Theoretical determination of the optimum number of parents for synthetics

A. Gallais

Station de Génétique Végétale, CNRS-INRA-UPS, Ferme du Moulon, F-91190 Gif sur Yvette, France and I.N.A. P-G, 16 rue Claude Bernard, F-75231 Paris-Cedex 05, France

Summary. From the consideration of the expressions of the mean and of the variances among k-parent synthetics, it is possible, in the absence of epistasis, to give an algebraic determination of the optimum number of parents to include in a synthetic. The knowledge of four components of variance of inbred populations is necessary. Such components can be replaced by four simple statistics for the plant breeder: variances of general and specific combining abilities, variance among S_1 families from the parents, and covariance between S_1 value and general combining ability (GCA). A numerical study shows that this optimum is rather broad for a number of parents greater than four. As expected, the optimum tends to be higher for greater inbreeding depression. With inbreeding depression greater than 0.30, the maximum gain, in comparison to the random mating population with realistic selection intensity, would be less than 5%. In such a situation it will be better to use as synthetic the population improved by recurrent selection.

Key words: Synthetics – Number of parents – Synthesizing ability – Varietal development

Introduction

From different experimental and theoretical studies, it is clear that there is generally an optimum number of parents for synthetic varieties (see, e.g., Kinman and Sprague 1945; Wricke and Weber 1978). However, a general theoretical demonstration of the optimum number of parents has never been given. From use of the expressions of mean and variances among synthetics, a simple approach to this problem is given.

The optimum will be determined algebraically, and it will be studied numerically to determine whether its profile is sharp or flat. Some consequences on the level of genetic advance from the development of synthetics will also be discussed.

Algebraic determination of the optimum

Consider the distribution of the genetic values of all k-parent synthetics. The values of the best varieties can be written:

$$\text{Syn Max} = \mu_S + i\,\sigma_S \tag{1}$$

where μ_S and σ_S^2 are, respectively, the mean and the genetic variance of the distribution of all k-parent synthetics, and i is the intensity of selection among synthetics.

Because the mean μ_S increases and the variance σ_S^2 decreases with an increasing number of parents, this suggests that an optimum number may exist. To solve this problem, two approaches of the mean and of variance among synthetics will be presented.

First approach

From the studies on means (restricting the approach on diploids), Gallais (1967) showed that with noninbred parents and in the absence of epistasis

$$\mu_S = \mu + \frac{1}{2k}\,\text{Do}, \tag{2}$$

where μ is the mean of the random mating population and Do is the maximum inbreeding depression: $\text{Do} = \mu_L - \mu$, μ_L being the mean of all lines that can be derived of the population; Do is negative.

In the absence of epistasis, the variance among synthetics can be written (Gallais 1974):

$$\sigma_S^2 = (1/k)\,\sigma_{A'_s}^2 + \left[\frac{2k-1}{4k^3}\right]\sigma_D^2\;; \tag{3}$$

σ_D^2 is the dominance variance in the random mating population. From Gallais (1979) and Gallais and Wright (1979)

$$\sigma_{A'_s}^2 = k\,2\,E({}_s\alpha_i^2)$$

with ${}_s\alpha_i$ being the additive effect for synthetic value:

$${}_s\alpha_i = \{\alpha_i + (1/4\,k)\,[\beta_{ii} - E\,(\beta_{ii})]\}/k\,, \tag{4}$$

where α_i is the classical additive effect, and β_{ii} is the residue for dominance in the homozygous genotype A_iA_i.

In the case of biallelism $(B, b; p, q)$ using the notations of Falconer (1960), the additive effect for synthetic value of the allele B could be written:

$${}_s\alpha_B = q\left[a - \left(1 - \frac{1}{2k}\right)d\,(p-q)\right]$$

as already shown by Wright (1974).

Then, from Eq. (4)

$$E\,({}_s\alpha_i^2) = \{E(\alpha_i^2) + (1/16\,k^2)\,E\,[\beta_{ii} - E\,(\beta_{ii})]^2$$
$$+ (1/2\,k)\,E\,(\alpha_i\,\beta_{ii})\}/k^2 \tag{5}$$

$$\sigma_{A's}^2 = 2\,k\,E({}_s\alpha_i^2) = (1/k)\,\sigma_A^2 + (1/k^2)\,\sigma_{ADo} + (1/8\,k^3)\,\sigma_{Do}^2\,,$$

where σ_{Do}^2 and σ_{ADo} are components of variances that appear under inbreeding (Gillois 1964; Harris 1964).

The variance among synthetics, therefore, is a function of the number of parents and of components of variance of an inbred population.

To maximize the value of the best synthetics, we have to maximize Eq. (1) when the number of parents is considered as variable. Therefore, we have to differentiate Eq. (1) with respect to k:

$$\frac{d\,(\mathrm{Syn})}{d\,(k)} = \frac{d\,(\mu_S)}{d\,(k)} + i\,\frac{d\,(\sigma_S)}{d\,(k)}$$

$$\frac{d\,(\mu_S)}{d\,(k)} = (1/2\,k^2)\,(\mu - \mu_L) \tag{6}$$

The derivative of σ_S is complex, but can be computed. The problem is to know the components of variance of an inbred population $[\sigma_A^2,\ \sigma_D^2,\ \sigma_{Do}^2,\ \sigma_{ADo}]$. The component can be estimated from a two-way mating design with self-fertilization of parents (Gallais 1988). If the design is a diallel (or a series of disconnected diallels), it is possible to formulate directly the variances in terms easy to understand by the breeder; this leads us to the second approach for the optimum.

Second approach

The structure of a synthetic at equilibrium can be considered as equivalent to the structure given by the complete diallel table among the parents. According to the well-known Sewall Wright formula (Wright 1922), the value of a synthetic is:

$$k\text{-Syn} = (1/k)\,\bar{S}_1 + (1-1/k)\,\bar{C}\,, \tag{7}$$

where $\bar{S}_1$ is the mean of S_1 progenies of the parents, and $\bar{C}$ is the mean of all crosses among the parents. The variance among k-parent synthetics is (Gallais 1975)

$$\mathrm{var}\,(k\text{-Syn}) = (1/k^2)\,\mathrm{var}\,\bar{S}_1 + 2\,(1/k)\,(1-1/k)\,\mathrm{cov}\,\bar{S}_1,$$
$$\bar{C} + (1-1/k)^2\,\mathrm{var}\,\bar{C}$$

with the following relationships:

$$\mathrm{var}\,\bar{S}_1 = (1/k)\,\sigma_{S_1}^2,\qquad (\sigma_{S_1}^2\!:\text{ genetic variance among }S_1)$$

$$\mathrm{cov}\,(\bar{S}_1,\,\bar{C}) = (2/k)\,\sigma_{S_1g},\qquad (\sigma_{S_1g}\!:\text{ covariance between }S_1$$
$$\text{value and GCA})$$

$$\mathrm{var}\,\bar{C} = \frac{4}{k}\,\sigma_g^2 + \frac{2}{k(k-1)}\,\sigma_s^2,\ (\sigma_g^2\!:\text{ variance of GCA,}$$
$$\sigma_s^2\!:\text{ variance of SCA})$$

where GCA and SCA are the general and specific combining abilities of the parents, respectively. Hence,

$$\mathrm{var}\,(k\text{-Syn}) = (1/k^3)\,\sigma_{S_1}^2 + \left[\frac{4\,(k-1)}{k^3}\right]\sigma_{S_1g}$$
$$+ \left[\frac{4\,(k-1)^2}{k^3}\right]\sigma_g^2 + \left[\frac{2\,(k-1)}{k^3}\right]\sigma_s^2\,. \tag{8}$$

Without epistasis and with noninbred parents:

$$\sigma_{S_1}^2 = \sigma_{As_1}^2 + (1/4)\,\sigma_D^2\,,$$

where

$$\sigma_{As_1}^2 = \sigma_A^2 + (1/2)\,\sigma_{ADo} + (1/8)\,\sigma_{Do}^2\,,$$

$$\sigma_{S_1g} = (1/2)\,\sigma_A^2 + (1/8)\,\sigma_{ADo}\,,$$

$$\sigma_g^2 = (1/4)\,\sigma_A^2\,,$$

and

$$\sigma_s^2 = (1/4)\,\sigma_D^2\,.$$

Insertion of these values in Eq. (8) gives Eq. (3).

Thus, knowing $\sigma_{S_1}^2,\,\sigma_{S_1g},\,\sigma_g^2$, and σ_s^2, it will be possible to compute the derivative of var $(k$-Syn) in terms of k.

As the mean of the k-parent synthetic can be predicted by

$$E\,(k\text{-Syn}) = (1/k)\,E\,(\bar{S}_1) + (1-1/k)\,E\,(\bar{C})\,,$$

with $\bar{E}\,(S_1)$ and $E\,(\bar{C})$ being replaced by their estimators, the optimum will be determined by the use of Eq. (6).

Note that

$$E\,(\bar{S}_1) = \mu + 1/2\,\mathrm{Do}\,,$$

and

$$E(\bar{C}) = \mu.$$

In Eq. (8) the coefficient of σ_s^2 is small in comparison to that of σ_g^2; if σ_s^2 is small in comparison to σ_g^2, σ_s^2 can be neglected. In this case it is only necessary to know three parameters: $\sigma_{S_1}^2$, $\sigma_{S_1 g}$, and σ_g^2. These parameters can be easily estimated by the simultaneous study for each evaluated parent of the population of S_1 and the GCA value (from to-pcross). In this case, prediction of synthetic value from S_1 and GCA is equivalent to the prediction from general synthesizing ability (GSA) (Gallais 1979).

These two approaches for determination of the optimum will not show how this optimum is pronounced, whether it is broad or narrow. The numerical study of the expected value of the best synthetic in terms of the number of parents is the best approach to study this problem.

Numerical study of the value of the best synthetics

To develop a numerical study of the optimum, Eq. (8) is transformed by the introduction of parameters, which allows the suppression of the scale of measurement:

ϱ = the genetic correlation between S_1 value and GCA,

$$q^2 = \sigma_{S_1}^2/\sigma_C^2$$

$$g^2 = 2\sigma_g^2/(2\sigma_g^2 + \sigma_s^2)$$

$$d = (\mu - \mu_{S_1})/\mu$$

The results:

$$\text{var}(k\text{-Syn}) = \sigma_C^2 \left\{ (1/k^3)\, q^2 + \left[\frac{4(k-1)}{k^3} \right] q\, g\, \sqrt{1/2} \right. \tag{9}$$

$$+ \left[\frac{2(k-1)^2}{k^3} \right] g^2 + \left[\frac{2(k-1)}{k^3} \right] (1 - g^2) \Big\}$$

$$= \sigma_C^2\, V_k$$

$$E(k\text{-Syn}) = \mu\, [1 - d/k]$$

Thus, Eq. (1) of the best synthetics can be written:

$$\text{Syn Max} = \mu\, [(1 - d/k) + i\, h\, c\, \sqrt{V_k}], \tag{10}$$

where $c = \sigma_C/\mu$ is the genetic coefficient of variation for hybrids, and h is the square root of the heritability among synthetics (if we consider the distribution of phenotypic values of synthetics instead of that of genetic values).

In Eq. (10), only μ is dependent on the scale of measurement. The other parameters are relative parameters $[\varrho, c, d, m, q^2, g^2]$ or fixed by the breeder $[i, k]$. Note that it is only necessary to study the combined effect of the product of three parameters (i, h, c). In addition to this product, it is only necessary to introduce d, q^2, g^2, ϱ.

Realistic range of parameters

A realistic rate of selection for a plant breeder is between 1% and 10%, giving a selection intensity between 2.66 and 1.75. However, the selection intensity to consider depends on the number of constituents: it decreases when the number of constituents increases. With selection on general synthesizing ability and with n parents selected among N candidates, the rate of selection will be of n/N. The resulting selection intensity has been considered in a first approximation as selection intensity among synthetics, with 100 and 300 candidates for selection.

Product $i \cdot h \cdot c$. The heritability h^2 is defined for means of varieties. With a low heritability at the individual level (0.10), such a heritability will be relatively high (at least 0.50 with three replications). The case where $h^2 = 1$ is interesting to consider, because it corresponds to the study with genetic values of synthetics. The genetic coefficient of variation for hybrids, according to data on corn (*Zea mays* L.) (Hallauer and Miranda 1981), can be considered to be between 0.05 and 0.10. Hence, a realistic range of the product $i \cdot h \cdot c$ is [0.04; 0.266].

d. d is the relative inbreeding depression from S_0 to S_1. In the absence of epistasis, such a depression is half the maximum inbreeding depression. If the maximum inbreeding depression is of the order of 0.70 (as is observed in corn), d will be 0.35. Therefore, a realistic range for d will be [0.10, 0.40].

ϱ. ϱ is the genetic correlation between the S_1 value and the GCA. With strict additivity such a correlation will be 1 and, in this case, there will be no inbreeding depression. The presence of dominance and overdominance (or its equivalent) will decrease such a correlation. A value of 0.50 is acceptable for different situations (Sampoux et al. 1989).

q^2. q^2 is the ratio $\sigma_{S_1}^2/\sigma_C^2$. With strict additivity this ratio will be 1. The examination of this ratio, in the case of biallelism (Gallais 1989) from partial to complete dominance, shows that this ratio is near 1 for high and low frequencies of the favorable gene and about 0.60 for intermediate values, although the variation is not symmetrical. It is necessary to conceive overdominance to have lower ratios.

g^2. g^2 is the relative importance of σ_g^2 in the variance among crosses. With additivity, this ratio will be 1. If we accept the same contribution of dominance as for additivity, this ratio will be 0.5.

Clearly all combinations of d, ϱ, q^2, g^2 are not possible because such parameters cannot be fixed independently. This is mainly true for extreme situations; e.g., with addi-

tivity $d=0$, $\varrho=1$, $q^2=1$, $g^2=1$. However, it will be assumed in the general case that for the range considered, the genetic situation is sufficiently complex to authorize independence among these parameters.

Results

The type of results achieved is shown in Fig. 1. First, consider the general effects of the parameters ϱ, q^2 and g^2. The effect of the correlation between the S_1 value and GCA, between 0.50 and 0.80, is low (less than 1% of the value of the synthetics in the best situations, i.e., with high heritability, high ratios q^2 and g^2). There is no effect for $k=1$. The effect of the ratio g^2, the proportion of GCA variance, is also relatively low (<0.01) and tends to be maximum for the optimum number of parents; there is no effect for $k=1$. The effect of the ratio q^2 is the greater for $k=1$ (up to 0.04) but it vanishes for $k>2$. Therefore, the main "genetic" parameters are inbreeding depression, heritability, and the genetic coefficient of variation. Such a result can be directly derived by the discussion of Eqs. (8) and (9).

As far as the optimum is concerned, what is striking is its aspect: it is generally very flat. The optimum tends to be more pronounced for a low ratio $q^2=\sigma_{S_1}^2/\sigma_C^2$ (e.g., 0.5 instead of 0.9). However if such a low value can be accepted with strong inbreeding depression, it is more difficult to accept for low inbreeding depression. Obviously, with sufficient inbreeding depression, the value of the best synthetics increases by increasing the number of parents from 1 to 4–6. But what is remarkable is that the increase of the number of parents further than the optimum leads to a very slow decrease in the performance of

the best synthetics. Between 6 and 10 parents, the variation in the performance of the best synthetics is small. This could explain why it is difficult to reach a clear conclusion about the optimum number from the experimental results.

As expected, the optimum (k opt) is higher for greater inbreeding depression (k opt $=2-3$ for $d=0.10$; k opt $=4-5$ for $d=0.20$; and k opt $=10-15$ for $d=0.40$). At the limit, with no inbreeding depression, the optimum will be 1, and the best variety will be a line. Obviously the level of the optimum, for a given inbreeding depression, increases with increasing selection intensity, heritabilities, and genetic coefficient of variation.

Another result, less encouraging for the plant breeder, relates to the level of the best synthetics in comparison to the mean of the random mating population. With inbreeding depression of 0.30 to 0.40, the maximum gain possible is less than 5%. For a depression of only 0.20, the gain is about 10%. Therefore, with strong inbreeding depression, there is little to gain by the use of the variance among synthetics (or it will be necessary to increase unrealistically the selection intensity). Thus, in such a situation it will be more efficient to increase first the tolerance to inbreeding of the breeding population by recurrent selection, and to use the population itself as a synthetic. However, a synthetic with a narrower base could be justified to increase genetic advance for less complex characters such as, e.g., disease resistance. Such a theoretical result could be another explanation for the fact that in forage grasses where synthetics are traditionally developed, the genetic advance for yield has been very low.

The relatively low genetic advance is mainly because of the strong inbreeding depression, which inhibits the expected gain by the use of the variance among synthet-
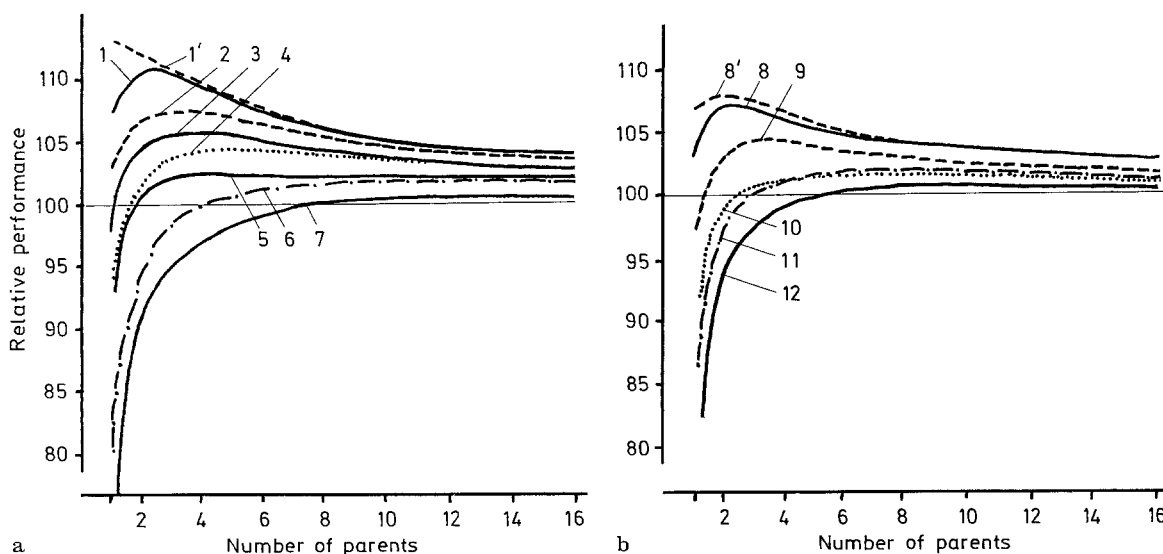

Fig. 1a and b. Effect of the number of parents on the relative performance of the best synthetics according to genetic situations (numbered from 1 to 12, defined in Table 1). **a** for $N=250$; **b** for $N=100$

ics. Furthermore, the examination of the variance among synthetics in the case of biallelism (Gallais 1989) shows that the variance among synthetics is always very low in comparison to the variance among single cross hybrids. Even with a low inbreeding depression and with the optimum number of parents, the genetic gain at the level of k-parent synthetics will be less than the genetic gain among lines or among single crosses. Synthetics must be considered either as a transitory type of variety, or as a means to use genetic variation of oligogenic traits without improvement of the more complex traits, in spite of the breeding effort to improve them.

References

Falconer DS (1960) Introduction to quantitative genetics. Oliver and Boyd, London

Gallais A (1967) Evolution de la vigueur des variétés synthétiques diploïdes au cours de générations de multiplication. I. Avec panmixie. Ann Amel Plant 17:291–301

Gallais A (1974) Selection among synthetics. Theor Appl Genet 44:24–30

Gallais A (1975) Prévision de la vigueur et sélection des parents d'une variété synthétique. Ann Amel Plant 25:233–264

Gallais A (1979) Application du concept d'aptitude variétale à la théorie de la sélection des variétés synthétiques chez les diploïdes et les autopolyploïdes. Ann Amel Plant 29: 111–129

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

Gallais A (1989) Application of the concepts of test value and varietal value to the study of genetic advance in recurrent selection: a synthesis. Euphytica (in press)

Gallais A, Wright AJ (1979) A general approach to the concept of varietal ability for synthetic varieties. Theor Appl Genet 57:81–87

Gillois M (1964) La relation d'identité en génétique. Dissertation, University of Paris

Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames/IA

Harris DL (1964) Genotypic covariance between inbreed relatives. Genetics 50:1319–1348

Kinman ML, Sprague GF (1945) Relation between number of parental lines and theoretical performance of synthetic varieties of corn. J Am Soc Agron 39:341–351

Sampoux JP, Gallais A, Lefort-Buson M (1989) Intérêt de la valeur propre des descendances S1 associée à la valeur en croisement avec un testeur pour la sélection du maïs fourrage. Agronomie 9:511–520

Wricke G, Weber WE (1978) Number of lines for synthetic varieties in rye. Ann Amel Plant 28:667–674

Wright AJ (1974) A genetic theory of general varietal ability for diploid crops. Theor Appl Genet 45:163–169

Wright S (1922) The effects of inbreeding and crossbreeding on guinea pigs. III. Crosses between highly inbred families. Bull US Dep Agric, Washington, No. 1121