

## Chapter 28

# *Recurrent Selection and Overdominance*

For many breeders, in considering problems that lie ahead and methods of meeting them, the main problem is whether to continue with varieties or breeds, or to work with inbred lines and  $F_1$  crosses. Behind this question are the problems dealing with the relative importance of general and specific combinability, or of prepotency and nicking:

Is the yield gain of hybrid corn due mainly to selection within and among inbred lines, or to selection among  $F_1$  crosses of inbred lines?

Is it due to improved frequencies of dominant favorable genes in elite inbred lines which are parents of elite-yield hybrids?

Is selection within and among inbred lines to accumulate higher frequencies of dominant favorable genes many times more powerful in one cycle without recurrence, than is selection without inbreeding through many recurring cycles?

To what extent may higher levels of specific combinability be reached by recurrent selection?

How may heritability of specific combinability be evaluated?

Why have the less favorable alleles of vigor genes been retained in such high frequencies?

May selection for general combinability and selection for specific combinability sometimes have counter effects on gene frequencies?

Does superiority of  $F_1$  crosses of inbred lines over varieties or breeds necessarily depend on overdominance?

If this choice of problems is approximately correct, the research emphasis may begin to shift from effects of inbreeding to effects of selection.

### EARLY EXPLANATIONS FOR HYBRID CORN

East and Emerson in an early paper considered the theoretical problem of recovering two traits together from a crossbreeding population in which the frequency of each trait was  $1/1000$ , and the two were independent. The authors offered two solutions: first to select at the rate of one per million in one generation, and second at the rate of one per thousand in two generations recurrently, first for one trait and then for the other. It is clear now that selec-

tion for both traits together each time, with normal distribution, would provide theoretical recovery in two generations at the rate of 1 per 400 or less. Multiplication of selection differentials in recurrent selection was sufficiently understood at the inception of hybrid corn. Nevertheless, hybrid corn has been developed with virtually no use or benefit from recurrence of selection. Hybrid corn is almost wholly an empirical development, but I think we may now consider applications of genetic science to improve the process.

*Recurrent selection* (Hull, 1945a) was meant to include reselection generation after generation, with interbreeding of selects to provide for genetic recombination. Thus, selection among isolates, inbred lines, or clones is not recurrent until selects are interbred and a new cycle of selection is initiated. Recurrent selection for specific combinability would seem to require a special breeding plan to provide heritability through successive cycles.

Shull's original plan for developing superior corn hybrids was designed for maximum immediate employment of specific combinability. Selection was mainly among specific  $F_1$  crosses of lines which had been isolated and stabilized by inbreeding, thus providing repeatability of crosses. This plan was consistent with a theory of heterozygosis of a degree here termed overdominance. Shull's plan did not involve recurrent selection to accumulate higher frequencies of favorable genes in successive cycles.

The apparent heterozygosis which Shull proposed to use was interpreted by Jones about ten years later as the expectation of repulsion phases of random linkages of dominant favorable factors and recessive less favorable alleles. This interpretation was particularly attractive because it seemed to eliminate any necessity of accepting overdominance. Overdominance is a contradiction of the time-honored principle that purity of blood is to be sought and maintained. Vigor was no exception to the old principles of *like begets like* and *breed the best to the best*. Moreover, the postulated linkage relations would appear to be inevitable where many loci are involved.

In the decade following appearance of the Jones hypothesis, most corn breeders began more intense selection for vigor within and among lines during the inbreeding process, and selection among lines for general combinability. Most of the very considerable success of hybrid corn came quickly after these modifications of Shull's method were adopted. Selection within and among inbred lines to improve frequencies of dominant favorable factors became the guiding principle for developing superior hybrids of corn, other crops, and of livestock. Selection among specific  $F_1$  crosses was retained as a final step, but with very little verbal emphasis.

Initial successes with hybrid corn (which so far have not been greatly surpassed) were obtained with inbred lines which were, for the most part, isolated directly from the open-pollinated varieties. Corn breeders then had at least two alternatives for further work. Empirically, the choice might well have been to continue isolation and testing of additional new lines from the

same sources, abandoning recurrent selection entirely. Usually, successful but mysterious processes are not modified on theoretical grounds alone. However, most of us, and myself most of all I suspect, chose the alternative course without question. New lines for a second cycle of selection were isolated from crosses of elite first-cycle lines. Since it was soon apparent that second-cycle lines as a group were a vast improvement over first-cycle lines, it was clear that we were on the right track. Recurrent selection for higher frequencies of dominant favorable genes was fulfilling expectation admirably. That it had failed in ear-to-row selection (progeny testing without inbreeding) meant that "selection within and among inbred lines" was the key. Apparently the protagonists of "early testing" have not fully appreciated this latter point.

### DISAPPOINTMENT WITH SECOND-CYCLE HYBRIDS

My first suspicion that all was not well was aroused by disappointing yield performance of second-cycle hybrids in 1941. The first reaction then was to conclude that heterosis might involve complex gene interactions to a greater extent than I had supposed. Cytoplasmic-nuclear interactions could not be ruled out entirely. But no thought of heterozygosis, of overdominance, was entertained at all, so thoroughly had I been weaned from it.

In 1942 we began the process of separating Florida inbred lines into two permanently distinct groups on the basis of combining values with two single cross testers which were thought to make a good double cross. Subsequent breeding operations after the initial separation were to consist of isolating new lines within each group from crosses of the older lines within the group. New lines were to be stabilized by at least three self-pollinations with accompanying selection for vigor and type, and then tested for combinability with the reciprocal group. This, of course, was reciprocal recurrent selection without *early testing*. I still adhered firmly to the efficacy of "selection within and among inbred lines."

Segregation of the breeding mass into two permanently distinct reciprocal groups, first of all, did not cost anything. A search for satisfactory substitutes for each of the four master tester lines was well in order. It seemed that the necessity of recovering specific combinability again as the last step of each breeding cycle might be avoided to some extent. Possibly higher levels of specific combinability might be accumulated.

Two years later, after interviews with a number of other corn breeders, it seemed that a still higher rating might be in order for specific combinability. Second and third-cycle hybrids were not much superior to first-cycle hybrids in yield of grain. Recurrent selection for general combinability was not proving to be very effective.

An early test of recurrent selection for specific combinability seemed desirable. One way to intensify the process already in operation was to adopt a more specific tester. This was done by abandoning the reciprocal feature of

the plan—by reducing one of the reciprocal groups to the single-cross tester alone. That tester is to be continued indefinitely. Another way to intensify the operation was to increase the frequency of recurrence of selection. This was done by adopting the general principle of early testing, by abandoning the inbreeding interphase of each cycle, by testing  $S_0$  plants rather than  $S_3$  lines or higher. Inbred lines, including the tester lines, of the second reciprocal group were intercrossed to provide one crossbred group of  $S_0$  plants. Repeated selection within this crossbred group for combinability with the permanent unrelated tester is the proposed plan. It is only for practical reasons that one homozygous line is not employed as the tester for field corn. With sweet corn a line tester might well be used.

The working definition of *specific combinability* employed in designing the foregoing breeding plans was about thus: that part of the genetic superiority of specific  $F_1$  crosses of homozygous lines which is not transmitted into or through general recombinations. The concurrent definition of *general combinability* then is: that part which is transmitted into and through general recombinations. That these definitions are perhaps inadequate for analyses of variance does not necessarily mean that they are not admirable for the other purpose.

Shull, East, and others who isolated inbred lines and crossed them discovered that inbreeding did little or no irreparable harm to the germ plasm. Gametes of inbred lines hardly differ basically from gametes of crossbred varieties. The inbreeding effect is very nearly or entirely a zygotic phenomenon. Vigor genes in both homozygous and heterozygous associations were obeying Mendel's first law of non-contamination. All of this was an important discovery.

Shull in addition invented *selection for specific combinability*, which was certainly something new under the sun; yet to be generally recognized as one of the great inventions. Shull was led, I suspect, to this invention by the empirical evidence before him, not by considering the more abstract concept of heterozygosis. Shull must have recognized very soon that reconciliation of his invention with his knowledge of genetics required heterozygosis, and perhaps the more inclusive *heterosis*.

#### RECURRENT SELECTION FOR SPECIFIC COMBINABILITY

A little more than thirty years later the inevitable invention of *recurrent selection for specific combinability* was made from matter-of-fact empirical considerations as outlined above. Again it seemed necessary soon afterwards to embrace some theory of heterozygosis for reconciliation with genetics. The breeding plan was presented (Hull, 1945a) with confusing emphasis on the abstract concept of overdominance, I fear, and too little emphasis on the actual motivation.

May it be said now that the first proposal was to determine with direct

tests if higher levels of specific combinability could be accumulated by recurrent selection. There is no need to await incontrovertible evidence of overdominance; indeed even if it were in hand the direct test would still be needed.

The second proposal was that if recurrent selection for specific combinability should be important, selection within and among inbred lines had been greatly over-emphasized. The inbreeding interphase could be abandoned. This would provide an enormous saving in time and otherwise, particularly with poultry and other livestock. Curiously, some reviewers have described the proposed breeding plan as a "laborious method."

Grain yield of corn depends appreciably on resistance to new and sporadic diseases, insects, and adverse environmental complexes. Here it would seem that overdominance is not likely, but that selection within and among inbred lines is yet of real value. Significant resistance where it exists will eventually be identified in continuing a stable line. Selected crosses will be generally superior insofar as the several resistances are dominant and matched combinations are found. Here again I am not certain that rapidly recurring progeny tests without inbreeding may not be equally or more effective in the main. One resistant line among some hundreds of susceptible ones in an epidemic provides a striking field illustration—perhaps a deceptive one.

Breeding plans to accumulate specific combinability may be designed in many ways, the better ones to be determined by actual tests. Testers might best be the male parent of the hybrid in some cases, or the female parent in others. The inbreeding interphase may be omitted or included in any practicable degree. It has been thought that the problem of the preceding paragraph might be met well enough by direct selection in the crossbred lot and selection among  $S_0$  testcrosses. But in some cases there might be an advantage with  $S_1$  or  $S_2$  testcrosses. With  $S_0$  or  $S_2$  some of the selection may be for general combinability, for higher frequencies of genes which are favorable in any combination.

The early view (Hull, 1945a, Proposition 7) was that where  $aA$  is generally intermediate to  $aa$  and  $AA$ ,  $A$  should be in high frequency, in improved varieties. Not much further opportunity for improving combinability would remain.

Crow's viewpoint, as he has presented it here, seems to be that without overdominance long continued selection in any form would have carried favorable alleles to high frequency in equilibrium with reverse mutation, where heterozygosity is infrequent and heterosis not large.

If recent shifts of environment or of emphasis in artificial selection should have provided important loci with intermediate gene frequencies, Crow's argument may not be germane. Here I may venture an opinion (Hull, 1945b) that without overdominance rapidly recurring mass or ear-row selection should continue to surpass contemporary selected  $F_1$  crosses of homozy-

gous lines. Or we may consider the more efficient technic of recurrent testing of controlled testcrosses of  $S_0$  plants with the parent variety and recombining the better ones into an improved variety. We know this will not work, although it has not been fairly tried. Finally, in modern corn breeding the same technic with  $S_4$  and higher lines has been extensively advanced through at least two cycles. Most corn breeders will admit that a general recombination or synthetic blend of parent lines of present elite-yield hybrids would hardly yield more than a random blend of parent varieties of today or of 50 years ago.

A few recombinations of lines selected wholly for general combinability have been reported with significantly higher yields than improved varieties. This result I will attempt to show later is a different matter, fully consistent with overdominance theory.

It seems likely that improvement of general combinability, accumulation of dominant favorable genes with respect to grain yield, in the field corn of our central Corn Belt in the past fifty years has been hardly significant except for that depending on disease resistance, resistance to lodging, to ear dropping, etc. Almost any one of the common breeding technics is quite effective with general improvement of morphology of the corn plant, or with oil and protein of the grain. Genetics of vigor would appear to differ in some important respect from genetics of the other traits.

Overdominance has seemed the more likely, but I have never meant to insist that the existence of every other alternative had been disproven. Refractory repulsion linkage has seemed insufficient alone to explain the apparent degree of overdominance in corn (Hull, 1945a).

The main point now is accumulation of general combinability with recurrent selection. It is axiomatic with most of us, including the corn breeders, that general combinability is the first consideration, despite the evidence cited here. This kind of evidence has been largely ignored and almost taboo.

Comstock *et al.* (1949) have proposed Reciprocal Recurrent Selection to obtain maximum utilization of general and specific combinability together. In this they have accepted that specific combinability might be accumulated in successive cycles, and that the inbreeding interphase could be abandoned entirely. This variation of the general plan was compared on theoretical considerations with selection in a crossbred for combinability with a homozygous tester. Now, since a homozygous tester is clearly impracticable in many cases and heterozygosity would impair efficiency of a tester except for reciprocal selection, there is an advantage in the reciprocal plan which the authors did not record.

It has never been my intent, however, to attempt to rule out judicious reciprocal selection. We have crossed each of the two tester lines of corn to a goodly number of unrelated strains, and have backcrossed in bulk to each tester line separately. The two lots are being held in separate crossbreeding

reserves with nominal selection for agronomic type. If either tester line should develop a serious fault, or if the present main selection for specific combinability should seem to reach a ceiling, reversal of selection would seem almost inevitable. A tester would be chosen from the current crossbred and the two bulk backcrosses would furnish a reciprocal crossbred to reverse the process, temporarily at least.

An accessory operation with bulk backcrosses is hardly practicable with livestock. But here the tester would be one inbred line which would need to be 50 per cent inbred for equal efficiency with the single cross of homozygous lines employed as the corn tester. The tester should be the male parent of the improved hybrid in livestock to avoid any impairment of the female function by inbreeding.

Beginning with a partly inbred or non-inbred stud flock or herd as the tester, and continuing mild inbreeding, it is inevitable that choices among young males for herd sires of the stud herd would depend partly on their testcross progeny. Sufficient vigor must be retained in this herd to provide satisfactory sires of commercial hybrids. The problem is real and obvious enough, but I have thought the details must await a demonstration that specific combinability can be accumulated in important amounts by recurrent selection. For an early test the more homozygous tester is probably to be preferred. If uniformity of the product is of some moment, the operator of reciprocal selection may expend considerable effort for it. Such expenditure might be avoided by partial inbreeding of one of the groups.

The two breeding plans, selection in a crossbred to a homozygous tester and reciprocal selection between two crossbreds, are the extremes of recurrent selection for specific combinability. Between these we may have any practicable degree of inbreeding of one of the groups at the start, or subsequently. Inbreeding restricts reciprocal selection but, aside from that, the reciprocal feature may be varied at will. I do not know what factors may determine the more efficient plans except that general combinability with respect to vigor is probably not an important one. Nor is it likely to be important to choose an inbred tester with above-average general combinability.

#### PHYSIOLOGICAL NATURE OF OVERDOMINANCE

Overdominance has been defined (Hull, 1946a) as  $aA > AA$ , which is a sufficient definition for present purposes. However, there may be some value in considering what the underlying physiology may be. Heterozygosis as considered by Shull and his early contemporaries is entirely or very nearly the same concept. Fisher (1918, 1932) has discussed this concept more generally as *super-dominance*. Some recent writers have employed *heterotic alleles* or *heterotic interaction of alleles* as a modern form of *heterozygosis*. But since any degree of dominance of the more favored allele is essentially a heterotic interaction, heterotic alleles does not necessarily imply  $aA > AA$ .

In the current sense that any interaction of alleles is dominance,  $aA > AA$  is overcomplete dominance, overdominance. In a similar sense all interactions of non-alleles are epistasis. Dominance and epistasis differ in distribution on chromosomes, but not necessarily in underlying physiology so far as I can see. Overepistasis would excite no particular comment.

Dominance and epistasis are no more fundamental properties of genes than is interaction a property of a unit of nitrogen or phosphorus. These fertilizer elements may exhibit an interaction in plant growth if made available to a living plant, or they may seem to act independently. One quantity of nitrogen may be adequate for the needs of the plant. Adding the same quantity again may produce no further effect. There is an interference or decreasing returns interaction.

East (1936) has discussed dominance as a decreasing returns or interference interaction of active alleles  $A_1$  and  $A_1$  in the homozygote. The amount by which the two together failed to do twice as much as either alone was  $\alpha$  the dominance effect—a loss which could not explain heterosis. East then proposed that if  $A_1$  should develop by successive steps to  $A_4$  (analogous to replacing successive parts of one bag of nitrogen above with phosphorus until there is one of phosphorus and one of nitrogen) of a different quality,  $A_1$  and  $A_4$  might interfere very little or none in  $A_1A_4$ . The principle as East states it is: "*The cumulative action of the non-defective allelomorphs of a given gene approaches the strictly additive as they diverge from each other in function.*"

The effect of the phosphorus and nitrogen together is the sum of their separate effects—no interference. Dominance by interference disappears when  $A_1$  and  $A_4$  are independent in functions, leaving  $A_1A_4$  superior to either  $A_1A_1$  or  $A_4A_4$ . Now it must be clear that any deviation of  $A_1A_4$  from the mid-point between the two homozygotes must be interpreted as dominance of  $A_1$  to  $A_4$  for the  $A_1$  function or dominance of  $A_4$  to  $A_1$  for the  $A_4$  function or both. If the primary dominance in each case is complete,  $A_1A_4$  will just equal the sum of  $A_1A_1$  and  $A_4A_4$  in total effect beyond a neutral  $aa$ .

Overdominance may occur when: (1)  $aa$  is neutral and  $aA$  is nearer to an optimum dose of  $A$  than is  $AA$ , (2)  $A'$  and  $A$  are both active for separate supplementary functions and each is dominant to the other for its own function (cf. East, 1936), (3)  $A'$  and  $A$  are both active for separate primary functions, and the primary functions interact to produce an effect greater than those of either  $A'A'$  or  $AA$  (Hull, 1945a).

Pseudo-overdominance may occur when  $A$  and  $B$  are linked: (1) with no epistasis,  $aB$  and  $Ab$  combinations simulate the second case above, (2) with positive epistasis  $aB$  and  $Ab$  simulate the third case.

If  $(aB \times Ab)$  is superior to both  $(ab \times AB)$  and  $(AB \times AB)$ , selection may tend to tighten the repulsion linkage until  $ab$  and  $AB$  disappear and the paired blocks are hardly distinguishable from alleles with primary overdominance.

It is clear enough that the frequency of heterozygotes is greater and of



homozygotes less for any locus with multiple alleles present in a crossbreeding population. If heterozygosity should be of general advantage, multiple alleles would provide more heterosis. East was at some pains to explain the development of  $A_4$  from  $A_1$  by successive steps to the end of a superior heterozygote. He apparently did not accept that heterozygote superiority might be general, with multiple alleles affecting vigor. I do not accept it either as a likely proposition.

It seems likely that production of grain, meat, eggs, or milk may consist of main effects and interactions of many, perhaps most, of the genes of the plant or animal. Main effects must be of many kinds and magnitudes. Where inbreeding depression and heterosis are evident there must be bias of positive dominance or interactions of alleles to provide a gain in heterozygotes over the arithmetic mean of homozygotes. Whether the interaction is basically a stimulation of unlike alleles in the heterozygote, an interfering depressing interaction in the top homozygote, or some other kind of interaction is an important problem in gene physiology. Present concern, however, is only with the magnitude and frequency of the effect without regard to its basic physiology.

Various writers have noted that dominance is not an absolute property. If the phenotype is fruit size, degree of dominance is hardly the same for both diameter and volume. The same genes might also affect stem length and exhibit a third degree of dominance there. Gene effects are often greatly subject to environmental fluctuations and to presence or absence of genes at other loci.

Within reasonable limits of soil fertility and climate, grain yield of selected homozygous corn is about 30 per cent of the yield of crossbred corn. Seventy per cent of the apparent yield of crossbred corn consists of dominance effects and perhaps of interactions of dominance with other gene effects. The 30 per cent yield of homozygous corn consists of main effects and epistatic interactions of main effects.

One difficulty in resolving the present situation without regard to how it may have evolved is that the absolute zero of the genetic yield range cannot be easily estimated. However, it might be assumed that it is less than zero on our data scale. More specifically, the homozygotes with more than two-thirds of the concerned loci  $aa$  or less than one-third  $AA$  may be inviable or have an average yield potential of zero. The 100 per cent of measured yield then would represent only the upper two-thirds of the total genetic range. With dominance of high yield complete at each locus and the foregoing assumption the present situation is adequately explained without resort to epistasis or overdominance.

#### LINEARITY OF INBREEDING DEPRESSION AND HETEROZYGOSITY

Any appreciable degree of interaction of dominance with other gene effects might be detectable in a non-linear relation of inbreeding depression to

predicted frequency of heterozygosity in succeeding generations of inbreeding. Since the considerable body of data on inbreeding effects on yield of corn fails to show any such non-linearity at all, I have been inclined to dismiss interaction of dominance with other gene effects. Since, in addition, backcrosses of  $F_1$ 's to homozygous parent lines fail to show significant non-linearity I have been inclined to dismiss epistasis in general as an appreciable part of the explanation of the disparity of yields of homozygous and cross-bred corn.

Overdominance alone is an adequate explanation of the disparity. Pseudo-overdominance from random linkage is not an adequate explanation by itself since the totals of gene effects are independent of linkage (Hull, 1945a).

### REGRESSION OF $F_1$ YIELD ON YIELDS OF PARENT LINES

Corn breeders have frequently chosen a small sample (usually 10) of inbred lines and have made all or most of the specific crosses. Comparable yield records on parent lines and  $F_1$ 's have become available now in 25 sets of data.  $F_2$  records are included with 3 of them. None of these data are mine. Some of them were analyzed in part by simple regression of yield of  $F_1$  on yield of parents, which would seem to provide the significant information from the general combinability viewpoint. Interaction of parents is mostly neglected.

Within each column or each row of a  $(10 \times 10)$  table as described are nine  $F_1$ 's or nine  $F_2$ 's with one common parent. The common parent is the tester of the other nine lines. Each line serves as the tester of one such group.

On the assumption that the partial regression of offspring on parent within a group having one common parent is a relative measure of heritability within the group, or of efficiency of the common parent as a tester, it has seemed worth while to calculate all of the regression coefficients for individual columns of the twenty-five  $F_1$  and three  $F_2$  tables. We tacitly accept that yield may be a heritable character. Beyond this we need no fine-spun theory nor any genetic theory at all to warrant direct regression analysis of the data. However, Mendel's final test of his theory was with backcrosses to  $aa$  and  $AA$  separately. He noted essentially that with completely dominant characters the expected regression of offspring phenotype on gene frequency of parent gamete was unity with the  $aa$  tester and zero with the  $AA$  tester. We may be dealing with multiple factor cases of such testcrosses and of course with different degrees of dominance at the several loci. The significant differentiation of our homozygous testers may be in relative frequencies of  $aa$  and  $AA$  at the  $a_1, a_2, a_3 \dots a_n$  loci.

Results with the first two examples are shown in Table 28.1. Yield of the tester parent ( $P$ ) is in bushels per acre. Directly below are the partial regression coefficients ( $b_p$ ) for the respective testers. Since there are apparently negative trends of  $b_p$  with respect to  $P$ , the second order regression ( $b_2$ ) of

$b_p$  on  $P$  has been calculated. The second order regression function has been solved for the special case  $b_p = 0$ , to obtain an estimate of  $P_c$  the critical value of  $P$  where the regression surface is level and heritability is zero.

The third summary in Table 28.1 is for average yields in six states of the

TABLE 28.1

REGRESSION OF YIELD OF  $F_1$  AND  $F_2$  CORN HYBRIDS ON YIELD OF  
INBRED PARENTS WITHIN GROUPS HAVING  
ONE COMMON PARENT

Yield of parents ( $P$ ) is recorded in bushels per acre, with the partial regression coefficient ( $b_p$ ) below each one for the group of which it is the tester. The second order regression  $b_2$  is regression of  $b_p$  on  $P$ . Critical  $\bar{P}$  ( $P_c$ ) is estimated value of  $P$  for  $b_p = 0$ .

*Stringfield, G. H. Unpublished. Ohio Agr. Exp. Sta. and USDA\**

$P$	14	28	30	46	51	55
$b_p(F_1)$	.68	.41	.31	.22	.07	.05
$b_p(F_2)$	.55	.45	.33	.24	.26	.17
Mean $b_p(F_1)$ .29, ( $F_2$ ) .33; $b_2(F_1)$ -.014, ( $F_2$ ) -.008; Mean $P$ 37; $P_c$ 58; Mean $F_1$ 97; Mean $F_2$ 70.						

*Kinman and Sprague, Agron. Jour. 1945\**

$P$	3	15	20	26	28	32	39	40	50
$b_p(F_1)$	.63	.75	.84	.69	.13	.30	.25	.39	.22
$b_p(F_2)$	.26	.36	.42	.69	.24	.29	.37	.58	.54
Mean $b_p(F_1)$ .42, ( $F_2$ ) .42; $b_2(F_1)$ -.016, ( $F_2$ ) +.005; Mean $P$ 29; $P_c$ 54; Mean $F_1$ 80; Mean $F_2$ 51.									

*USDA and State Regional Tests, Midseason 1943; Iowa, Kans., Ill., Ind., Ohio, Penn.  $P$  values from Kinman and Sprague above; their  $F_1$ 's included here\**

$b_p(F_1)$	-.05	+.11	+.08	-.13	-.20	-.11	+.12	-.01	-.18
Mean $b_p$ -.01; $b_2$ -.004; $P_c$ 25.									

\* Sources of data.

same  $F_1$ 's as those of Kinman and Sprague in Iowa. The Iowa test included parent lines and  $F_2$ 's as well as  $F_1$ 's. The third summary has been made with Iowa records on parent lines. An analysis was made also of the  $F_1$  records for each state separately with the same values of  $P$ . Regression trend was positive for the Indiana data, thus failing to support any theory of dominance of high yield. Regression trends for the other four states were negative with estimates of  $P_c$  all lower than the one for Iowa.

The eighteen other sets of data not summarized in the table are from Minnesota, Iowa, Illinois, Ohio, New York, and North Carolina. They are believed to be generally independent genetically and ecologically. Regression trends are positive in eight cases. Taking the five cases summarized together in Table 28.1, as five separate ones, we have seventeen with negative regression trend to eight with positive. Estimates of  $P_c$  for the seventeen negative trends are near to or within the range of data as in Table 28.1 for each case but one. With one of the least extensive tests the estimate of  $P_c$  is roughly 12 times the top inbred line, thus agreeing nicely with incomplete dominance.

Insofar as regression trends are due to heterozygosity they may be expect-

ed to disappear with inbreeding of the crosses. The first two examples in Table 28.1 are the two more extensive of the three cases which include  $F_2$ , and it is apparent that the negative trend of  $F_1$  has decreased or become positive in  $F_2$ . It is positive in  $F_2$  of the third case also with a strong negative trend in  $F_1$ .

The regularity of regression trends apparent in the first two examples in Table 28.1 is by no means so readily apparent in any of the other twenty-three cases. The eight cases with positive trends do not appear worse in this respect than the others.

The possibility that the 10 inbred lines of Kinman and Sprague do not comprise a representative sample has been tested by dividing the 10 into two groups of 5 each in various ways. This provides a  $5 \times 5$  table in each case with a unique sample of 25  $F_1$ 's from the total of 45. These  $5 \times 5$  tables do not have vacant cells which arise when one parent line is included on both margins of a table. Each tester in one group is rated with the same five lines in the other group. Estimates of  $b_2$  and  $P_c$  from such  $5 \times 5$  tables have consistently substantiated those reported in Table 28.1, for the  $10 \times 10$  table.

Analyses of six of the twenty-five cases have been done also with logarithms of  $P$  and  $F_1$  records, with results generally in agreement with those of the original data.

Most or all of the individual values of  $b_p$  and  $b_2$  are not statistically significant. The distribution of the twenty-five  $b_2$ 's is distinctly bi-modal. Eight are positive indicating dominance of low yield, one is negative and small enough to indicate intermediate dominance of high yield. Sixteen are negative and decidedly in the overdominance range. No explanation of the bi-modality is apparent now. The eight positive values of  $b_2$  are in some degree suspect since they are inconsistent with so many facts. All of these tests could be repeated with the same unique samples of genotypes insofar as the parent lines were homozygous and are still available. We need more comprehensive and precise data.

Present evidence from regression analysis is slightly in favor (2 to 1) of the conclusion that a zone of nearly level regression, nearly zero heritability, exists near the upper end of the range of present data. This conclusion would be more consistent with the failure of selection for general combinability if it should be that selection for specific combinability should favor  $aA$  over  $AA$ , and thus tend to degrade gene frequencies below that equilibrium where heritability and regression change from positive to negative.

#### GENETIC INTERPRETATION OF THE REGRESSIONS

The problem of genetic interpretations of  $b_p$  and  $b_2$  may be approached first with the simpler case of no epistasis. Consider the multiple gene set  $a_1A_1$  to  $a_nA_n$ . Let  $(1 - v)$  and  $v$  be relative frequencies of  $a$  and  $A$  in the gametes of  $P_i$  with respect to the  $n$  loci, and  $w$  similarly for  $P_j$ . The product

of the two gametic arrays provides expected frequencies of  $aa$ ,  $aA$ ,  $AA$  in  $F_1(P_i \times P_j)$  with respect to  $n$  loci.

$$F_1 \cong n(1-v)(1-w)aa + n[v(1-w) + w(1-v)]aA + nvwAA$$

Define<sup>1</sup> phenotypes:

$$\begin{aligned} a_1a_1 a_2a_2 \dots a_na_n &= T \\ aA &= T + d + kd \\ AA &= T + 2d \end{aligned}$$

$$F_1 = n(v+w)(d+kd) - n2vw(kd) + T \quad (1)$$

This is the regression of  $F_1$  phenotype on gene frequencies of parents and is independent of degree of inbreeding of parents.

If  $P_i$  is homozygous it has  $n(1-v)aa$  and  $nvwAA$  loci.

$$\begin{aligned} P_i &= nv(2d) + T, & v &= (P_i - T) / n2d \\ P_j &= nw(2d) + T, & w &= (P_j - T) / n2d \end{aligned} \quad (2)$$

Substituting for  $v$  and  $w$  in (1)

$$F_1 = (1+k+KT/nd)(P_i+P_j)/2 - (k/n2d)(P_iP_j) - kT^2/n2d - kT \quad (3)$$

This is the regression of  $F_1$  phenotype on phenotypes of homozygous parents, the equation of a surface in three dimensions,  $F_1$ ,  $P_i$ ,  $P_j$ . The surface is a plane if  $b_2 = k/n2d$  is zero, if  $k = 0$ , if there is no dominance, no interaction of  $P_i$  with  $P_j$ . Then,  $F_1 = (P_i + P_j)/2$ , by setting  $k = 0$  in (3).

Taking  $P_i$  constant as the common tester of one column of the regression table,

$$F_1 = [\frac{1}{2}(1+k) - k(P_i - T)/n2d]P_j + C' \quad (4)$$

$b_p$  is the coefficient of  $P_j$ , within brackets,

$$b_p = (-k/n2d)P_i + \frac{1}{2}(1+k) + kT/n2d. \quad \text{If } k = 0, \quad b_p = \frac{1}{2}$$

Regression of  $b_p$  on  $P_i$  is  $b_2 = -k/n2d$ . Since  $P_i = nv(2d) + T$ ,  $bp = \frac{1}{2}(1+k) - kv$ . If  $bp = 0$ ,

$$v = (1+k)/2k \quad (5)$$

With no dominance,

$$k = 0, \quad v = 1/0 \text{ at equilibrium}$$

1.  $T$ ,  $d$ , and  $kd$  are defined here in units of bushels/acre or pounds/plot, for example. Then,  $k = kd/d$  is in units of (bu./A)/(bu./A), likewise  $b_p$ , but  $b_2$  is in units of 1/(bu./A), making the whole term  $b_2P_iP_j$  in bu./A.

In terms of selective values it is convenient to define  $d_s$  in terms of number of progeny surviving to breed. Then,  $k_s$  may be greater or less than  $k$ , depending on artificial breeding plan. If roan in Shorthorn cattle is intermediate,  $k$  is essentially zero, but if roan is favored in artificial selection over red and over white,  $k_s > 1$  and there is overdominance with respect to artificial selective values.

With corn yield no single locus is identified, no heterozygote may be favored to provide  $k_s > 1$ , except that  $k > 1$ . Then,  $k_s$  may depend on gene frequency and on rate of culling.

## Complete dominance

$$k = 1, \quad v = 1 \quad \text{at equilibrium}$$

$$k = 2, \quad v = 3/4 \quad \text{at equilibrium}$$

For the more general case where  $P_i$  and  $P_j$  are (not inbred) individuals in a crossbreeding population, equations paralleling (2), (3), and (4) are second, fourth, and second degree, respectively. The simplification obtained with homozygous parents is reduction of the three functions to first, second, and first degree, respectively, by removing dominance effects (allelic interactions) from parent phenotypes  $P_i$  and  $P_j$ . Mendel found the simplification obtained with homozygous parents to be of considerable value in his early studies of monogenic inheritance.

The Mendelian model (2), (3), (4) may be complicated with innumerable kinds of interactions (epistasis) by simple, compound or complex transformations (log, anti-log, exponential, etc.) of (2), (3), and (4). It is not intended to imply, however, that interactions of alleles must precede interactions of non-alleles in living organisms.

The estimate of  $b_p$  for any tester parent line is independent of gene frequencies of the other parent lines with respect to dominance interactions. If obtained estimates of  $b_p$  for the same tester with samples of weak and strong lines respectively should differ significantly, the necessary interpretation would seem to require some kind of interaction other than between alleles, or that the lines were not strictly homozygous.

Interpretation of  $b_p$ ,  $b_2$ , etc., by the Mendelian model presented here will not be biased by linkage of two loci if frequencies of  $ab$ ,  $aB$ ,  $Ab$ , and  $AB$  do not deviate significantly from expectation from random recombination of gene frequencies of the two loci with respect to all of the parent lines. Free assortment of the two loci is then effectively simulated. But any union of two unlike gametes must contain some cases of repulsion linkage close enough to retain the  $aB$  and  $Ab$  combinations in high frequency through several generations. A sample of lines all derived directly by selfing from one heterozygous parent plant may well contain many cases of repulsion linkage to simulate overdominance. This effect would not be counterbalanced by high frequencies of coupling linkage of other pairs of loci. Lines within each of the 25 samples reported here are in most or all cases no more closely related than plants within one or more varieties.

Variations of  $d$  and  $k$  from locus to locus would contribute to total variance, but would not seem to impair seriously the validity of the estimates of regression coefficients, nor of  $P_e$  when  $b_p = 0$ .

When all loci are  $aa$  or all loci are  $AA$ ,  $P_i = P_j = F_1 = F_2 = x$ . With this restriction (3) becomes a quadratic with roots equal to the phenotypes at the two limits. The difference is  $n(2d)$ , the genetic range, the denominator of  $b_2 = -k/n2d$ . Values of  $k$ , calculated thus, for the nine cases where parent

and  $F_1$  yields are strictly comparable and  $b_2$  is negative, are: 2.25, 1.50, 1.88 (2.18), 1.83, 1.78, 2.45, 1.41, 2.25, 1.69. The 1.88 (2.18) entry is  $F_1$  and  $F_2$  respectively of Stringfield's example, Table 1. The value 1.09 from  $F_2$  data was doubled to correct for the effect of inbreeding.

If these independent estimates of  $k$  should be unbiased operationally, we must still be cautious in attempting any unique physiological interpretation. All of the several types of apparent overdominance listed here and others too may be operating in corn yield.

Estimates of backcrosses  $B_i$  and  $B_j$  may be written by inspection of (1) and (2).  $F_1$  is transformed to  $F_2$  (by selfing  $F_1$ ) by multiplying the coefficient of each  $k$  term in (1) by  $\frac{1}{2}$ . This provides three linear sets  $F_1$ ,  $F_2$ , bar  $P$ ;  $F_1$ ,  $B_i$ ,  $P_i$ ; and  $F_1$ ,  $B_j$ ,  $P_j$ , on the assumption of no epistasis.  $F_1$ ,  $F_2$ , and  $\bar{P}$  are alike in gene frequency. They differ only in frequency of heterozygosity. Differences in the backcross comparisons arise from both gene frequency and frequency of heterozygosity.

### GRAPHIC TRANSFORMATIONS TO REMOVE EPISTASIS

Where the two intervals in any one of the three comparisons are not equal, epistasis may be suspected and a transformation of data may help to eliminate some of its effects. No transformation of the corn yield data would be warranted by all of the considerable amount of published data I have found, since the data fit the linear hypothesis very closely with  $F_2$  and backcross comparisons.

Where transformation is clearly indicated, I may suggest a graphical determination of the best function. Plot the data,  $P_i$ ,  $P_j$ ,  $\bar{P}$ ,  $F_1$ ,  $F_2$ ,  $B_i$ ,  $B_j$ , and  $\bar{B}$  on the vertical axis, and the same values on the horizontal axis linearly with no dominance, with any arbitrary scale. If the plotted points do not seem to provide a smooth curve, move  $F_1$  to the right a trial distance. Move  $F_2$ ,  $B_i$ ,  $B_j$ , and  $\bar{B}$  the same direction one-half as far. Move to the right or left ( $F_1$  twice as far as the others each move) until the best fit visually to a smooth curve is found as the best transformation function. The only excuse for suggesting such a crude process is that if it is carefully carried out with good data the function is so much more refined than any arbitrarily chosen function for the purpose of correcting a complex of different kinds of epistasis together.

The transforming function determined by the above process with all available data on grain yield of corn would not differ sensibly from a straight line. From this I have said earlier that epistasis is unimportant in corn yield. Considerable amounts of increasing and decreasing returns types of epistasis may be effectively balanced, of course. In that case, epistasis would provide no explanation of the disparity of inbred and crossbred yields.

### MAXIMUM YIELDS FROM CROSSING HIGH BY LOW?

For four loci with  $v$  and  $w = \frac{3}{4}$ , the gametes are  $a_1A_2A_3A_4$ ,  $A_1a_2A_3A_4$ ,  $A_1A_2a_3A_4$ ,  $A_1A_2A_3a_4$ . Equations (1) or (3) with appropriate substitutions

calculate the mean of the 16  $F_1$  combinations of four gametes of equal gene frequency. Deviations of the individual  $F_1$ 's from the mean are not predictable from parent phenotypes. They are due to specific combinability arising from varying frequencies of heterozygosity. No more than two loci can be heterozygous in this example. But if  $v = \frac{3}{4}$ ,  $w = \frac{1}{4}$ , six of the sixteen  $F_1$ 's are heterozygous at all four loci. In the event of overdominance  $\frac{6}{16}$  of high  $\times$  low combinations may exceed the best high  $\times$  high combination. If  $1 < k < 2$ , and  $v = \frac{3}{4}$ , the mean of high  $\times$  high is greater than the mean of high  $\times$  low. From the general combinability viewpoint we see only the difference of means. Selection of the very few elites among specific  $F_1$ 's would, however, find them more frequently in high  $\times$  low combinations. Hayes and Immer (1942, Table 21) present data of Johnson and Hayes which seem to agree with this interpretation in that the mean of high  $\times$  high is best, but the highest specific combination is more likely in high  $\times$  low.

### EQUILIBRIUM FREQUENCIES OF GENES

We may substitute for  $v$  in equation (1) the mean gene frequency of a group of lines or of a variety, a general tester, to be held constant. Then if  $v$  is less than  $(1 + k)/2k$ , and  $k > 1$ , regression of  $F_1$  on  $w$  is positive. Selection for general combinability with the same tester should continue to fixation of  $A$  except for reverse mutation. But if selected lines are recombined for each cycle and the recombination is the tester for the next cycle, selection comes to equilibrium when gene frequency of the tester reaches  $(1 + k)/2k$ , short of fixation if  $k > 1$ .

If concurrent with the foregoing process there should be selection of the high specific combinations (high  $\times$  low) with lower gene frequencies, the combined effect on gene frequency may be nil. It may even be to degrade gene frequency when gene frequency is so near the equilibrium that heritability of general combinability is weak. From this view we may expect in the event of overdominance to find the equilibrium zone nearer the upper end of the range of data, providing some degree of positive heritability, some degree of positive regression of  $F_1$ 's on inbred parents.

Ear-to-row selection should have progressed toward equilibrium gene frequencies except for the counter effect of selection of superior plants within ear-rows and within recombinations, selections of elite specific combinations of two gametes with above-average heterozygosity and lower gene frequency.

Modern corn breeding is failing largely beyond the first cycle for the same reasons that caused the failure of ear-to-row selection, except that inbred lines provide for a more efficient identification of elite specific combinations which may have the lower gene frequencies.

The whole of the evidence fits the generalized Mendelian model neatly enough if we may accept overdominance and otherwise proceed without prejudice to those conclusions more consistent with the data.



In familiar theory, selective advantage of a heterozygote leads to an equilibrium gene frequency in natural selection, where every individual leaves progeny (no culling) in proportion to fitness or where fitness is in fact fertility or more specifically, number of offspring surviving to breed. We must distinguish now between  $k$  for a physical trait,  $k_{ns}$  for natural selection of the same trait, and  $k_s$  for artificial selection. Since there is little apparent difference between bushels per acre and potential number of offspring surviving to breed, it may be supposed that  $k$  and  $k_{ns}$  are about the same for yield factors in corn. But if  $k > 1$ , artificial selection including strong culling may make  $k_s$  appreciably greater than  $k$ , and  $(1 + k_s)/2k_s$  appreciably less than  $(1 + k)/2k$ . The expected effect of any single cycle of artificial selection is to shift gene frequency towards  $(1 + k_s)/2k_s$ , if  $k > 1$ . The operator's success (measured by  $k_s$ ) in culling out homozygotes will improve as gene frequency approaches  $\frac{1}{2}$  and frequency of  $aA$  approaches maximum. The limit is reached when  $k_s$  is infinite, and gene frequency is  $(1 + \infty)/2\infty$  or  $\frac{1}{2}$ ; e.g., as when saving only roan Shorthorns for breeding stock. The roans then have infinitely more progeny than whites or reds, which have none.

It does not seem likely that the limit equilibrium of  $q = \frac{1}{2}$  can be reached or maintained with multigenic complexes such as corn yield, because of inability to cull absolutely all homozygotes. On this theory, strong selection will seem to degrade vigor. Relaxation of selection may allow vigor of the corn variety to improve. But there may be important loci where overdominance does not obtain, which tend to obscure the overdominance effect.

If artificial control should maintain fertility continually proportional to the physical trait where  $k > 1$ , gene frequency should progress to equilibrium at  $(1 + k)/2k$ ; cf. recurrent selection for general combinability for corn yield. The population mean is maximum for the physical trait when  $q = (1 + k)/2k$ .

If overdominance should be important in vigor of cattle at a number of un-fixed loci and a herd is close to  $(1 + k)/2k$  for those loci, mild culling of females would tend to raise gene frequencies above  $(1 + k)/2k$ . Strong culling of males might have the opposite effect. Founding an *elite* herd with choice females from many herds and an expensive bull might be more likely to degrade gene frequency below optimum in the event of overdominance. The offspring of the choice animals might be disappointing aside from expected regression towards the mean of the breed.

### EFFECTIVENESS OF RECURRENT SELECTION

Most of the selection practiced with plants and animals is recurrent. Exceptions are selection among homozygous lines or among clones. Inbreeding may curtail the efficiency of recurrent selection by lengthening the cycle. Selection within inbred lines during the process of inbreeding is recurrent but inefficient to the extent that freedom of recombination is curtailed. I have

suggested before that breeders of self-fertilized crops might find greater efficiency in more frequent recombinations. It was to emphasize these considerations that the term *recurrent selection* was introduced. The sense of recurring back to the same tester was never intended.

Breeders of open-pollinated corn need to save no more than 1 ear from 500 or more to plant the same acreage again. If selection is only 20 per cent effective, the net effect in ten years is  $(\frac{1}{10})^{19}$ . The number of corn plants grown in the world in one year is roughly  $(10)^{11}$ . In 100 million times the world acreage of corn there might be one plant as good as the farmer's whole field after he has done 10 to 12 years of recurrent<sup>2</sup> selection. That this seemingly fantastic theoretical concept is essentially correct is supported very well, I think, by results of selection for oil and protein of the corn kernel in the well-known Illinois experiments and in many other less well documented cases with animals, too. East has proposed that selection for oil and protein in corn might be more efficient with inbred lines. However, East proposed that  $S_1$  lines from the selected ears after chemical analyses be recombined for another cycle of selection. He employed inbreeding only to avoid open-pollination of the ears to be analyzed. It is unthinkable that East meant to propose that selection within and among inbred lines for oil or protein without recurrence of selection should be the more effective process.

Open-pollinated corn varieties of 50 or 30 years ago were actually pretty good, in yield and in many other respects. The selection differentials by which they were isolated were probably enormous. Nevertheless, specific combinations of inbred lines are sometimes 20 to 30 per cent above the varieties in yield. That this gain is mainly due to higher frequencies of dominant favorable genes in the elite inbred lines isolated from only a few hundred without recurrence of selection is really inconceivable.

A single corn plant in the variety is a product of two gametes. An  $F_1$  of two homozygous lines is a product of two gametic types. The plant and the  $F_1$  are genetically the same in mean, variance, and expectation of homozygosity in advanced generations as well as the first. It should not be difficult, if asexual propagation were possible, to isolate from the single plants clones that are easily superior to the present elite  $F_1$ 's. That the reservoir of specific combinability in corn is far from exhaustion by present hybrids is evident in comparisons of  $F_1$ 's with the range of individual plants in varieties. The animal breeder may look upon a family of full sibs (from four grand-parental gametes) as a double cross of unselected but homozygous lines, for a rough estimate of possibilities with hybrids. But, aside from that, the breeder of open-pollinated corn was selecting among specific combinations of two gametes the same as in selection among  $F_1$ 's. Continued selection within varieties might have degraded gene frequency below  $(1 + k)/2k$  at any locus

2. Cf. Huxley, *Genetics in the 20th Century*, p. 595. "Recurrent selection," natural or artificial, is designed to multiply improbabilities; requires heritability in the strictest sense. Selection among inbred lines may go on and on without "recurrence."

where  $k > 1$ , thus providing the positive mild regression of offspring on parent, the heritability which so many have taken as strong evidence against  $k > 1$ .

Many traits of the corn plant are mostly independent of genes concerned with yield. Many others may be optimum for yield at intermediate points genotypically as well phenotypically. It should hardly seem surprising if, subsequent to intense selection for yield, we should find evidence of intermediate gene frequencies and very little inbreeding depression or heterosis with such characters. An intermediate optimum may place some premium on  $aA$ , but hardly to the extent of explaining the evident heterosis of corn yields, so far as I can see.

Evidence cited here of overdominance in the genetics of grain yield of corn consists of:

1. Failure of mass selection and ear-to-row selection beyond the level of the adapted variety.
2. Crossbreeding recombinations of parent lines of elite hybrids yield little more than the original varieties.
3. Hybrids of second-cycle and third-cycle lines yield little more than those of the first cycle.
4. Homozygous corn yields 30 per cent as much as heterozygous corn.
5. No evidence of epistasis in corn yield.
6. Regression analyses of yields of  $F_1$ 's and inbred parents indicate a zone of nearly level regression near the upper end of the range of present data, where it might be predicted with the kind of artificial selection which has been practiced, and in the event of overdominance.
7. There is some evidence that selection for general combinability alone with respect to yield is effective and this too is consistent with the expectation of overdominance theory.
8. The fact of hybrid corn is hardly to be explained as other than a result of selection for specific combinability, which in turn is manifestly dependent on heterozygosity of corn yield genes.

My proposal (Hull, 1945a) that recurrent selection for specific combinability be given a trial was made on the assumption that recurrent selection for general combinability or for accumulation of dominant favorable genes had been fairly tried in mass selection and subsequently. The tentative conclusion was that varieties (and breeds perhaps) were near equilibrium, with mean gene frequencies approximately at  $(1 + k)/2k$ . Regression analyses a little later indicated that the corn samples were below equilibrium. Since then it has been proposed orally many times that two parallel breeding plans restricted respectively to specific and to general combinability might well be run with corn and with small laboratory animals as pilot experiments. I have later come to believe that recurrent selection among homozygotes might also provide results of considerable theoretical interest.

Present-day corn breeding is done in three steps: selection among inbreds based on their own phenotypes; selection among inbreds for general combinability; selection among specific  $F_1$ 's of the remaining inbreds. These steps are the three processes of the preceding paragraph. The corn breeder applies the three processes in the order named to the same stock, then recombines the elite lines and begins the cycle again. The present proposition is to apply the three processes separately to parallel stocks, and thus attempt to learn which ones are primarily responsible for superior hybrids.

#### RECURRENT SELECTION AMONG HOMOZYGOTES

This process can be done effectively enough in corn, perhaps with  $S_2$  lines. Two selfings would amount statistically to reducing the degree of dominance to one-fourth of the original value. One-half of the  $S_1$  lines could be discarded in the first comparison. About fifty  $S_2$  lines should be retained in the recombination. Selection within ear-rows should be rigidly excluded.

There is no reason to suppose that a physiological barrier would be reached short of the level of elite hybrids. Recurrent selection towards an extreme has been very effective with many characters where not much dominance is apparent. In noted cases no limit of genetic variance has been reached. What genetic limit might be reached with vigor or yield genes of corn when the confusion of dominance is artificially eliminated is to be explored. Theoretically, this process of recurrent selection should be much superior to any non-recurrent selection among gametes or doubled haploids.

#### RECURRENT SELECTION FOR GENERAL COMBINABILITY

Strictly, the tester should be the variety.  $S_0$  plants or  $S_n$  lines are to be testcrossed with several plants of the variety. The  $S_0$  plant must be selfed at the same time. Parents of elite testcrosses are recombined into an improved variety which becomes the tester for the next cycle. If gene frequency of the variety is improved to approach  $(1 + k)/2k$ , where  $k > 1$ , heritability will approach zero and the variety mean its maximum. If pseudo-overdominance from repulsion linkage is important the equilibrium may advance to higher levels as recombinations occur. But, aside from that, we have now no experimental verification of a selection equilibrium, and a test would seem desirable. Concurrent selection for specific combinability should be strictly avoided in this test.

#### RECURRENT SELECTION FOR SPECIFIC COMBINABILITY

This process has been adequately described both here and earlier (Hull, 1945a). From the theoretical viewpoint it would be best to use a homozygous tester and avoid selection within the crossbred except that based on testcross performance. The purpose is to determine first how much specific combinability may be accumulated in early cycles and eventually to determine where this process may reach physiological or genetic limits.

Now if we are convinced that overdominance is not very important and that, perhaps for other reasons too, selection for general combinability will eventually win, or at least not lose, we may proceed at once with recurrent selection for general combinability to render hybrid corn obsolete. Some of us may find it necessary to include an inbreeding interphase between cycles. Breeders of livestock may as well return to improvement of pure breeds by progeny testing. We will run these pilot tests merely for the sake of verification.

But if it should seem likely that recurrent selection for specific combinability may win, the breeder of livestock may begin now with recurrent reciprocal selection for specific combinability. For my part, I would choose two crossbreds for the start and would begin mild inbreeding in one of them which would become the stud herd. On one side of this is the Comstock plan with no inbreeding in either herd. On the other side we might choose a line with 50 per cent inbreeding at the start and practice reciprocal selection along with continued mild inbreeding. Evaluation of these alternatives of the reciprocal plan with small laboratory animals, along with the other two main plans, would be of considerable interest theoretically. The cost might be minute in comparison with the total of wasted effort in current breeding practices.

Recurrent selection for general combinability without the inbreeding interphase is a fairly obvious technic which has been employed and described variously. The first discussion of it from the overdominance viewpoint with the restriction against selection for specific combinability was that of Hull (1946b). Since then I have continued to urge parallel tests with fast breeding species as pilot experiments. Recurrent selection for superior homozygotes is proposed here for the first time, I believe.

Reciprocal selection for specific combinability was a counter proposal to me of several corn breeders in 1944 and later, when I proposed selection in a crossbred for combinability with a fixed tester, a homozygous line or  $F_1$  of two homozygous lines.

For simplicity of illustration we may consider a 4-factor example with gene frequency in a homozygote or gamete ( $v$  or  $w$ ) taking values, 0,  $\frac{1}{4}$ ,  $\frac{2}{4}$ ,  $\frac{3}{4}$ ,  $\frac{4}{4}$ . Gene frequencies intermediate to these values may occur in heterozygotes and in whole populations. Let us take  $k = 2$  for the degree of dominance as suggested roughly for corn yield by estimates reported here. Then regression of offspring phenotype on gene frequency of parent in any column of the  $(5 \times 5)$  Mendelian checkerboard is  $b_p = \frac{1}{2}(3) - 2v$ , where  $v$  is gene frequency of the common parent of the column. Substituting the five values of  $v$  provides the five values of  $b_p$ ,  $1\frac{1}{2}$ , 1,  $\frac{1}{2}$ , 0,  $-\frac{1}{2}$ , for the five columns. Heritability changes from positive to negative where  $v = (1 + k)/2k = \frac{3}{4}$ . These values of  $b_p$  for the given values of  $v$  are the same for any number of loci. In any case the zone of near-zero heritability for one locus is relatively broad on both sides of the critical value of zero. Reciprocal selection between two crossbreds

is at equilibrium for one locus when gene frequencies are  $(1 + k)/2k$  in both, and  $k > 1$ . It is conceivable that gene frequencies of the two crossbreds may wander in the zone of low heritability through many cycles of reciprocal selection, but they must eventually separate on opposite sides to approach *aa* and *AA* respectively with increasing velocities. When the two gene frequencies are on opposite sides of the equilibrium initially, reciprocal selection will tend to drive them farther apart. If they are on the same side both will tend to approach equilibrium. Comstock's statement here that the one nearest equilibrium may approach it more rapidly and continue beyond to reverse the trend of the other, thus obtaining a quick separation, seems good. I had overlooked this point and hope it may be experimentally verified.

Gametes with critical gene frequencies in the present model are *aAAA*, *AaAA*, *AAaA*, *AAAA*. A general tester composed of the four homozygous lines producing these four gametes respectively will provide zero heritability. So also will a crossbred tester for every locus where gene frequency is  $\frac{3}{4}$ . One of the homozygous lines alone as a specific tester provides mean  $b_p = 0 = [\frac{3}{2} + 3(-\frac{1}{2})]/4$ . But here the individual values of  $b_p$  for each locus are at maximum,  $\frac{3}{2}$  for the *aa* locus, and  $-\frac{1}{2}$  for each *AA* locus, providing maximum heritability in selection to a homozygous tester.

Defining phenotypes of *aa*, *aA*, *AA* alternatively as  $1 - s$ ,  $1 - hs$ ,  $1$ , provides  $b_p = 1 - h - (1 - 2h)v$ . Then with  $h = -\frac{1}{2}$  for the same degree of dominance as the present model,  $b_p = \frac{1}{2}(3) - 2v$  again. The only inconsistency between the two systems of defining phenotypes which may be encountered here, I think, is failure to distinguish between physical values and selective values, e.g., body weight and number of offspring surviving to breed.

It seems fairly clear that overdominance of the degree considered here may provide considerable variation of heritability within a finite sample, a herd or a variety on one farm. Mean  $b_p$  may be positive and fairly large, yet  $b_p = 0$  near the upper range of gene frequency in the sample. Moreover, the degree of dominance for selective values might be appreciably greater than for the physical trait. For these reasons, selection indexes made up with average heritabilities of physical traits could be misleading.

Parallel operations of the foregoing breeding plans with heavy dosages of mutagenic agents in addition might provide significant information on progressive improvement, where the objectives respectively are the superior homozygote, the mean of the population, and the superior heterozygote. This proposal will be subject to criticism by those who are convinced that it is only in gene-by-gene analysis that real advances in knowledge of genetics can be obtained. I have no quarrel with that viewpoint except that where many genes with minute effects may be involved the gene-by-gene approach still seems fairly remote.

Recurrent selection in prolific species such as corn, chickens, mice, and *Drosophila* may soon build up very large selection intensities, perhaps to re-

cover high frequencies of rare natural or mutant alleles. Chemists have employed high pressures and temperatures to obtain reactions of great interest. They have concentrated rare elements and rare isotopes by various ingenious processes. With selection intensities and mutation rates well above natural values it might be possible to obtain estimates of the minimum ratio of selection to mutation for survival or improvement of the variety or breed.