Euphytica 4 (1955): 58-66

THE EFFICIENCY OF SOME SELECTION METHODS IN POPULATIONS RESULTING FROM CROSSES BETWEEN SELF-FERTILIZING PLANTS

F. K. VAN DER KLEY [1]
Agricultural University, Wageningen

Received 23 Sept. 1954

1. INTRODUCTION

It is the ideal of the breeder to produce good varieties in as short a time as possible and with the aid of a minimum of land, capital and labour. In order to realize this objective it is necessary to develop efficient selection methods, which separate the desirable genotypes from the undesirable ones reliably and at an early stage. The theoretical aspects of this problem with some important factors affecting the selection possibilities in different generations are dealt with in this article. For self fertilizing crops, the merits of the principal selection methods are discussed and indications of a means to improve their efficiency are given.

2. GENETICS AND YIELD

The breeding objective of ariable crops is as a rule the production of plants that give a high yield per unit area of land. An approximately ideal highly productive plant can be considered as a physiological system which is carefully balanced by quantitatively harmonizing effects of many genetic factors. This great number of genes that affect yield necessitates working with large numbers of plants in order to obtain every possible gene combination. Because of the gene interaction mentioned, the cultural value (phenotype) of a heterozygous plant is not a reliable guide to the value of the lines which may be derived from it. In fact plants that might segregate valuable gene combinations after selfing, could easily get lost, if plants, dominant or intermediate for some detrimental genes, are eliminated from highly heterozygous populations by means of phenotypic selections (1, 4, 9).

3. HETEROZYGOSITY IN THE F_2

In the F_2, obtained by selfing an n-fold heterozygous F_1, 3^n different genotypes may occur. Among every 4^n plants (in the absence of linkage) 2^n different homozygous genotypes may be expected to occur each once. The greater part of the plants in the F_2 viz. $\left(\frac{2^n - 1}{2^n}\right) \times 100\,\%$ is therefore more or less heterozygous.

The relative numbers of plants which possess a certain *degree of heterozygosity*

---

[1] The author wishes to thank Prof. Dr J. C. DORST and Mr M. KEULS for their helpful criticism. To the former he is also grateful for kind permission to publish this part of a M. Sc. thesis and to the latter for valuable mathematical suggestions and aid.

or homozygosity respectively, comply with the coefficients resulting from the working of $(1 + 1)^{n}$*.

For increasing n, the number of heterozygous alleles tends to be normally distributed with mean $\frac{1}{2}n$ and variance $\frac{1}{4}n$. At the same time all genotypes that deviate from the average heterozygosity tend to be rare.

Valuable genotypes which are heterozygous for about half of the number of segregating genes may, therefore, easily get lost by selecting in small F_2 populations when breeding in crops which are difficult to cross. This will be the more so in the presence of linkage (2,5).

4. The attainment of homozygosity in later generations

When selfing is continued, the relative number of homozygotes increases in the population. For the F_x of a crop with 100 % selffertilization and no natural selection, the following formula can be deduced with the aid of Mendel's Laws:

$$\frac{\text{Expected number of n-fold homozygotes}}{\text{Total number of plants}} \text{ in the } F_x = \left(\frac{2^x - 2}{2^x}\right)^n$$

Using this formula, the well-known S-shaped homozygous curves indicating the percentage of homozygotes in consecutive generations can be constructed. These, inter alia, are given by Banga (2, p. 280). In practice, however, the attainment of homozygosity is less complete and is delayed by spontaneous intercrossing (2, 7, 8). Calculations of the degree of heterozygosity with a formula of Löhner (6, p. 61), however, show that these disadvantages have often been over-estimated. It may be concluded also that when some spontaneous intercrossing occurs in a population, the percentage of highly heterozygous plants decreases considerably.

5. The frequency of the best homozygous genotype

In the F_2 population mentioned, 2^n different homozygote genotypes can occur.

From the formula given it follows that:

$$\frac{\text{Expected number of plants of the best homozygous genotype}}{\text{Total number of plants}} \text{ in the } F_x = \left(\frac{2^{x-1} - 1}{2^x}\right)^n$$

With the aid of this formula, the frequency of the best homozygous genotype in the F_2 (column 2) and in the $F\infty$ was calculated and is given in Table 1. (See p. 60).

The figures in column 4 indicate the number of generations after which the frequency amounts to 90 % of that in the $F\infty$.

It appears from column 2 that it is improbable (even in the most favourable case when alle desired genes are dominant) that the breeder would succeed in isolating the best plant from the F_2 in a reasonable degree of homozygosity, assuming that the

---

*) In this and all following formulae n always represents the number of *independently* segregating genetic factors in the population. The X occurring in the formulae indicates the number of the generation to which the formula applies (F_X). The p, if used, means the number of genes on which selection takes place.

grandparents, differ in more than 7 independently segregating genetic factors. In practice this number of genes is usually more than 7.

TABLE 1. THE FREQUENCY OF THE BEST N-FOLD HOMOZYGOUS DOMINANT GENOTYPE IN THE F_2 (COLUMN 2) AND IN THE ∞ (COLUMN 3). THE VALUES INDICATED IN COLUMN 3 ARE THE LIMIT VALUES WHICH WILL BE OBTAINED FOR 90%, AFTER THE NUMBER OF GENERATIONS INDICATED IN COLUMN 4. (NATURAL SELECTION AND SPONTANEOUS INTERCROSSING IS PRESUMED NOT TO OCCUR IN THE POPULATION.)

The number of indep. segreg. factors	The frequency of the best homozygous genotype in the		The number of generations after which the limit value indicated in column 3 is reached for 90%
	F_2	$F\infty$	
n	$100/4^n$ %	$100/2^n$ %	
5	0.098%	3.1%	6
7	0.006%	0.8%	7
10	0.000.06%	0.1%	8
20	0.000.000.000.04%	0.000.1%	9
30	0.000.000.000.000.000.1%	0.000.000.1%	10

Apparently the chance (column 3) of isolating the best type from a segregated population (column 4) in a homozygous state is, although considerably increased, also insufficient. In fact it may be concluded that neither a final selection made in F_2 or F_3, nor a selection made within a segregated population in which no pre-selection was effected by man (§ 6 and § 7) or by nature (§ 8), will be efficient.

6. PRE-SELECTION IN CONSECUTIVE GENERATIONS

The breeder might start in the F_2 with pre-selection by removing all plants that are homozygous for one up to a maximum of p recessive detrimental genes. Repeating this removal in every generation it can be calculated that

$$\left(\frac{2^{x-1}+1}{2^{x-1}+2}\right)^{p} \times 100\,\% \text{ of all plants are retained in the } F_x.$$

The amount of this percentage for different values of x and p is indicated in Table 2.

TABLE 2. THE EXPECTED PERCENTAGE OF PLANTS RETAINED IN THE F_x, AFTER REMOVING IN THE F_2 AND SUBSEQUENT GENERATIONS THE UNDESIRABLE PLANTS, HOMOZYGOUSLY RECESSIVE FOR AT THE LEAST ONE AND AT THE UTMOST P DETRIMENTAL GENES

	p = 1	p = 2	p = 3	p = 4	p = 5	p = 6	p = 7	p = 8	p = 9	p = 10	p = 11	p = 14	p = 20	p = 30
x = 2	75	56	42	32	24	18	13	10	7.5	5.6	3.2	1.7	0.3	0.02
x = 3	83	69	58	48	40	33	28	23	19	16	11	7.8	2.6	0.4
x = 4	90	81	73	66	59	53	48	43	37	35	28	23	12	4.2
x = 5	94	89	84	80	75	71	67	63	60	56	50	45	32	18
x = 6	97	94	91	89	86	84	81	79	76	74	70	66	55	41

The frequency improvements resulting from the above outlined selections have been calculated. After continuous selection in F_2 up to F_x (F_x not included) for p independently segregating factors $\left(\frac{2^{x-1}-1}{2^{x-1}+2}\right)^{p}$ x 100 % of plants will be homozygously dominant for these factors in the F_x. If no pre-selection had been practised, this percentage would have been $\left(\frac{2^{x-1}-1}{2^{x}}\right)^{p}$ x 100 %, according to § 5. When the latter formula is divided into the former we obtain the numbers $\left(\frac{2^{x-1}}{2^{x-2}+1}\right)^{p}$ of Table 3.

It is apparent from Table 3 that errorless pre-selection during some consecutive generations increases the frequency of the genotypes homozygously dominant for all p factors and, consequently, the frequency of the best genotype in the F_x population to a considerable degree.

TABLE 3. BY SELECTION IN THE F_2 AND UP TO AND EXCLUDING THE F_X ON P INDEPENDENTLY SEGREGATING GENETIC FACTORS, THE FREQUENCY OF THE BEST HOMOZYGOUSLY DOMINANT GENOTYPE FROM THE POPULATION IS EXPECTED TO BE MULTIPLIED IN THE F_X BY THE VALUES INDICATED IN THE TABLE

	P	P=1	P=2	P=3	P=4	P=5	P=6	P=7	P=8	P=9	P=10	P=11	P=14	P=20	P=30
X = 3	$(^4/_3)^p$	1.3	1.8	2.4	3.2	4.2	5.6	7.5	10	13	18	24	56	313	$5.5=10^3$
X = 4	$(^8/_5)^p$	1.6	2.6	4.1	6.6	10	17	27	43	69	110	176	721	1.2×10^3	1.8×10^5
X = 5	$(^{16}/_9)^p$	1.8	3.2	5.6	10	18	32	56	100	180	316	561	3,152	9.9×10^3	1.0×10^6
X = 6	$(^{32}/_{17}{}^p$	1.9	3.5	6.7	12	24	44	84	157	296	557	1,048	6,988	3.1×10^4	3.1×10^7
X = 7	$(^{64}/_{33})^p$	1.9	3.8	7.3	14	27	53	103	200	388	752	1,457	10,630	5.6×10^4	1.7×10^8
X = ∞	$(2)^p$	2	4	8	16	32	64	128	256	512	1,024	2,048	16,384	1.1×10^6	1.1×10^9

From Table 2 it follows, however, that using the described system of pre-selection, in consecutive generations an increasing percentage of plants will be retained. Thus applying the strongest selection in the most heterozygous generations has the disadvantage that valuable gene combinations may be lost (§ 2 and § 3). This disadvantage can be avoided if the breeder selects in the F_2 for a few well-known recessive detrimental genes and if he increases gradually in later generations this number of genes.

7. GRADUAL SELECTION

Gradual selection may be defined as a selection for a number of genes gradually increasing in subsequent generations and corresponding with a constant percentage of retained plants in every generation. The gradual increase of this number of genes in every generation can be calculated with the aid of the equations:

$$R^*_2 = S_2^{p_2};\quad R^*_3 = R_3^{p_2} \text{ x } S_3^{p_3};\quad \text{in general:}$$
$$R^*_x = [R_x^{p_2+p_3+\dots\dots\dots p_{x-1}}] \text{ x } S_x^{px}$$

In these formulae:

100 R^*_x = the chosen percentage of plants, retained in the F_x after removing in F_2, F_3, F_4, ...and F_x all plants which are homozygous for one up to p_2, $p_2 + p_3$, $p_2 + p_3 + p_4$, ... and $p_2 + \dots\dots p_x$, recessive detrimental genes respectively;

100 $R_x^{p_2+\dots\dots p_{x-1}}$ the percentage $\left(\frac{2^{x-1}+1}{2^{x-1}+2}\right)^{p_2+\dots\dots p_{x-1}}$ of plants retained

in F_x after removing in F_2 up to and including F_x all plants that are homozygous for one up to a constant number $p_2 + \ldots\ldots . p_{x-1}$ of detrimental recessive genes. The values of the different terms $R_x^{p_2 + \ldots\ldots p_{x-1}}$ may be seen from Table 2.

$100\ S_x^{p_x}$ the percentage of plants $\left(\frac{2^{x-1}+1}{2^x}\right)^{p_x}$ retained in F_x after removing in F_x all plants that are homozygous for one up to $p_2 + \ldots\ldots . p_x$ detrimental recessive genes on which no selection had been practised in previous generations. The results of these calculations are shown in Table 4.

TABLE 4. INCREASES P_2, P_3, ... P_6, (COLUMN 2) IN THE GENERATIONS F_2 UP TO AND INCLUDING F_6 RESPECTIVELY OF THE NUMBERS (COLUMN 3) OF DETRIMENTAL DOMINANT GENES, ON WHICH PRE-SELECTION MAY BE APPLIED, IF ONE DECIDES TO RETAIN (AT SELECTION) AT LEAST 100 R^*_x% (COLUMN 1) OF PLANTS IN EVERY GENERATION

100 R^*_x greater than:	Increase of genes					Total number of genes in generation				
	p_2	p_3	p_4	p_5	p_6	F_2	F_3	F_4	F_5	F_6
50	2	0	0	0	0	2	2	2	2	2
25	4	1	1	1	1	4	5	6	7	8
12.5	7	1	2	2	2	7	8	10	12	14
6.25	9	2	2	3	3	9	11	13	16	19
3.125	12	2	3	3	4	12	14	17	20	24
1.5625	14	3	4	4	5	14	17	21	25	30

It is obvious that the frequency improvements, obtained after gradual selection according to Table 4 are the same as those shown in Table 3. Gradual selection is, therefore, as efficient as the selection in consecutive generations for a constant number $p_2 + \ldots\ldots . p_x$ genes as discussed in § 6. Gradual selection, however, has obvious practical advantages and may be concluded to be the more satisfactory method of pre-selection. Since pre-selection has been shown to be necessary (§ 5) it may be recommended to utilize a pre-selection for a gradually increasing number of well-known genes, thus considerably enhancing the chances of a successful selection in later generations.

The pre-selection which is effected by nature (§ 8) in every generation does not always take place in a desirable direction.

8. NATURAL SELECTION

In natural selection certain genotypes are eliminated from a population which is propagated in the absence of special breeding measures, while the percentage of other genotypes increases. The principal interacting causes of natural selection will now be discussed.

8.1. *Natural selection through influence of environment*

Distinct genotypes often respond to a similar location in a different way; especially environment may offer better reproductive possibilities to one genotype than to another. A noticeable movement of genotypes in a population due to this cause is called natural selection by environment. The breeder can give direction to this movement in so far as he can control the environment. By propagating a population under different climatic and soil conditions, a bias in one direction or another can be prevented and consequently an undesired natural selection by environment can be delayed. A *favourable natural selection* can be encouraged if the population is increased in an environment closely related to its future cultivation. Promoting the detrimental characteristics of the environment will give the natural selection the best opportunity of removing the undesired genotypes from the population.

Some well-known applications of this procedure are as follows:

a. The encouragement of natural infection and the application of inoculation methods to test disease resistance.

b. The cultivation of the populations under unfavourable circumstances to test the drought resistance, etc.

8.2. *Natural selection due to differences in reproductive capacity*

The reproductive capacity of a certain genotype is defined by the number of seed bearing plants which is produced by that genotype under optimal conditions.

The number of plants with a high reproductive capacity will increase at the expense of plants with lesser reproductive capacity in a population, provided that there are no differences in aggressiveness and in reaction to environment. In this case it is called natural selection by difference in reproductive capacity.

In crops mainly cultivated for their seeds (the majority of self-fertilizing arable crops), the high yielding varieties are generally more favoured by this form of natural selection. There are, however, exceptions to this statement. Under natural conditions there is a tendency to produce a large number of seeds per plant and this property is often positively correlated with that for small seeds. The selection of plants giving a large number of seeds per individual can even run parallel with a population with a low yield of seeds. This adverse type of natural selection has been reported in wheat by PALMER (7).

The characteristic for the production of small seeds is a desirable one in green fodder and green manure producing crops. Apart from these, the influence of natural selection in this direction with crops which are principally grown for their vegetative parts is largely unfavourable. If the yield of marketable vegetative parts is negatively correlated with the seed yield, as may occur in flax, the fine and few-seeds-producing types are the most desirable ones.

The unfavourable effects of this form of natural selection can be partly eliminated. In theory, this is possible by dividing the population into sections, each containing plants with approximately the same reproductive capacity. Apart from the large amount of work involved, this is usually inapplicable because the breeder has insufficient knowledge of his material. The elimination of strongly branched oil flax types from fibre flax populations can be mentioned as one of few practical applications. It is also possible to prevent natural selection by differences in reproductive capacity by harvesting an equal number of seeds per plant. This procedure also involves much labour and can result in a loss of genes.

8.3. *Natural selection due to differences in aggressiveness*

In a bulked population, individuals with a rapid germination capacity, rapid youth development and tillering ability will be favoured in the competition for space, food, moisture, light and air.

Their reproduction possibilities are more favourable, so they will oust other less aggressive plants from the population. The most aggressive genotypes, however, are not always the most desirable ones. Some genotypes which make high demands on the environment and are much less aggressive under less favourable conditions, yet they appear to be valuable varieties. Apart from many Dutch examples, this experience has been cited by PALMER (7), who showed that from an artificial population obtained by mixing cultivated barley varieties, the highest yielding but least aggressive varieties had disappeared from the population after 16 years.

9. SELECTION METHODS

In § 8 was deduced that natural selection has favourable as well as unfavourable features which can predominate, depending on breeding objective, crop and environment. Valuable genotypes can easily be lost when applying the *bulk selection method* in a crop where an unfavourable natural selection dominates. Under such circumstances the pedigree selection method should be advised.

In applying the *pedigree method* as reported by HARRINGTON (3) the best plants are selected annually and are given much attention. Due to this individual treatment, the detrimental effects of natural selection caused by environment and differences in aggressiveness may be avoided for the greater part. This method is laborious and therefore expensive, however, due to the large numbers of separate lines that must be maintained from year to year. Confining this number of lines by annually discarding many plants has the drawback that valuable genotypes may easily be lost. Moreover, it is difficult to carry out a fixed annual programme of selecting the best plants since circumstances change from year to year.

The last difficulty is largely encountered in the *mass pedigree method* designed and advocated by HARRINGTON (3) who maintains that the method can make the best use of the natural selection possibilities because of its flexibility. He increases the F_2 in bulk (with or without the application of mass selection) until an opportunity occurs of reducing the population by about half in the F_3, F_4 or F_5 by a selection in a favourable year.

The remaining plants are then harvested separately; the various progenies are compared in the following year; the majority are eliminated and the remainder are combined to give a new population. After some generations a final selection programme is initiated. The detrimental effects of unfavourable natural selection in many crops, however, are a serious drawback to this method, while a loss of valuable genotypes during the selection in a still highly heterozygous F_3, F_4 or F_5 is not excluded.

10. CONCLUSIONS AND SUGGESTIONS

All the methods mentioned in § 9 have the drawback that valuable genotypes, or plants capable of producing these genotypes, may be rejected. Theoretically, it is clear that this could be avoided by giving each plant individual treatment (in this way preventing natural selection by competition) and by retaining a sufficient number of plants in every generation. In practice, however, the breeder is obliged at an early stage to remove plants, due to the often considerable number of seeds produced by a single plant. This removal may be carried out by selection as well as by dividing a population into equivalent parts after carefully mixing all the seeds which have been harvested. It is obvious, however, that a *reliable selection* will always be preferred to removing at random. The first may be effected the best by conducting resistance tests in the material successively or simultaneously. Since pre-selection for dominant detrimental genes cannot be recommended in the first generations (see page 61), the investigated susceptibility must be inherited simply and recessively. This condition is satisfied in a sufficient number of diseases. According to § 7 it is recommended to increase the number of applied resistance tests gradually in later generations, thus retaining in every generation (theoretically spoken) a constant percentage of plants. The actual value of this percentage primarily chosen may depend on the number of viable seeds on the average produced by a single plant, on the number of F_2 plants and the area of land available, and according to the breeder's view. Variations on the theme proposed, e.g. retaining in every generation a successively decreasing percentage of plants, thus applying an *increased gradual selection*, are also possible.

After (increased) gradual selection during several generations the population may be supposed to be suitable for selection. Consequently the remaining plants are harvested separately. The next year, in considering visible characteristics, the breeder makes a rough estimate of the probable rate of heterozygosity in the separate rows.

If this rate is not too high, yield estimations are carried out and the elimination of less valuable lines can then be considered. The prospect of combining certain lines can also receive consideration. It is better, as a rule, to carry on the work with separate drills and lines. It is the author's opinion that this procedure eliminates the drawbacks of the bulk-method, the pedigree method and the mass pedigree method mentioned in § 9.

SUMMARY

The question which selection method is most suitable for application in populations resulting from crosses of self-fertilizing plants is discussed.

Some factors affecting the value of selecting methods in successive generations of a segregating population are reported. After extensive calculations (§ 2–§ 7), the author concludes that the pedigree method, the bulk method and the mass pedigree method are not the most efficient ones, due to the small frequency of desired genotypes and to the possible loss of valuable genotypes by a wrongly directed natural or human selection. He suggests (§ 10) that the best method of making populations more suitable for selection lies in a pre-selection on a gradually increasing number of recessive detrimental genes during consecutive generations, which may be effected by conducting various resistance tests. The efficiency of the „*gradual selection*" *procedure* proposed is calculated to be high when compared to the above mentioned methods.

SAMENVATTING

Over de efficiency van enige in kruisingspopulaties van zelfbevruchters gebruikte selectiemethoden

Na een bespreking van enige factoren die van invloed zijn op de selectie-mogelijkheden in verschillende generaties van een uitmendelende populatie berekent schrijver, dat zowel de lijnselectie in jonge populaties (pedigree method) als de lijnselectie in oudere populaties (bulk method) en het bijeen voegen van door lijnselectie in jonge populaties verkregen materiaal als uitgangspunt voor latere lijnselectie (mass pedigree method) onvoldoende waarborgen bieden dat het best mogelijke genotype door de kweker ook werkelijk uit de populatie geïsoleerd kan worden. Bij toepassing van bewerkelijke en daarom kostbare lijnselectie *in jonge populaties* wordt reeds in sterk heterozygoot materiaal op opbrengst geselecteerd. Afgeleid wordt (§ 2 en § 3), dat dientengevolge gemakkelijk dragers van belangrijke genencombinaties verloren kunnen gaan. Bij toepassing van de lijnselectie *in oudere populaties* kunnen door verkeerd gerichte natuurlijke selectie (§ 9) eveneens waardevolle genotypen uit de populatie verdwijnen. Beide gevaren bestaan in verminderde mate, indien het door lijnselectie verkregen materiaal bijeengevoegd wordt tot uitgangsmateriaal voor latere lijnselectie (§ 9).

Afgeleid wordt dat bovenstaande bezwaren het best vermeden kunnen worden door de jonge populaties te onderwerpen aan een beperkt aantal resistentie-tests en daarop geen andere selectiemaatregelen toe te passen. Daar een voorselectie op dominante schadelijke genen in de eerste generaties niet aanbevelenswaardig is (§ 2) dienen de op aanwezigheid onderzochte vatbaarheden recessief te vererven. Dit is met het merendeel der resistenties het geval. Aanbevolen wordt deze voorselectie in latere generaties trapsgewijs uit te breiden over een groter aantal factorenparen (“*gradual selection*”, § 7 en § 10). Blijkens tabel 4 is dit mogelijk door in iedere generatie een constant of afnemend percentage planten aan te houden. Bij toepassing van de geschetste werkwijze is de populatie reeds na enkele jaren rijp voor lijnselectie, zodat haar efficiency in vergelijking met die der gecritiseerde methoden hoog genoemd mag worden.

LITERATURE CITED

1. ATKINS, R. E. and MURPHY, H. C., Evaluation of yield potentialities of oat crosses from bulk hybrid tests. Agr. J. **41** (1949): 41–45.
2. BANGA, O., Inleiding tot de plantenveredeling. Tjeenk Willink, Zwolle. 1953: 275–311.
3. HARRINGTON, J. B., The mass-pedigree method in the hybridization improvement of cereals. J. Am. Soc. Agron. **29** (1937): 379–384.
4. KALTON, R. R., Breeding behaviour at successive generations following hybridization in soybeans. Iowa Agr. Exp. Sta. Res. Bull. **358** (1948): 671–732.
5. KEULS, M., The non-importance of linkage in breeding self fertilizing crops. Euphytica 4 (1955): in the press.
6. LÖHNER, L., Die Inzucht. Freising-München. 1929: 61.
7. PALMER, T. P., Populations and selection studies in a *Triticum* cross. Heredity 6 (1952): 171–185.
8. SIRKS, M. J., Die Verschiebung genotypischer Verhältniszahlen innerhalb Populationen laut mathematischer Berechnung und experimenteller Prüfung. Med. Landbouw Hogeschool Wageningen, **26** (1923) nr 4: 1–40.
9. SUNESON, C. A. and STEVENS, H., Studies with bulked hybrid populations of barley. Techn. Bull. **1067** (1953) U.S.D.A., 14 pp.

DRUK: H. VEENMAN EN ZONEN, WAGENINGEN