Heredity **59** (1987) 105–115

Received 2 October 1986

Expectation of means and variances of testcrosses produced from F_2 and backcross individuals and their selfed progenies

A. E. Melchinger

Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, P.O. Box 70 05 62, D-7000 Stuttgart 70, F.R.G.

A biometrical genetic model is presented for the testcross performance of genotypes derived from a cross between two pure-breeding lines. The model is applied to obtain the genetical expectations of first and second degree statistics of testcrosses established from F_2, first and higher backcross populations and their selfing generations. Theoretically, the testcross mean of these populations is expected to be a linear function of the percentage of germplasm from each parent line in the absence of epistasis. In both F_2 and backcross populations, the new arising testcross variance between sublines is halved with each additional generation of selfing. Special consideration is given to the effects of linkage and epistasis, and tests for their presence are provided. The results are discussed with respect to implications in "second cycle" breeding. It is concluded that the choice of base populations between F_2 and first backcrosses can be made on the distributions of testcrosses from the first segregating generation. Schnell's (1983) "usefulness" criterion is recommended for choosing the optimum type of base population.

INTRODUCTION

In an advanced stage of hybrid breeding *e.g.*, in maize (*Zea mays* L.), new lines are predominantly developed either from advanced populations undergoing recurrent selection or from *ad hoc* synthesized populations between elite lines. According to a survey by Bauman (1977), the majority of the most widely used public inbreds in commercial U.S. maize hybrids were "second cycle" inbreds of the latter type selected from F_2 or backcross populations. Moreover, this author estimated that about 80 per cent of the effort in private maize breeding programmes is devoted to the improvement of established lines.

As in any breeding programme, the choice of base materials is critical for success in "second cycle" breeding. This holds true no matter whether the pedigree or the single seed descent systems are used to extract new inbred lines with superior performance. In the case of "second cycle" breeding, the choice of the base materials involves (a) choice of the parent lines and (b) a decision about the type of base population to be established from these lines. In practice, F_2 and backcrosses are most frequently used.

Concerning the first problem, Dudley (1984) recently presented a theoretical basis and possible solution to the problem of choosing the most promising lines for improving the parents of a single cross. Theoretical results for comparing F_2 and backcross populations with regard to *per se* performance were provided by Mather, Jinks and co-workers (see Mather and Jinks, 1982). However, no such theory is available with respect to the testcross performance of these populations.

In this paper a biometrical genetic analysis is presented for the testcross means and testcross variances of F_2 and arbitrary backcross populations and their selfing generations developed from a cross between two pure-breeding lines. Special consideration is given to the effects of epistasis and linkage and tests for their presence based on first and second degree statistics are provided. In addition, the results are discussed with regard to implications for "second cycle" breeding.

ASSUMPTIONS AND NOTATION

The two initial parents $P1$ and $P2$ of the populations subsequently considered are assumed to be homozygous lines. The F_1 cross between $P1$ and $P2$ is backcrossed to each parent to produce the $B1$ and $B2$ (also referred to as $B1_1S_0$ and $B2_1S_0$) generations, respectively. Backcrossing to parent

line Pp, $p = 1$ or 2, is continued by bulking pollen from a large (conceptionally infinite) number of Bp_bS_0 plants producing generation $Bp_{b+1}S_0$ ($b \geqq 0$). The population obtained from Bp_bS_0 after n generations of self-pollination is designated as Bp_bS_n. For reasons of generality, the F_1 is considered as backcross generation zero and the F_n is referred to as Bp_0S_n.

The tester T is assumed to be a homozygous line. In practice, the tester is often an elite line unrelated to $P1$ and $P2$ and with promising potential as a parent in hybrid combinations with newly developed lines. The testcrosses are established by mating the candidates with the tester.

In producing the backcrosses, selfed progenies, and testcrosses, the following assumptions are made:

(a) Normal diploid behaviour at meiosis
(b) No gametic or zygotic selection
(c) No mutation
(d) Recombination frequencies are the same in the male and female gametogenesis and independent of the cytoplasm and genetic background.

In the genetic model, it is assumed that epistatic interactions among three or more loci are absent.

THE GENETIC MODEL

Since the tester was assumed to be a homozygous line, differences in the genotypic values among testcross individuals are solely attributable to the genotype of the gametes received from the test candidate. Consider two loci, j and k, with two alleles, A-a and B-b, present in parents $P1$ and $P2$. Employing the statistical model proposed for hybrid populations (Griffing, 1962; Schnell, 1965) for the present case, the genotypic values of testcrosses resulting from the four possible types of gametes produced by the candidate population can be sub-divided in the following manner:

$$\begin{aligned} T \times AB &= m^T + d_j^T + d_k^T + i_{jk}^T \\ T \times Ab &= m^T + d_j^T - d_k^T - i_{jk}^T \\ T \times aB &= m^T - d_j^T + d_k^T - i_{jk}^T \\ T \times ab &= m^T - d_j^T - d_k^T + i_{jk}^T. \end{aligned} \qquad (1)$$

Here the parameters on the right-hand side are statistically defined using as a base of reference the gene-orthogonal population (Schnell, 1965) between tester T and the F_2-derived population of the cross $P1 \times P2$ in linkage equilibrium. Their dependence on the tester genotype is indicated by a superior T. Parameters d_j^T and d_k^T refer to half the average effect of a gene substitution at the loci j and k in the gamete from the candidate, respectively. Similarly, i_{jk}^T relates to the additive by additive epistatic effect between loci j and k.

The above model is identical in form to the model equation for the *per se* performance of diploid genotypes corresponding to the gamete of the candidate (see Mather and Jinks, 1982). In contrast to the parameters in equation (1), however, those of the latter model are defined with respect to the F_∞ metric (Van der Veen, 1959). Despite this fundamental difference, the formal analogy implies that part of the results given in this paper correspond to previous results in literature concerning the *per se* performance of homozygous lines. Conversely, new results given here for the testcross performance apply directly to the *per se* performance of homozygous lines.

The above model equation also holds true for the expected testcross performance whatever the population structure of the tester may be (*e.g.*, a single cross, synthetic or population), even if it is not in linkage equilibrium. However, heterogeneity in the gametic array produced by the tester would cause genetic variation within testcross progenies not considered here.

LINKAGE DISEQUILIBRIUM

With regard to expressions for the means and variances given in subsequent sections, it is convenient to provide general formulae describing the linkage disequilibrium in the gametic arrays produced by the populations investigated.

Using the conventional definition (Falconer, 1981, p. 19), the linkage disequilibrium D_{jk} between loci j and k in the gametic array produced from cross $P1 \times P2$ after n generations of selfing is given by the following expression (Cockerham and Weir, 1973):

$$\begin{aligned} {}_nD_{jk} = &\frac{\lambda_{jk}}{4(2-\lambda_{jk})} \\ &\times \{1 + (\lambda_{jk}/2)^n (1-\lambda_{jk})\} \Delta_{jk}, (n \geqq 0); \end{aligned} \qquad (2)$$

where

$$\Delta_{jk} = \begin{cases} +1 & \text{for coupling linkage and} \\ -1 & \text{for repulsion linkage in the parents, and} \end{cases}$$

λ_{jk} denotes the linkage values between loci j and k (Schnell, 1961).

In terms of the recombination frequency p_{jk} between loci j and k, λ_{jk} is equal to $1-2p_{jk}$ and hence becomes zero in the absence of linkage and unity with complete linkage.

Extending this result to backcrossing with subsequent selfing, the linkage disequilibrium ${}_{b,n}D_{jk}$ in the gametic array produced by generation Bp_bS_n ($p=1,2$; $b, n \geqq 0$) is obtained as

$$ {}_{b,n}D_{jk} = (1/4)^{b+1}\{(1+\lambda_{jk})^b(1+4|{}_nD_{jk}|)-1\}\Delta_{jk}. \quad (3) $$

Because of symmetry reasons, the linkage disequilibrium is independent of p, *i.e.*, identical for the two backcross series. Furthermore, if $\lambda_{jk}=0$ then ${}_{b,n}D_{jk}=0$ for every b, $n \geqq 0$, *i.e.*, the gametic array of any backcross generation is in linkage equilibrium for unlinked loci.

OVERALL MEANS OF TESTCROSSES

Table 1 shows the frequencies of individual genotypes in the bth backcross (Bp_bS_0) and those obtained from each backcross genotype after n generations of selfing (Bp_bS_n) using results of Jennings (1917). In addition, the testcross mean of Bp_bS_0- and Bp_bS_n-derived lines (adopting the terminology of Wricke and Weber, 1986, p. 73) is given for testcrosses produced from generation

Table 1. Genotype frequencies in populations Bp_bS_0 and Bp_bS_n and genotypic testcross means of lines derived from them when testcrosses are produced in generation S_t ($t \geqq n$) on a digenic interaction model with linkage; the recurrent backcross parent is assumed to have genotype AB/AB

Genotype in Bp_bS_n	Parent genotype in Bp_bS_0: AB/AB Frequency $\left(\frac{2^b-1}{2^b}\right)^2$ $+\tau_b$†	AB/Ab $\frac{2^b-1}{4^b}$ $-\tau_b$	AB/aB $\frac{2^b-1}{4^b}$ $-\tau_b$	AB/ab $\frac{1}{4^b}$ $+\tau_b$	Testcross mean* of Bp_bS_n-derived lines when testcr. are produced in S_t: Coefficient of d_j^T	d_k^T	i_{jk}^T
		Frequency of descendants in Bp_bS_n					
AB/AB	1	$\frac{2^n-1}{2^{n+1}}$	$\frac{2^n-1}{2^{n+1}}$	$\frac{2^{n-1}-1}{2^{n+1}}+\frac{V^n+Un}{4}$§	1	1	1
AB/aB	0	$\frac{1}{2^n}$	0	$\frac{1}{2^{n+1}}-\frac{V^n}{2}$	1	0	0
AB/aB	0	0	$\frac{1}{2^n}$	$\frac{1}{2^{n+1}}-\frac{V^n}{2}$	0	1	0
AB/ab	0	0	0	$\frac{V^n+W^n}{2}$¶	0	0	$4_{t-n}D_{jk}$ ††
Ab/Ab	0	$\frac{2^n-1}{2^{n+1}}$	0	$\frac{2^{n-1}-1}{2^{n+1}}+\frac{V^n-U_n}{4}$	1	−1	−1
Ab/aB	0	0	0	$\frac{V^n-W^n}{2}$	0	0	$-4_{t-n}D_{jk}$
Ab/ab	0	0	0	$\frac{1}{2^{n+1}}-\frac{V^n}{2}$	0	−1	0
aB/aB	0	0	$\frac{2^n-1}{2^{n+1}}$	$\frac{2^{n-1}-1}{2^{n+1}}+\frac{V^n-U_n}{4}$	−1	1	−1
aB/ab	0	0	0	$\frac{1}{2^{n+1}}-\frac{V^n}{2}$	−1	0	0
ab/ab	0	0	0	$\frac{2^{n-1}-1}{2^{n+1}}+\frac{V^n+U_n}{4}$	−1	−1	1
Testcr. mean* of Bp_bS_0 derived lines when testcr. are prod. in S_t	$d_j^T+d_k^T+i_{jk}^T$	d_j^T	d_k^T	$4_tD_{jk}i_{jk}^T$ ††			

* Genotypic deviation from m^T.
† $\tau_b=\{(1+\lambda_{jk})^b-1\}/4^b$.
§ $V=(1+\lambda_{jk}^2)/4$; $U_n=\{1-(\lambda_{jk}/2)^n\}\lambda_{jk}/(2-\lambda_{jk})$.
¶ $W=\lambda_{jk}/2$.
†† See equation (2).

$S_t (t \geqq n)$. Using these results, the following general formula is obtained for the expected mean of the testcrosses produced from population Bp_bS_t:

$$\overline{T \times Bp_bS_t} = m^T + (-1)^{p+1}\{1-(1/2)^b\}[d^T] + \{1-(1/2)^b\}^2[i^T] + 4\sum_{j<k} |_{b,t}D_{jk}|\Delta_{jk}i^T_{jk}, \quad (b, t \geqq 0;\ p = 1, 2); \quad (4)$$

where

$$[d^T] = \sum_j \theta_j d^T_j,$$

$$\theta_j = \begin{cases} +1 & \text{if } P1 \text{ contains the favourable} \\ & \text{allele at locus } j \text{ and} \\ -1 & \text{otherwise,} \end{cases}$$

and

$$[i^T] = \sum_{j<k} \Delta_{jk} i^T_{jk}, \quad \text{with } \Delta_{jk} \text{ as defined above.}$$

The coefficient of the different terms on the right-hand side for the testcross means of various populations are presented in table 2. For F_1-derived generations, the results are analogous to those for the *per se* performance of homozygous lines developed by dihaploidy or single seed descent (Jinks and Pooni, 1981; Snape and Simpson, 1981). In the absence of epistasis, the testcross mean of the various populations is a linear function of $[d^T]$, *i.e.*, depends linearly on the percentage of germplasm of the two parents. With epistasis, this relation may become non-linear, the curvature depending on the sign of $[i^T]$ (fig. 1).

Assuming epistasis but no linkage, the averaged testcross mean of the parents $\overline{T \times P} = \overline{(T \times P1 + T \times P2)/2}$ differs from the testcross mean $\overline{T \times F_1} = \overline{T \times F_2}$ by $[i^T]$, the balanced sum of digenic epistatic effects. Similarly, the average testcross mean of the backcrosses $\overline{T \times B} = \overline{(T \times B1 + T \times B2)/2}$ differs from $\overline{T \times F_2}$ by $[i^T]/4$. Comparisons between these means therefore provide tests for the contribution of epistatic effects to the testcross performance. Net coupling linkage with complementary epistasis ($i^T_{jk} > 0$) or net repulsion linkage with duplicate epistasis ($i^T_{jk} < 0$) results in positive values for $[i^T]$. Alternatively,

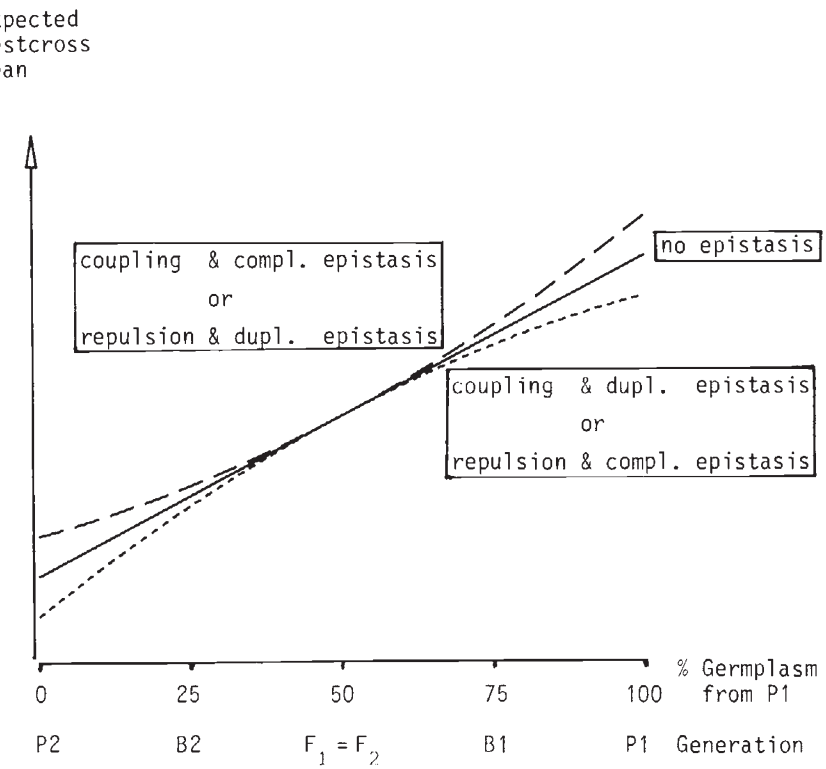

Figure 1 Graphical representation of the expected testcross means for various populations. The solid and the broken lines refer to the absence and presence of epistasis respectively. Calculations are based on $d^T_j = d^T_k = 1$ and $i^T_{jk} = \pm\frac{1}{2}$ assuming coupling phase in the parents and no linkage.

Table 2 Expected testcross means of various populations on a digenic interaction model in the presence of linkage calculated from equation (3); for explanation of symbols see text

Generation	m^T	Coefficient of $(-1)^{p+1}[d^T]$	Coefficient of $[i^T]$	Coefficient of $4\Delta_{jk}i^T_{jk}$
$T \times Pp$	1	1	1	0
$T \times Bp_0S_0 = T \times F_1$	1	0	0	λ_{jk}
$T \times Bp_0S_1 = T \times F_2$	1	0	0	$\lambda_{jk}(1+\lambda_{jk})/2$
$T \times Bp_0S_2 = T \times F_3$	1	0	0	$\lambda_{jk}(4+\lambda^2_{jk}-\lambda^3_{jk})/4(2-\lambda_{jk})$
$T \times Bp_0S_\infty = T \times F_\infty$	1	0	0	$\lambda_{jk}/(2-\lambda_{jk})$
$T \times Bp_1S_0 = T \times Bp$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\lambda_{jk}(2+\lambda_{jk})/4$
$T \times Bp_1S_1$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\lambda_{jk}(3+2\lambda_{jk}+\lambda^2_{jk})/8$
$T \times Bp_1S_\infty$	1	$\frac{1}{2}$	$\frac{1}{4}$	$3\lambda_{jk}/4(2-\lambda_{jk})$
$T \times Bp_2S_0$	1	$\frac{3}{4}$	$\frac{9}{16}$	$\lambda_{jk}(3+3\lambda_{jk}+\lambda^2_{jk})/16$
$T \times Bp_2S_\infty$	1	$\frac{3}{4}$	$\frac{9}{16}$	$\lambda_{jk}(5+2\lambda_{jk})/16(2-\lambda_{jk})$

For $p = 1, 2$.

for situations of coupling linkage with duplicate epistasis or repulsion linkage with complementary epistasis, $[i^T]$ becomes negative. Since only the net effect can be observed, balancing positive and negative contributions may result in epistasis not being detected even when it is present.

With net epistasis and no linkage among the genes involved, equations (3) and (4) imply that the testcross means of higher selfing generations do not differ from the testcross mean of the original backcross generation Bp_bS_0. In particular, the values for $\overline{T \times F_1}$, $\overline{T \times F_2}$, etc. and $\overline{T \times F_\infty}$ should be identical (table 2). With linkage, however, these means may differ from each other. The size of the differences depends on the linkage values and is greatest for intermediate rather than extreme λ_{jk} values. A comparison of $\overline{T \times F_1}$ and $\overline{T \times F_\infty}$ provides the most sensitive test for linked epistatic effects.

GENETIC VARIANCES OF TESTCROSSES

Utilising the results in table 1 and the appropriate general formula, the expectation of the total (between and within line) genetic variance among testcrosses produced from population Bp_bS_t becomes ($p = 1, 2$; $b, t \geqq 0$):

$$\begin{aligned}\text{Variance } (T \times Bp_bS_t) &= [1-\{1-(1/2)^b\}^2]\sum_j d_j^{T^2} + 4\sum_{j\neq k} {}_{b,t}D_{jk}d_j^T d_k^T \\ &+ (1-\{1-(1/2)^b\}^4)\sum_{j<k} i_{jk}^{T^2} \\ &- \{1-(1/2)^b\}^2 8\sum_{j<k} |{}_{b,t}D_{jk}| i_{jk}^{T^2} \\ &- 16\sum_{j<k} {}_{b,t}D_{jk}^2 i_{jk}^{T^2} \\ &+ (-1)^{p+1}2\{1-(1/2)^b\} \\ &\times \left[(1-\{1-(1/2)^b\}^2)\sum_{j\neq k}\theta_j\Delta_{jk}d_j^T i_{jk}^T \right. \\ &\left. -4\sum_{j\neq k}\theta_{j\,b,t}D_{jk}d_j^T i_{jk}^T\right]. \end{aligned} \quad (5)$$

An analogous formula applies to the genetic variance for the *per se* performance of doubled haploids derived from Bp_bS_t. For $b = 0$, *i.e.*, generations derived from the F_1, it agrees with the results of Snape and Simpson (1981).

Let ${}_gV(p_b; n)$ denote the total heritable testcross variance between Bp_bS_n-derived lines. This variance is given by the summation

$${}_gV(p_b; n) = \sum_{r=0}^{n}\sigma^2(p_b; r); \quad (6)$$

where $\sigma^2(p_b; r)$ denotes the genetic testcross variance between Bp_bS_0-derived lines ($r = 0$) or between Bp_bS_r-derived sublines within Bp_bS_{r-1}-derived lines ($r > 0$).

Making use of the results in table 1, these components of variance have the following expectations ($p = 1, 2$; $b \geqq 0$):

Model 1: No epistasis and no linkage

$$\sigma^2(p_b; 0) = (1/2)^b\{1-(1/2)^b\}\sum_j d_j^{T^2}; \quad (7)$$

$$\underline{r \geq 0}: \quad \sigma^2(p_b; r) = (1/2)^{b+r}\sum_j d_j^{T^2}; \quad (8)$$

Model 2: No epistasis but with linkage

$$\sigma^2(p_b; 0) = (1/2)^b\left[\{1-(1/2)^b\}\sum_j d_j^{T^2} + \sum_{j\neq k}\frac{(1+\lambda_{jk})^b-1}{2^b}\Delta_{jk}d_j^T d_k^T\right]; \quad (9)$$

$r>0$:

$$\sigma^2(p_b;r)=(1/2)^{b+r}\left[\sum_j d_j^{T^2}+\sum_{j\neq k}\lambda_{jk}^r\{(1+\lambda_{jk})/2\}^b\Delta_{jk}d_j^T d_k^T\right]; \quad (10)$$

Model 3: With epistasis and no linkage

$$\sigma^2(p_b;0)=(1/2)^b\left[\{1-(1/2)^b\}\sum_j d_j^{T^2}+\left(\frac{2^{3b+1}-5\cdot 2^{2b}+2^{b+2}-1}{8^b}\right)\sum_{j<k} i_{jk}^{T^2}+(-1)^{p+1}\{1-(1/2)^b\}^2 2\times\sum_{j\neq k}\theta_j\Delta_{jk}d_j^T i_{jk}^T\right]; \quad (11)$$

$r>0$:

$$\sigma^2(p_b;r)=(1/2)^{b+r}\left[\sum_j d_j^{T^2}+\left(\frac{2^{b+r+1}-3}{2^{b+r}}\right)\sum_{j<k} i_{jk}^{T^2}+(-1)^{p+1}\{1-(1/2)^b\}2\times\sum_{j\neq k}\theta_j\Delta_{jk}d_j^T i_{jk}^T\right]; \quad (12)$$

Model 4: With epistasis and linkage, assuming that the testcrosses were produced from selfed progenies in S_∞

$$\sigma^2(p_b;0)=\text{formula (9)}+(1/2)^b\times\left\{\sum_{j<k}\left\{2-(1/2)^{b-2}\times\left[1+\frac{\{(1+\lambda_{jk})/2\}^{2b}}{(2-\lambda_{jk})^2}\right]+\left(\frac{1+\lambda_{jk}}{2}\right)^b\left[1+\left(\frac{\lambda_{jk}}{2-\lambda_{jk}}\right)^2-\left(\frac{4}{2-\lambda_{jk}}\right)\{1-(1/2)^{b-1}\}\right]\right\} i_{jk}^{T^2}+(-1)^{p+1}\sum_{j\neq k}2\left\{1-(1/2)^{b-1}+\left(\frac{1+\lambda_{jk}}{2}\right)^b\left(\frac{(1/2)^{b-1}-\lambda_{jk}}{2-\lambda_{jk}}\right)\right\}\times\theta_j\Delta_{jk}d_j^T i_{jk}^T\right\}; \quad (13)$$

$r>0$:

$$\sigma^2(p_b;r)=\text{formula (10)}+(1/2)^{b+r}\times\left\{\sum_{j<k}\left\{2-\frac{(3-\lambda_{jk}^2)(2-2\lambda_{jk}+\lambda_{jk}^2)}{(2-\lambda_{jk})^2}\times\left(\frac{1+\lambda_{jk}^2}{2}\right)^{r-1}\left(\frac{1+\lambda_{jk}}{2}\right)^b\right\} i_{jk}^{T^2}+(-1)^{p+1}\times\sum_{j\neq k}2\left\{1-\left(\frac{1+\lambda_{jk}}{2}\right)^b\right\}\theta_j\Delta_{jk}d_j^T i_{jk}^T\right\}. \quad (14)$$

Under models 1 to 3 the formulae for $\sigma^2(p_b;r)$ do not depend on the selfing generation S_t used for producing the testcrosses. With both epistasis and linkage, however, the testcross means of Bp_bS_n-derived lines and consequently their variance may change with additional selfing due to the opportunity for further recombination as can be seen from table 1. For reasons of simplification, model 4 was restricted to testcrosses produced from S_∞.

For the selfing series obtainable from the $F_1(b=0)$ the above formulae concur with those of Jinks and Pooni (1984) for the *per se* performance of homozygous lines apart from some misprints. Furthermore, since F_1 individuals and their selfing progenies do not vary in their genotypic testcross means, we obtain $\sigma^2(p_0;0)=0$ under all four models.

Under model 2, $\sigma^2(p_b;r)$ is independent of p, *i.e.*, $\sigma^2(1_b;r)=\sigma^2(2_b;r)$ for $b>0$, $r\geqq 0$. Thus, both backcross series release the same genetic testcross variance in corresponding generations. With epistasis (models 3 and 4) these variances may differ on account of cross product terms in d_j^T and i_{jk}^T. These terms cancel by averaging over both backcross series ($p=1,2$).

Under model 1 it follows from equation (8) that $\sigma^2(p_b;r)=\sigma^2(pb';r')$ if $b+r=b'+r'$ for b, $b'\geqq 0$ and r, $r'>0$. This signifies that in the absence of epistasis and linkage the testcross variance originating from segregation in a later selfing generation depends solely on the sum of $b+r$.

F_2 AND FIRST BACKCROSS DERIVED SELFING GENERATIONS

Relationships among genetic variances

Let us now consider the results of the previous section in greater detail for the most important special cases $b=0$ and $b=1$, *i.e.*, testcrosses established from F_2 and first backcross populations and their selfed progenies. For reasons of simplicity, the abbreviations $p_0:=0$ and $p_1:=p$ ($p=1,2$) are subsequently used.

With no epistasis and no linkage (model 1), the new arising testcross variance between sublines in F_2 populations is halved with each further generation of selfing, *i.e.*, the ratio $\sigma^2(0;r):\sigma^2(0;r+1)=2:1$ for $r>0$ (table 3). In

Table 3 Sub-division of the genetic testcross variances between homozygous lines derived from F_2 and first backcross (Bp; $p = 1, 2$) populations; for symbols see text

Segregating generation	Source of variation	Designation of genetic variance	Coefficient of $\sum_j d_j^{T^2}$ $M. 1, 2, 3, 4$†	$\Delta_{jk} d_j^T d_k^T$ model 2, 4	$\sum_{j<k} i_{jk}^{T^2}$ model 3	$\sum_{j \neq k} \theta_j \Delta_{jk} d_j^T i_{jk}^T$ model 3	$i_{jk}^{T^2}$ model 4	$\theta_j \Delta_{jk} d_j^T i_{jk}^T$ model 4
1	Between F_2-derived lines	$\sigma^2(0; 1)$	$\frac{1}{2}$	$\lambda_{jk}/2$	$\frac{1}{4}$	0	$(2-2\lambda_{jk}+\lambda_{jk}^2-2\lambda_{jk}^3+\lambda_{jk}^4)/2(2-\lambda_{jk})^2$	0
2	Between F_3-derived sublines within F_2-derived lines	$\sigma^2(0; 2)$	$\frac{1}{4}$	$(\lambda_{jk}/2)^2$	$\frac{5}{16}$	0	$(1/2)-(1+\lambda_{jk}^2)(3-\lambda_{jk}^2) \times(2-2\lambda_{jk}+\lambda_{jk}^2)/8(2-\lambda_{jk})^2$	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$n+1$	Between F_{n+2}-derived sublines within F_{n+1}-derived lines	$\sigma^2(0; n+1)$	$(1/2)^{n+1}$	$(\lambda_{jk}/2)^{n+1}$	$(2^{n+2}-3)/4^{n+1}$	0	$(1/2)^n-\{(1+\lambda_{jk}^2)/4\}^n\{(3-\lambda_{jk}^2) \times(2-2\lambda_{jk}+\lambda_{jk}^2)/2(2-\lambda_{jk})^2\}$	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	Total genetic var. between SSD lines developed from F_2	${}_gV(0; \infty)$	1	$\lambda_{jk}/(2-\lambda_{jk})$	1	0	$1-\{\lambda_{jk}/(2-\lambda_{jk})\}^2$	0
1	Between Bp_1S_0-derived lines	$\sigma^2(p; 0)$	$\frac{1}{4}$	$\lambda_{jk}/4$	$\frac{3}{16}$	±§1/4	$(1+\lambda_{jk})(3-5\lambda_{jk}+2\lambda_{jk}^2)/4(2-\lambda_{jk})^2$	±§$(1-\lambda_{jk}^2)/2(2-\lambda_{jk})$
2	Between Bp_1S_1-derived sublines within Bp_1S_0-derived lines	$\sigma^2(p; 1)$	$\frac{1}{4}$	$(\lambda_{jk}/2)(1+\lambda_{jk})/4$	$\frac{5}{16}$	±1/4	$(1/2)-\{(1+\lambda_{jk})(3-\lambda_{jk}^2) \times(2-2\lambda_{jk}+\lambda_{jk}^2)/8(2-\lambda_{jk})^2\}$	$\pm(1-\lambda_{jk})/4$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$n+1$	Between Bp_1S_n-derived sublines within Bp_1S_{n-1}-derived lines	$\sigma^2(p; n)$	$(1/2)^{n+1}$	$(\lambda_{jk}/2)^n(1+\lambda_{jk})/4$	$(2^{n+2}-3)/4^{n+1}$	$\pm(1/2)^{n+1}$	$(1/2)^n-\{(1+\lambda_{jk}^2)/4\}^{n-1}\{(1+\lambda_{jk}) \times (3-\lambda_{jk}^2) \times(2-2\lambda_{jk}+\lambda_{jk}^2)/8(2-\lambda_{jk})^2\}$	$\pm(1/2)^{n+1}(1-\lambda_{jk})$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	Total genetic var. between SSD lines developed from Bp	${}_gV(p; \infty)$	$\frac{3}{4}$	$(3/4)\lambda_{jk}/(2-\lambda_{jk})$	$\frac{15}{16}$	±3/4	$\{15-6\lambda_{jk}/(2-\lambda_{jk}) -9\lambda_{jk}^2/(2-\lambda_{jk})^2\}/16$	$\pm(3/2)(1-\lambda_{jk})/(2-\lambda_{jk})$

† Model 1: No epistasis and no linkage; model 2: no epistasis but with linkage; model 3: with epistasis and no linkage; model 4: with epistasis and linkage.
§ The + sign refers to $p = 1$ and the − sign to $p = 2$.

the absence of epistasis (model 2), the preponderance of coupling or repulsion linkages enhances or reduces this ratio, respectively. The quantitative effect depends on the linkage values. With extremely tight linkage, the ratio is hardly affected. With extremely loose linkage, the contributions of the cross product terms $\Delta_{jk} d_j^T d_k^T$ to $\sigma^2(0; r)$ and ${}_gV(0; \infty)$ are rather small. The greatest influence on deviations from the ratio of 2:1 is exerted by gene pairs with intermediate to low linkage values, particularly during the first few generations of selfing.

Epistasis between unlinked loci (model 3) leads to an increase in the secondary ($r = 2$) and tertiary ($r = 3$) variances relative to the primary ($r = 1$) variance. In later selfing generations, epistasis causes practically no deviation from the ratio 2:1. Linkage hardly affects the coefficients of $i_{jk}^{T^2}$ for lower linkage values. With tighter linkage, the contributions of epistatic effects to $\sigma^2(0; r)$ and ${}_gV(0; \infty)$ decrease and finally vanish for completely linked loci ($\lambda_{jk} = 1$).

Under model 1, the genetic testcross variance between homozygous lines derived from F_2, ${}_gV(0; \infty)$, is twice the testcross variance between F_2 individuals, σ^2 (0; 1). Hence, the latter can be used to predict the former. With significant epistasis, however, this predictor is biased. Considerable bias can also be generated by linkage as was demonstrated by Kearsey (1985) for the *per se* performance of lines.

The results for the backcrosses $B1$ and $B2$ (table 3) show that under model 1, $\sigma^2(p; 0)$, the testcross variance between Bp_1S_0-derived lines, is equal to $\sigma^2(p; 1)$, the testcross variance between sublines after the first selfing generation. The testcross variances between sublines in successive selfing generations comply with a ratio of 2:1 as might be anticipated from the above results for F_2-derived progenies.

Under model 2, $\sigma^2(p; 0)$ is greater than $\sigma^2(p; 1)$ when the parents are predominantly in coupling and smaller when they are in repulsion. As with the initial selfing generations of F_2 populations, loci pairs with intermediate linkage values exert the strongest influence on the ratio. Epistasis between unlinked genes (model 3) causes an increase in $\sigma^2(p; 1)$ relative to $\sigma^2(p; 0)$. Epistasis and linkage affect the testcross variances between sublines from Bp in later selfing generations in an analogous manner as specified above for F_2-derived progenies.

The genetic testcross variance between homozygous lines from Bp, ${}_gV(p; \infty)$, can either be predicted by $3\sigma^2(p; 0)$ or $3\sigma^2(p; 1)$. The first predictor is simpler and more rapidly obtainable but is biased in the presence of epistasis. The second predictor is not biased by epistasis between unlinked loci and also yields good approximation in the presence of linkage.

Under model 2, the individual variance components $\sigma^2(p; r)$ and consequently ${}_gV(p; \infty)$ are identical for both backcrosses ($p = 1, 2$) as pointed out in the preceding section. Differences between corresponding testcross variances in $B1$ and $B2$ are attributable to the sum of cross product terms $\theta_j \Delta_{jk} d_j^T i_{jk}^T$ and therefore supply additional evidence of epistasis. For dispersed gene association, $\theta_j \Delta_{jk} d_j^T i_{jk}^T$ and $\theta_k \Delta_{jk} d_k^T i_{jk}^T$ have opposing signs and hence their contributions cancel each other at least partly in the total sum. With coupling linkage, both products have the same sign. Altogether, it can be concluded that the components of the testcross variance in $B1$ (the backcross to the higher performing parent with $\theta_j = 1$ at the majority of loci) are greater than those in $B2$ when there is predominantly complementary epistasis between gene pairs in coupling. Conversely, the variances in $B1$ are smaller than in $B2$ with a preponderance of duplicate epistasis among loci in coupling.

Table 3 enables us to compare the testcross variances of F_2 and backcross populations in corresponding generations. Under model 2, we have

$$\sigma^2(0; 1) = 2\sigma^2(p; 0), \qquad (p = 1, 2). \tag{15}$$

With epistasis among unlinked genes (model 3), we obtain

$$\sigma^2(1; 0) + \sigma^2(2; 0) - \sigma^2(0; 1) = \sum_{j<k} i_{jk}^{T^2}/8. \tag{16}$$

This relationship therefore provides a test for epistasis and allows the variance of epistatic effects, $I^T = \sum_{j<k} i_{jk}^{T^2}$, to be estimated. It takes into consideration both complementary and duplicate types of epistasis. With lower linkage values, this relationship also applies approximately for epistasis between linked loci.

Under model 3, we obtain from table 3 for $r > 0$:

$$\{\sigma^2(1; r) + \sigma^2(2; r)\}/2 = \sigma^2(0; r+1), \tag{17}$$

i.e., the average testcross variance between sublines in the rth selfing generation of $B1$ and $B2$ equals the testcross variance between sublines from the F_2 with one additional generation of selfing. In the presence of linkage, however, the left- and right-hand side may differ. Equation (16) thus provides a test for linkage which is neither conditioned by

epistatic interactions nor biased by epistasis between unlinked genes. It is most sensitive for intermediate linkage values ($\lambda_{jk} = 0{\cdot}5$). However, because positive and negative cross product terms $\Delta_{jk} d_j^T d_k^T$ for coupling and repulsion linkages respectively cancel at least partly in the total sum, only the net linkage effect can be detected.

Under model 2, the testcross variance between homozygous lines extracted from the F_2 and $B1$ or $B2$ populations (${}_gV(0;\infty)$, ${}_gV(1;\infty)$, and ${}_gV(2;\infty)$, respectively) are in a ratio of 4:3. In the presence of epistasis (model 3), we obtain

$$2\{{}_gV(1;\infty) + {}_gV(2;\infty)\}/3 - {}_gV(0;\infty) = I^T/4. \tag{18}$$

This relationship therefore gives a further test for epistasis and another estimate of I^T unaffected by linkage. It is more sensitive than the test presented in equation (15) but requires more time and effort.

Implications for "second cycle" breeding

Optimum allocation of test resources. In developing improved inbred lines by pedigree selection, the breeder usually selects for testcross performance in several generations. With respect to maximising the total gain from selection, an important question in this context concerns the optimum number of entries to be tested in each generation. Knowledge of the genetic variances in segregating generations provides a basis for optimising this multi-generation selection problem.

Utz (1984) gave an approximation for this optimum, assuming that the total number of plots in the selection process is constant and that the total selection response is obtained from the sum of the selection responses in the individual generations. Accordingly, the number of entries to be tested in each generation should be chosen in proportion to the genetic variability released in the various generations. Generally, the ratio of the genetic standard deviations or genetic variances should be employed for traits with a high or low heritability respectively.

Using these results and those given in the previous section, the ratio of the number of testcrosses to be evaluated in any two successive generations with respect to F_2 populations should range between $\sqrt{2}$:1 and 2:1 according to the heritability. With the $B1$ and $B2$ populations, an equal number of testcrosses should be tested in the first two segregating generations. In the subsequent selfing generations, one should proceed as stated for F_2 progenies.

Choice of the type of base population. The theoretical results in the sections headed "Overall means of testcrosses" and "Relationships among genetic variances" also provide information concerning under which conditions F_2 or backcrosses offer greater potential for extracting new lines with superior testcross performance. As already discussed in connection with table 3, the F_2 and $B1$ or $B2$ populations release the same testcross variances between sublines in each generation except for the first one, ignoring epistasis and minor departures due to linkage. Consequently, the selection response for testcross performance within F_2- and Bp_1S_0-derived lines in later generations is expected to be almost identical. The decision between these two types of base populations can therefore be made on the basis of the distribution properties of the first segregating generation.

Fig. 2 shows the assumed genotypic frequency distributions of the testcross means of random individuals from the F_2, $B1$ and $B2$ populations for any quantitative trait in the absence of epistasis (model 2). According to the results in section 5, the overall testcross mean $\overline{T \times B1}$ falls between $\overline{T \times P1}$ and $\overline{T \times F_2}$ and thus exceeds the latter. On the other hand, the genetic standard deviation between testcross progenies is in F_2 $\sqrt{2}$ times greater than in $B1$ or $B2$ (see equation (15)). Altogether, the merits of F_2 vs. backcross base populations mainly depend on the differences in their testcross means relative to the size of their genotypic testcross variances.

Following Jinks and Pooni (1976), the proportion of recombinants with a genotypic testcross performance falling beyond a certain standard x (*e.g.*, the testcross performance of the better parent $P1$) might be considered as an objective criterion for choosing the optimum type of base population. Let $\pi_0(x)$ and $\pi_1(x)$ denote the probabilities of obtaining such recombinants in the F_2 and $B1$ populations respectively. These probabilities can be calculated by integrating the corresponding genotypic density functions as outlined by Mather and Jinks (1982, p. 344). Ignoring epistasis (model 2) and approximating the genotypic distribution by a standard normal distribution, these integrals may be replaced by the one tail normal probability integrals corresponding to the abscissa values of $(x - m^T)/\sqrt{\sigma^2(0;1)}$ and $(x - m^T - [d^T]/2)/\sqrt{\sigma^2(1;0)}$ for F_2 and $B1$ respectively. Using equation (15) we obtain that $\pi_0(x) > \pi_1(x)$ if and only if $x > m^T + [d^T](1 + 1/\sqrt{2})$. This signifies that the probability of obtaining recombinants with a genotypic testcross performance greater than x is higher for F_2 than for $B1$ if $x >$

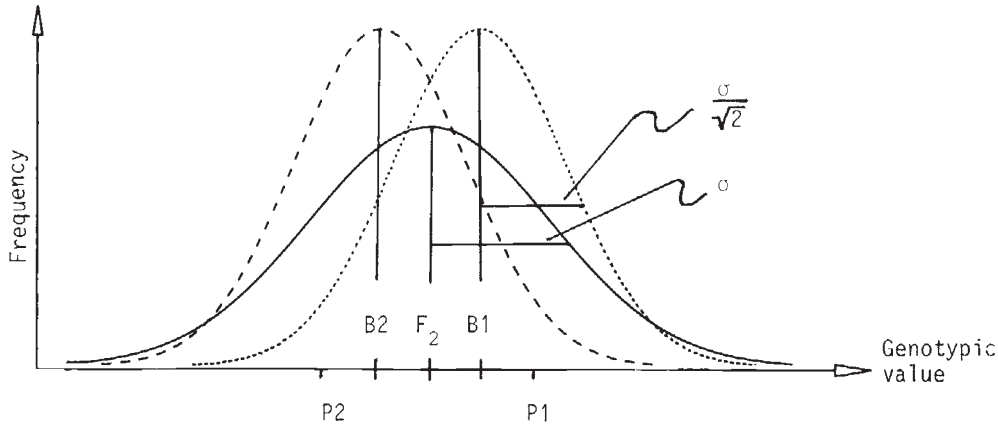

Figure 2 Assumed genotypic frequency distributions for the means of testcross progenies of random individuals from the F_2 and backcross populations $B1$ and $B2$ in the absence of epistasis.

$m^T + [d^T](1 + 1/\sqrt{2})$ and vice versa. We thus obtain the remarkable result that the sign of $\pi_0(x) - \pi_1(x)$ depends solely on m^T and $[d^T]$, *i.e.*, the testcross means of F_2 and $B1$ but not on their testcross variances. Due to this property, the probabilities $\pi_0(x)$ and $\pi_1(x)$ provide only limited information about the relative merits of F_2 vs. backcross populations.

As an alternative criterion, I suggest employing Schnell's (1983) concept of "usefulness" ("Brauchbarkeit") for assessing the breeding prospects of base populations in "second cycle" breeding. He defined "usefulness" $U(\alpha)$ as

$$U(\alpha) = \mu + R(\alpha) = \mu + i_\alpha \sigma h \qquad (19)$$

where $\mu \ldots$ denotes the population mean, $R(\alpha) =$ the expected selection response when selecting the upper α % phenotypes, $\sigma =$ the genetic standard deviation, $h =$ the square root of the heritability of the considered trait, and, $i_\alpha =$ the selection intensity when the upper α% phenotypes are selected.

In other words, "usefulness" denotes the expected genotypic mean of the upper α% phenotypes. Here, all quantities refer to testcross performance.

Obviously, "usefulness" accounts for differences in both means and genetic variances. Furthermore, it demonstrates that the heritability should also be considered for assessing the prospects of obtaining and identifying superior genotypes. As follows from equation (15), h for the testcrosses from F_2 is about $\sqrt{2}$ times as great as h for those from $B1$ if the masking variances (genotype by environment and error variances) are large, compared to the genetic variances. On the other hand, h is of a similar size in both types of populations if the masking variances are relatively unimportant. The dependence of $U(\alpha)$ on i_α indicates that the amount of resources allotted to a breeding programme also influences the "usefulness".

With "usefulness" as a criterion, the choice of F_2 vs. backcross base populations is reduced to the question whether the disadvantage of a lower population mean for the F_2 testcrosses can be offset by their greater selection response. This question cannot be answered in general but depends on the specific conditions of the materials and trait(s) considered. In summary, it can be concluded that the F_2 is likely to have superior "usefulness" than $B1$ if

—the differences in the testcross means of the F_2 and $B1$ populations are small compared to the pertinent genetic standard deviations.
—the heritability of the character is high, and
—a high selection intensity can be applied.

In practice, the choice of F_2 vs. backcross base populations in "second cycle" breeding is complicated by the fact that the breeder regards not only a single trait but several characters simultaneously. Despite this complication, the principles outlined above should help in clarifying the various aspects to be considered for this decision. Experimental results related to the theory in this paper are in progress and will be published elsewhere.

Acknowledgements The author is indebted to Professor H. H. Geiger for his advice and helpful criticism in preparing the manuscript; to Drs H. C. Becker, U. Posselt and H. F. Utz for critical reading of the manuscript; to M. C. Dawson, M. A., for his correction of the English text; and to Mrs Schwarz and Mr W. Daum for their assistance in typing the manuscript and preparing the figures.

A simplified version of this paper was read at the sixth meeting of the EUCARPIA Section "Biometrics in Plant Breeding" held in Birmingham, U.K., July 28-August 1, 1986.

REFERENCES

BAUMAN, L. F. 1977. Improvement of established maize inbreds. *Maydica, XXII*, 213-222.

COCKERHAM, C. C. AND WEIR, B. S. 1973. Descent measures for two loci with some applications. *Theor. Pop. Biol., 4*, 300-330.

DUDLEY, J. W. 1984. A method of identifying lines for use in improving parents of a single cross. *Crop Science, 24*, 355-357.

FALCONER, D. S. 1981. *Introduction to Quantitative Genetics* (2nd Edn.), Longman Press, London.

GRIFFING, B. 1962. Prediction formulae for general combining ability selection methods utilizing one or two random mating populations. *Australian J. Biol. Sci., 15*, 650-665.

JENNINGS, H. S. 1917. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. *Genetics, 2*, 97-154.

JINKS, J. L. AND POONI, H. S. 1976. Predicting the properties of recombinant inbred lines derived by single seed descent. *Heredity, 36*, 253-266.

JINKS, J. L. AND POONI, H. S. 1981. Properties of pure-breeding lines produced by dihaploidy, single seed descent and pedigree breeding. *Heredity, 46*, 391-395.

JINKS, J. L. AND POONI, H. S. 1984. Comparison of inbred lines produced by single seed descent and pedigree inbreeding. *Heredity, 53*, 299-308.

KEARSEY, M. J. 1985. The effect of linkage on additive genetic variance with inbreeding an F_2. *Heredity, 55*, 139-143.

MATHER, K. AND JINKS, J. L. 1982. *Biometrical Genetics*, 3rd ed. Chapman and Hall, London.

SNAPE, J. W. AND SIMPSON, E. 1981. The genetical expectation of doubled haploid lines derived from different filial generations. *Theor. Appl. Genet., 60*, 123-128.

SCHNELL, F. W. 1961. Some general formulations of linkage effects in inbreeding. *Genetics, 46*, 947-957.

SCHNELL, F. W. 1965. Die Covarianz zwischen Verwandten in einer gen-orthogonalen Population. I. Allgemeine Theorie. *Biometr. Zeitschr., 7*, 1-49.

SCHNELL, F. W. 1983. Probleme der Elternwahl—Ein Überblick. *Arbeitstagung der Arbeitsgemeinschaft der Saatzuchtleiter in Gumpenstein*, pp. 1-11.

UTZ, H. F. 1984. Calculating and maximizing the gain from selection. *Proceedings of the Fifth Meeting of the Biometrics Section of Eucarpia.* Hohenheim, Stuttgart. 30-41.

VAN DER VEEN, J. H. 1959. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. *Genetica, XXX*, 201-232.

WRICKE, G. AND WEBER, W. E. 1986. *Quantitative Genetics and Selection in Plant Breeding.* Walter de Gruyter Press, Berlin.