Theor Appl Genet (1988) 75:625–627

Prediction of mean and variance of hybrids and of lines that can be derived from a random mating population

A. Gallais
CNRS-INRA-UPS, Ferme du Moulon, F-91190 Gif-sur-Yvette, France
and INA-PG, 16 rue Claude Bernard, F-75231 Paris Cedex 06, France

Received July 15, 1986; Accepted May 23, 1987
Communicated by R. Riley

Summary. A method is presented to estimate, from a two-factor crossing design including self-fertilization, mean and variance of lines and hybrids that can be derived from a random mating population. The derivation is only valid in the absence of epistasis. From such an estimation, it is possible to derive the expected value of the best lines and of the best hybrids that can be derived from a population.

Key words: Varietal development – Hybrid breeding – Line breeding – Quantitative genetics

Introduction

The theory of plant breeding is classically developed separately for cross-fertilized plants and for self-fertilized plants. For cross-fertilized plants, if heterosis is important the breeder has to develop hybrids; in self-fertilized plants, he frequently develops pure lines. However, the possibility of control of cross-fertilization in autogamous species, using male sterility or gametocides, gives the expectation of development of hybrids. In one generation of crossing, hybrids between lines allow for the accumulation in one genotype of dominant favorable genes, resulting in heterosis. This does not mean that all heterosis results from such a mechanism. In this case, it is conceivable to derive long term inbred lines similar to single crosses. The use of doubled haploids in recurrent selection may favor such development in some situations.

To answer the question of choice between lines or hybrids from a statistical point of view and for a given breeding material, we have to compare the expected value of the best hybrids to the expected value of the best lines. We will consider here only the case of hybrids and lines that can be derived from a random mating population. If we consider the distribution of the phenotypic values of all single crosses that can be derived from the population of lines, the expectation of the best possible single crosses could be:

$$\mu^*_{SC} = \mu_{SC} + i\, h_{SC}\, \sigma_{SC} \qquad (1)$$

i representing the theoretical selection intensity in standard units, μ_{SC} the mean of all single crosses without selection, σ^2_{SC} the genetic variance among all single crosses, and h^2_{SC} the broad sense heritability defined at the level of single crosses.

If we consider the population of lines that can be derived from the random mating population, the best possible lines will have the expected mean:

$$\mu^*_{L} = \mu_{L} + i\, h_{L}\, \sigma_{L} \qquad (2)$$

μ_L being the mean of all lines that can be derived from the population, or the line value of the population (Gallais 1978, 1979). The genetic variance among all lines that can be derived from the considered random mating population is σ^2_L and h^2_L is the broad sense heritability defined at the level of lines.

If the difference Δ between (1) and (2) is positive, it would be better to develop hybrids; if the difference Δ is negative, it would be justified to develop lines;

$$\Delta = (\mu_{SC} - \mu_L) + i\,(h_{SC}\,\sigma_{SC} - h_L\,\sigma_L) \qquad (3)$$

$\mu_{SC} - \mu_L$ represents the maximum inbreeding depression. In order to have an advantage of lines, the variance among lines must be greater than the variance among single crosses, particularly if the environmental effects are greater for lines than for hybrids. We consider the same selection intensity for both types of varieties. However, it could be different, according to the cost of production of the variety.

To answer the question of the choice between lines and hybrids we must estimate the six parameters μ_{SC}, σ_{SC}, h_{SC}, μ_L, σ_L and h_L. We consider here only the prediction of the means and of the variances. Since μ_{SC} and σ_{SC} can be easily estimated from a two-factor mating design between non-inbred plants, we will develop a formula to predict the mean and variance of all lines which can be derived from a random mating population. We will consider a two-factor mating design between non-inbred plants and progenies from one generation of self-fertilization (S_1) of parents used in the mating design.

Prediction formula for the distribution of single crosses

In the absence of selection during inbreeding or development of doubled haploids and with the assumption of linkage equilibrium, the mean of all single crosses is equal to the mean of the random mating population. The sample of gametes from the lines is the same as the sample of gametes from the random mating population:

$$\mu_{SC} = \mu$$

μ being the mean of the random mating population. Using the same assumptions, the genetic variance among single crosses is equal to the genotypic variance σ^2_G of the random mating population (Griffing 1956):

$$\sigma^2_{SC} = \sigma^2_G$$

and without epistasis:

$$\sigma^2_G = \sigma^2_A + \sigma^2_D$$

σ^2_A being the variance of additive effects and σ^2_D the variance of dominance effects defined in the random mating population (Kempthorne 1957). For estimating the variance among single crosses we therefore have to estimate the components σ^2_A and σ^2_D of the genotypic variance. This can be achieved through a two-factor mating design (hierachical or design I-NC, factorial or design II-NC and diallel). Such two-factor mating designs allow for estimation of covariances between full-sibs and between half-sibs from which σ^2_A and σ^2_D can be estimated as follows:

$$\text{cov FS} = 1/2\,\sigma^2_A + 1/4\,\sigma^2_D = 2\,\sigma^2_g + \sigma^2_s,$$

$$\text{cov HS} = 1/4\,\sigma^2_A = \sigma^2_g,$$

then:

$$\sigma^2_A = 4\,\sigma^2_g$$

$$\sigma^2_D = 4\,\sigma^2_s$$

σ^2_g and σ^2_s representing the variance of general and specific combining abilities.

Prediction formula for the distribution of lines

We consider the previous two-factor mating design with self-fertilization of the common parent in design I, and of all parents in design II or in diallel. We have two types of progenies: (1) progenies from crossing $S_0 \times S_0$ plants with full-sib and half-sib families, and (2) S_1 progenies. We will assume that S_1 and crosses are evaluated in the same experimental design with caution; due to border effects, S_1 can be grouped in each repetition. Their analysis of variance will be separated from that of crosses.

Prediction formula for the mean of all lines. The consideration of the mean of all S_1 plants allows for an estimation of the line value of the population; in the absence of epistasis there is a linear relationship between the mean of a population and its coefficient of inbreeding (Kempthorne 1957). Since the coefficient of inbreeding of S_1 plants is 1/2, the mean of all S_1 can be:

$$\mu_{S_1} = \mu + 1/2\,E(\beta_{ii}) \qquad (4)$$

$E(\beta_{ii})$ representing the expected value of the dominance residue for all homozygotes over the set of involved loci. The mean of all lines can also be deduced from the formula given by Kempthorne (1957), with a coefficient of inbreeding of 1:

$$\mu_L = \mu + E(\beta_{ii}). \qquad (5)$$

The result of (4) and (5) is:

$$\hat{\mu}_L = 2\,\hat{\mu}_S - \hat{\mu}$$

which is similar to the formula $2\,F_2 - F_1$ to predict the line value of a cross (Gallais 1978; Jinks 1976). Fundamentally, the basis is the same.

It is then possible to estimate the line value of a random mating population using S_1 families. This result is not new, but should be reconsidered in the present frame.

Prediction formula for the variance among lines. The genetic variance among lines can be expressed in terms of parameters being defined at the level of the random mating population, using the general formula of covariances between inbred relatives developed by Gillois (1964) and Harris (1964), and specified by Gallais (1970, 1974) and Cockerham (1983) for the case of self-fertilization. The variance can also be derived by using the classical genetical model. At one locus we can write for the value of a line $A_i A_i$:

$$Y_{ii} = \mu + 2\,\alpha_i + \beta_{ii}$$

α_i being the additive effect and β_{ii} the residue of dominance for the homozygous genotype $A_i A_i$.

The variance among lines will be the genotypic variance of a completely inbred population:

$$\begin{aligned}\sigma_L^2 &= \mathrm{E}\,[\mathrm{Y}_{ii} - \mathrm{E}\,(\beta_{ii})]^2 \\ &= 4\,\mathrm{E}\,(\alpha_i^2) + 4\,\mathrm{E}\,(\alpha_i\,\beta_{ii}) + [\mathrm{E}\,(\beta_{ii}^2) - (\mathrm{E}\,(\beta_{ii}))^2] \\ &= 2\,\sigma_A^2 + 4\,\sigma_{AD_0} + \sigma_{D_0}^2\end{aligned}$$

$\sigma_{AD_0} = \mathrm{E}\,(\alpha_i\,\beta_{ii})$ is the covariance additivity and dominance, $\sigma_{D_0}^2$ is the variance of quantities β_{ii}.

Without epistasis the summation is over the set of involved loci and the structure of the expression does not change, so we will omit this summation in the following; σ_A^2 is estimated by the crossing plan; it remains to estimate σ_{AD_0} and $\sigma_{D_0}^2$.

Consider the expression of the covariance between general combining ability effects and S_1 effects. It could also be deduced from the general expression of the covariance between inbred relatives of Gillois (1964) and Harris (1964). However, we have derived it directly. At one locus the value Y_{ij} of a genotype $A_i\,A_j$ in the random mating population can be decomposed according to Kempthorne's (1957) model:

$$Y_{ij} = \mu + \alpha_i + \alpha_j + \beta_{ij}\,.$$

The general combining ability of this genotype is:

$$g_{ij} = 1/2\,(\alpha_i + \alpha_j)$$

(half of the additive part of the genotypic value).

The mean of all S_1 plants from this genotype would be:

$$S_{1\,(ij)} = \mu + \alpha_i + \alpha_j + 1/2\,\beta_{ij} + 1/4\,(\beta_{ii} + \beta_{jj}) \tag{6}$$

$$\mathrm{cov}\,(g, S_1) = 1/2\,\sigma_A^2 + 1/4\,\sigma_{AD_0}\,.$$

Because it is possible to estimate cov (g, S_1) according the crossing plan, having already estimated σ_A^2, it results in an estimation of σ_{AD_0}:

$$\hat{\sigma}_{AD_0} = 4\,\widehat{\mathrm{cov}}\,(g, S_1) - 8\,\hat{\sigma}_g^2$$

Now it remains to estimate $\sigma_{D_0}^2$.

Consider the genetic variance among S_1, var (S_1); it can be estimated from the analysis of variance of the S_1 trial. According to (6):

$$\mathrm{var}\,(S_1) = \sigma_A^2 + 1/4\,\sigma_D^2 + 1/8\,\sigma_{D_0}^2 + \sigma_{AD_0}$$

σ_A^2, σ_D^2, σ_{AD_0} have already been estimated, so $\sigma_{D_0}^2$ can be estimated.

Finally, the result is an estimation of genetic variance of all lines that can be derived from the random mating population:

$$\hat{\sigma}_L^2 = 8\,[\mathrm{var}\,S_1 - 2\,\widehat{\mathrm{cov}}\,(g, S_1) + \hat{\sigma}_g^2 - \hat{\sigma}_s^2]\,.$$

Clearly the accuracy of such an estimation will depend on the accuracy of each parameters $\widehat{\mathrm{var}}\,S_1$, $\widehat{\mathrm{cov}}\,(g, S_1)$, $\hat{\sigma}_g^2$ and $\hat{\sigma}_s^2$. Such parameters have to be estimated with accuracy to have acceptable accuracy for $\hat{\sigma}_L^2$. Such a statistical problem will not be discussed here.

Conclusion

Using a two-factor mating design including S_1 progenies, with the assumptions of absence of epistasis and absence of competition, it appears possible to simultaneously predict the mean and the variance of all single crosses and of lines that can be derived from a random mating population. From such estimations, knowing the broad sense heritability defined for each type of variety, it would be possible to compare the value of the best lines and of the best hybrids that can be derived from the same population. The prediction formula for lines can be very useful when doubled haploids cannot be used or when a single seed descent system is too expensive for the derivation of the population of lines from the random mating population.

References

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

Gallais A (1970) Covariances entre apparentés quelconques avec linkage et épistasie. I. Expression générale. Ann Genet Sel Anim 2:281–310. II. Evolution en régime d'autofécondation. Ann Genet Sel Anim 2:417–427

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

Gallais A (1978) The concept of varietal ability in plant breeding. Euphytica 28:811–823

Gallais A (1979) Le concept de valeur en lignées d'un génotype et son utilisation possible en sélection. Ann Amelior Plant 29:1–22

Gillois M (1964) La relation d'identité en génétique. Thèse Fac Sci Paris, 205 pp

Griffing (1956) A generalized treatment of the use of diallel crosses in quantitative inheritance. Heredity 10:31–50

Harris DL (1964) Genotypic covariances between inbred relatives. Genetics 50:319–348

Jinks JL, Pooni HS (1976) Predicting the properties of recombinant inbred lines derived by single seed descent. Heredity 36:253–266

Kempthorne O (1957) An introduction to genetic statistics. Wiley, New York, 450 pp