

THE CONFLICT BETWEEN INBREEDING AND SELECTION

I. SELF-FERTILIZATION

By J. B. S. HALDANE

Department of Biometry, University College, London

(With One Text-figure)

On the whole, inbred populations become more homozygous in successive generations. But occasionally heterozygosis is preserved, either at a locus or for a pair of chromosomal orders, permanently or for longer than would be expected unless heterozygotes were at a great advantage. Hollingsworth & Maynard Smith (1955) give an example.

Hayman & Mather (1953) have discussed this question. However, their principal results appear to me to be incorrect, and they have only considered a very few of the ways in which selection may act. It is therefore desirable to deal with the matter afresh.

In any system of complete inbreeding, as opposed to inbreeding varied by an occasional outcross, the population is divided into a number of lines. In each generation a line is represented by one or more individuals, one in the case of self-fertilization, two in that of sib-mating, and so on. It is convenient to consider a population consisting of a fairly large and constant number of lines. In fact this number is often far from constant in practice. Thus Hollingsworth & Maynard Smith inform me that in their O line the number of pairs set up per generation between F_2 and F_{17} varied irregularly from 13 to 55, with mean 24.6.

It is theoretically possible to keep the number of lines constant, breeding from only one individual, pair, trio, etc., of each line in each generation, provided the depression due to inbreeding is not too severe. In this case there is no selection between lines. More usually some lines die out owing to inbreeding depression, or are discarded either because they show signs of weakness or sterility, or because they do not conform to some desired standard (Fig. 1). In this case there is voluntary or involuntary selection between lines.

In this series of papers I shall mainly consider selection acting on a single pair of alleles at one locus, or on a pair of alternative chromosome orders. Either may be denoted by A and a . It is not assumed that a is recessive. Selection is supposed to favour Aa at the expense of one or both homozygotes, but this selection is assumed not to be complete; as it can be when heterozygotes can be picked out with certainty, or in the case of balanced lethals. I shall also neglect mutation, which I have considered earlier (Haldane, 1936).

It follows that in every line there is in each generation a finite chance that the next generation will consist entirely of like homozygotes. This is not necessarily so if numerous alleles or chromosome orders are present in a population. For example, the mating $a^1a^2 \times a^3a^4$ cannot give rise to homozygotes. But a line consisting of like homozygotes will always give rise to like homozygotes. Hence no amount of selection within lines can prevent the population from ultimately consisting of homozygotes in the case of self-fertilization, and of sets of like homozygotes in the more complicated cases. It may of course slow this process down considerably.

If, however, there is selection between lines, there are two possibilities. The homozygous, or like homozygous, lines may die out or be discarded so rapidly, and the heterozygous, or partly heterozygous, lines expanded so rapidly, that heterozygosis never disappears. Or the heterozygotes may ultimately disappear, as with selection within lines. There will be a critical value of selection for which heterozygotes just disappear, as Hayman & Mather pointed out.

However, this critical value is a little deceptive. Supposing that we were keeping twenty-five lines in each generation, some being discarded, and others split into two or more, and that calculation showed that in a large number, N , of lines $0.08N$ would be expected to remain wholly or partly heterozygous. We should expect two out of twenty-five to remain so. But if so the probability of finding no heterozygous or partly hetero-

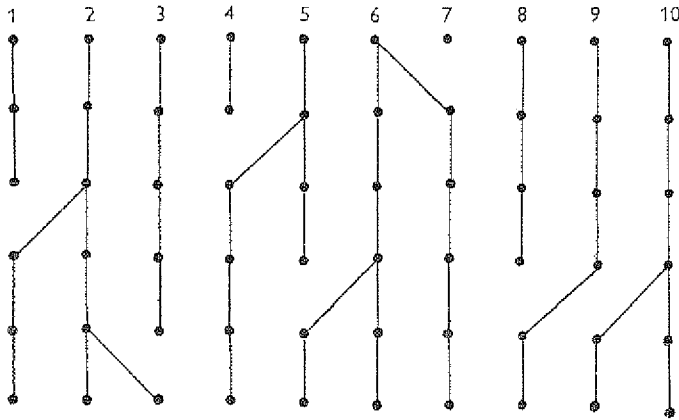


Fig. 1. Hypothetical diagram of selection between lines. Ten selfed individuals or mated sib pairs are grown in each generation, represented by the points. But lines 1, 3, 4, 7 and 8 are extinguished, while line 10 is finally represented by two lines, and lines 2 and 5 by three lines each.

zygous lines would be $(0.92)^{25}$, or 0.12. Thus in the course of time heterozygotes would be bound to disappear. In fact, therefore, homozygosis is ultimately achieved with intensities of selection well below the critical values calculated. I shall neglect this complication in what follows. In consequence the results are only valid in the case of selection between lines if the number of lines is pretty large.

The present paper deals entirely with self-fertilization. I suppose that, in the n th generation, the breeding population consists of: $Nx_n AA$, $Ny_n Aa$, $Nz_n aa$, with $x_n + y_n + z_n = 1$, where N is large, but not necessarily constant. As measures of selection I use k and l , which correspond with Hayman & Mather's parameters $1 - x$ and $1 - y$.

SELECTION WITHIN LINES ONLY

Suppose that in a progeny from Aa selfed, AA and aa are at disadvantages given by k and l . Then the probabilities that AA , Aa and aa will be the genotype of the next parent in a heterozygous line are not $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$, but

$$\frac{1-k}{4-k-l}, \quad \frac{2}{4-k-l} \quad \text{and} \quad \frac{1-l}{4-k-l}.$$

On the other hand, AA and aa lines are at no disadvantage. k and l are assumed to be positive, but cannot exceed unity. Then

$$\left. \begin{aligned} x_{n+1} &= x_n + \frac{(1-k)y_n}{4-k-l}, \\ y_{n+1} &= \frac{2y_n}{4-k-l}, \\ z_{n+1} &= z_n + \frac{(1-l)y_n}{4-k-l}. \end{aligned} \right\} \quad (1)$$

Clearly $y_n = \left(\frac{2}{4-k-l}\right)^n y_0$. So since it is assumed that both k and l are not unity, that is to say, that homozygotes of at least one sort have a chance, heterozygotes ultimately disappear. Also

$$\begin{aligned} x_n &= x_0 + \frac{1-k}{4-k-l} (y_0 + y_1 + y_2 + \dots + y_{n-1}) \\ &= x_0 + \frac{1-k}{2-k-l} \left[1 - \left(\frac{2}{4-k-l}\right)^n \right] y_0. \end{aligned}$$

The population therefore tends to an equilibrium given by

$$\left. \begin{aligned} X &= x_0 + \frac{(1-k)y_0}{2-k-l}, \\ Y &= 0, \\ Z &= z_0 + \frac{(1-l)y_0}{2-k-l}. \end{aligned} \right\} \quad (2)$$

And in the n th generation

$$\left. \begin{aligned} x_n &= X - \frac{1-k}{2-k-l} \left(\frac{2}{4-k-l}\right)^n y_0, \\ y_n &= \left(\frac{2}{4-k-l}\right)^n y_0, \\ z_n &= Z - \frac{1-l}{2-k-l} \left(\frac{2}{4-k-l}\right)^n y_0. \end{aligned} \right\} \quad (3)$$

SELECTION BETWEEN LINES ONLY

I next suppose that there is no selection within each progeny of a self-fertilized heterozygote, but that there is selection between lines. This could occur, for example, if A and a had no effect on viability, but the Aa plants were, on an average, somewhat superior to AA and aa as regards an economically important character. I suppose that an AA line has only $1-k$ times the chance of a segregating line of being perpetuated, and an aa line $1-l$ times the chance. Some homozygous lines are discarded in each generation, and some heterozygous lines split. This is what happens in the early stages of wheat breeding, though there is usually selection within lines also. In fact k and l probably tend to

increase with time, since as other loci become homozygous, it is easier to pick out Aa lines. I neglect this complication. We find

$$\left. \begin{aligned} x_{n+1} &= \frac{(1-k)x_n + \frac{1}{2}y_n}{1-kx_n - lz_n}, \\ y_{n+1} &= \frac{\frac{1}{2}y_n}{1-kx_n - lz_n}, \\ z_{n+1} &= \frac{(1-l)z_n + \frac{1}{2}y_n}{1-kx_n - lz_n}. \end{aligned} \right\} \quad (4)$$

Now suppose that at equilibrium the frequencies are X , Y and Z . Then if $Y \neq 0$, $1-kX-lZ = \frac{1}{2}$, from the second of equations (4), or $kX+lZ = \frac{1}{2}$. Hence the first of equations (4) becomes $X = 2(1-k)X + \frac{1}{2}Y$, or $X = \frac{Y}{2(2k-1)}$. Similarly $Z = \frac{Y}{2(2l-1)}$. This implies that $k > \frac{1}{2}$, $l > \frac{1}{2}$. Hence

$$\begin{aligned} 1 &= X + Y + Z \\ &= \left[\frac{1}{2(2k-1)} + 1 + \frac{1}{2(2l-1)} \right] Y. \end{aligned}$$

The equilibrium frequencies are thus

$$\left. \begin{aligned} X &= \frac{2l-1}{2(4kl-k-l)}, \\ Y &= \frac{(2k-1)(2l-1)}{4kl-k-l}, \\ Z &= \frac{2k-1}{2(4kl-k-l)}. \end{aligned} \right\} \quad (5)$$

Next suppose that $l > \frac{1}{2}$, but $k < \frac{1}{2}$, that is to say, that there is strong selection against aa , but not very strong against AA . The only possible equilibrium is $X=1$, $Y=0$, $Z=0$, that is to say, the whole population comes to consist of AA . If $k < \frac{1}{2}$ and $l < \frac{1}{2}$, heterozygotes first disappear, and then the less fit homozygote.

The progress towards homozygosis can readily be calculated in the artificially simple case where $l=k$. If so

$$1 - kx_n - kz_n = 1 - k + ky_n.$$

Hence from (4)

$$y_{n+1} = \frac{y_n}{2(1-k) + 2ky_n}. \quad (6)$$

This is a non-linear equation, and we cannot hope, save in exceptional cases, to obtain linear equations describing selection. It is, however, readily soluble. For

$$\frac{1}{y_{n+1}} = \frac{2(1-k)}{y_n} + 2k.$$

So
$$\frac{1}{y_{n+1}} + \frac{2k}{1-2k} = 2(1-k) \left(\frac{1}{y_n} + \frac{2k}{1-2k} \right),$$

and
$$\frac{1}{y_n} + \frac{2k}{1-2k} = (2-2k)^n \left(\frac{1}{y_0} + \frac{2k}{1-2k} \right).$$

Hence

$$\left. \begin{aligned}
 y_n &= \frac{(1-2k)y_0}{(2-2k)^n(1-2k+2ky_0)-2ky_0} & \text{if } k < \frac{1}{2}, \\
 &= \frac{(2k-1)y_0}{2ky_0 + (2-2k)^n(2k-1-2ky_0)} & \text{if } k > \frac{1}{2}, \\
 &= Y - \frac{(2k-1)(2-2k)^n(2k-1-2ky_0)}{2k[2ky_0 + (2-2k)^n(2k-1-2ky_0)]} & \text{if } k > \frac{1}{2}.
 \end{aligned} \right\}$$

Another simple case occurs if $l=1$, that is to say, no aa plants are bred from. In practice the progeny of plants segregating for aa , that is to say, of Aa plants, would probably be discarded also. However, the case is perhaps worth working out. Equations (4) become $z_n=0$, $x_n+y_n=1$, and $y_{n+1} = \frac{y_n}{2(1-k)+2ky_n}$. Hence equation (7) holds in this case also.

The general case is only a little more difficult. I write equations (4) as

$$\left. \begin{aligned}
 4u_n x_{n+1} &= 4(1-k)x_n + y_n, \\
 2u_n y_{n+1} &= y_n, \\
 4u_n z_{n+1} &= 4(1-l)z_n + y_n, \\
 u_n &= 1 - kx_n - lz_n.
 \end{aligned} \right\}$$

From the last of these equations

$$\begin{aligned}
 u_n &= 1 - kx_n - l(1-x_n-y_n) \\
 &= 1-l - (k-l)x_n + ly_n,
 \end{aligned}$$

or

$$\begin{aligned}
 (k-l)x_n &= ly_n + 1-l - u_n \\
 &= ly_n + 1-l - \frac{y_n}{2y_{n+1}}.
 \end{aligned}$$

Substituting in the first equation

$$(k-l)y_n + 4(1-k) \left(ly_n + 1-l - \frac{y_n}{2y_{n+1}} \right) - \frac{2y_n}{y_{n+1}} \left(ly_{n+1} + 1-l - \frac{y_{n+1}}{2y_{n+1}} \right) = 0,$$

or $(k+l-4kl)y_{n+2}y_{n+1}y_n + 4(1-k)(1-l)y_{n+2}y_{n+1} - 2(2-k-l)y_{n+2}y_n + y_{n+1}y_n = 0$.

Hence $\frac{1}{y_{n+2}} - \frac{2(2-k-l)}{y_{n+1}} + \frac{4(1-k)(1-l)}{y_n} + k+l-4kl = 0$.

If $\frac{1}{y_n} = t_n + \frac{4kl-k-l}{(2k-1)(2l-1)}$,

then $t_{n+2} - 2(2-k-l)t_{n+1} + 4(1-k)(1-l)t_n = 0$.

Hence $t_n = A(2-2k)^n + B(2-2l)^n$.

Substituting the values of t_0 and t_1 we find

$$A = \frac{x_0}{y_0} + \frac{1}{2(1-2k)},$$

$$B = \frac{z_0}{y_0} + \frac{1}{2(1-2l)},$$

$$\begin{aligned}
 y_n &= 2(2k-1)(2l-1)y_0 [2(4kl-k-l)y_0 - (2-2k)^n(2l-1)\{y_0 - 2(2k-1)x_0\} \\
 &\quad - (2-2l)^n\{y_0 - 2(2l-1)z_0\}]^{-1}.
 \end{aligned}$$

The expressions for x_n and z_n are readily found from (8). The difference from the equilibrium value soon approximates to a geometric series whose common ratio, if $l > k$, is $(2-2k)$ if $k > \frac{1}{2}$, and $(2-2k)^{-1}$ if $k < \frac{1}{2}$. I have given this calculation in what is perhaps needless detail, as others have clearly found it difficult.

SEED SELECTION

Let us next suppose that seeds are sown at random, and the relative fitnesses of AA , Aa and aa are $1-k:l:1-l$. This appears to correspond to Hayman & Mather's hypothesis, and is also an approximation to what might happen in nature, though rarely under artificial conditions. We find

$$\left. \begin{aligned} u_n x_{n+1} &= (1-k)(x_n + \frac{1}{4}y_n), \\ u_n y_{n+1} &= \frac{1}{2}y_n, \\ u_n z_{n+1} &= (1-l)(z_n + \frac{1}{4}y_n), \\ u_n &= 1-k(x_n + \frac{1}{4}y_n) - l(z_n + \frac{1}{4}y_n). \end{aligned} \right\} \quad (11)$$

At equilibrium, provided Y is not zero, $U = \frac{1}{2}$. So

$$2(2k-1)X = (1-k)Y \quad \text{or} \quad X = \frac{(1-k)Y}{2(2k-1)},$$

and similarly

$$Z = \frac{(1-l)Y}{2(2l-1)}.$$

This implies that $k > \frac{1}{2}$, $l > \frac{1}{2}$. Hence

$$\frac{1}{2} = 1 - \frac{kY}{4(2k-1)} - \frac{lY}{4(2l-1)},$$

from the last of equations (11); and at equilibrium, if $k > \frac{1}{2}$, $l > \frac{1}{2}$,

$$\left. \begin{aligned} X &= \frac{(1-k)(2l-1)}{4kl-k-l}, \\ Y &= \frac{2(2k-1)(2l-1)}{4kl-k-l}, \\ Z &= \frac{(1-l)(2k-1)}{4kl-k-l}. \end{aligned} \right\} \quad (12)$$

If k or $l < \frac{1}{2}$, only the fitter homozygote survives. If $k=l=\frac{1}{2}$ there is an unstable equilibrium, and in the long run one homozygote or the other will prevail.

Again the solution of (11) is simple if $k=l$, or $l=1$. If $l=k$,

$$y_{n+1} = \frac{y_n}{2(1-k) + ky_n},$$

and solving as before

$$\left. \begin{aligned} y_n &= \frac{(1-2k)y_0}{(2-2k)^n(2-2k+ky_0) - ky_0} & \text{if } k < \frac{1}{2}, \\ y_n &= \frac{(2k-1)y_0}{ky_0 + (2-2k)^n(2k-1 - ky_0)} & \text{if } k > \frac{1}{2}. \end{aligned} \right\} \quad (13)$$

If $l=1$, so that aa is eliminated at once,

$$y_{n+1} = \frac{y_n}{2(1-k) + \frac{1}{2}(3k-1)y_n}.$$

So

$$\left. \begin{aligned} y_n &= \frac{2(1-2k)y_0}{(2-2k)^n[2-4k+(3k-1)y_0] - (3k-1)y_0} & \text{if } k < \frac{1}{2}, \\ y_n &= \frac{2(2k-1)y_0}{(3k-1)y_0 + (2-2k)^n[4k-2-(3k-1)y_0]} & \text{if } k > \frac{1}{2}. \end{aligned} \right\} \quad (14)$$

In the general case we find, by the same method as before,

$$\frac{1}{y_{n+2}} - \frac{2(2-k-l)}{y_{n+1}} + \frac{4(1-k)(1-l)}{y_n} = \frac{1}{2}(4kl - k - l).$$

Hence

$$y_n = \frac{2(2k-1)(2l-1)y_0}{(4kl - k - l)y_0 + (2-2k)^n(2l-1)[2(2k-1)x_0 - (1-k)y_0] + (2-2l)^n(2k-1)[2(2l-1)z_0 - (1-k)y_0]} \quad (1)$$

The corresponding values of x_n and z_n can easily be calculated. Again the difference from the equilibrium value falls off in an approximately geometrical series.

If we think of this situation in terms of lines, we see that there is selection both between lines and within lines.

A CONSIDERATION OF HAYMAN AND MATHER'S ANALYSIS

Hayman & Mather (1953, pp. 167-70) have dealt with the case when a generation consisting of a fraction p of homozygotes and q of heterozygotes, is selfed. Denoting the corresponding values in the next generation by p' and q' , the upper matrix on their p. 168 is equivalent to

$$\begin{aligned} p' &= p + \frac{1}{2}q, \\ q' &= \frac{1}{2}q. \end{aligned}$$

which is correct. They then suppose that 'all the heterozygotes have the same viability, and... only x of any of the homozygotes survive for each one of the heterozygotes'. They give a second matrix which is equivalent to

$$\begin{aligned} p' &= xp + \frac{1}{2}xq, \\ q' &= \frac{1}{2}q. \end{aligned}$$

This is incorrect. For

$$p' + q' = xp + \frac{1}{2}(1+x)q = 1 - (1-x)(p + \frac{1}{2}q).$$

They state in a footnote that this 'need cause no trouble'. However the correct equations are

$$\begin{aligned} p' &= \frac{x(p + \frac{1}{2}q)}{xp + \frac{1}{2}(1+x)q}, \\ q' &= \frac{\frac{1}{2}q}{xp + \frac{1}{2}(1+x)q}. \end{aligned}$$

These equations, which are equivalent to my (11) with $k=l=1-x$, are non-linear, and cannot be solved by matrix methods, though they are readily solved by the methods given in this paper. All Hayman & Mather's subsequent calculations both as regards

self-fertilization and other forms of inbreeding seem to be based on similar errors. However, they give the correct critical values of $\frac{1}{2}$ for k and l , or x . This is because when y_n is approaching zero in equations (11) the value of u_n becomes almost constant.

On the other hand, Hayman's (1953) calculations as to the equilibrium reached under a mixed system of self-fertilization and random mating, which are not based on matrix algebra, appear to be correct.

DISCUSSION

I hope that I have proved two points. First, if there is any such complication as that referred to in the footnote of Hayman & Mather's p. 168 it is always desirable to write out the equations to be solved in full, rather than as a matrix. In my experience it is quite unusual to obtain a set of linear recurrence equations such as my equations (1) in connexion with a problem of selection. Secondly, it is desirable to specify the conditions of selection as carefully as possible. If so, it will often turn out that, as in this paper, such an expression as 'a disadvantage of homozygotes equal to $1-k$ ' may have a number of different meanings.

Thus I agree with Hayman & Mather that the critical fitness of homozygotes is half that of heterozygotes if there is selection between lines. But if x (in their terminology) is the relative fitness of homozygotes, they find that the frequency of heterozygotes at equilibrium is $\frac{1-2x}{1-x}$, whereas I find that in different circumstances it may be zero

(equations (2)), $\frac{1-2x}{2(1-x)}$ (equations (5)), or $\frac{1-2x}{1-x}$ (equations (12)).

I have not considered the situation where selection of unequal intensity against homozygotes occurs within lines and between lines. It can be discussed by the methods developed here, but I doubt if enough biological data exist to warrant its discussion. I also postpone a general discussion of inbreeding as it occurs in practice until the publication of calculations of the same type as those here set out, on sib mating.

I must thank Dr B. I. Hayman and Dr E. Reeve for correcting an error in my algebra.

SUMMARY

Reason is given for doubting the validity of Hayman & Mather's results on this question. Selection in favour of heterozygotes within lines slows down the onset of homozygosis, but cannot prevent it, if any homozygotes are allowed to breed. Similar selection between lines leads to different equilibria in different circumstances.

REFERENCES

- HALDANE, J. B. S. (1936). The amount of heterozygosis to be expected in an approximately pure line
J. Genet. 32, 375-91.
- HAYMAN, B. I. (1953). Mixed selfing and random mating when homozygotes are at a disadvantage.
Heredity, 7, 185-92.
- HAYMAN, B. I. & MATHER, K. (1953). The progress of inbreeding when homozygotes are at a disadvantage.
Heredity, 7, 165-83.
- HOLLINGSWORTH, M. J. & MAYNARD SMITH, J. (1955). The effects of inbreeding on rate of development and on fertility in *Drosophila subobscura*. *J. Genet.* 53, 295.