yy}$-$Q_{xx})/2-\sigma^2_A$
$\sigma^2_{Do}=\Sigma_l\,[E(\beta_{ii}-E(\beta_{ii}))^2_l]$	$\sigma^2_\infty-2\,C+2\,\sigma^2_A$	D^*_2	Q_{xx}
$\sigma^2_D=\Sigma_l\,[E(\beta_{ij})^2_l]$	σ^2_D	σ^2_D	σ^2_D
$\Sigma_l\,Do^2=\Sigma_l\,[E(\beta_{ii})_l]^2$	μ_∞	H^*	$(Q_{yy}+Q_{xx})/2$ -$Q_{xy}-\sigma^2_A$

[a] Cornelius 1975; Cornelius and Dudley 1975 1976

[b] Cockerham 1983

[c] Cornelius and Van Sanford 1988

Definitions and relationships of parameters in models for covariances of inbred relatives. α_i is the additive effect, β_{ii} is the interaction between two identical genes (homozygous dominance); Σ_l means the summation on the involved loci, and E( ) means expectation.

References

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859–882

Cockerham CC (1983) Covariances of relatives from self-fertilization. Crop Sci 23:1177–1180

Cornelius PL (1975) Identity coefficients and covariances between relatives in a parent-offspring inbreeding system. Theor Appl Genet 46:201–212

Cornelius PL (1988) Properties of components of covariance of inbred relatives and their estimates in a maize population. Theor Appl Genet 75:701–711

Cornelius PL, Dudley JW (1975) Theory of inbreeding and covariances between relatives under full-sib mating in diploids. Biometrics 31:169–187

Cornelius PL, Dudley JW (1976) Genetic variance components and predicted response to selection under selfing and full-sib mating in a maize population. Crop Sci 16:333–339

Cornelius PL, Van Sanford DA (1988) Quadratic components of covariance of inbred relatives and their estimation in naturally self-pollinated species. Crop Sci 28:1–7

Gallais A (1970) Covariances entre apparentés quelconques avec linkage et épistasie. Ann Genet Sél Anim 2:281–310

Gallais A (1974) Covariances between arbitrary relatives with linkage and epistasis in the case of linkage disequilibrium. Biometrics 30:429–446

Gallais A (1976) An experimental check of quantitative genetics on an autotetraploid plant, *Medicago sativa*, L., with special reference to the identity by descent relationship. In: Proc Int Conf Quant Genet, Ames, Iowa, Pollak E, Kempthorne O, Bacley TB, Iowa Univ Press, pp 519–540

Gallais A (1988) Prediction of mean and variance of hybrids and of lines that can be derived from a random mating population. Theor Appl Genet 75:625–627

Gallais A (1990a) Théorie de la sélection en amélioration des plantes. Masson Ed, Paris

Gallais A (1990b) Application of the concepts of test value and varietal value to the study of genetics in recurrent selection: a synthesis. Euphytica 48:197–207

Gallais A (1990c) Quantitative genetics of doubled haploid populations and application to the theory of line development. Genetics 124:199–206

Gillois M (1964) La relation d'identité en génétique. Thèse de la Faculté des Sciences, Paris

Harris DL (1964) Genotypic covariances between inbred relatives. Genetics 50:1319–1348

Hayman BI (1958) The separation of epistatic from additive and dominance variation in generation means. Heredity 12: 371–390

Kempthorne O (1957) An introduction to quantitative genetics. John Wiley and Son, NewYork

Wright AJ, Cockerham CC (1986) Covariances of relatives and selection response in generations of selfing from an outcrossed base population. Theor Appl Genet 72:689–694