Quantitative Genetics of Doubled Haploid Populations and Application to the Theory of Line Development

A. Gallais

CNRS-INRA-UPS, Ferme du Moulon, 91190 Gif sur Yvette, France, and INA-PG, 16 Rue Claude Bernard, 75231 Paris cedex 05, France

ABSTRACT

The line value of a genotype is defined as the expected value of all lines that can be derived from this genotype. Specific genetic effects are defined for this value: only additive and additive by additive epistatic effects are necessary. There is no dominance effect for such a value. A general expression for the covariances between related lines is given. From a design with several lines per haplodiploidized plant taken at random from a population it is possible to estimate the additive variance for line value and the variance of additive by additive epistasis for line value. Variances of higher order epistasis can be estimated with a two-factor mating design in which a cross is replaced by the population of lines that can be derived from it. With a diadial or a factorial design a direct test for the presence of homozygous by homozygous epistasis is possible. The application of the concept of line value to the theory of line development leads to simple expressions of genetic advance in one cycle of recurrent selection according to the testing system. A brief consideration of these expressions leads to the conclusion that single doubled haploid descent recurrent selection will be one of the most efficient methods for low heritabilities and with a rapid development of doubled haploid lines.

Doubled haploids (DH) are increasingly being used in plant breeding. They can be derived by another culture (in several solanaceae, cruciferous as in Brassica napus) or by interspecific crosses (as in Hordeum with Hordeum bulbosum). Derivation of DH can fundamentally change the procedures of plant breeding. Rapid derivation of homozygous lines eliminates the phase of pedigree selection and allows a direct evaluation of the “line value” of a cross (GALLAIS 1979a, b; 1988). Hybrids can be developed directly and recurrent selection methods for line development can also be changed (GALLAIS 1988, 1989).

To formulate a general theory for the use of DH it is necessary to develop quantitative genetics theory of doubled haploid populations derived from random mating populations. CHOO (1980, 1981a, b), CHOO, CHRISTIE and REINBERGS (1979), and CHOO et al. (1986) have made a significant contribution to this problem. However the derivation of their formulas is based upon a specific approach with biallelism and the meaning of the parameters is not explained in terms of classical effects defined for random mating populations. Furthermore, they have used the same notation for components of the genetic variance of doubled haploid lines as for those of random mating populations. What they have called additive variance is not the classical additive variance, but what I call in the following additive variance for line value.

This paper will show that the introduction of the concept of line value allows a very general approach of DH populations in the case of multiallelism with a general formulation of the covariances between related lines. In what follows the process of derivation of DH lines from a heterozygous plant will be considered as a system of mating which does not induce new variability and does not modify Mendelian inheritance. Such a system of mating will be called haplodiploidization.

After introducing the concept of line value and defining the corresponding genetic effects, I will consider how to formulate the variances and covariances among lines derived by haplodiploidization from a random mating population. Finally, I will consider an application of the concept to the theory of line development through population improvement.

THE CONCEPT OF LINE VALUE OF A GENOTYPE

The line value of a genotype can be defined as the expected value of all lines which can be derived from this genotype. Consider the case of a genotype reduced to one locus, with two alleles A_i and A_j, its line value will be

\[ L(A_iA_j) = 1/2(y_{ii} + y_{jj}) \]

where y_{ii} and y_{jj} represent the values of the two lines A_iA_i and A_jA_j which can be derived from A_iA_j. Considering the case of a genotype A_iA_jB_kB_l with two loci A and B, in the absence of linkage the line value will be

\[ L(A_iA_jB_kB_l) = 1/4(y_{iik} + y_{jil} + y_{ijl} + y_{jil}) \]
where $y_{ijkl}$ represents the value of the line $A_iA_jB_kB_l$.
The extension and meaning of the line value for a genotype with $n$ loci is straightforward.

The line value $\mu_L$ of a population will be the expected value of all lines which can be derived from the population. So at one locus

$$\mu_L = E[L(A_iA_j)] = E(y_{ij})$$

and at two loci:

$$\mu_L = E(y_{ijkl})$$

**GENETIC EFFECTS FOR LINE VALUE**

The genetic effects for line value are defined with the classical model of the decomposition of the genotypic value (KEMPThORNE 1957). The value $y_{ij}$ for a genotype $A_iA_j$ can be written:

$$y_{ij} = p + a_i + a_j + p_{ij},$$

where $p$ is the mean of the random mating population, $a_i$ and $a_j$ are the additive effect of the gene $A_i$ and $A_j$, and $p_{ij}$ is the interaction effect (dominance) between genes $A_i$ and $A_j$.

So for the line $A_iA_i$

$$y_{ii} = p + 2a_i + p_{ii},$$

where $p_{ii}$ is the interaction between two like alleles. Then the line value of the genotype $A_iA_j$ will be

$$L(A_iA_j) = p + a_i + a_j + 1/2(p_{ii} + p_{jj}) + \mu_L$$

and

$$\mu_L = E[L(A_iA_j)] = p + E(\beta_{ii})$$

So it is possible to write

$$L(A_iA_j) = \mu_L + L\alpha_i + L\alpha_j,$$

where $L\alpha_i$ is the additive effect for line value.

As expected there are no dominance effects for line value. In the absence of epistasis, the line value of a two-locus genotype $A_iA_jB_kB_l$ can be written

$$L(A_iA_jB_kB_l) = \mu_L + L\alpha_i + L\alpha_j + L\alpha_k + L\alpha_l$$

where index 1 is for genes at locus $A$, and index 2 for genes at locus $B$. With epistasis this value can be written

$$L(A_iA_jB_kB_l) = \mu_L + L\alpha_i + L\alpha_j + L\alpha_k + L\alpha_l + L(\alpha\alpha)_{iikl} + L(\alpha\alpha)_{ikl} + L(\alpha\alpha)_{iikl} + L(\alpha\alpha)_{ikk} + L(\alpha\alpha)_{iikl} + L(\alpha\alpha)_{iikl}$$

with

$$\mu_L = \mu + E(\beta_{iikl}) + E(\beta_{ikk}) + E(\beta_{iikl})$$

and

$$E[L(A_iA_j)] = \sum_j p_j L(A_iA_j)$$

with $E[L(A_iA_j)] = \sum_j p_j L(A_iA_j)$, $p_j$ being the frequency of allele $A_j$, and

$$L(\alpha\alpha)_{iikl} = (\alpha\alpha)_{iikl} + 1/2(\alpha\beta)_{iikl} + 1/2(\beta\beta)_{iikl} + 1/4(\alpha\alpha)_{iikl}$$

with $E(\beta\beta)_{iikl} = (\alpha\beta)_{iikl}$ because $E(\alpha\beta)_{iikl} = E(\alpha\beta)_{iikl} = 0$ and, $(\beta\beta')_{iikl} = (\beta\beta')_{iikl} - E[(\beta\beta')_{iikl}].$ So $L(\alpha\alpha)$ can be considered as an homozygote by homozygote epistasis. However, I will call it additive by additive epistasis for line value. In the following, to simplify the notation in the case of epistasis, the indices 1 and 2 for the loci will be suppressed.

Another and more direct way to define the additive effect for line value is to consider the diallel table for the matching of all alleles at one locus (Figure 1).

Each cell of this table determines a unique genotype. Associating with each allele $A_i$ its frequency $p_i$, such a table allows the determination of the genotypic structure of a random mating population. With the corresponding genotypic value in each cell, the classical additive effect is defined as the main effect of one row or one column of this table. Now, if each cell contains the associated line value, the additive effect for line value can be defined as the main effect of one row or one column of this new table:

$$E[L(A_iA_j)] = \sum_j p_j L(A_iA_j)$$

Figure 1.—Definition of the additive effects for line value (see text). $A_1, A_2, A_3, \ldots$ are alleles at the locus considered. $L(A_iA_j)$ is the line value of the genotype $A_iA_j$.

$$L(\alpha\alpha)_{iikl} = (\alpha\alpha)_{iikl} + 1/2(\alpha\beta)_{iikl} + 1/4(\beta\beta')_{iikl}$$

with

$$E(\beta\beta')_{iikl} = (\alpha\beta)_{iikl}$$

because $E(\alpha\beta)_{iikl} = E(\alpha\beta)_{iikl} = 0$ and, $(\beta\beta')_{iikl} = (\beta\beta')_{iikl} - E[(\beta\beta')_{iikl}].$ So $L(\alpha\alpha)$ can be considered as an homozygote by homozygote epistasis. However, I will call it additive by additive epistasis for line value. In the following, to simplify the notation in the case of epistasis, the indices 1 and 2 for the loci will be suppressed.

Another and more direct way to define the additive effect for line value is to consider the diallel table for the matching of all alleles at one locus (Figure 1). Each cell of this table determines a unique genotype. Associating with each allele $A_i$ its frequency $p_i$, such a table allows the determination of the genotypic structure of a random mating population. With the corresponding genotypic value in each cell, the classical additive effect is defined as the main effect of one row or one column of this table. Now, if each cell contains the associated line value, the additive effect for line value can be defined as the main effect of one row or one column of this new table:

$$E[L(A_iA_j)] = \sum_j p_j L(A_iA_j)$$

with

$$E(\beta\beta')_{iikl} = (\alpha\beta)_{iikl}$$

because $E(\alpha\beta)_{iikl} = E(\alpha\beta)_{iikl} = 0$ and, $(\beta\beta')_{iikl} = (\beta\beta')_{iikl} - E[(\beta\beta')_{iikl}].$ So $L(\alpha\alpha)$ can be considered as an homozygote by homozygote epistasis. However, I will call it additive by additive epistasis for line value. In the following, to simplify the notation in the case of epistasis, the indices 1 and 2 for the loci will be suppressed.

Another and more direct way to define the additive effect for line value is to consider the diallel table for the matching of all alleles at one locus (Figure 1). Each cell of this table determines a unique genotype. Associating with each allele $A_i$ its frequency $p_i$, such a table allows the determination of the genotypic structure of a random mating population. With the corresponding genotypic value in each cell, the classical additive effect is defined as the main effect of one row or one column of this table. Now, if each cell contains the associated line value, the additive effect for line value can be defined as the main effect of one row or one column of this new table:

$$E[L(A_iA_j)] = \sum_j p_j L(A_iA_j)$$

with

$$E(\beta\beta')_{iikl} = (\alpha\beta)_{iikl}$$

because $E(\alpha\beta)_{iikl} = E(\alpha\beta)_{iikl} = 0$ and, $(\beta\beta')_{iikl} = (\beta\beta')_{iikl} - E[(\beta\beta')_{iikl}].$ So $L(\alpha\alpha)$ can be considered as an homozygote by homozygote epistasis. However, I will call it additive by additive epistasis for line value. In the following, to simplify the notation in the case of epistasis, the indices 1 and 2 for the loci will be suppressed.
$p_j$ being the frequency of allele $A_j$,

$$E_j[L(A,A_j)] = 1/2 y_j + 1/2 \mu_L.$$

Since $L(A,A_j) = \mu_L + L \alpha_i + L \alpha_j$, it is clear that there is no dominance effect.

Note that it is possible to write

$$\alpha_i = 1/2 (y_i - \mu_L) - 1/2 \beta_i.$$

This expression shows how classical additive effect is related to "homozygous" effects.

Such a diallel approach can be extended to define epistatic effects. With two loci 1 and 2 the additive effect for line value of the allele $A_1$ is defined as

$$L(\alpha_i) = \sum_{j_1} \sum_{j_2} p_{i1} p_{i2} L_{i1,j_1,j_2} - \mu_L,$$

$L_{i1,j_1,j_2}$ being the line value of a genotype with allele $i_1$, $j_1$ at locus 1 and $i_2$, $j_2$ at locus 2. The additive by additive epistasis for line value (or homozygous by homozygous epistasis) for two alleles $i_1$ and $i_2$ is then

$$L(\alpha_{i1,j_2}) = \sum_{j_1} \sum_{j_2} p_{i1} p_{i2} L_{i1,j_1,j_2} - \alpha_{i1} - \alpha_{i2} - \mu_L.$$

**Application to the case of biallelism:** the consideration of the case of biallelism can be useful to clarify the definition of the additive effects for line value. Considering one locus, with two alleles $B$ and $b$ in frequencies $p$ and $q$, using the notation of Falconer (1961) for genotypic values ($y_{BB} = a$, $y_{Bb} = d$, $y_{bb} = -a$) the line value of each genotype can be written:

$$L(BB) = a, \quad L(Bb) = 0, \quad L(bb) = -a.$$

Then, the line value of the population is:

$$\mu_L = E[L_{ij}] = (p - q)a,$$

and the genotypic values as deviation from the mean are

$$L'(BB) = 2qa, \quad L'(Bb) = (q - p)a,$$

$$L'(bb) = -2pa.$$

From such results, or from the definition of the additive effect for line value

$$L \alpha_B = qa \quad \text{and} \quad L \alpha_b = -pa.$$

Considering the well known expression of the additive effect for per se value of a genotype

$$a_B = qa \quad \text{and} \quad a_b = -pa$$

with $\alpha = a - (p - q)d$,

it is clear that:

$$\alpha_B = L \alpha_B - q(p - q)d$$

and

$$\alpha_b = L \alpha_b + p(p - q)d.$$
The variance among lines from the same plant is the residual, or:
\[ \sigma_{DLW}^2 = \sigma_{DLT}^2 - \sigma_{DLB}^2 = \sigma_L^2 + 3\sigma_{L3L}. \]  
(9)

In the absence of epistasis genetic variance among and within plants is the same
\[ \sigma_{DLB}^2 = \sigma_{DLW}^2 = \sigma_L^2. \]

Note that, as a consequence of the expression (3) of the additive effects for line value, the variance \( \sigma_L^2 \) can be expressed in terms of components defined for an inbred population (to simplify summation over loci is omitted)
\[ \sigma_L^2 = \sigma_L^2 + \sigma_{AD0} + 1/2 \sigma_{D0}^2, \]
where
\[ \sigma_L^2 = 2E(\alpha_i^2) \]
\[ \sigma_{AD0} = 2E(\alpha_i\beta_0) \]
\[ \sigma_{D0}^2 = E(\beta_0^2) - [E(\beta_0)]^2 \]
are components of variance specific to inbred populations defined by GALLAIS (1964) and HARRIS (1964).

According to expressions (6) the epistatic component \( \sigma_{L3L} \) can be related to the epistatic component specific to an inbred population as defined by GALLAIS (1970, 1974):
\[ \sigma_{L3L} = \sigma_{L3L} + 1/2 \sigma_{D0}^2 + 1/4 \sigma_{D0}^2, \]
where
\[ \sigma_{L3L} = 4E(\alpha_0^2) \]
\[ \sigma_{D0}^2 = 2E(\alpha_i\beta_0) + E(\beta_i\beta_i) \]
\[ \sigma_{D0}^2 = E(\beta_i\beta_i) + [E(\beta_i)]^2 \]
\[ \text{cov}(AA, AD0) = 8E[(\alpha_0\alpha_0)(\alpha_i\beta_i)] \]
\[ + E[(\alpha_0\alpha_0)(\beta_i\beta_i)] \]
\[ \text{cov}(AA, D0D0) = 4E[(\alpha_0\alpha_0)(\beta_i\beta_i)] \]
\[ \text{cov}(AD0, D0D0) = 4E[(\alpha_0\alpha_0)(\beta_i\beta_i)] \]
\[ \text{cov}(AD0, D0A) = 2E[(\alpha_0\alpha_0)(\beta_i\beta_i)] \]
\[ + E[(\beta_i\beta_i)(\beta_i\beta_i)] \]
As \( (\alpha_0\beta_i) = (\alpha_0\beta_i), \) and \( E(\alpha_0\alpha_0) = 0, \) \( \beta_i \) can be replaced by \( \beta \) in the expectations where there is \( (\alpha_0\beta_i) \) or \( (\alpha_i\alpha_i). \)

One of the uses of the concept of line value is to reduce the number of parameters by condensing the parameters as introduced by GALLAIS (1970, 1974).

Covariances between lines from relatives in a random mating population: Consider next the covariances among lines that can be derived from relatives in a random mating population. A full-sib family of lines will be the set of lines derived from individuals within a classical full-sib family. In the same manner, a half-sib family of lines will be the set of lines derived from individuals within a classical half-sib family.

In the absence of epistasis the covariance between lines from two related zygotes, \( X \) and \( Y \), can be deduced from the study of the covariance between two related lines reduced to one locus. Let \( A_iA_j \) be the genotype of \( X \), and \( A_i' A_j' \) the genotype of \( Y \).

The covariance will be
\[ \text{cov}(X_L, Y_L) = 1/4 \text{cov}(y_{ii}, y_{ii'}) + \text{cov}(y_{ii'}, y_{ii'}) \]
\[ = \text{cov}(y_{ii}, y_{ii'}) \]
\[ = 4E(\alpha_i\alpha_i) \]
If \( i \) and \( i' \) are not identical by descent, the covariance is zero. If they are identical by descent, with probability \( \varphi \), the coefficient of kinship, the covariance is not zero, so that
\[ \text{cov}(X_L, Y_L) = 2\varphi \sigma_{L3L}. \]

With epistasis, restricted to two loci
\[ \text{cov}(X_L, Y_L) = \text{cov}(y_{iikk}, y_{i'ik'}). \]

Thus, in the absence of linkage,
\[ \text{cov}(X_L, Y_L) = 2\varphi \sigma_{L3L} + 4\varphi^2 \sigma_{L3L}. \]

This expression is very similar to the classical expression of covariances between relatives in a random mating population, with no dominance effects. It remains valid for covariances between relatives derived from crosses of inbred but unrelated parents, and in this case it is well known (KEMPThORNE 1957) that: \( \varphi = (1 + F)\varphi_0 \), \( F \) being the inbreeding coefficient of the parents and \( \varphi_0 \) the coefficient of kinship for non-inbred parents.

Application to some classical situations (with non-inbred parents): 1. Covariances between parent and offspring (\( \varphi = 1/4 \)): According to expression (10)
\[ \text{cov}(P_L, O_L) = 1/2 \sigma_{L3L} + 1/4 \sigma_{L3L}. \]

This is very similar to the classical expression of covariance between parent and offspring. Without epistasis a noninbred parent has line value \( A_L \), and transmits half of this value to its offspring, so the covariance between parent and offspring is \( 1/2 \sigma_{L3L} \).

2. Covariances between half-sibs (\( \varphi = 1/8 \)). From expression (10)
\[ \text{cov}(HS_L) = 1/4 \sigma_{L3L} + 1/16 \sigma_{L3L}. \]

3. Covariances between full-sibs (\( \varphi = 1/4 \)). From expression (10)
\[ \text{cov}(FS_L) = 1/2 \sigma_{L3L} + 1/4 \sigma_{L3L}. \]
4. Covariances of an individual with itself (\(\varphi = 1/2\))
\[
\text{cov}(X_L, X_L) = \sigma_{L\!L}^2 = \sigma_L^2 + \sigma_{AA_L}^2.
\]

This gives the genotypic variance for line value at the level of the population, and shows the generality of the reasoning.

**Estimation of variance components:** The aim is to estimate the components of variance for line value at the level of a random mating population. Two types of designs will be briefly considered: the “one-way” design in which the DH lines are directly derived from the plants of the population and the two-way mating designs (nested, factorial or diallel) in which the cross between two plants (inbred or not) is replaced by a set of DH lines that can be derived from it.

Without epistasis and with a sufficiently large number of DH lines per plant, the additive variance for line value can be estimated directly by the genetic variance among plants: \(\sigma_{L\!L}^2 = \sigma_L^2\).

If there is epistasis or a restricted number of lines per plant, a simple nested design can be used. Suppose that there are \(l\) lines for each plant, with \(p\) plants, the components of variance can be deduced from the analysis of variance of such an experiment (Table 1). In this case
\[
\sigma_{L\!L}^2 = \sigma_L^2 + \sigma_{AA_L}^2
\]
and the variance within plants is:
\[
\sigma_{L\!W}^2 = \sigma_L^2 + 3\sigma_{AA_L}^2.
\]

Then,
\[
\sigma_L^2 = 1/2(3\sigma_{L\!W}^2 - \sigma_{L\!L}^2)
\]
\[
\sigma_{AA_L}^2 = 1/2(\sigma_{L\!L}^2 - \sigma_{L\!W}^2) \quad (11)
\]
This is a very simple experiment to detect the presence of epistasis.

In comparison to the previous situation, with a two-factor mating design between inbred parents, a third level is added. Then it is possible to estimate three components of variance. The case of a diallel with completely inbred parents has been already considered by CHOO et al. (1986) from another approach. However it must be noted that they have used a diallel design with selfs (parents) (GRIFFING’s method 2, model 1) (GRIFFING 1956) which is not appropriate for estimating variance components from the random mating base population. The case of a diallel with non-inbred plants, without reciprocals or selfs (GRIFFING’s method 4, model 2) is considered Table 2. If the lines from each haplodiploidized plant are pooled, so there is a variance among plants within a cross, three components of variance can also be estimated. The variance of general combining ability (GCA) is
\[
\sigma_{L:\!G}^2 = \text{cov}(H_S)_L = 1/4\,\sigma_{AA_L}^2 + 1/16\,\sigma_{AA_A}^2 + 1/64\,\sigma_{AA_A}^2
\]
The variance of specific combining ability (SCA) is
\[
\sigma_{L:\!S}^2 = \text{cov}(FS)_L - 2\,\text{cov}(HS)_L = 1/8\,\sigma_{AA_L}^2 + 3/32\,\sigma_{AA_A}^2
\]
(it is zero without epistasis, as shown, by CHOO (1981a)). The variance among mother plants within crosses is \(\sigma_{L:\!C}^2\) and is the difference between the total variance among mother plants \(\sigma_{L:\!G}^2\) and the variance among crosses \(\text{cov}(FS)_L\):
\[
\sigma_{L:\!C}^2 = \sigma_{L:\!G}^2 - \text{cov}(FS)_L
\]
\[
= 1/2\,\sigma_{AA_L}^2 + 3/4\,\sigma_{AA_A}^2 + 7/8\,\sigma_{AA_A}^2.
\]
If the lines from each haplodiploidized plant are tested separately, it will be possible to estimate variance components associated with epistasis between four loci \(\sigma_{AA_A}^2\) (Table 2). Restricting epistasis to the first degree, such a design gives two degrees of freedom to test the effect of higher order of epistasis by the least square analysis.

Due to the absence of dominance in the line value, it is interesting to note that the test of “specific combining ability” in a diallel or of the “male by female” interaction in a factorial design, will be a test of the presence of homozygous by homozygous epistasis, even with a fixed set of parents. Thus haplodiploidization appears as a powerful tool to show homozygous by homozygous epistasis from a random or fixed point of view.

### Table 1

Analysis of variance of an experiment with \(b\) replications, \(l\) DH lines per plant and \(p\) haplodiploidized plants

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Mean square</th>
<th>Expected mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between plants</td>
<td>((p-1))</td>
<td>(M_b)</td>
<td>(\sigma_L^2 + b\sigma_{L!L}^2 + bl\sigma_{L!W}^2)</td>
</tr>
<tr>
<td>Within plants</td>
<td>((l-1))</td>
<td>(M_w)</td>
<td>(\sigma_L^2 + b\sigma_{L!L}^2)</td>
</tr>
<tr>
<td>Residual</td>
<td>(r)</td>
<td>(M_{rs})</td>
<td>(\sigma_{L!L}^2)</td>
</tr>
</tbody>
</table>

### Table 2

Analysis of variance of a diallel table for line value with \(n\) haplodiploidized plants per cross and \(l\) lines per plant, with \(b\) replications, and \(p\) parents

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Expected mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCA ((p-1))</td>
<td>(\sigma_L^2 + b\sigma_{L!L}^2 + bl\sigma_{L!W}^2 + bln\sigma_{L!W}^2 + b(n-1)\sigma_{L!W}^2)</td>
<td></td>
</tr>
<tr>
<td>SCA ((p-3)/2)</td>
<td>(\sigma_L^2 + b\sigma_{L!L}^2 + bl\sigma_{L!W}^2 + bln\sigma_{L!W}^2)</td>
<td></td>
</tr>
<tr>
<td>Plants within cross ((n-1))</td>
<td>(\sigma_{L!L}^2 + b\sigma_{L!W}^2 + b\sigma_{L!W}^2)</td>
<td></td>
</tr>
<tr>
<td>Lines within plant ((l-1))</td>
<td>(\sigma_{L!L}^2 + b\sigma_{L!W}^2)</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>(r)</td>
<td>(\sigma_{L!L}^2)</td>
</tr>
</tbody>
</table>

The total genotypic variance among lines is:
\[
\sigma_{L\!L}^2 = \text{cov}(H_S)_L = 1/4\,\sigma_{AA_L}^2 + 1/16\,\sigma_{AA_A}^2 + 1/64\,\sigma_{AA_A}^2
\]
APPLICATION TO THE STUDY OF RECURRENT SELECTION FOR LINE VALUE

Recurrent selection for line value is an important step towards the development of lines. To simplify the expression of genetic advance expected from recurrent selection the formulation is restricted to the case that assumes the absence of epistasis. According to a general expression given by Gallais (1976) the expected genetic advance from one cycle of recurrent selection is

\[ \Delta G = \theta \frac{\text{cov}(TM)}{\sqrt{\text{var} \ T}} \]

where \( T \) is the value of the candidate to selection according to the testing system, and \( M \) the value of its progeny after intercrossing, according to the testing system used to evaluate genetic advance; \( \theta = 1 \) or 2 according to the control of selection on one or two sexes. As only additive effects are transmitted by gametes, \( \text{cov}(TM) \) is in fact a covariance between the parents evaluated according to the system of test (\( T \)) and their offspring evaluated according to the system of test \( M \).

So with non-inbred parents:

\[ \text{cov}(TM) = \frac{1}{2} \sigma_{A_iA_j}, \]

\( A_T \) being the additive part of phenotypic value in the test and \( A_M \) the additive part of varietal value (Gallais 1979b).

For line development,

\[ \text{cov}(TM) = 1/2 \sigma_{A_iA_j}. \]

More generally it is possible to write

\[ \text{cov}(TM) = 2\varphi \sigma_{A_iA_j}, \]

with

\[ \varphi = 1/4 \text{ for parent-offspring kinship.} \]

However, grand-parent-offspring relationship (\( \varphi = 1/8 \)) is involved in phenotypic family selection.

The covariance \( \sigma_{A_iA_j} \) involves two types of additive effects: those defined for line value (\( A_L \)) and those defined for the testing system of the parents (\( A_T \)).

In what follows seven specific systems of testing are considered. The first four methods are classical and described in books of plant breeding. They do not involve the production and testing of DH lines but are interesting to consider for their genetic advance in line value. The last three were described by Gallais (1988, 1989).

Method 1: Phenotypic mass selection: individual mass selection. Individuals are evaluated on the basis of their phenotypic value \( P \), with

\[ P = A + D + E. \]

\( A \) for additive effects, \( D \) for dominance effects, \( E \) for environmental effects.

So:

\[ A_T = A \]

\[ \text{cov} \ TM = 1/2 \sigma_{A_iA_j}, \]

\[ \sigma_{A_iA_j} = 2E(a_La_i) = \sigma_A^2 + 1/2 \sigma_{AD} \]

Method 2: Phenotypic half-sib family selection: Half-sib families are evaluated and selected families are intercrossed. In this case we have to consider covariances between grand-parent and offspring, so, with \( \varphi = 1/8 \),

\[ \text{cov} \ TM = 1/4 \sigma_{A_iA_j} = 1/8 \sigma_{A_iA_j} \]

and

\[ A_T = 1/2 A \]

(only half of the additive value of the selected plant contributes to the value in test).

Method 3: Half-sib progeny selection (selection on General Combining Ability (GCA)).

Half-sib families are evaluated as previously but in this case mother plants of the best families are intercrossed. So:

\[ A_T = (1/2 A) \]

as previously, but

\[ \text{cov}(TM) = 1/2 \sigma_{A_iA_j} \]

because it is a covariance between parent and offspring, so

\[ \text{cov}(TM) = 1/4 \sigma_{A_iA_j}. \]

Method 4: SI progeny selection: The plants are evaluated from their SI progeny (derived by one generation of self-fertilization) and the best plants are intercrossed. Cov(TM) is again a covariance between parent and offspring (\( \varphi = 1/4 \)) and \( A_T = A_S \) (the additive value in SI) (Gallais 1979a), so

\[ \text{cov}(TM) = 1/2 \sigma_{A_iA_j} \]

The additive value in SI is defined analogously to the additive value in line, putting the SI value in each cell of the diallel table (Figure 1) in place of genotypic value. As

\[ S_1(A_iA_j) = \mu + \alpha_i + \alpha_j + 1/2 \beta_{ij} + 1/4 (\beta_u + B_y), \]

it results

\[ S_1(A_iA_j) = \mu_S + s_i \alpha_i + s_j \alpha_j + s_i \beta_{ij} \]

and

\[ s_j \alpha_i = \alpha_i + 1/4 \beta_u' \]

So

\[ \text{cov}(TM) = 1/2 \sigma_{A_iA_j} \]
and
\[ \sigma_{A_i A_i} = 2E(z_i a_i a_i) = \sigma_A^2 + 3/8 \sigma_{AD} + 1/8 \sigma_P^2. \]

**Method 5:** Selection on line value of a plant (GALLAIS 1979a); plants are evaluated directly for their line value and the best plants are intercrossed. This is equivalent to progeny testing for line value. Thus, in this case \( \text{cov}(TM) \) is a covariance between parent and offspring evaluated according to their line value; so \( A_T = A_L \) and

\[ \text{cov}(TM) = 1/2 \sigma_{A_L}^2. \]

**Method 6:** Single doubled haploid descent recurrent selection (GALLAIS 1988); only one DH line is derived per plant and the best lines are intercrossed. In this case we have an inbred plant with a coefficient of inbreeding of 1 so

\[ \varphi = 1/2, \quad \text{and} \quad \text{cov}(TM) = \sigma_{A_L}^2. \]

Such a result can be derived directly by noting that the value of a homozygous line \( A_A \) is \( \mu + 2 \sigma_A \) and that half of this additive value is transmitted to offspring, so

\[ \text{cov}(TM) = 2E(z_i a_i^2). \]

**Method 7:** Combined selection on phenotypic value of the lines: all lines are derived per plant, and the best lines within the best plants are selected and intercrossed. In this case the genetic advance has two parts and we have already shown (GALLAIS 1988, 1989) that:

\[ \Delta G = i_s \frac{\sigma_{A_i}^2}{\sqrt{\text{var} \ P_{LB}}} \sigma_L + \frac{\sigma_{A_i}^2}{\sqrt{\text{var} \ P_{LW}}} \sigma_W \]

where \( \text{var} \ P_{LB} \) is the phenotypic variance between mother plants, \( \text{var} \ P_{LW} \) is the phenotypic variance within mother plant (among lines of the same mother plant) and \( \sigma_L = \sqrt{(l - 1)/l}, \ sigma_W = \sqrt{(l - 1)/l} \) (GALLAIS 1989). Note that the case \( l = 1 \) gives the method 6.

With the experimental structure of method 7, the best genetic advance will be with the use of an index of line/family and of familial values.

Table 3 gives a summary of the expected genetic advances for the six breeding methods considered with \( \theta = 2 \). Note that another useful expression of the genetic advance in one cycle of recurrent selection is:

\[ \Delta G = i \theta \rho_{TL} h_T \sigma_A, \]

in which \( \rho_{TL} \) is the correlation between additive value in test and the additive value in line, \( h_T \) is the square root of the heritability for the testing system

\[ h_T^2 = \frac{\sigma_T^2}{\sigma_P^2}, \]

(\( \sigma_T^2 \) being the phenotypic variance of units tested).

### Table 3

<table>
<thead>
<tr>
<th>Breeding methods</th>
<th>Genetic advance in line value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Phenotypic mass selection</td>
<td>( i_s \frac{\sigma_{A_L}^2}{\sqrt{\text{var} \ P}} )</td>
</tr>
<tr>
<td>2. Family selection with half sibs</td>
<td>( i_s \frac{1/4 \sigma_{A_L}^2}{\sqrt{\text{var} \ P_{NS}}} )</td>
</tr>
<tr>
<td>3. Selection on GCA</td>
<td>( i_s \frac{1/2 \sigma_{A_L}^2}{\sqrt{\text{var} \ P_{NS}}} )</td>
</tr>
<tr>
<td>4. Selection on line value</td>
<td>( i_s \frac{\sigma_{A_L}^2}{\sqrt{\text{var} \ P_{LW}}} )</td>
</tr>
<tr>
<td>5. Selection on line value</td>
<td>( i_s \frac{\sigma_{A_L}^2}{\sqrt{\text{var} \ P_{LW}}} )</td>
</tr>
<tr>
<td>6. Single DH descent recurrent selection</td>
<td>( i_s \frac{2 \sigma_{A_L}^2}{\sqrt{\text{var} \ P_{L}}} )</td>
</tr>
</tbody>
</table>

\( \text{Var} \ P \) is the phenotypic variance of the tested units, \( i_s \) is the selection intensity for mass selection, and \( i_s \) is the selection intensity for family or progeny testing.

It is not the aim of this paper to compare the relative merits of the different methods. However, according to the expressions given, and knowing the parameters, and in particular \( \sigma_{A_L}, \rho_{TL} \) and \( h_T^2 \), this will be possible. GALLAIS (1989) showed that method 6 is expected to be better than method 5. Without calculation it appears that method 6, with the same amount of resources for testing, the same selection intensity and the same length of cycle, is more efficient than methods 3, 4, and 5, mainly for low heritabilities due to a large effect of environment on the value in test.

With a greater length of the cycle for the single DH descent recurrent selection (method 6), it would be necessary to study the genetic advance per unit of time. In this case, the same conclusion holds as in the GRIFFING's (1975) study of the use of DH in recurrent selection to improve the value per se or the general combining ability: the key of the efficiency of methods of improvement for line value using DH is the development of rapid doubled-haploid extraction procedures. However for the plant breeder, for line development there remains a great advantage to methods 3, 4, and 5, mainly for low heritabilities due to a large effect of environment on the value in test.

With a greater length of the cycle for the single DH descent recurrent selection (method 6), it would be necessary to study the genetic advance per unit of time. In this case, the same conclusion holds as in the GRIFFING's (1975) study of the use of DH in recurrent selection to improve the value per se or the general combining ability: the key of the efficiency of methods of improvement for line value using DH is the development of rapid doubled-haploid extraction procedures. However for the plant breeder, for line development there remains a great advantage to methods 3, 4, and 5, mainly for low heritabilities due to a large effect of environment on the value in test.

**CONCLUSION**

The introduction of the concept of line value allows a general description of the populations of lines derived by haplodiploidization from a random mating population. From a quantitative genetics point of
view, such populations seem to provide a powerful tool for investigating epistasis of the homozygote \( \times \) homozygote type. CHOO (1981a) has also shown that such populations can be very useful to explore the effect of linkage. From a plant breeding point of view, the theory introduced leads to the conclusion that the single doubled haploid descent recurrent selection will be one of the most efficient methods to improve the line value of a population.

The author thanks very much M. TURELLI, B. S. WEIR and the three anonymous reviewers for their helpful suggestions.

**LITERATURE CITED**


Communicating editor: B. S. WEIR