

A STUDY OF INBREEDING IN SOME MODELS OF ARTIFICIAL SELECTION

V. ARUNACHALAM

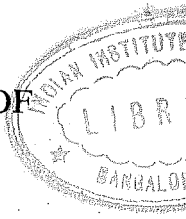
Division of Genetics, Indian Agricultural Research Institute, New Delhi 110012

The effects of artificial selection have been studied in finite and infinite populations through one locus models in discrete and continuous time (Haldane, 1931; Robertson, 1960). In most of these investigations the changes in the mean and the variance of the gene frequency in the population and the rates of advance under selection were studied by the 'diffusion' approach employed by Kimura (1957, 1958). Problems of artificial selection in finite populations in which individuals are chosen as parents for the next generation on a ranking of their phenotypic values for a quantitative character have recently received the attention of several workers (Kojima, 1961; Pike, 1969; Hill, 1969).

Some forms of artificial selection result in sib mating giving rise to inbreeding, and the population becomes homozygous if the selection is continued for many generations. Fisher (1965) in his book, *The Theory of Inbreeding*, visualised 'the possibilities the (inbreeding) process opens out for the practical improvement of domestic plants and animals' and observed that 'a theoretical and practical study of the subject will form an essential part of future programmes either of genetic or of agricultural research'.

Inbreeding for a population with impressive yield, a breeder is faced at some stage, with the problem of fixing the genes controlling the components of yield at the highest frequency. The aim is to obtain essentially a population of homozygotes with a high mean value of the metric character. Usually, if not always, the process known as 'mass selection' is employed wherein the seeds of the plants which are phenotypically uniform and which possess good yield attributes are bulked. The same process is continued in the progeny populations. However, it is also not uncommon to choose one or two single plants which possess the most desirable characters from a small segregating population and to raise the subsequent progeny generations from them. These methods of selection result in inbreeding for the practical improvement of the population. It will, therefore, be interesting to examine theoretically the speed of approach to homozygosity when one or two or more individuals are selected on the basis of their phenotypic performance and the progeny generations bred from them.

We consider in this paper a population consisting of three genotypes resulting from a single locus with two alleles only. The models used are briefly described below.



MODEL 1

Let the population consist of the three genotypes AA , Aa and aa numbered 1, 2 and 3 for convenient reference. We assume discrete generations throughout the study. The individuals in the population are ranked on the basis of the phenotypic values of a metric; the best male and the best female are mated. Let this mating produce N viable offsprings of each sex. Again the best male is mated to the best female from this progeny population and the process is continued. We assume that the progeny population contains $2N$ viable individuals in each generation. The initial mating can be any one of the six numbered in the following order : 1- $AA \times AA$; 2- $AA \times Aa$; 3- $AA \times aa$; 4- $Aa \times Aa$; 5- $Aa \times aa$; 6- $aa \times aa$.

Let a_{ij} be the conditional probability of occurrence of the genotype j given the parental mating i ($i = 1, 6$; $j = 1, 3$). For example, $a_{41} = .25$, $a_{63} = 1$ under random mating.

Let $\mu_{i\sigma}$ be the genotypic value of the genotype i ($i = 1, 3$). The phenotypic values of the genotype i are considered to be normally distributed with mean $\mu_{i\sigma}$ and a common variance σ^2 .

We shall now evaluate s_{ij} , the probability that the best genotype is j in the progeny population given that the parental mating is i .

Let

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

and

$$F(x) = \int_{-\infty}^x f(t) dt$$

so that

$$F'(x) = f(x).$$

By the definition of the a_{ij} s, it follows that

$$\sum_{j=1}^3 a_{ij} = 1$$

for each i ($i = 1, 6$).

Let $x_j = X_{j\sigma}$ represent the phenotypic value of the genotype j so that

$$t_j = \frac{x_j - \mu_{j\sigma}}{\sigma} = X_j - \mu_j$$

becomes the standard normal variate ($j = 1, 3$). Under the mating i , the probability that a member of the progeny population has a phenotypic value not exceeding Y is

$$\sum_{j=1}^3 a_{ij} F(Y - \mu_j) \quad (1)$$

where

$$Y = x/\sigma.$$

Hence the probability that all the N individuals have phenotypic values not exceeding Y is

$$P_N(Y) = \left[\sum_j a_{ij} F(Y - \mu_j) \right]^N.$$

If Y is now the phenotypic value of the best of N individuals, its distribution function is $P_N(Y)$.

Hence the probability that the best individual has its phenotypic value in $(Y, Y + dY)$ and is of the genotype j

$$= N \left[\sum_j a_{ij} F(Y - \mu_j) \right]^{N-1} a_{ij} f(Y - \mu_j) dY.$$

Hence the probability that the best of N progeny is of genotype j given the parental mating i , is

$$s_{ij} = \int_{-\infty}^{\infty} N a_{ij} f(X - \mu_j) \left[\sum_{r=1}^3 a_{ir} F(X - \mu_r) \right]^{N-1} dX, \quad X = x/\sigma. \quad (2)$$

Thus, given N , a_{ij} and μ_j one can evaluate s_{ij} by numerical integration. Simple methods like Simpson's or Weddle's rule were not found to be efficient in evaluating s_{ij} to the desired degree of accuracy. The method used is hence summarised in Appendix I.

Hill (1969) has provided a formula for selecting a best sub-sample of N individuals containing N_1 individuals of the genotype AA , N_2 of Aa , and N_3 of aa ($N_1 + N_2 + N_3 = N$) from a finite sample of M individuals available for selection containing M_1 of AA , M_2 of Aa and M_3 of aa ($M_1 + M_2 + M_3 = M$). The formula given here can also be derived from the one given by him.

Values of N were chosen as the geometric progression (G.P.) from 1 to 1024 with a common ratio equal to 2; values of σ^2 , the phenotypic variance formed another G.P. from 0.25 to 2.0 with the same common ratio. Two levels of dominance (h) for the values of the metric were considered—(i) no dominance (additive model) where $\mu_1 = 1$, $\mu_2 = 0$, $\mu_3 = -1$ and (ii) dominance towards AA where $\mu_1 = 1$, $\mu_2 = 0.25$, $\mu_3 = -1$. The matrix (s_{ij}) was computed for all combinations of N , σ^2 and h .

It is evident that under the parental mating 1, $s_{11} = 1$, $s_{12} = 0$, $s_{13} = 0$ irrespective of N , σ^2 and h . Similar observations hold for the matings 3 and 6. The computed values of s_{ij} upheld these observations thus establishing the accuracy and adequacy of the numerical methods employed. As expected, $s_{21} = s_{52}$; $s_{22} = s_{53}$; $s_{22} = s_{51} = 0$ for every combination of N , σ^2 and h . The results for the matings $AA \times Aa$ and $Aa \times Aa$ are shown in Figs. 1 and 2. Since the results for the mating $Aa \times aa$ are a reflection of those for $AA \times Aa$, they are not shown separately in a figure.

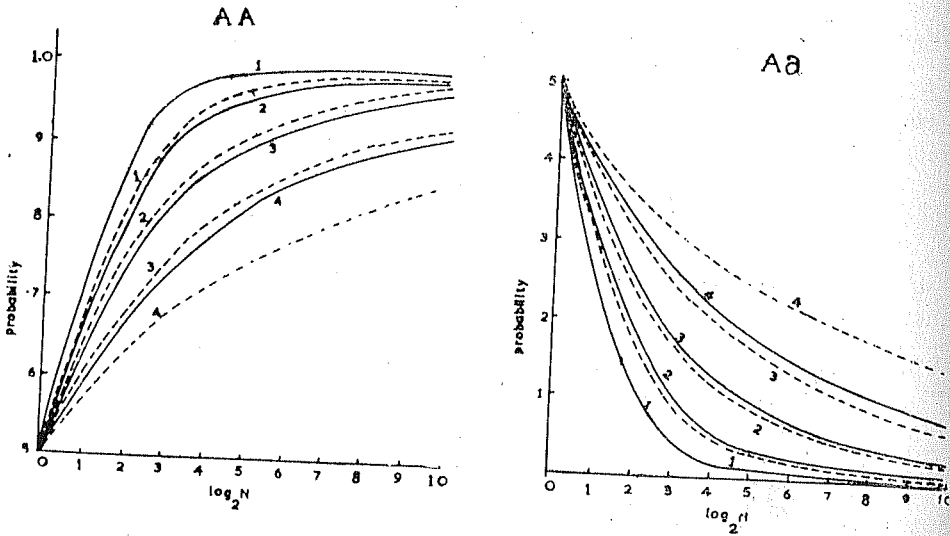


Fig. 1. Probabilities that a selected individual is of a given genotype-Mating AA X Aa

———— Additive Model $X_{AA} = 1; X_{Aa} = 0; X_{aa} = -1$
 - - - - - Dominance Model $X_{AA} = 1; X_{Aa} = .25; X_{aa} = -1$
 $1 - \sigma^2 = .25; 2 - \sigma^2 = .5; 3 - \sigma^2 = 1.0; 4 - \sigma^2 = 2.0$

Speed of Approach to Homozygosity: Let now the frequency that the parental mating is i be m_i before selection and m_i' after selection and mating. As already observed, the matrix $((s_{ij}))$ gives the probabilities of selecting genotype j given the parental mating i with $\sum_j s_{ij} = 1$. For convenient reference, the matrix is given in Table 1.

After selection of the best genotype as described earlier, the probabilities of mating types k , ($k = 1, 6$) given the parental mating i , ($i = 1, 6$) in the previous generation are given by the matrix $((c_{ki}))$ of the order 6×6 as set out in Table 2. It is then easy to see that

$$m_k' = \sum_k c_{ki} m_i \quad (3)$$

Table 1. The matrix (s_{ij})

Mating	1	2	3	4	5	6
AA	1	s_{21}	0	s_{41}	0	0
Aa	0	s_{22}	1	s_{42}	s_{52}	0
aa	0	0	0	s_{43}	s_{53}	1

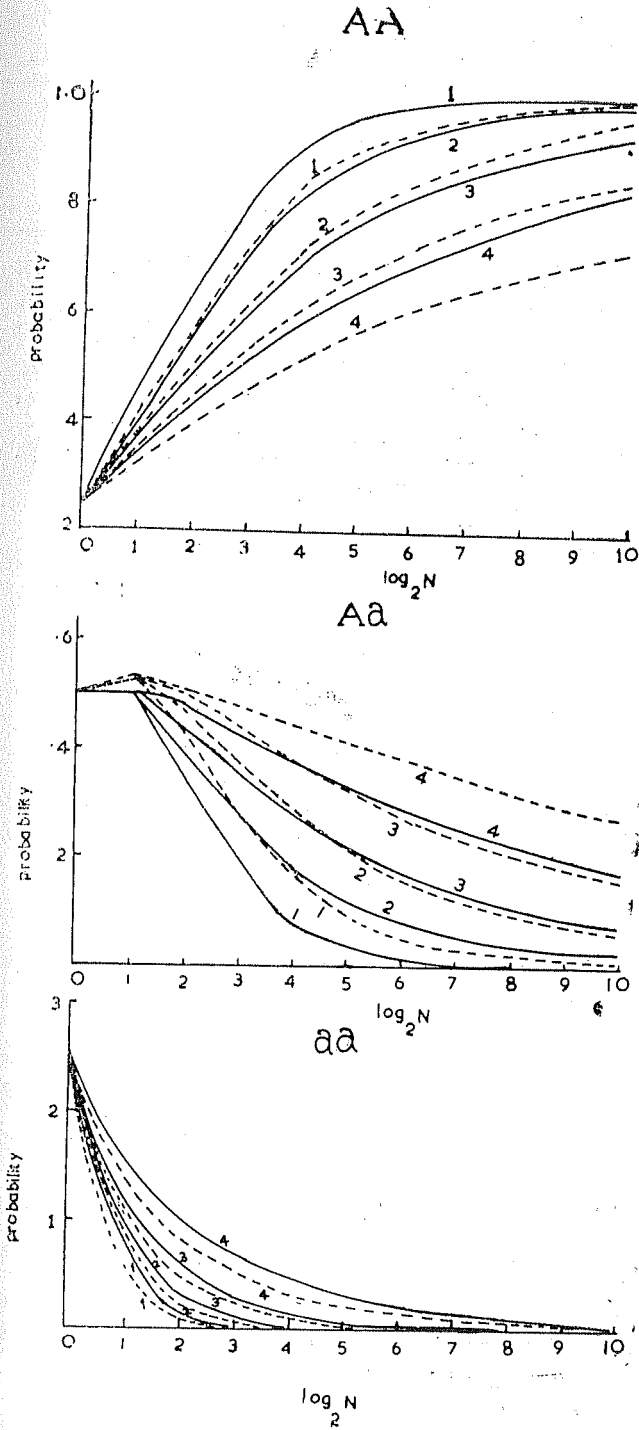


Fig. 2. Probabilities that a selected individual is of a given genotype - Mating $Aa \times Aa$

— Additive Model $X_{AA} = 1; X_{Aa} = 0; X_{aa} = -1$

..... Dominance Model $X_{AA} = 1; X_{Aa} = .25 X_{aa} = -1$

1 - $\sigma^2 = .25; 2 - \sigma^2 = .5; 3 - \sigma^2 = 1.0; 4 - \sigma^2 = .20;$

Table 2. The matrix (c_{ki})

$k \backslash i$	1	2	3	4	5	6
1	1	s_{21}^2	0	s_{41}^2	0	0
2	0	$2s_{21}s_{22}$	0	$2s_{21}s_{42}$	0	0
3	0	0	0	$2s_{41}s_{43}$	0	0
4	0	s_{22}^2	1	s_{42}^2	s_{52}^2	0
5	0	0	0	$2s_{42}s_{43}$	$2s_{52}s_{53}$	0
6	0	0	0	s_{43}^2	s_{53}^2	1

As shown by Fisher (1965) the process given by (3) is dominated by the largest eigenvalue λ less than unity of the matrix (c_{ki}) . An examination of the structure of this matrix shows that two of its eigenvalues are equal to unity. The speed of approach to homozygosity is then given by $S = -\log_e \lambda$ and an approximate estimate of the number of generations required for the frequency of gene A to reach 0.95 starting from a very low frequency in the population is given by $n_{0.95} = -\log_e (0.05)/S$. The speeds of approach to homozygosity computed for different levels of N, σ^2 and h are presented in Fig. 3 (i).

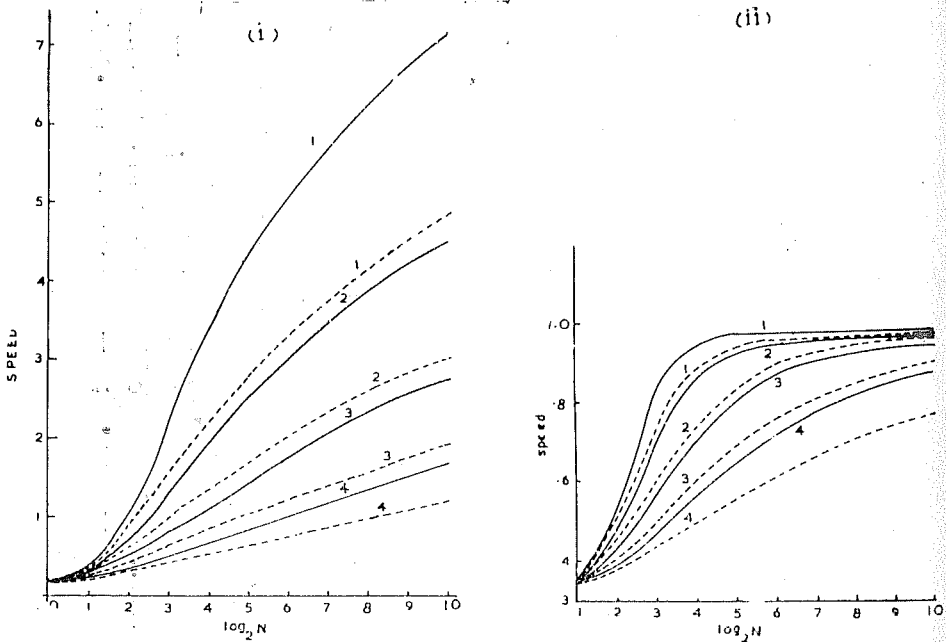


Fig. 3. Speed ($= -\text{Log}_e \lambda$ the dominant latent root) of Approach to Homozygosity by Selection

(i) Selection of one best

(ii) Selection of a best pair

———— Additive Model $x_{AA} = 1; x_{Aa} = 0; x_{aa} = -1$

..... Dominance Model $x_{AA} = 1; x_{Aa} = .25; x_{aa} = -1$

1 — $\sigma^2 = .25; 2 - \sigma^2 = .5; 3 - \sigma^2 = 1.0; 4 - \sigma^2 = 2.0$

MODEL 2

This model is essentially the same as model 1 except that the best pair of individuals is selected instead of the best individual in each sex on the basis of the phenotypic values of the metric. We do not impose the matings to be (best male \times best female) and the (next high male \times the next high female) only, and so the mating between the selected pair is at random. This random mating results in N offspring of each sex. The process is repeated until a desired level of homozygosity is attained in the population.

Let the progeny pair be specified by a number j as follows:

$$\begin{aligned} j=1: AA, AA (1, 1); j=2: AA, Aa (1, 2); j=3: AA, aa (1, 3); j=4: Aa, Aa (2, 2); \\ j=5: Aa, aa (2, 3); j=6: aa, aa (3, 3). \end{aligned} \quad (4)$$

The types of mating can equally be specified by j ($j = 1, 6$) for convenience.

We shall now find u_{ij} , the probability that the selected pair is j [$= (k, t)$, $k, t = 1, 3$, $k > t$] given that the parental mating is i ($i, j = 1, 6$). The probability that out of N progeny there will be $(N-2)$ individuals having a phenotypic score less than x_1 , an individual of genotype k , $k=1, 3$, between x_1 and x_2 and another of genotype t , $t=1, 3$, greater than x_2 is given by

$$\frac{N!}{(N-2)!} p_1^{N-2} p_2 p_3$$

where

$$p_1 = \sum_{r=1}^3 \alpha_{ir} F(y_1 - \mu_r)$$

$$p_2 = \alpha_{ik} [F(y_2 - \mu_k) - F(y_1 - \mu_k)]$$

and

$$p_3 = \alpha_{it} [1 - F(y_2 - \mu_t)]$$

where

$$y_i = x_i/\sigma \text{ as shown in (1).}$$

If now we consider the interval (x_1, x_2) as $(x, x + dx)$, the probability that the two top individuals are of genotypes k and t , given the parental mating i is

$$\begin{aligned} u_{ij} = N(N-1) \alpha_{ik} \alpha_{it} \int_{-\infty}^{\infty} \left[\sum_{r=1}^3 \alpha_{ir} F(y - \mu_r) \right]^{N-2} \\ \times [1 - F(y - \mu_t)] F(y - \mu_k) dy \end{aligned}$$

where j represents the pair whose genotypes are (k, t) .

Since the order of ranking in the selected pair is irrelevant, the probability of selecting the pair j given the parental mating i = Probability of selecting (k, t) + Probability of selecting (t, k) .

Hence,

$$u_{ij} = N(N-1) a_{ik} a_{it} \int_{-\infty}^{\infty} \left[\sum_r a_{ir} F(y - \mu_r) \right]^{N-2} \\ \times \{ [1 - F(y - \mu_k)] f(y - \mu_k) + \\ [1 - F(y - \mu_t)] f(y - \mu_t) \} dy. \quad (5)$$

The same method of numerical integration as in model 1 was employed to evaluate (5) (See Appendix 1).

Values of u_{ij} at different levels of N , σ^2 and h have thus been evaluated numerically. The results are parallel to those under model 1. Hence the results for $\sigma^2 = 0.25$ and $\sigma^2 = 2.0$ for the parental matings $AA \times Aa$ and $Aa \times Aa$ only are summarised in Tables 5 and 6.

Table 3. The 21 parental pair combinations

s	Genotypes of the parental pairs	Code number	s	Genotypes of the parental pairs	Code number
1	$AA, AA; AA, AA$	1; 1	12	$AA, aa; AA, aa$	3; 3
2	$AA, AA; AA, Aa$	1; 2	13	$AA, aa; Aa, Aa$	3; 4
3	$AA, AA; AA, aa$	1; 3	14	$AA, aa; Aa, aa$	3; 5
4	$AA, AA; Aa, Aa$	1; 4	15	$AA, aa; aa, aa$	3; 6
5	$AA, AA; Aa, aa$	1; 5	16	$Aa, Aa; Aa, Aa$	4; 4
6	$AA, AA; aa, aa$	1; 6	17	$Aa, Aa; Aa, aa$	4; 5
7	$AA, Aa; AA, Aa$	2; 2	18	$Aa, Aa; aa, aa$	4; 6
8	$AA, Aa; AA, aa$	2; 3	19	$Aa, aa; Aa, aa$	5; 5
9	$AA, Aa; Aa, Aa$	2; 4	20	$Aa, aa; aa, aa$	5; 6
10	$AA, Aa; Aa, aa$	2; 5	21	$aa, aa; aa, aa$	6; 6
11	$AA, Aa; aa, aa$	2; 6			

Speed of approach to Homozygosity.—The inbreeding process in this model is a bit complicated and can be specified by an equation similar to (3). The process involves random mating of the best male pair with the best female pair. We thus have 21 parental pair combinations as indicated in Table 3. We shall use in the following r and s as the suffixes to denote the parental and selected pairs, $r, s = 1, 21$.

Table 4. Probabilities that the selected best individual be of a given genotype under the parental mating $Aa \times Aa$ (Additive Model)

σ^2	N	Deterministic			Simulation			λ	x	p
		AA	Aa	aa	AA	Aa	aa			
0.25	1	.2500	.5000	.2500	.2333	.4817	.2850	.8090	14	50.0
	2	.4175	.5000	.0825	.3933	.5017	.1050	.6609	7	92.3
	4	.6219	.3627	.0154	.6200	.3717	.0083	.3305	2	80.7
	8	.8064	.1917	.0019	.8050	.1917	.0033	.1105	1	65.3
	16	.9203	.0795	.0002	.9133	.0867	0	.0344
	32	.9690	.0310	0	.9717	.0283	0	.0137
	64	.9866	.0134	0	.9817	.0183	0	.0066
	128	.9934	.0066	0	.9900	.0100	0	.0035
0.50	1	.2500	.5000	.2500	.2667	.4600	.2733	.8090	14	45.7
	2	.3950	.5000	.1050	.3817	.5017	.1166	.7070	8	89.3
	4	.5592	.4066	.0342	.4933	.4700	.0367	.4698	3	81.7
	8	.7085	.2819	.0096	.7483	.2483	.0034	.2620	2	90.3
	16	.8189	.1784	.0027	.8400	.1600	0	.1407	1	70.7
	32	.8890	.1102	.0008	.8750	.1250	0	.0779	1	76.7
	64	.9304	.0693	.0003	.9350	.0650	0	.0490
	128	.9548	.0451	.0001	.9583	.0417	0	.0219
1.00	1	.2500	.5000	.2500	.2750	.4817	.2433	.8090	14	50.0
	2	.3677	.5000	.1323	.3733	.5000	.1267	.7467	10	93.3
	4	.4915	.4440	.0645	.5217	.4217	.0566	.5981	5	92.3
	8	.6044	.3648	.0308	.6000	.3583	.0417	.4418	3	87.7
	16	.6971	.2878	.0151	.6683	.3150	.0167	.3177	2	84.7
	32	.7686	.2237	.0077	.7783	.2133	.0084	.2297	2	93.0
	64	.8219	.1739	.0042	.8133	.1817	.0050	.1694	1	66.0
	128	.8615	.1362	.0023	.8600	.1350	.0050	.1278	1	72.7
2.00	1	.2500	.5000	.2500	.2433	.5283	.2284	.8090	14	48.0
	2	.3405	.5000	.1595	.3583	.5117	.1300	.7741	11	86.7
	4	.4300	.4690	.1016	.4200	.4750	.1050	.6896	8	93.3
	8	.5116	.4236	.0648	.5167	.4150	.0683	.5907	5	89.0
	16	.5823	.3749	.0428	.5850	.3767	.0383	.4977	4	90.3
	32	.6422	.3289	.0289	.6567	.3083	.0350	.4176	3	88.7
	64	.6922	.2877	.0201	.6933	.2900	.0167	.3514	2	47.7*
	128	.7340	.2517	.0143	.7350	.2500	.0150	.2972	2	54.3*

λ — Greatest eigenvalue < 1 ; for x and p see text.

* — Slight underestimate.

Let m_s be the probability that the selected pairs are given by s ($s = 1, 2, 1$) and m'_s the corresponding probability in the next generation.

$$\begin{aligned}
 m'_s &= \sum_{r=1}^{21} Pr(\text{the selected pairs are given by } s | \text{the parental pairs are } r) \\
 &\quad \times Pr(\text{the parental pairs are } r) \\
 &= \sum_{r=1}^{21} c_{rs} m_r.
 \end{aligned}
 \tag{6}$$

For different values of r and s, the values of c_{rs} form a matrix ((C)) of the order 21×21 . The method of constructing the matrix ((C)) is detailed in Appendix II.

The speed with which the gene A tends to get fixed in the population obtained by random mating of the selected best pair can now be obtained as explained under model 1 as $-\log_e \lambda$ where λ is the largest latent root of the matrix (C) less than unity in absolute value. The speeds of approach to homozygosity for different values of N , σ^2 and h are shown graphically in Fig. 3 (ii).

Simulation of Models 1 and 2.—A computer programme was written to simulate the models 1 and 2 using Monte Carlo methods. The methods used were essentially those outlined by Fraser (1957 a, b). For each mating, N progeny of each sex were obtained and their genotypes identified. Their phenotypic values were obtained from the normal distribution with its mean at the genotypic value of the metric and its variance equal to σ^2 which was given. The genotypes were ranked on the basis of their phenotypic values and the top one or two individuals (as the case required) of each sex were selected to form the parents of the first generation. They were mated again to produce N progeny of each sex and the process repeated. The process was stopped at the generation when the frequency of gene A in the population containing N individuals of each sex reached 0.95 and the number of generations taken to reach this stage was recorded. This constituted one run of the experiment. Simulation of the process was done in each of the 300 replicate runs to the number of generations given by deterministic calculation, required for the frequency of gene A to reach 0.95. The computer also recorded those runs in which gene A did and did not reach the frequency of 0.95 within the specified number of generations. The number of times one best individual of a given genotype or one best pair of individuals of a given type (as defined earlier) were selected, was counted over 300 replicate runs. On this basis, the probability that the selected individual is of a particular type was calculated. This was done for both the sexes and averaged to give the required probabilities.

Table 4 summarises the probabilities that the selected individual is of a given genotype (denoted hereafter by PG for convenience) obtained by simulation, for the mating $Aa \times Aa$ and for the additive model. For purposes of comparison, the probabilities obtained by numerical integration and by simulation are given side by side. The dominant latent root of the process and on this basis, the approximate number of generations x , for gene A to reach 0.95 in the population are also shown. The last column shows the percentage of replicate runs in which gene A reached 0.95 in $(x + 1)$ generations or less.

RESULTS

Model 1.—The probabilities that the selected individual is AA , Aa or aa are plotted against $\log_2 N$ in Figs. 1 and 2 for the matings $AA \times Aa$ and $Aa \times Aa$ respectively. Under an additive model, the probability of selecting AA increases when the proportion selected is reduced. The probability is much higher (almost unity) when one best genotype is selected out of 1024 progeny than when one is selected out of 1 or 2 progenies per mating. The trend of this increase in the probabilities slows down when the variance of the distribution of phenotypes is increased

This is because the variance increases the spread of the distribution which, in turn, increases the area of selection, reducing the chances of AA alone getting selected. Thus we find a sharp increasing curve for $\sigma^2 = 0.25$ compared to that for $\sigma^2 = 2.0$. Incidentally we note that the $(p - \log_2 N)$ graph for Aa in Fig. 1 is the reflection of the graph for AA about the x -axis as either AA or Aa has got to be selected under the mating $AA \times Aa$.

In the dominance model we have considered, the dominance deviation h is equal to 0.25. Due to this dominance towards AA , the chance that Aa genotypes are selected is increased, thus bringing down the probability of selecting the AA genotype. The level of dominance chosen appears to be sufficient to modify the probabilities to be very near those under the next variance level with no dominance. Thus in general, the $(p - \log_2 N)$ graph for dominance with $\sigma^2 = 0.25$ is very near to the graph for no dominance with $\sigma^2 = 0.5$ and so on. An interesting pattern of modification by dominance is revealed in Figs. 1 and 2. Under the mating $AA \times Aa$, the $(p - \log_2 N)$ graph for $\sigma^2 = 0.25$ with dominance is very near but above the graph for $\sigma^2 = 0.5$ with no dominance, and so on for the genotype AA . In the case of Aa , the situation is reversed and the $(p - \log_2 N)$ graph for $\sigma^2 = 0.25$ with dominance is very near but below the graph for $\sigma^2 = 0.5$ with no dominance, and so on. Similar results hold for the genotypes AA and Aa under the mating $Aa \times Aa$. An intermediate situation is revealed in the case of selection of Aa (Figure 2). For example, the graph for $\sigma^2 = 0.25$ with dominance shows that the rate of decrease of probability is slower than for $\sigma^2 = 0.5$ with no dominance when the proportion of selected individuals is approximately greater than or equal to $\frac{1}{8}$ and is faster when the proportion is below $\frac{1}{8}$. Similar observations hold for the other cases. There is thus an optimum proportion of selection (which is between $1/8$ and $1/32$ in the case considered here) below which dominance with $\sigma^2 = 0.25$ favours more the selection of AA genotypes than additivity with $\sigma^2 = 0.5$, under the mating of a heterozygote \times heterozygote. This is true for the other levels of variance as well.

Figure 3 (i) shows the speeds of approach to homozygosity under this model. We see that the speed of approach is much higher when $\sigma^2 = 0.25$ than when $\sigma^2 = 2.0$ as expected. Dominance modification of the speed of approach are in line with the observations made earlier.

Model 2.—Tables 5 and 6 summarise the probabilities of selecting one of the possible six pairs under the mating $AA \times Aa$ and $Aa \times Aa$ respectively. The general trends of variation in the $(p - \log_2 N)$ graphs are similar to those for model 1 except that six possibilities of selecting a pair of genotypes exist in this model corresponding to the three possibilities of selecting one genotype in Model 1. The modifications due to dominance are also very similar to those in Model 1. As one expects, the speed of approach to homozygosity by selecting a pair of individuals is in almost all cases lower than the speed by selecting one best (Fig. 3). The speeds of approach to homozygosity under these models and the relative speed under Model 1 as compared to Model 2 are presented in Table 7.

Table 5. *Probabilities of a selected pair of individuals being of a given type*
 Parental Mating : AA × Aa

σ^2	Selected pair	Additive or dominance	N									
			2	4	8	16	32	64	128	256	512	1024
0.25	1	A	.250	.591	.843	.946	.979	.990	.995	.997	.998	.999
		D	.250	.522	.734	.858	.920	.952	.970	.980	.986	.990
	2	A	.500	.333	.145	.052	.021	.010	.005	.003	.002	.001
		D	.500	.385	.239	.136	.078	.047	.030	.020	.014	.010
	4	A	.250	.076	.012	.002	0	0	0	0	0	0
		D	.250	.092	.027	.006	.002	.001	0	0	0	0
2.00	1	A	.250	.379	.482	.565	.633	.688	.732	.769	.799	.825
		D	.250	.346	.420	.483	.536	.582	.622	.655	.686	.713
	2	A	.500	.468	.421	.371	.324	.282	.246	.216	.189	.167
		D	.500	.481	.453	.422	.391	.361	.333	.308	.284	.263
	4	A	.250	.153	.097	.064	.043	.030	.022	.015	.012	.008
		D	.250	.173	.127	.095	.073	.057	.045	.037	.030	.024

Table 6. *Probabilities of a selected pair or individuals being of a given type*
 Parental Mating : Aa × Aa

σ^2	Selected pair	Additive or dominance	N									
			2	4	8	16	32	64	128	256	512	1024
0.25	1	A	.063	.226	.493	.752	.900	.959	.981	.990	.994	.997
		D	.063	.201	.403	.607	.760	.855	.910	.942	.962	.973
	2	A	.250	.414	.374	.217	.095	.041	.019	.010	.006	.003
		D	.250	.433	.433	.332	.220	.138	.088	.057	.038	.027
	3	A	.125	.033	.005	.001	0	0	0	0	0	0
		D	.125	.024	.002	0	0	0	0	0	0	0
	4	A	.250	.265	.122	.029	.005	0	0	0	0	0
		D	.250	.292	.159	.061	.020	.007	.002	.001	0	0
	5	A	.250	.057	.006	.001	0	0	0	0	0	0
		D	.250	.045	.003	0	0	0	0	0	0	0
	6	A	.062	.005	0	0	0	0	0	0	0	0
		D	.062	.005	0	0	0	0	0	0	0	0
2.00	1	A	.063	.132	.209	.288	.365	.436	.500	.557	.608	.651
		D	.063	.121	.179	.235	.288	.338	.384	.427	.466	.502
	2	A	.250	.354	.411	.433	.430	.413	.387	.359	.329	.300
		D	.250	.361	.422	.452	.463	.461	.452	.439	.422	.404
	3	A	.125	.097	.075	.057	.043	.033	.025	.019	.014	.011
		D	.125	.085	.060	.043	.031	.023	.017	.013	.010	.008
	4	A	.250	.250	.214	.171	.132	.101	.078	.059	.045	.035
		D	.250	.277	.256	.223	.190	.160	.135	.114	.097	.082
	5	A	.250	.145	.083	.047	.028	.017	.010	.006	.004	.003
		D	.250	.138	.077	.044	.027	.018	.012	.007	.005	.004
	6	A	.062	.022	.008	.004	.002	0	0	0	0	0
		D	.062	.018	.006	.003	.001	0	0	0	0	0

Table 7. *Relative speed of approach to homozygosity by selection of one or two best*

N		$\sigma^2 = .25$		$\sigma^2 = .5$		$\sigma^2 = 1.0$		$\sigma^2 = 2.0$	
		A	D	A	D	A	D	A	D
2	S1	.414	.381	.347	.319	.292	.273	.256	.244
	S2	.349	.349	.349	.349	.349	.349	.349	.349
	RS	1.19	1.09	0.99	0.91	0.84	0.78	0.73	0.70
4	S1	1.107	.914	.755	.626	.514	.440	.372	.331
	S2	.535	.498	.476	.447	.426	.407	.392	.381
	RS	2.07	1.84	1.59	1.40	1.21	1.08	0.95	0.87
8	S1	2.203	1.590	1.339	1.002	.817	.648	.526	.442
	S2	.826	.737	.704	.613	.573	.510	.475	.439
	RS	2.67	2.16	1.91	1.63	1.43	1.27	1.11	1.01
16	S1	3.369	2.239	1.961	1.371	1.147	.860	.698	.561
	S2	.945	.880	.859	.751	.712	.611	.566	.502
	RS	3.57	2.54	2.28	1.83	1.61	1.41	1.23	1.12
32	S1	4.292	2.813	2.527	1.713	1.471	1.064	.873	.678
	S2	.971	.938	.925	.841	.811	.696	.651	.562
	RS	4.42	3.00	2.73	2.04	1.81	1.53	1.34	1.21
64	S1	5.021	3.314	3.015	2.026	1.776	1.255	1.046	.792
	S2	.977	.960	.952	.893	.871	.763	.722	.616
	RS	5.14	3.45	3.17	2.27	2.04	1.64	1.45	1.29
128	S1	5.644	3.758	3.442	2.313	2.058	1.436	1.213	.902
	S2	.979	.969	.964	.922	.907	.814	.778	.664
	RS	5.77	3.88	3.57	2.51	2.27	1.76	1.56	1.36
256	S1	6.197	4.159	3.823	2.578	2.318	1.605	1.374	1.007
	S2	.980	.973	.970	.940	.927	.850	.821	.706
	RS	6.32	4.27	3.94	2.74	2.50	1.89	1.67	1.43
512	S1	6.699	4.526	4.171	2.824	2.559	1.765	1.528	1.108
	S2	.980	.976	.974	.951	.943	.877	.853	.743
	RS	6.85	4.64	4.28	2.97	2.71	2.01	1.79	1.49
1024	S1	7.162	4.867	4.493	3.055	2.784	1.917	1.676	1.204
	S2	.981	.978	.976	.959	.952	.897	.877	.774
	RS	7.30	4.98	4.60	3.19	2.92	2.14	1.91	1.56

S1—Speed under Model 1. S2—Speed under Model 2. RS = S1/S2, the relative speed.

Simulation.—The probabilities PG obtained by simulation of Model 1 are close to those obtained by numerical integration (Table 4). Allowing a fixed standard error of one generation, the percentage of runs p in which the gene A reaches a value of 0.95 in $(x + 1)$ generations or less is calculated and is shown in the last column. An examination of these p values indicate that a good degree of accuracy is obtained by simulation when $N > 1$ and $N < 32$. One of the ways to achieve close accuracy at $N = 1$ is to increase the number of replicate runs, and this will tend to even out the deficiencies of the pseudo-random number generators generally used for simulation. Since the estimates of the probabilities were close to the estimates obtained by numerical integration (the difference being less than or of the order of 10% in most of the cases) when the replicate runs were 300, no attempt to increase it beyond 300 is made. Though the results of simulation are shown only for the mating

Table 8. *The matrix (β_{sj}) $s=1, 21, j=1, 6$*

s/j	1	2	3	4	5	6
1	1
2	9/16	3/8	..	1/16
3	1/4	1/2	..	1/4
4	1/4	1/2	..	1/4
5	1/16	3/8	..	9/16
6	1
7	81/256	27/64	9/128	9/64	3/64	1/256
8	9/64	3/8	3/32	1/4	1/8	1/64
9	9/64	3/8	3/32	1/4	1/8	1/64
10	9/256	15/64	9/128	25/64	15/64	9/256
11	9/16	3/8	1/16
12	1/16	1/4	1/8	1/4	1/4	1/16
13	1/16	1/4	1/8	1/4	1/4	1/16
14	1/64	1/8	3/32	1/4	3/8	9/64
15	1/4	1/2	1/4
16	1/16	1/4	1/8	1/4	1/4	1/16
17	1/64	1/8	3/32	1/4	3/8	9/64
18	1/4	1/2	1/4
19	1/256	3/64	9/128	9/64	27/64	81/256
20	1/16	3/8	9/16
21	1

$Aa \times Aa$, a few sample estimates were made for the other matings which have shown parallel results. However, not less than 600 runs were found necessary to obtain results as good as those from numerical integration for Model 2.

DISCUSSION

This study has brought out some salient features of simple selection models which promote a certain amount of inbreeding in the population. The main aim of this study is to formulate theoretically the probability PG of obtaining a given genotype by selecting on the basis of measurements of a quantitative character and to use this in deriving a measure for the speed of approach to homozygosity. This has a certain bearing on the artificial selection experiments conducted on living organisms or crop plants, since it is usual to select the top few individuals and to use them as parents of the next generation. It is admitted that the models studied here are simplified though the theoretical formulation of the process is not quite so.

The probability PG is found to depend on $1/N$, the proportion selected ($2/N$ in Model 2), and on the variance of the distribution of the phenotypes, since they are the determinants of these models. A close examination of the probabilities suggest that the probability can be expressed as an algebraic function of σ^2 , $1/N$ and

of the value of the metric for the particular genotype. Initial attempts were not successful. In the literature, we do find approximations relating the effect of a gene on a quantitative character with the consequent selective advantage caused by artificial selection (see Falconer, 1964, Chapter 11). An approximation for i , the selection intensity was also given as $i = 0.8 + 0.4 \ln(N-1)$ (Smith, 1969). But the approximation desired here is for the probability that a selected genotype be a specified one say AA (or that the selected pair of genotypes be specified ones AA, Aa say). The approximation, if found, will be theoretically advantageous in order to calculate easily the probabilities that a number k ($k > 2$) of selected individuals based on their phenotypic scores be of desired genotypes and hence to obtain the rate of approach to homozygosity.

The speed of approach to homozygosity slows down considerably when a best pair of individuals is selected instead of the best single individual (Table 7). When each mating produces only 2 or 4 individuals, Model 2 appears to be better than Model 1 above certain levels of the variance of the phenotypic distribution. This situation is, however, unlikely in practice; otherwise Model 1 promotes homozygosity more quickly than Model 2 as expected. It is again interesting to observe that dominance (with $h = 0.25$) modifies the speed of approach such that the relative speed when $\sigma^2 = 0.25$ with dominance approximately equals that when $\sigma^2 = 0.5$ with no dominance for all N , and so on. As observed earlier, such a pattern was evident in the probabilities PG as well.

Reeve (1955) has discussed the effects of inbreeding when both the homozygotes are at a selective value s in a single locus model. He has considered the speed of approach to homozygosity in selfing, half-sib and double first cousin matings. Within-line selection was considered and selection was through differential survival of the homozygotes and heterozygotes, closely approximating to the process of natural selection. Though we have considered here only the effects of artificial selection, the results suggest that a given level of homozygosity is obtained much faster in the models considered in this paper than in the matings considered by Reeve. This is so because the selection process discussed here is very intensive.

Projecting the results of this study, it seems desirable to say, in a plant breeding programme (i) to select from a population which has already approached a desired level of yield, a small number of plants showing superior individual performance, and (ii) to obtain from it, a progeny population fixed for all the components of yield. The results of Robertson (1970) that greater selection advance can be obtained by varying the proportion selected over generations, the proportion selected being made small as selection proceeds, are of direct relevance.

This study also reveals the necessity of exercising caution in interpreting the results by simulation of any artificial selection experiment. All simulation experiments depend ultimately to a large extent on the random number generators that are used. Although very efficient random number generators are available, they are nevertheless pseudo-random. This makes it all the more necessary to have

as many replicate runs as possible. There seems to be no yardstick to measure how many such runs are an absolute necessity before one may rely on the results, though it can be agreed that the validity of the simulation is checked on a 'null' or a 'trial' run of the model with known parameters and known results. In this study, simulation of simple selection models whose results can, of course, be obtained exactly by analytical methods, was done because it would be then be possible to find out the number of replicate runs that are needed to match the results obtained by analytical methods. Even in a simple situation as we had, 300 replicate runs or more were needed to get close estimates.

In complicated experiments, the problem of ascertaining the number of replicate runs needed to obtain reliable results assumes crucial importance.

These points are not made to argue that the simulation of genetic models in general, should be dispensed with, but to stress the need to handle the problems of simulation with caution. When possible, it is fruitful to solve the model by mathematical or other means to the extent possible and use those results to further simulate the model. However, when problems arise for which simulation is the only solution, it is profitable to base the results on a large number of replicate runs and to interpret them with care and caution.

ACKNOWLEDGEMENT

I am grateful to Dr. A. R. G. Owen, formerly of the Department of Genetics, University of Cambridge, where this work was done for his keen interest in this study and for many illuminating discussions on this subject.

SUMMARY

An attempt is made to study the progress of inbreeding in a population by simple models of artificial selection. One or two top ranking individuals are selected based on the phenotypic value of a metric from the N progenies resulting from any particular mating (denoted as Model 1 and Model 2). The generation matrix method is employed to study the progress towards homozygosity. The probability that a selected individual is of a given genotype or that a selected pair consists of specified genotypes has been computed by numerical integration as well as by simulation, assuming the phenotypic distribution to be normal with the mean at the genotypic value and a given variance. It is observed that Model 1 promotes homozygosity more quickly than Model 2 as expected. Modifications due to dominance in the progress towards homozygosity are also discussed. The study throws some light on simulation studies in genetics.

REFERENCES

- FALCONER, D. S. (1964). *Introduction to Quantitative Genetics*. Oliver and Boyd, Edinburgh and London.
- FISHER, R. A. (1965). *The Theory of Inbreeding*. Oliver and Boyd, Edinburgh, Second Edition.
- FRASER, A. S. (1957a). Simulation of genetic systems on automatic digital computers. I. Introduction. *Aust. J. Biol. Sci.*, **10**, 484-491.

- FRASER, A. S. (1957b). Simulation of genetic systems on automatic digital computers. II. Effects of linkage on rates of advance under selection. *Aust. J. Biol. Sci.*, **10**, 492-499.
- GOODWIN, E. T. (1949). The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(x) e^{-x^2} dx$. *Proc. Camb. Phil. Soc.*, **45**, 241-245.
- HALDANE, J. B. S. (1931). A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. *Proc. Camb. Phil. Soc.*, **27**, 131-136.
- HASTINGS, C. (1955). *Approximations for Digital Computers*. Princeton University Press, London: Geoffrey Cumberlege.
- HILL, W. G. (1969). On the theory of artificial selection in finite populations. *Genet. Res.*, **13**, 143-163.
- KIMURA, M. (1957). Some problems of stochastic processes in genetics. *Ann. Math. Stat.*, **28**, 882-901.
- KIMURA, M. (1958). On the change of population fitness by natural selection. *Heredity*, **12**, 145-167.
- KOJIMA, K. (1961). Effects of dominance and size of population on response to mass selection. *Genet. Res.*, **2**, 177-188.
- PIKE, D. J. (1969). A comparison of two methods for predicting changes in the distribution of gene frequency when selection is applied repeatedly to a finite population. *Genet. Res.* **13**, 117-26.
- REEVE, E. C. R. (1955). Inbreeding with the homozygotes at a disadvantage. *Ann. Hum. Genet.*, **19**, 332-346.
- ROBERTSON, A. (1960). A theory of limits in artificial selection. *Proc. Royal Soc.*, **153 B**, 234-249.
- ROBERTSON, A. (1970). Some optimum problems in individual selection. *Theoret. Population Biol.*, **1**, 120-127.
- SMITH, C. (1969). Optimum selection procedures in animal breeding. *Anim. Prod.*, **11**, 433-442.

APPENDIX I

Numerical method of calculating s_{ij} given in (2) of the text

Let

$$\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Hastings (1955) has given a valid numerical approximation to $\text{Erf}(x)$. It is

$$\text{Erf}(x) = 1 - [1/(1 + a_1x + a_2x^2 + \dots + a_6x^6)^{16}]$$

where

$$\begin{aligned} a_1 &= \cdot 0705230784 & a_2 &= \cdot 0422820123 & a_3 &= \cdot 0092705272 \\ a_4 &= \cdot 0001520143 & a_5 &= \cdot 0002765672 & a_6 &= \cdot 0000430638, \end{aligned}$$

when $0 \leq x < \infty$. Noting that

$$\text{Erf}(x) = -\text{Erf}(y)$$

when $x = -y$, we can show that

$$F(X - \mu_r) = \frac{1}{2} + \frac{1}{2} \text{Erf} \left(\frac{X - \mu_r}{\sqrt{2}} \right) = \beta_r \text{ (say),}$$

where

$$X = x/\sigma.$$

Hence

$$\begin{aligned} s_{ij} &= \int_{-\infty}^{\infty} N a_{ij} f(X - \mu_j) \left(\sum_{r=1}^3 a_{ir} \beta_r \right)^{N-1} dX \\ &= (1/\sqrt{2\pi}) N a_{ij} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(X-\mu_j)^2} \left(\sum_r a_{ir} \beta_r \right)^{N-1} dX \end{aligned}$$

Let

$$t = \frac{X - \mu_j}{\sqrt{2}} = \frac{x}{\sigma} - \frac{\mu_j}{\sqrt{2}},$$

i.e.,

$$x = (t\sqrt{2} + \mu_j)\sigma$$

so that

$$\frac{X - \mu_r}{\sqrt{2}} = t + \frac{\mu_j - \mu_r}{\sqrt{2}} = \eta_r \text{ (say).}$$

Hence

$$\beta_r = \frac{1}{2} + \frac{1}{2} \text{Erf}(\eta_r)$$

so that

$$s_{ij} = (1/\sqrt{\pi}) N a_{ij} \int_{-\infty}^{\infty} e^{-t^2} \left(\sum_r a_{ir} \beta_r \right)^{N-1} dt$$

which is now of the form $\int_{-\infty}^{\infty} e^{-t^2} q(t) dt$ where

$$q(t) = (1/\sqrt{\pi}) N a_{ij} \left(\sum_r a_{ir} \beta_r \right)^{N-1}$$

Goodwin (1949) has given an approximation to integrals of this form. Using his approximation,

$$s_{ij} \approx r \sum_{k=-\infty}^{\infty} q(kr) e^{-k^2 r^2} \text{ where } r \text{ is the step length. This gives a very}$$

good approximation to the integral so long as r is kept reasonably small. In this investigation, r was taken to be $= 0.2$ and the summation was carried between $k = -30$ and $k = 30$. These values of k and r were found to be adequate for s_{ij} 's to be accurate to 4 places of decimal.

Numerical method of calculating u_{ij} given in (5) of the text

Rewriting (5) we get,

$$u_{ij} = N(N-1) a_{ik} a_{it} (A + B)$$

where

$$A = \int_{-\infty}^{\infty} \left[\sum_r a_{ir} F(y - \mu_r) \right]^{N-2} [1 - F(y - \mu_t)] f(y - \mu_k) dy$$

and

$$B = \int_{-\infty}^{\infty} \left[\sum_r a_{ir} F(y - \mu_r) \right]^{N-2} [1 - F(y - \mu_k)] f(y - \mu_t) dy.$$

Splitting up A into two integrals and following the procedure outlined above, it is easy to see that

$$A = \int_{-\infty}^{\infty} e^{-z^2} \left[\sum_r a_{ir} \left\{ \frac{1}{2} + \frac{1}{2} \operatorname{Erf} \left(z + \frac{\mu_k - \mu_r}{\sqrt{2}} \right) \right\} \right]^{N-2} \times \left[\frac{1}{2} - \frac{1}{2} \operatorname{Erf} \left(z + \frac{\mu_k - \mu_t}{\sqrt{2}} \right) \right] dz,$$

where

$$x = \sigma(z\sqrt{2} + \mu_j).$$

We can evaluate B in a similar manner and hence u_{ij} .

APPENDIX II

Method of calculation of matrix ((C))

Let i, j denote the order of pairs as mentioned in (4). Let u_{ij} be the probability that the pair j is selected from the parental mating i ($i, j = 1, 6$). The matrix U of the order 6×6 formed by u_{ij} is given below:

$i \backslash j$	1	2	3	4	5	6
1	1	u_{21}	0	u_{41}	0	0
2	0	u_{22}	0	u_{42}	0	0
3	0	0	0	u_{43}	0	0
4	0	u_{24}	1	u_{44}	$u_{54} = u_{21}$	0
5	0	0	0	u_{54}	$u_{55} = u_{22}$	0
6	0	0	0	u_{64}	$u_{56} = u_{23}$	1

Let α_{sh} be the conditional probability that a progeny is of the genotype h ($h = 1$ for AA ; $= 2$ for Aa ; $= 3$ for aa) given the parental pairs s .

The process of random mating among the parental pairs can completely be specified by the random union of the gametes whose frequencies in the parental pairs are as follows:

Parental pair, i	Gametic Frequency	
	A	a
1- AA, AA	1	0
2- AA, Aa	$\frac{3}{4}$	$\frac{1}{4}$
3- AA, aa	$\frac{1}{2}$	$\frac{1}{2}$
4- Aa, Aa	$\frac{1}{2}$	$\frac{1}{2}$
5- Aa, aa	$\frac{1}{4}$	$\frac{3}{4}$
6- aa, aa	0	1

The distribution of α_{sh} can now be completely specified. For example, for the parental pairs 9, i.e., for the pair AA, Aa ; Aa, Aa (see Table 3), $\alpha_{91} = \frac{3}{8}$, $\alpha_{92} = \frac{1}{2}$ and $\alpha_{93} = \frac{1}{8}$. Thus the matrix of values for α_{sh} of the order 21×3 can be written down. From this matrix, it is possible to calculate β_{sj} , the probability of a (progeny) mating of type j given the parental pairs s , $s = 1, 21$; $j = 1, 6$. For the above example, we would have $\beta_{91} = 9/64$, $\beta_{92} = 3/8$, $\beta_{93} = 3/32$, $\beta_{94} = 1/4$, $\beta_{95} = 1/8$ and $\beta_{96} = 1/64$. The matrix of values β_{sj} is presented in Table 8.

From the u_{ij} 's, we can now get v_{jr} , the probability of two pairs given by $r (=k, t \text{ say})$ getting selected given the parental pair $j, r=1, 21; j, k, t=1, 6$.

$$\begin{aligned} v_{jr} &= 2u_{jk}u_{jt} \quad \text{if } k \neq t \\ &= u_{jk}^2 \quad \text{if } k = t. \end{aligned}$$

Thus it follows that

$$\begin{aligned} c_{rs} &= Pr(\text{the two selected pair is } s | \text{parental pair is } r) \\ &= \sum_{j=1}^6 Pr(\text{the two selected pair is } s | \text{mating } i) \times \\ &\quad Pr(\text{mating } j | \text{Parental pair } r). \\ &= \sum_{j=1}^6 v_{js} \beta_{rj}. \end{aligned}$$

The matrix (C) is thus determined.