Response to recurrent selection under small effective population size

Cláudio Lopes de Souza Jr., Isaias Olívio Geraldi and Roland Vencovsky

Abstract

A formula was derived for the prediction of the response to recurrent selection when the effective population size (N_e) is small. Usually, responses to selection have been estimated by Rs = icσ_A^2/σ_y, where i, c, σ_A^2, and σ_y stand for standardized selection differential, parental control, additive variance, and phenotypic standard deviation, respectively. This expression, however, was derived under the assumption of infinite population size. By introducing the effects of finite population size, the expression derived was Rs = [ic(σ_A^2 + ΔFD)/σ_y] - ΔFD, where ΔF, ID and D_i are the changes in the inbreeding coefficient, the inbreeding depression, and the covariance of additive and homozygous dominance effects, respectively. Thus, the predicted responses to selection based on these expressions will be smaller than those based on the standard procedures for traits with a high level of dominance such as yield. Responses to five cycles of half-sib selection were predicted for maize by both expressions, considering that 100 progenies were evaluated and 10 S_1 progenies were recombined, which corresponds to N_e = 10 for each cycle. The accumulated response to selection estimated with the new expression was about 47 and 28% smaller than that based on the standard expression for yield and plant height, respectively. Thus, the expression usually used overestimates the responses to selection, which is in agreement with reported results, because it does not take into account the effective population size that is generally small in recurrent selection programs.

INTRODUCTION

The improvement of populations through recurrent selection is a common procedure in breeding programs designed to develop hybrids from inbred lines in maize and other allogamous species. Several recurrent selection methods have been reported (Hallauer et al., 1988), and formulas for the expected response to selection for each method have been derived. These expressions are linear functions of the intensity of selection and of the additive genetic variance, and are inverse functions of the phenotypic variance (Empig et al., 1972; Cockerham and Matzinger, 1985).

Genetic drift is expected to occur in recurrent selection irrespective of the method, because of the small number of selected progenies recombined to give rise to the improved populations. Furthermore, to increase the genetic gain and to lower the genetic load of the populations, inbred progenies have been used for recombination. These factors lead to a decrease in the effective size of the population under selection in which genetic drift is likely to take place, as has been reported for maize populations (Smith, 1979a,b, 1983; Helms et al., 1989; Eyerabide and Hallauer, 1991; Keeratinijakal and Lamkey, 1993). Expected and observed responses to intrapopulation selection usually show large discrepancies (Burton et al., 1971; Penny and Eberhart, 1971; Crossa and Gardner, 1989; Arriel et al., 1993). A possible reason for these discrepancies may be that the expressions of the expected responses to selection have been derived for a random mating population under the assumption of infinite population size (Kempthorne, 1957; Falconer and Mackay, 1996). The objective of this study was to derive a general expression for the expected response to intrapopulation selection by taking into account the effects of finite population size and present experimental results for a maize population.

MATERIAL AND METHODS

Genetic model

Consider a random mating population in both Hardy-Weinberg and linkage equilibrium under intrapopulation recurrent selection. Let p_0 and q_0 be the frequencies of the favorable (B) and unfavorable (b) alleles, and a and d be half the difference of the homozygous effects and the dominance effects, respectively (Falconer and Mackay, 1996). After one cycle of recurrent selection the frequencies of B and b alleles will be p_1 and q_1, respectively. The changes in these frequencies are functions of both selection (Δp) and drift (δp). Then, ε(p_1) = ε(p_0 + Δp + δp), and ε(q_1) = ε(q_0 - Δp - δp), where ε(Δp) = icp_0 q_0 σ_y/σ_p (Empig et al., 1972), and ε(δp) = ε(p δp) = ε(q δp) = ε(Δp δp) = 0, and ε(δp)^2 = p_0 q_0 / 2N_e (Falconer and Mackay, 1996). In these expressions i is the standardized selection differential, c is the parental control, σ_y is the phenotypic standard deviation, N_e is the effective size of the population, a = a + (q_0 - p_0) id is the average effect of gene substitution and ε denotes expectation. We also assumed that changes in gene frequencies through recurrent selection are small, so one could assume that ε(Δp)^2 ≡ 0, and this term will therefore be neglected in the following derivations.

Following the genetic model, the mean of the origi-
nal (unselected) population for one locus is $u_0 = u + (p_0 - q_0)a + 2p_0q_0d$ and the mean of the population after one cycle of selection is $u_1 = u + (p_1 - q_1)a + 2p_1q_1d$. Now, substituting $p_1$ and $q_1$ in $u_1$, by their expectations and after algebraic manipulations we will have $u_1$ as a function of the original mean, the changes due to selection and drift. Then, the expected response to selection was derived from the contrast $R_s = u_1 - u_0$.

Inbreeding depression (ID) is usually estimated by the expression: $ID = 2(u_{S0} - u_{S1})$, where $u_{S0}$ is the mean of the original population ($u_o$) and $u_{S1}$ is the mean of this population after one generation of selfing, i.e., $u_{S1} = u + (p_0 - q_0)a + p_0q_0d$. Then, inbreeding depression is $ID = 2p_0q_0d$. The inbreeding coefficient ($F_t$) as a function of the effective population size in the $t$th generation is $F_t = (1/2N_e) + [1 - (1/2N_e)]F_{t-1}$, where $F_t$ and $F_{t-1}$ are the inbreeding coefficients in the $t$th and $(t-1)^{th}$ generations, respectively. Then, the inbreeding depression due to the small population size is estimable by ∆FID for each cycle of selection, where ∆F is the change in the inbreeding coefficient (Falconer and Mackay, 1996).

Experimental procedures

Experimental data from ESALQ-PB1, an open-pollinated maize population, were used for illustration. One hundred pairs of $S_1$ and half-sib progenies were obtained in two-ear plants in a lower plant density field (25,000 plants/ha) by selfing the lower ears and crossing the upper ears with a pollen mixture from 50 random plants. These pairs of progenies were evaluated in a randomized complete block design, with the half-sib and $S_1$ progenies in a split-block arrangement to avoid competition due to inbreeding depression. Each sub-plot was one row 4.0 m long with 1.0 m between rows, with 20 plants per sub-plot after thinning (50,000 plants/ha). The experiments were carried out at two locations in Piracicaba (SP) and one location in Uberlândia (MG), with three replications per location. Data on yield as the mean of unhusked ear weight, and on plant height as the mean of five competitive plants per sub-plot were recorded.

Analysis of variance was performed according to the experimental design. Subsequently, analyses of variance were computed for each progeny type for each environment, and then combined across environments. Inbreeding depression was estimated as $ID = 2(S_{S0} - S_{S1})$ and $ID%$ as $(ID/S_{S0})100$, which corresponds to the expected total inbreeding depression, i.e., $ID = S_{S0} - S_{S1}$. In these expressions $S_{S0}$ and $S_{S1}$ refer to the general means of half-sibs and $S_1$ progenies, respectively (Table I). From the half-sib analysis, estimates of genetic and phenotypic parameters were obtained as follows: additive variance: $\sigma^2_A = 4(M_p - M_e)/RL$; phenotypic variance among progeny means: $\sigma^2 = M_p/RL$; and heritability ($h^2_{S1}$) = $(\sigma^2_A/4\sigma^2_{S1})100$, where $M_p$ and $M_e$ stand for progeny and error mean squares, and $R$ and $L$ are the number of replications and of locations, respectively.

The covariance of additive and homozygous dominance effects was estimated as $D_1 = 4[Cov_G(HS,S_1) - (1/2) \sigma^2_A]$, where $Cov_G(HS,S_1)$ is the genetic covariance of half-sib and $S_1$ progenies (Table II). For estimation of the expected response to selection, an intrapopulation half-sib selection was considered, where half-sib progenies and $S_1$ progenies are the selection and recombination units, respectively. Since the effective size of an $S_1$ progeny is 1, the different effective population sizes considered in this paper refer to the number of recombined $S_1$ progenies.

RESULTS AND DISCUSSION

Theoretical response to selection

The expected mean value of a population following one cycle of selection is:

$$u_1 = u + (p_1 - q_1)a + 2p_1q_1d,$$

and after algebraic manipulations, the expected mean of a population after one cycle of selection can be expressed as:

$$u_1 = u_o + 2p_0a - 2(\delta p)d.$$

For a random sample from a base population, $\epsilon(\delta p)^2 = p_0q_0/2N_e$ (Falconer and Mackay, 1996). However, to ac-

Table I - Mean values of grain yield (Y) and plant height (PH) for ESALQ-PB1 maize population in the $S_0$ and $S_1$ generations, inbreeding depression (ID), and coefficient of experimental variation (CV%).

<table>
<thead>
<tr>
<th>Generation</th>
<th>Y (g/plant)</th>
<th>PH (cm/plant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_0$</td>
<td>201.88 ± 8.42</td>
<td>213.31 ± 5.54</td>
</tr>
<tr>
<td>$S_1$</td>
<td>118.91 ± 9.37</td>
<td>156.16 ± 6.82</td>
</tr>
<tr>
<td>ID</td>
<td>165.94 ± 15.2</td>
<td>114.30 ± 10.4</td>
</tr>
<tr>
<td>ID%</td>
<td>82.2</td>
<td>53.6</td>
</tr>
<tr>
<td>CV% ($S_0$)</td>
<td>10.5</td>
<td>4.6</td>
</tr>
<tr>
<td>CV% ($S_1$)</td>
<td>16.9</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Table II - Estimates of additive variance ($\sigma^2_A$), phenotypic variance ($\sigma^2$), Dominance effects ($D_1$), and heritability ($h^2_{S1}$) from maize half-sib progenies, for yield (Y) and plant height (PH).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Y (g/plant)</th>
<th>PH (cm/plant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_A$</td>
<td>197.41 ± 67.91</td>
<td>184.39 ± 42.48</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>111.63 ± 15.79</td>
<td>70.88 ± 10.02</td>
</tr>
<tr>
<td>$D_1$</td>
<td>-50.32 ± 19.39</td>
<td>-34.72 ± 11.30</td>
</tr>
<tr>
<td>$h^2_{S1}$</td>
<td>44.21 ± 15.21</td>
<td>65.04 ± 14.98</td>
</tr>
</tbody>
</table>
commodate the situation in which changes in gene frequency occur both due to selection and drift, the set of selected genotypes was considered as the reference population for measuring the effect of sampling on allelic frequencies. Taking $p' = p_0 + \Delta p$ as the gene frequency in this set then

$$\varepsilon(\delta p')^2 = p'_0 q'_0 / 2N_e.$$  

Then, for the selected sample we have:

$$\varepsilon(\delta p')^2 = p_0 q_0 + \Delta \rho(q_0 - p_0) / 2N_e,$$

where $\rho$ is the inbreeding coefficient, $q_0$ the frequency of allele $b$, and $\Delta$ is the difference between the allele frequencies of the same locus. Then, if $\varepsilon(\delta p')^2 = p_0 q_0 + \Delta \rho(q_0 - p_0) / 2N_e$, and

$$\varepsilon(\delta p')^2 = [p_0 q_0 + \Delta \rho(q_0 - p_0)] / 2N_e.$$

Now, including $\varepsilon(\Delta p)$ in this expression leads to:

$$\varepsilon(\delta p')^2 = [(p_0 q_0 / 2N_e) + ic q_0 (q_0 - p_0)] / 2N_e.$$

Then, substituting $\varepsilon(\Delta p)$ and $\varepsilon(\delta p')^2$ in $u_i$, we have:

$$u_i = u_0 + \alpha \varepsilon(\delta p')^2,$$

and the expression of the expected response to one cycle of selection ($Rs = u_1 - u_0$) taking into account the effective population size is:

$$Rs = [\alpha \varepsilon(\delta p')^2(\sigma_{\gamma}^2 + \Delta F_D) / \sigma_{\gamma}^2] - \Delta F D,$$

and the accumulated response after $t$ cycles of selection is:

$$Rs = [\alpha \varepsilon(\delta p')^2(\sigma_{\gamma}^2 + \Delta F_D) / \sigma_{\gamma}^2] - \Delta F D$$

These results were derived for one locus-two alleles model, but the extension for all loci is straightforward by letting $\sigma_{\gamma}^2$, $D_1$, and ID be the sum for all loci, as follows:

$$\sigma_{\gamma}^2 = \sum_k \sigma_{\gamma k}^2, D_1 = \sum_k D_k,$$

and $\mathrm{ID} = \sum_k \mathrm{ID}_k$.

The response to selection ($Rs$) has usually been estimated by using the following expression (Falconer and MacKay, 1996):

$$Rs = \alpha \varepsilon(\delta p')^2 / \sigma_{\gamma}^2$$

Notice that Equation 2 was derived on the assumption of infinite effective population size, whereas Equation 1 took into account the effective population size as finite. Thus, the expression derived in this study is less biased than Equation 2 because of the small number of progenies that has usually been recombined in recurrent selection programs (Smith, 1983). Comparing Equation 1 with 2, we note that the difference is a function of the last two terms, i.e., the $D_1$ and ID terms, related to genetic drift caused by small population size.

For most of the traits the term $-F D$ will reduce Rs, because it represents the decrease in the heterozygosity of the selected population. However, $D_1$, as a covariance, may be either negative or positive (Cockerham, 1984), and then could reduce or increase Rs. The effects of genetic drift increase as selection cycles proceed. Then, $N_e$ should be maintained at a level such that this effect could be reduced, and, therefore, the response to selection could be maintained as high as possible.

It is important to note that genetic drift may also lead to the fixation of undesirable alleles. Even without dominance $(d = 0)$, i.e., no inbreeding depression $(ID = 0)$ and $D_1 = 0$, genetic drift may reduce the response to selection (Robertson, 1960).

It should be emphasized that genetic drift is affected by selection, because $\varepsilon(\delta p')^2$ plus the effects of selection on it. The second term is a function of the selection intensity $(i)$, the selection procedure used $(c)$, and the allele frequencies $(p_0$ and $q_0)$. Note that this term may be either positive or negative, because it depends on the differences between the allele frequencies of the same locus. Then, if this term is positive, $(\delta p')^2 > (\delta p)^2$, and vice-versa.

The effective population size $(N_e)$ is also affected by a long-term selection, because as selection proceeds the individuals of the population under selection become genetically related, and, consequently, the inbreeding coefficient $(F)$ and the rate of inbreeding $(\Delta F)$ increases (Wray and Thompson, 1990; Wray et al., 1994). Therefore, for a long-term selection, the increase in $F$ and in $\Delta F$ should be added to the formula derived in this paper; otherwise the response to selection would be overestimated.

Wei et al. (1996) derived a formula to predict response to selection in finite populations by jointly considering the effects of genetic drift, linkage disequilibrium, inbreeding depression, and mutational variance. However, they did not consider the effects of selection on genetic drift and, consequently, the component $D_1$. Nonetheless, their results showed that mutational variance and effective population size could affect considerably the response to a long-term selection.

### Numerical evaluation

Values of expected responses $(Rs)$ for five cycles of selection were computed from parameter estimates of ESALQ-PB1 maize population for yield and plant height. Although theoretical results have shown that genetic variances change with selection and small population sizes (Nei, 1963; Cockerham, 1984), and that inbreeding depression is also expected to change with selection (Souza Jr., 1985), we considered $\sigma_{\gamma}^2$, $D_1$, ID and $\sigma_{\gamma}^2$ as constant for all
cycles. We assumed that these restrictions are stringent because experimental results have shown small changes in these parameters following selection (Stucker and Hallauer, 1992; San Vicente and Hallauer, 1993; Schnicker and Lamkey, 1993; Benson and Hallauer, 1994).

Estimates of response to one cycle of selection (Rs%) for yield and plant height were 3.9 and 6.2%, respectively, for finite size (Ne = 10), and 8.1 and 9.0%, respectively, for infinite Ne. Thus, the decreases in response when finite population size was taken into account were about 52 and 31% for yield and plant height, respectively. The greater decrease in yield occurred because of the higher level of dominance for this trait, which leads to a greater inbreeding depression. Note that these results refer to a half-sib selection scheme, where 10 S1 progenies were used for recombination. The expected accumulated decrease in yield due to icFtD/σPh and FtID following five cycles of selection for several Ne’s (Table III) showed that as Ne decreases and cycles of selection proceed, the effects of ID and D1, related to the genetic drift, increase as expected. For example, at the 5th selection cycle, the accumulated decrease from Ne = 30 (13.71 g/plant) to Ne = 10 (38.47 g/plant) increased about 181%. Also, for Ne = 10 the accumulated decrease from the 1st (8.51 g/plant) to the 5th (38.47 g/plant) cycle of selection increased about 352%. Note that the contributions of D1 and ID to the decrease were different, with the FtID component being about 40 times greater than the icFtD/σPh component.

Expected responses to selection for yield, considering the intensity of selection (10%) to be constant, and varying the number of progenies evaluated, i.e., varying the effective population size (Figure 1), showed an increase in the response to selection as Ne increased. Nevertheless, this increment was not linear, increasing sharply up to Ne = 20 and decreasing thereafter. For instance, the response to selection for Ne = 20 was about 42% greater than that for Ne = 10, whereas the response to selection for Ne = 30 was only 11% greater than that for Ne = 20. Thus, for yield in maize it seems worthwhile keeping the effective size number at least at 20. Considering the method and the intensity of selection used in this example, where S, progenies are used for recombination, the number of progenies to be evaluated should be around 200 to lower the inbreeding depression caused by the small population size.

Another situation, where the number of progenies to be evaluated is constant (100) but not the intensity of selection, i.e., varying Ne’s from 10 to 30, for yield was also considered (Figure 2). Results showed that the expected response to selection increased as the intensity of selection increased from 30 to 20%, but decreased thereafter (from 20 to 10%). This occurred because of the decrease in the effective population size that offsets the response to selection as shown in Table III. Despite the small differences in responses to selection, it would be important to maintain effective size as high as possible because genetic variance is expected to become smaller due to small population size.

General results for five cycles of selection for plant height, considering constant intensity of selection (10%) but varying effective population size by increasing the number of progenies evaluated (Figure 3), showed the same pattern as for yield (Figure 1), but with smaller differences among varying Ne’s. The level of dominance for plant height was lower than that for yield, and, consequently, its inbreeding depression was smaller than for yield (Hallauer et al., 1988). Thus, the differences among

---

### Table III - Expected decreases in the response to five cycles of half-sib selection due to icFtD/σPh (A), FtID (B), and the sum (A + B) for three population sizes (Ne), for maize yield (g/plant) at 10% selection intensity.

<table>
<thead>
<tr>
<th>Cycles</th>
<th>icFtD/σPh (A)</th>
<th>FtID (B)</th>
<th>A + B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ne</td>
<td>Ne</td>
<td>Ne</td>
<td>Ne</td>
</tr>
<tr>
<td>1</td>
<td>0.21</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>0.41</td>
<td>0.21</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>0.60</td>
<td>0.31</td>
<td>0.21</td>
</tr>
<tr>
<td>4</td>
<td>0.77</td>
<td>0.40</td>
<td>0.27</td>
</tr>
<tr>
<td>5</td>
<td>0.94</td>
<td>0.50</td>
<td>0.34</td>
</tr>
</tbody>
</table>

*i = 1.755; c = 0.50.*
the responses to selection by varying $N_e$’s for plant height were smaller than for yield because of the lower level of dominance of this trait.

Discrepancies between the expected and observed responses to selection can also be explained by sampling errors, precision of estimates, and genotype-environment interaction. Sampling errors could occur during the recombination of selected progenies; estimates of genetic and phenotypic variances generally have low precision, and genotype-environment interaction may reduce the response to selection, mainly in atypical years. Also, errors can occur in the expected response by using the standardized selection differential if the progeny means do not fit a normal distribution (Arriel et al., 1993). Recently, the effect of the gametic-phase disequilibrium was reported as another important source of discrepancies (Mackay and Gibson, 1993). Simulation studies by these authors showed reductions attributed to linkage disequilibrium between 16 and 32%, when comparing observed and expected responses to selection.

It was shown in this study that effective population size ($N_e$) is an important source of discrepancy between the expected and observed responses to selection, mainly when effective size is very small and the level of dominance of the trait is high. We have considered, for the sake of simplicity, only half-sib selection, but deriving expressions for other selection schemes are straightforward by letting parental control (c), effective population size ($N_e$), and the phenotypic standard deviation ($\sigma_p$) be those related to the selection scheme considered.

ACKNOWLEDGMENTS

Publication supported by FAPESP.

REFERENCES


(Received July 14, 2000)