Genet. Res., Camb. (1980), **36**, *pp.* 277-287
With 4 text-figures
Printed in Great Britain

Combining superior alleles from two homozygous populations in a cross-fertilizing species

By Y. T. HO* and R. E. COMSTOCK

Department of Genetics and Cell Biology, University of Minnesota, St Paul, Minnesota 55108, U.S.A.

(*Received* 6 *August* 1979 *and in revised form* 27 *May* 1980)

SUMMARY

The issue of incorporating the good alleles from two homozygous populations of a cross-fertilizing species into a single improved population was investigated assuming independent assortment, no epistasis, and either complete dominance (of the favourable or the unfavourable alleles) or additive genes. The selection limit in the foundation stock is a function of the effective population size (N), the proportion (x) contributed by the better source population (P_1), the difference in relative fitness between single locus homozygotes and the proportion of loci (that will segregate in the foundation stock) fixed favourably in P_1. In real life the last two of these are never known. We therefore focused on the response limits given $x = 0{\cdot}5$, $0{\cdot}75$, $0{\cdot}875$ or the optimum value of x (which is a function of the other three parameters). Our general finding was that in situations where N is large enough so that a major portion of the potential can be achieved, the F_2 population should be used as the foundation stock when the two source populations do not differ greatly in performance; but when one population performs considerably better than the other, the first backcross (but not second backcross) would be the choice.

1. INTRODUCTION

The majority of plant and animal breeders believe that even the best populations in economic species do not possess the total allele resource of the species. They believe that to produce the best possible genotypes, the alleles required must be brought together from different populations. This view has been advanced by Lerner & Donald (1966) and Harland (1975).

Aggregation of useful alleles from separate sources calls for crossing followed by recurrent selection to increase frequencies of all useful alleles introduced from any of the populations involved. Assuming a selection criterion such that all desired alleles are continuously favoured by the selection practised, the goal of such a programme would be fully achieved in a population having infinite effective size. On the other hand, in finite populations alleles may be lost through random drift despite being favoured by selection. Kimura (1957) illuminated the way in which the probability of fixation of an allele is conditioned by selection and effective

* Present address: Department of Biology, University of Wisconsin-Eau Claire, Eau Claire, Wisconsin 54701, U.S.A.

0016-6723/80/2828-8230 $01.00

population size when the gene assorts independently from all other segregating genes that contribute to the selective value of the total genotype. Hill & Robertson (1966) explored the effects of linkage between two genes on the probabilities of fixation of the alleles of both.

In recurrent selection programmes aimed at maximizing total value of the genotype, the distribution of fixation probabilities for the whole genome is at issue. This was dealt with by Robertson (1960, 1970). In the first of these contributions he assumed independent assortment and in the second he explored linkage effects given linkage equilibrium when selection is initiated. He did not focus on the special questions that arise in connexion with attempts to assemble all useful alleles available from two or more populations into one genotype.

The positive potential of allele introgression, by crossing, into high performance, plateaued populations has been demonstrated by Roberts (1967) and Osman & Robertson (1968) working with growth in mice and bristle numbers in Drosophila, respectively. The latter authors presented approximate expectations for the special case in which a plateaued population is crossed to the unimproved foundation stock from which the plateaued population was derived.

The work reported in this paper focused on the problem of combining the useful alleles contained in two homozygous lines of a cross-fertilizing species. It deals with the special case where all those alleles segregate independently.

2. METHOD AND SCOPE OF INVESTIGATION

The genetic assumptions included those made by Kimura (1957), Robertson (1960) and Osman & Robertson (1968). These were two alleles per locus, effective population size the same in each generation, relative fitness of genotypes constant through time, no gametic selection, random mating except as conditioned by variation in relative fitness, diploid organisms and that each gene considered assorts independently from all others with effects on fitness. Our additional assumption was that the differential effects of alleles were the same at all loci.

Primary objectives of the research were as follows:

(A) To determine the optimum proportions of separate source populations to be used when attempting to incorporate favourable alleles from two source populations into one improved population, and the factors that influence them.

(B) To discover whether that information would have practical value for breeders.

There are cost arguments for using a larger proportion of the genetically best of two source populations than the one that maximizes the selection limits. To do so usually means a higher initial performance level in the selection population and therefore higher average performance and lower net cost over the period of the development programme. In addition there are certain easily achieved proportions. The easiest way to get started is to cross the two populations and use the F_2 as the base for the selection population in which case the source populations are used in equal proportions. It the two populations differ considerably in performance the cost argument outlined above will be favourable to using a first or second back-

cross to the better source population, P_1, as the base population in which case the proportion of P_1 would be 0·75 or 0·875. With these things in mind, special attention was given to the difference in outcome to be expected when the proportion of P_1 is 0·5, 0·75 or 0·875 instead of the optimum, whatever that might be.

These issues were investigated assuming (1) no dominance, (2) complete dominance of the more favourable allele, and (3) complete dominance of the less favourable allele.

Mathematical analysis, based on Kimura's equation for the probability of fixation of a specific allele (Kimura, 1957; Crow & Kimura, 1970), was employed to answer these questions. For two alleles, let s and hs be the selective advantage of the favourable homozygote and the heterozygote, respectively, over the less favourable homozygote. Then the probability of fixation of a superior allele with initial frequency, q, is given by

$$\mu(q) = \frac{\int_0^q e^{-2Ns(2h-1)v(1-v)-2Nsv}dv}{\int_0^1 e^{-2Ns(2h-1)v(1-v)-2Nsv}dv}, \tag{1}$$

where N is the effective population size and e is the base of the natural system of logarithms. Assumptions involved were listed above.

If the alleles have additive effects on selective advantage, we have, by putting $h = \frac{1}{2}$ into equation (1),

$$\mu(q) = \frac{\int_0^q e^{-2Nsv}dv}{\int_0^1 e^{-2Nsv}dv}$$

$$= \frac{1-e^{-2Nsq}}{1-e^{-2Ns}}. \tag{2}$$

For dominance or recessivity of the favourable allele ($h = 1$ or 0), explicit solutions cannot be obtained and we use numerical integration of (1).

3. RESULTS

(i) *Optimum proportions contributed by two homozygous source populations*

Let x be the proportion of the foundation stock coming from P_1. The foundation stock will have the four classes of loci identified in Table 1. Selection can only be effective at loci of classes II and III. Let w be the proportion of loci (that will segregate in the foundation stock) that are fixed favourably in P_1.

Given x and w, we can derive the selection limit, following Robertson (1960), as the expected percentage, ϕ_x, of all segregating loci fixed for the favourable alleles. This will be the weighted sum of the probabilities of fixation for the favourable alleles at each of the two classes of segregating loci. Thus

$$\phi_x = w\mu(x) + (1-w)\,\mu(1-x), \tag{3}$$

Table 1. *Initial frequency of favourable alleles in the foundation stock*

Class of loci	Fixed in	Initial frequency of favourable alleles
I	P_1 and P_2*	1
II	P_1 but not P_2	x†
III	P_2 but not P_1	$1-x$
IV	Neither P_1 nor P_2	0

* P_1 and P_2 indicate the better and poorer source populations respectively.
† x is the proportion of the foundation stock coming from P_1.

where $\mu(q)$, the probability of fixation of an allele with initial frequency q, is given by equation (1).

For two alleles with additive effects on selective advantage, i.e. when $h = \frac{1}{2}$, we have, by substituting $\mu(q)$ given in equation (2) into equation (3),

$$\phi_x = w\left[\frac{1-e^{-2Nsx}}{1-e^{-2Ns}}\right]+(1-w)\left[\frac{1-e^{-2Ns(1-x)}}{1-e^{-2Ns}}\right].$$

The optimum proportion, $\hat{x}$, contributed by P_1, obtained by differentiation but remembering that x cannot exceed 1·0, is

$$\hat{x} = \tfrac{1}{2}\left[1+\frac{1}{2Ns}\ln\frac{w}{1-w}\right]$$

$$\text{or } 1{\cdot}0 \text{ when } \left[\frac{1}{2Ns}\ln\frac{w}{1-w}\right] > 1{\cdot}0. \tag{4}$$

Equation (4) may be written:

$$\hat{x}-\tfrac{1}{2} = \frac{1}{4Ns}\ln\frac{\frac{1}{2}+(w-\frac{1}{2})}{\frac{1}{2}-(w-\frac{1}{2})}$$

$$= \frac{1}{4Ns}[4(w-\tfrac{1}{2})+\text{terms in } (w-\tfrac{1}{2})^3]$$

$$\sim (w-\tfrac{1}{2})/Ns$$

$\hat{x}$ is then a function of the effective population size (N), the selection advantage of the better homozygote (s), and the proportion of loci (that will segregate in the foundation stock) that are fixed favourably in P_1 (w).

Figure 1 gives the fraction of segregating loci expected to be fixed favourably for different values of x. The following facts, illustrated by this figure, deserve attention:

(a) An increase in Ns increases the percentage of segregating loci fixed favourably.

(b) When w is one-half, $\hat{x}$ is one-half.

(c) When Ns is large, $\hat{x}$ approaches one-half.

(d) ϕ_x may be less than w for every value of $x < 1{\cdot}0$. This is illustrated in the case when $Ns = 1$ and $w = 0{\cdot}9$.

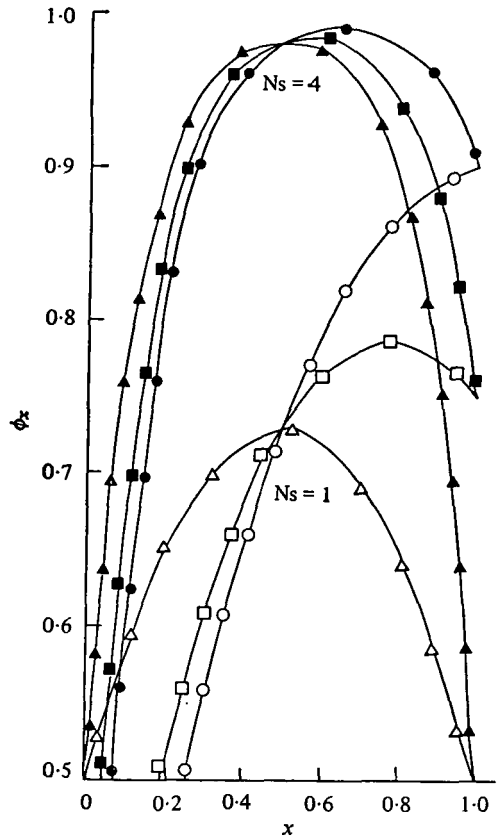

Fig. 1. Expected fraction of segregating loci fixed for the favourable alleles given a set of values for w, Ns and x when dominance is absent ($h = \frac{1}{2}$). ● or ○, $w = 0{\cdot}9$; ■ or □, $w = 0{\cdot}75$; ▲ or △, $w = 0{\cdot}5$.

Clearly when $\hat{x} = 1{\cdot}0$, a genetic loss must be expected when any use is made of the inferior source population. By rearranging equation (4), we can derive the upper limit of w for which there is any expectation of gain from crossing. We find that $\hat{x} < 1{\cdot}0$ requires

$$w < \frac{e^{2Ns}}{1+e^{2Ns}}. \tag{5}$$

Some of such critical values of w are shown in Table 2. When Ns is moderately large, superior populations can be improved by combining with inferior ones even though there is a considerable difference in performance between them.

When selection is for either a completely recessive ($h = 0$) or a completely dominant allele ($h = 1$), the optimum proportion $\hat{x}$ contributed by P_1 cannot be given explicitly in terms of Ns and w. The approximate value of $\hat{x}$ for particular cases can be found graphically from Figs 2 and 3. Figures 1–3 show that in general ϕ_x is largest when favourable alleles are dominant, and smallest when favourable alleles are recessive. The reverse is true, though differences are trivial, when both

https://doi.org/10.1017/S0016672300019893 Published online by Cambridge University Press

Table 2. *Maximum value of w that allows* $\phi_x \geqslant w$*

Ns	*w*
0·5	0·731
1·0	0·881
1·5	0·953
2·0	0·982
2·5	0·993
3·0	0·998
3·5	0·999
4·0	0·999

* ϕ_x is the expected percentage of all segregate loci fixed favourably.
w is the proportion of loci (that will segregate in the foundation stock) that are fixed favourably in the better source population.

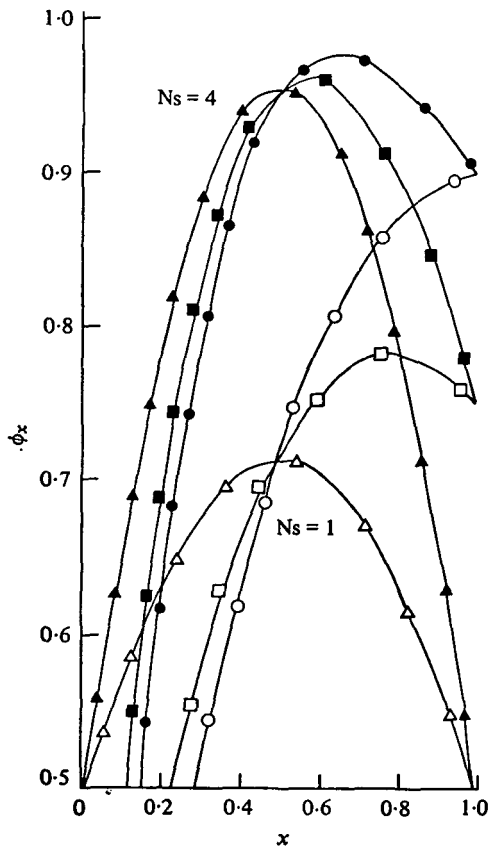

Fig. 2

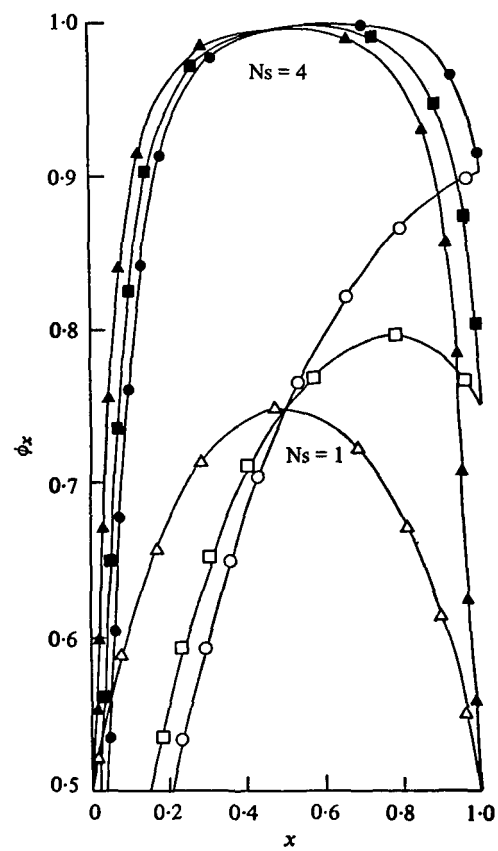

Fig. 3

Fig. 2. Expected fraction of segregating loci fixed for the favourable alleles given a set of values for w, Ns and x when favourable alleles are recessive ($h = 0$). ● or ○, $w = 0{\cdot}9$; ■ or □, $w = 0{\cdot}75$; ▲ or △, $w = 0{\cdot}5$.

Fig. 3. Expected fraction of segregating loci fixed for the favourable alleles given a set of values for w, Ns and x when favourable alleles are dominant ($h = 1$). ● or ○, $w = 0{\cdot}9$; ■ or □, $w = 0{\cdot}75$; ▲ or △, $w = 0{\cdot}5$.

https://doi.org/10.1017/S0016672300019893 Published online by Cambridge University Press

x and w are close to 1·0 and Ns is small. These facts are consistent with results reported by Carr & Nassar (1970).

Table 3. *Relative response,* R_x (%), *at selected values of* x,* w,† Ns *and different levels of dominance*

		Favourable alleles are recessive ($h = 0$)				No dominance ($h = \frac{1}{2}$)				Favourable alleles are dominant ($h = 1$)			
w	Ns	$R_{\hat{x}}$	$R_{0\cdot5}$	$R_{0\cdot75}$	$R_{0\cdot875}$	$R_{\hat{x}}$	$R_{0\cdot5}$	$R_{0\cdot75}$	$R_{0\cdot875}$	$R_{\hat{x}}$	$R_{0\cdot5}$	$R_{0\cdot75}$	$R_{0\cdot875}$
0·75	4	85·0	81·8	67·5	38·2	94·5	92·8	85·8	63·0	98·0	97·6	95·2	81·7
	5	91·5	89·9	73·4	42·4	98·0	97·3	91·6	71·3	99·5	99·4	98·1	88·7
	6	95·0	94·3	77·9	46·0	99·5	99·0	95·0	77·7	99·9	99·8	99·2	93·0
0·90	4	76·0	54·6	65·9	37·9	89·0	82·0	84·6	62·7	96·0	94·1	94·3	81·3
	5	87·0	74·7	72·9	42·3	96·0	93·3	91·3	71·3	99·0	98·7	98·0	88·7
	6	92·5	85·7	77·7	46·0	99·0	97·5	94·9	77·7	99·9	99·7	99·2	93·0

* See footnote in Table 1. † See footnote in Table 2.

(ii) *Effects of deviations from the optimum for the proportion contributed by the better homozygous source population*

$\phi^{\max}$, the maximum value of ϕ_x that can be attained is realized only when the favourable allele is fixed at all loci segregating in the foundation stock. Numerically, $\phi^{\max} = 1$. We define relative response as

$$R_x = \frac{\phi_x - w}{1 - w}(100)\%. \tag{6}$$

Figures 4*a*, 4*b* and 4*c* show R_x as a function of x for various combinations of Ns, w and h. The highest point on any curve is the relative response corresponding to the optimum value of x. Table 3 lists values of R_x at selected values of x and w and for values of Ns not treated in the figures. The following facts illustrated in these figures and tables should be noted.

(*a*) $\hat{x}$, the optimum for x, is not much affected by level of dominance.

(*b*) In every instance relative response is greatest when $h = 1$ and least when $h = 0$. Moreover, the decline in relative response as x deviates from $\hat{x}$ is least when $h = 1$ and greatest when $h = 0$. Thus the actual x employed would be less critical if a high proportion of favourable alleles were dominant.

(*c*) In all instances relative response is substantially greater when $Ns = 4$ than when $Ns = 2$. With $Ns = 4$, the expected relative response is $> 70\%$ unless x is too far from $\hat{x}$, as indicated when $h = 0$ and $w = 0{\cdot}9$.

(*d*) When $Ns = 4$ and $w \geqslant 0{\cdot}75$, the reductions in relative response when $x = 0{\cdot}75$ instead of $\hat{x}$ are quite small (ranging from 2 to 20%). On the other hand, the reductions when $x = 0{\cdot}875$ are much larger (ranging from 16 to 55%).

4. DISCUSSION

If any justification were needed for study of the problem of aggregating favourable alleles in the very special case of the alleles present in just two pure lines, it is

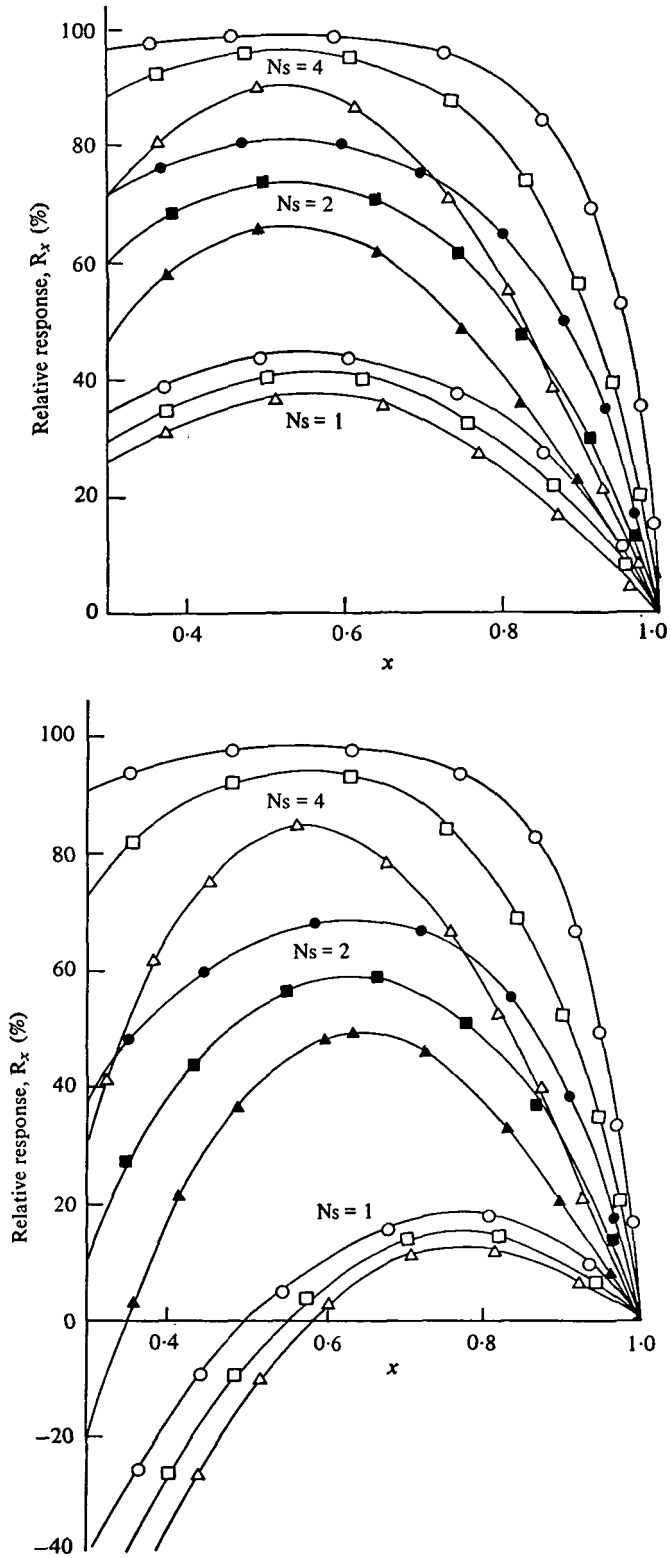

Fig. 4(*a*) and (*b*). For legend see opposite.

https://doi.org/10.1017/S0016672300019893 Published online by Cambridge University Press

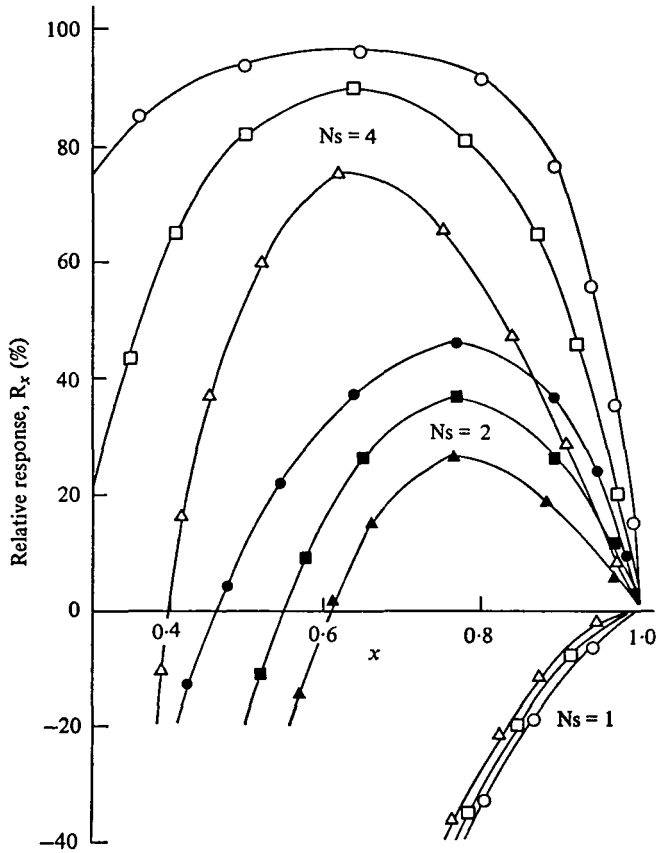

Fig. 4*a*. Relationship between relative response, proportion coming from better source population (P_1), Ns and different levels of dominance when the proportion of loci fixed favourably in P_1 is 0·55. ○ or ●, $h = 1$; □ or ■, $h = \frac{1}{2}$; △ or ▲, $h = 0$.

Fig. 4*b*. Relationship between relative response, proportion coming from better source population (P_1), Ns and different levels of dominance when the proportion of loci fixed favourably in P_1 is 0·75. ○ or ●, $h = 1$; □ or ■, $h = \frac{1}{2}$; △ or ▲, $h = 0$.

Fig. 4*c*. Relationship between relative response, proportion coming from better source population (P_1), Ns and different levels of dominance when the proportion of loci fixed favourably in P_1 is 0·90. ○ or ●, $h = 1$; □ or ■, $h = \frac{1}{2}$; △ or ▲, $h = 0$.

provided by the fact that some of the very best commercial maize hybrids are single crosses of homozygous lines. Moreover, insights concerning the more general problem of aggregating useful alleles from two or more non-homozygous populations should follow from study of the special case presented by a pair of pure lines.

Two prime problems to be dealt with by the breeder have been identified: (*a*) Choice of x, the proportion of genes in the foundation stock that are derived from the better homozygous source population and (*b*) the effective population size (N) that must be employed if Ns is to be large enough so that expected relative response will approach 100 per cent. These are problems because explicit information on s values and w is never available, and s will vary from gene to gene.

https://doi.org/10.1017/S0016672300019893 Published online by Cambridge University Press

The most reasonable *a priori* alternatives relative to x seemed to be random breeding populations initiated from the F_2 ($x = 0{\cdot}5$), the first backcross to P_1, the superior pure line ($x = 0{\cdot}75$) or the second backcross to P_1 ($x = 0{\cdot}875$). Based on expected relative response, which we have shown substantially reduced when x is increased from 0·75 to 0·875, we now drop the second backcross from further consideration. When the two pure lines are comparable in performance, the F_2 will always be the appropriate choice because (*a*) w, and therefore $\hat{x}$, should be close to 0·5 and (*b*) there would be very little performance level advantage in use of the first backcross as the starting point. When there is a substantial difference in performance between the pure lines, the backcross should usually be chosen because there would be a worthwhile performance level advantage and the change in relative response when $x = 0{\cdot}75$ instead of 0·5 is small if $Ns \geqslant 4$ and $h \geqslant \frac{1}{2}$. Fortunately, any case involving h near 0·0 at a majority of loci would be indicated by F_1 performance close to that of the poorer parent line.

The effective population size required if Ns is to be equal to or larger than any chosen value, say 4, obviously depends on the magnitude of s. Though s varies within limits depending on the selection system and intensity, it is basically a biological reality over which the breeder lacks control. Moreover, firm knowledge concerning the range of s values associated with the gene segregations that are significant relative to potential for genetic improvement is not available. However, some insights are available. Let $2u$ be the genetic effect on the selection criterion of substituting the favourable for the less favourable homozygous genotype at a single locus, σ_p^2 be the phenotypic variance of the selection criterion and k be the standardized selection differential. Then, following Falconer (1960),

$$s = 2ku/\sigma_p$$

when selection is among individuals on the basis of their own phenotypes and the gene in question segregates independently from others that affect the selection criterion. On the average, the effective value of s is decreased by genetic linkage (see Robertson, 1970; Ho, 1973). The appropriate approximation to s varies with the selection system employed (see Hill, 1970; Comstock, 1978). However, in every case k and $2u/\sigma_p$ are critical ingredients. Of these, $2u/\sigma_p$ is the more troublesome because it is ordinarily not measurable. Its evaluation when σ_p^2 is variance among individuals has been discussed by Falconer (1960, 1971) and Comstock (1974, 1978). Falconer (1971) inferred values in the 0·2–0·5 range for litter size in mice and Comstock (1974) concluded that for net worth, values in the 0·05–0·125 range are probable.

REFERENCES

Carr, R. N. & Nassar, R. F. (1970). Effects of selection and drift on the dynamics of finite populations. I. Ultimate probability of fixation of a favourable allele. *Biometrics* **26**, 41–49.

Comstock, R. E. (1974). Consequences of genetic linkage. *Proceedings, 1st World Congress on Genetics Applied to Animal Production*, pp. 353–364.

COMSTOCK, R. E. (1978). Quantitative genetics in maize breeding. *Proc. International Maize Symposium, Genetics and Breeding* (ed. David B. Walden), pp. 191–206. New York: Wiley-Interscience.

CROW, J. F. & KIMURA, M. (1970). An introduction to population genetics theory. New York: Harper and Row Publishers.

FALCONER, D. S. (1960). Introduction to quantitative genetics. New York: The Ronald Press Co.

FALCONER, D. S. (1971). Improvement of litter size in a strain of mice at a selection limit. *Genetical Research* **17**, 215–235.

HARLAND, J. R. (1975). Our vanishing genetic resources. *Science* **188**, 618–621.

HILL, W. G. & ROBERTSON, A. (1966). Effects of linkage on the limits to artificial selection. *Genetical Research* **8**, 269–294.

HO, YUI TIM (1973). Genetic improvement by combining superior alleles from different populations in a naturally cross-fertilizing species. Ph.D. thesis, University of Minnesota.

KIMURA, M. (1957). Some problems of stochastic processes in genetics. *Annals of Mathematical Statistics* **28**, 882–901.

LERNER, I. M. & DONALD, H. P. (1966). Modern developments in animal breeding. New York: Academic Press.

OSMAN, H. E. S. & ROBERTSON, A. (1968). The introduction of genetic material from inferior into superior strains. *Genetical Research* **12**, 221–236.

ROBERTS, R. C. (1967). The limits to artificial selection for body weight in the mouse. IV. Sources of new genetic variance – irradiation and outcrossing. *Genetical Research* **9**, 87–98.

ROBERTSON, A. (1960). A theory of limits in artificial selection. *Proceedings of the Royal Society* B **153**, 234–249.

ROBERTSON, A. (1970). A theory of limits in artificial selection with many linked loci. In *Mathematical Topics in Population Genetics* (ed. K. Kojima), pp. 246–288. Berlin and New York: Springer-Verlag.