THE THEORY OF INBREEDING WITH FORCED HETEROZYGOSIS.

BY M. S. BARTLETT AND J. B. S. HALDANE.
(*University College, London, and John Innes Horticultural Institution.*)

(With Three Text-figures.)

FOR a large number of purposes, both theoretical and practical, it is desirable to produce a population as homozygous as possible. In the course of this process it may be necessary to keep a certain gene or group of completely linked genes heterozygous. In these cases it is desirable to calculate the probability that the other pairs of genes linked with the gene kept heterozygous should still be heterozygous after n generations.

The principal gene may be heterozygous for two reasons. In the first place the object of the experiment may be to introduce it into a pure line, so as to study its effects against a standard genetic background. This is done by repeated back crossing, and if it is not lethal a homozygous stock can finally be obtained. Or the heterozygosis may be enforced by genetical considerations, as when the gene in question is lethal when homozygous, or when it is responsible for sex, heterostylism or self-sterility. In these cases self-fertilisation or brother-sister mating leads to a population homozygous for most genes, but still heterozygous for genes closely linked to the permanently heterozygous pair.

Examples of the first type of heterozygosis occur in the work of Timoféeff-Ressowsky (1933) on *Drosophila* and Fisher (unpublished) on poultry. Examples of the second type are found in the selfing of *Oenothera Lamarckiana* or the double-throwing *Matthiola incana*, the inbreeding of yellow mice by Little and McPheters (1932), or that of self-sterile plants by numerous workers.

I. INTRODUCTION OF NEW GENES INTO A PURE LINE.

(*a*) *Autosomal dominant.*

Here and throughout the gene pair kept permanently heterozygous will be denoted by **Z**, **z**, the pair linked with it, whose heterozygosis is to be investigated, by **A**, **a**, and the probability of recombination, or cross-over value, by c, while c' is the corresponding probability in the opposite sex.

In this case an organism containing the dominant gene is crossed to the pure line, and the progeny carrying the dominant back-crossed again, for n generations. If crossing-over is confined to one sex, the dominant gene must of course be carried, after the first generation, by that sex only. Otherwise all genes linked with it will be introduced along with it. If the original cross is **ZZAA** × **zzaa**, let x_n be the proportion of **ZA.za** in F_n, $1-x_n$ being the proportion of **Za.za**.

Clearly $x_{n+1}=(1-c)\,x_n$, and $x_1=1$, so $x_n=(1-c)^{n-1}$. So after n generations the chance of introducing an unlinked gene is 2^{1-n}, that of introducing a linked gene $(1-c)^{n-1}$. Further, the mean genetical length of chromosome introduced along with **Z**, on each side of it, if it is not terminal or nearly so, is

$$\int_0^{\frac{1}{2}}(1-c)^{n-1}dc \quad \text{or} \quad \frac{1-2^{-n}}{n},$$

which approximates to $1/n$ when n is large.

(*b*) *Sex-linked dominant.*

The calculation is exactly the same as in the last case, provided that (in the case of *Drosophila*) males of the pure line are used in each cross.

(*c*) *Autosomal recessive.*

Suppose that the procedure of Fig. 1 is adopted, and repeated m times, *i.e.* $2m+1$ generations in all are used, or n generations of crossing, where $n=2m$. Let x_m be the proportion of **za** gametes produced by the **zz** females of F_{2m}, where **a** was originally associated with **z**. Then F_{2m+1} consists of $(1-x_n)$ **ZA.zA** : x_n **ZA.za**, giving **z** gametes:

$$(1-x_m+cx_m)\,\mathbf{zA} \;:\; (1-c)\,x_m\,\mathbf{za},$$

and

$$(1-x_m+c'x_m)\,\mathbf{zA} : (1-c')\,x_m\,\mathbf{za},$$

in the two sexes. Hence

$$x_{m+1}=\left(1-\frac{c+c'}{2}\right)x_m;\ \text{and}\ x_0=1.$$

So

$$x_m=\left(1-\frac{c+c'}{2}\right)^m.$$

If $c'=0$, as in *Drosophila*, then for all linked genes $x_m=\left(1-\frac{c}{2}\right)^m$; for unlinked genes $x_m=2^{-m}$, while for linked genes with approximately 50 per cent. crossing-over $x_m=(\frac{3}{4})^m$.

When $c'=c$ the mean length of chromosome introduced on each side of the locus **Z** is $\frac{1-2^{-m-1}}{m+1}$, or approximately $2/n$ when n is large. When $c'=0$ this length is

$$\frac{2}{m+1}\left[1-\left(\tfrac{3}{4}\right)^{m+1}\right],$$

or approximately $4/n$ when n is large.

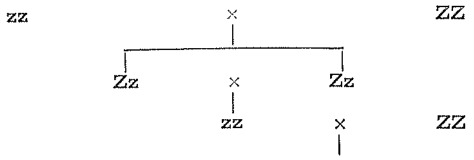

Fig. 1.

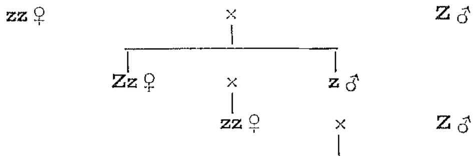

Fig. 2.

(*d*) *Sex-linked recessives.*

Suppose the procedure of Fig. 2 is adopted, as it was by Timoféeff-Ressowsky, the **Z** ♂♂ being taken throughout from a pure line **ZA**, and the original **zz** ♀ from a **za** line. Let x_m be the proportion of **a** genes in the **zz** ♀♀ of F_{2n}. Then

$$x_{m+1}=\left(1-\frac{c}{2}\right)x_m \text{ and } x_0=1.$$

So $$x_m=\left(1-\frac{c}{2}\right)^m \text{ just as in case II } (c).$$

Timoféeff-Ressowsky generally carried out his procedure for twenty or more generations, *i.e.* $m=10$. The mean length of foreign chromosome introduced into the pure line was thus slightly under 200/11 or 18 units on each side of the locus of the introduced gene. Even after forty generations the length would only be reduced to 9 units on each side. While, therefore, we do not wish to criticise Timoféeff-Ressowsky's general conclusions, the possibility that important linked genes were introduced along with the principal gene is by no means excluded.

In particular, allelomorphs of white were studied. A fairly large group of genes, including yellow and scute, are at a distance of 1·5 units ($c=0{\cdot}015$) from white. After twenty generations the probability that genes at these loci were introduced with white was $0{\cdot}9925^{10}$ or 0·926, *i.e* in 93 per cent. of cases these genes would not have been eliminated. Even after forty generations they would still be present in 86 per cent. of cases.

We believe that these difficulties could largely be overcome by the following procedure. The gene to be introduced, *e.g.* white, is put into the same chromosome with two closely adjacent genes on each side of it, *e.g.* broad and facet. In the course of inbreeding males are picked out which contain X-chromosomes that have been formed by two successive cross-overs, one on each side of the gene introduced. Only after this is systematic inbreeding begun.

We now pass to cases of inbreeding in organisms which are necessarily heterozygous on genetical grounds. There are three groups of cases, according as the heterozygosis is of the same character in all the organisms, confined to one member of each mating pair, or different in the two mates.

The first group, our group II, is found where lethal genes or deficiencies, or genes causing sterility or suppression of one class of gamete, are found balancing one another in a homologous pair of chromosomes. This is the case in *Oenothera Lamarckiana*, the ever-sporting double-throwing races of *Matthiola incana*, and the yellow mouse. The second, our group III, occurs when one partner is homozygous, *e.g.* a female *Lebistes reticulatus* or a pin *Primula obconica*, while the other, a male or thrum, is necessarily heterozygous. The third, our group IV, occurs in plants where self-sterility is determined as in *Nicotiana*. Here both partners are heterozygous, but not for the same pair of genes.

II. INBREEDING WITH BALANCED LETHALS.

(*a*) *Self-fertilisation.*

Consider a plant **ZA.za**, where **ZZ** and **zz** are inviable or sterile, or cannot be obtained because one class of gametes are not formed or do not function. Here on selfing **ZA.za** we obtain

$$\tfrac{1}{2}(c+c'-2cc')\,\mathbf{ZA.zA} : (1-c)(1-c')\,\mathbf{ZA.za} : cc'\,\mathbf{Za.zA} : \tfrac{1}{2}(c+c'-2cc')\,\mathbf{Za.za}.$$

Hence if x_n be the proportion of **Aa** heterozygotes in the nth generation, $x_{n+1}=(1-c-c'+2cc')\,x_n$ and $x_0=1$. So $x_n=(1-c-c'+2cc')^n$, or if $c'=c$, $x_n=(1-2c+2c^2)^n$, or putting $c(1-c)=k$, $x_n=(1-2k)^n$.

The mean length of chromosome remaining heterozygous after n generations is thus, when n is large, about $1/2n$ on each side of every gene or group of genes which is kept heterozygous, *e.g.* after twenty-five generations, 2 units. Thus in *Oenothera*, where cross-over values are very small, of the order of 1 or 2 per cent., heterozygosis due to mutation will be eliminated very slowly. In a self-fertilised plant it is normally halved in each generation. In the case of a gene giving 1 per cent. of crossing-over with the complex it is reduced to 0·98 per generation, *i.e.* the number of generations needed to halve it is about thirty-four. Thus the amount of heterozygosis due to mutation in a selfed population of *Oenothera* may be expected to be of the order of thirty times that of a normal plant.

The problem of creating a pure line of a heterothallic haplont in which sex-linked genes occur, *e.g. Sphaerocarpus Donnellii* or *Neurospora sitophila*, is clearly similar. Here since the haploids mated are derived by meiosis from the same diploid, the process is analogous to self-fertilisation.

(*b*) *Brother-sister mating with autosomal lethal, linkage equal in two sexes.*

In what follows we shall not give the somewhat complicated formulae arising when the cross-over values are finite and different in both sexes, but consider the cases in which they are the same in both sexes, as is nearly true in mammals, or zero in one sex, as in *Drosophila*.

In a population where all individuals are **Zz**, there are four genotypes, **ZA.zA**, **Za.za**, **ZA.za** and **Za.zA**. There are ten types of mating. Let them occur with frequencies:

$$p_n\begin{cases}\mathsf{ZA.zA}\times\mathsf{ZA.zA}\\ \mathsf{Za\ .za}\ \times\mathsf{Za\ .za}\end{cases}\quad q_n\ \ \mathsf{ZA.zA}\times\mathsf{Za.za}\quad r_n\begin{cases}\mathsf{ZA.zA}\times\mathsf{ZA.za}\\ \mathsf{Za\ .za}\ \times\mathsf{Za\ .zA}\\ \mathsf{Za\ .za}\ \times\mathsf{ZA.za}\\ \mathsf{ZA.zA}\times\mathsf{Za\ .zA}\end{cases}$$

$$s_n\begin{cases}\mathsf{ZA.za}\times\mathsf{ZA.za}\\ \mathsf{Za\ .za}\times\mathsf{Za\ .zA}\end{cases}\qquad t_n\ \ \mathsf{ZA.za}\times\mathsf{Za.zA}$$

where $p_n+q_n+r_n+s_n+t_n=1$, and reciprocal crosses are equally frequent, but matings of the types bracketed together are not in general equally frequent, though they are so if we start from a mating **ZA.zA** × **Za.za**, *i.e.* $q_0=1$. It is clear that the proportion of heterozygotes

$$x_n=\tfrac{1}{2}r_n+s_n+t_n.$$

Then the frequencies of the mating types in the next generation are given by the following set of simultaneous linear finite difference equations, on the supposition that $c' = c$, and $k = c - c^2$.

$$\begin{aligned}
p_{n+1} &= p_n + \tfrac{1}{4} r_n + 2k^2 s_n + \tfrac{1}{2}(1-2k)^2 t_n,\\
q_{n+1} &= 2k^2 s_n + \tfrac{1}{2}(1-2k)^2 t_n,\\
r_{n+1} &= \tfrac{1}{2} r_n + 4k(1-2k)\, s_n + 4k(1-2k)\, t_n,\\
s_{n+1} &= \tfrac{1}{2} q_n + \tfrac{1}{4}(1-2k)\, r_n + (1-4k+2k^2)\, s_n + 2k^2 t_n,\\
t_{n+1} &= \tfrac{1}{2} q_n + \tfrac{1}{2} k r_n + 2k^2 s_n + 2k^2 t_n.
\end{aligned}$$

The equations are derived as follows. The **Zz** progeny of the mating **ZA.zA** × **ZA.za** are in the proportions:

$$\tfrac{1}{2}\,\mathbf{ZA.zA} : \tfrac{1}{2}(1-c)\,\mathbf{ZA.za} : \tfrac{1}{2}c\,\mathbf{Za.zA}.$$

Hence matings between them occur in the proportions:

$$\tfrac{1}{4}\,\mathbf{ZA.zA}\times\mathbf{ZA.zA} : \tfrac{1}{2}\begin{cases}\mathbf{ZA.zA}\times\mathbf{ZA.za}\\ \mathbf{ZA.zA}\times\mathbf{Za.zA}\end{cases} : (1-2k)\begin{cases}\mathbf{ZA.za}\times\mathbf{ZA.za}\\ \mathbf{Za.zA}\times\mathbf{Za.zA}\end{cases}$$
$$: \tfrac{1}{2}k\ \mathbf{ZA.za}\times\mathbf{Za.zA}.$$

Hence r_n contributes to p_{n+1}, r_{n+1}, s_{n+1} and t_{n+1} in the proportions $\frac{1}{4} : \frac{1}{2} : \frac{1}{4}(1-2k) : \frac{1}{2}k$, and similarly with the other mating types. The last four of the above five equations are an independent group. Hence the values of q_n, r_n, s_n, t_n are of the form

$$a_1\lambda_1{}^n + a_2\lambda_2{}^n + a_3\lambda_3{}^n + a_4\lambda_4{}^n,$$

where a_1, a_2, a_3, a_4 are constants depending on the initial values q_0, r_0, s_0, and t_0, and λ_1, λ_2, λ_3, λ_4 are the roots of the determinant

$$\begin{vmatrix} -\lambda & 0 & 2k^2 & \tfrac{1}{2}(1-2k)^2 \\ 0 & \tfrac{1}{2}-\lambda & 4k(1-2k) & 4k(1-2k) \\ \tfrac{1}{2} & \tfrac{1}{4}(1-2k) & 1-4k+2k^2-\lambda & 2k^2 \\ \tfrac{1}{2} & \tfrac{1}{2}k & 2k^2 & 2k^2-\lambda \end{vmatrix} = 0.$$

Clearly x_n is an expression of the same form.

If we put $\mu = 2\lambda$, $l = 2k$ the determinant becomes

$$\mu^4 - (3-4l+2l^2)\,\mu^3 + (1-4l+4l^2-4l^3)\,\mu^2$$
$$+ (3-10l+14l^2-12l^3+8l^4)\,\mu - 2\,(1-5l+10l^2-10l^3+4l^2) = 0.$$

The four roots are always real, the values of λ lying between ± 1. The actual values of x_n for a series of values of k are given in Table I and Fig. 3. When n is sufficiently large we have approximately $x_n = a_1\lambda_1{}^n$, where λ_1 is the largest of the four roots. We can obtain the value of a_1 with sufficient accuracy from the value of x_{10}. Values of a_1 and λ_1 are given in Table I. The exact solutions when $k=0$ (complete linkage) and $k = 0{\cdot}25$ (no linkage) are

$$x_n = \tfrac{2}{3}\left[1-(-2)^{-n}\right], \quad \text{and} \quad x_n = \frac{2}{\sqrt{5}}\left[\left(\frac{\sqrt{5}+1}{4}\right)^n - \left(\frac{1-\sqrt{5}}{4}\right)^2\right],$$

the latter being of course the formula for the amount of heterozygosis at a normal locus in brother-sister mating. It will be seen that in this case, as in all subsequent cases, the values of x_n finally approximate to a diminishing geometric series whose common ratio lies between 1 and 0·8090, the value found in the absence of linkage.

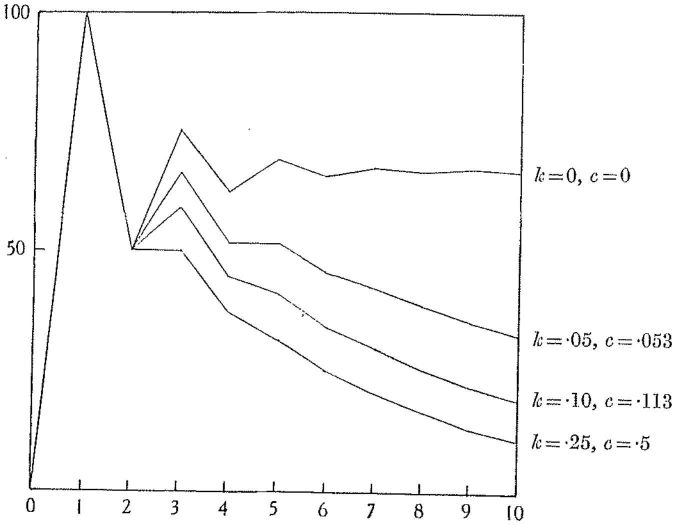

Fig. 3. Abscissa: time in generations. Ordinates: Percentages of **Aa** heterozygotes, starting from **ZA.zA** × **Za.za**, when **ZZ** and **zz** are eliminated and brothers and sisters inbred, linkage equal on both sides.

TABLE I.

x_n, in case II (b).

Generations ...	1	2	3	4	5	6	7	8	9	10
$k=0$	1	0·5	0·75	0·625	0·687	0·656	0·672	0·664	0·668	0·666
0·05	1	0·5	0·66	0·519	0·516	0·454	0·424	0·384	0·353	0·322
0·1	1	0·5	0·59	0·447	0·410	0·342	0·298	0·254	0·219	0·188
0·15	1	0·5	0·54	0·403	0·350	0·284	0·237	0·195	0·162	0·134
0·2	1	0·5	0·51	0·381	0·321	0·257	0·210	0·171	0·139	0·113
0·25	1	0·5	0·5	0·375	0·313	0·250	0·203	0·164	0·133	0·107

k ...	0	0·05	0·1	0·15	0·2	0·25
a_1	0·667	0·790	0·861	0·890	0·893	0·894
λ_1	1·0	0·9143	0·8585	0·8274	0·8129	0·8090

When c is very small, $\lambda_1 = 1 - 2k = 1 - 2c$, approximately. The mean length of chromosome remaining heterozygous on each side of the heterozygous locus, if $\lambda_1 = 1 - \alpha c$, is

$$\int_0^{\frac{1}{2}} a_1 (1-\alpha c)^n \, dc = \frac{a_1}{(n+1)\,\alpha} \left[1 - \left(1 - \frac{\alpha}{2}\right)^{n+1} \right],$$

or approximately $\dfrac{a_1}{(n+1)\,\alpha}$. In this case the length on each side is

$\frac{1}{3(n+1)}$. Thus after twenty generations of brother-sister mating of yellow mice we may expect to find a mean length of 200/63 or 3·2 units still heterozygous. The probability of finding heterozygosis at a locus 1 unit distant from that of yellow is $\frac{2}{3}\,0{\cdot}98^{20}$, or 0·44. It is worth noticing how much more completely such a population of mice approximates to a pure line than a population of *Drosophila* where the procedure for introducing a recessive gene has been carried out for twenty generations.

(*c*) *Inbreeding with an autosomal lethal, no crossing-over in one sex.*

Here reciprocal crosses give quite different results. If the frequencies of mating types are as in the last case, we arrive at the equations:

$$\begin{aligned}
p_{n+1} &= p_n + \tfrac{1}{4}r_n + \tfrac{1}{2}c^2 s_n + \tfrac{1}{2}(1-c)^2 t_n,\\
q_{n+1} &= \tfrac{1}{2}c^2 s_n + \tfrac{1}{2}(1-c)^2 t_n,\\
r_{n+1} &= \tfrac{1}{2}r_n + 2ks_n + 2kt_n,\\
s_{n+1} &= \tfrac{1}{2}q_n + \tfrac{1}{4}(1-k)\,r_n + (1-c)^2 s_n + c^2 t_n,\\
t_{n+1} &= \tfrac{1}{2}q_n + \tfrac{1}{4}kr_n.
\end{aligned}$$

So

$$\begin{vmatrix} -4\lambda & 0 & c^2 & (1-c)^2 \\ 0 & 1-2\lambda & k & k \\ 1 & 1-k & (1-c)^2-\lambda & c^2 \\ 1 & k & 0 & -\lambda \end{vmatrix} = 0,$$

or if $\mu = 2\lambda$,

$$\mu^4 - (3-2k)\,\mu^3 + (1-2k)\,\mu^2 + \{(1-2k)\,[1+2\,(3-2k)\,(1-2c)] \\ -4\,(1-2c)\,(1-c)^4\}\,\mu - 2\,(1-k)\,(1-2k)\,(1-2c) = 0.$$

The values of x_n for a series of values of c are given in Table II. When n is large we have, as before, $x_n = a_1\lambda_1{}^n$.

TABLE II.

x_n, *in case II* (*c*).

Generations ...	1	2	3	4	5	6	7	8	9	10
$c=0$	1	0·5	0·750	0·625	0·687	0·656	0·672	0·664	0·668	0·666
0·1	1	0·5	0·664	0·523	0·521	0·460	0·431	0·391	0·361	0·330
0·3	1	0·5	0·558	0·416	0·367	0·300	0·253	0·211	0·176	0·147
0·5	1	0·5	0·5	0·375	0·313	0·25	0·203	0·164	0·133	0·107

c ...	0	0·1	0·2	0·3	0·4	0·5
a_1	0·667	0·779	—	0·886	—	0·894
λ_1	1·0	0·9177	0·8671	0·8358	0·8192	0·8090

Values of a_1 and λ_1 are also given in Table II. The exact solutions when $c=0$ and $c=\frac{1}{2}$ are the same as in the last case.

For very small values of c, $\lambda = 1-c$, so that $x_n = \frac{2}{3}(1-c)^n$ and the mean length of chromosome heterozygous on each side of the locus of **Z** is $\frac{2}{3(n+1)}$.

(*d*) *Inbreeding with a sex-linked lethal.*

Let us suppose that **z** is lethal to the male (supposed heterogametic) while **ZZ** females are eliminated. This occurs in a culture of such a deficiency as Notch in *Drosophila melanogaster*. Thus all matings are **Zz** ♀ × **Z** ♂. Let mating types occur with the frequencies

$$p_n \begin{cases} \mathbf{ZA.zA} \times \mathbf{ZA} \\ \mathbf{Za.za} \times \mathbf{Za} \end{cases} \qquad q_n \begin{cases} \mathbf{ZA.zA} \times \mathbf{Za} \\ \mathbf{Za.za} \times \mathbf{ZA} \end{cases}$$

$$r_n \begin{cases} \mathbf{ZA.za} \times \mathbf{ZA} \\ \mathbf{Za.zA} \times \mathbf{zA} \end{cases} \qquad s_n \begin{cases} \mathbf{ZA.za} \times \mathbf{Za} \\ \mathbf{Za.zA} \times \mathbf{ZA} \end{cases}$$

where $p_n+q_n+r_n+s_n=1$, and the proportion of **Aa** females, $x_n=r_n+s_n$.

Then
$$\begin{aligned} p_{n+1} &= p_n + c(1-c)\,r_n + c(1-c)\,s_n, \\ q_{n+1} &= c^2 r_n + (1-c)^2 s_n, \\ r_{n+1} &= (1-c)^2 r_n + c^2 s_n, \\ s_{n+1} &= q_n + c(1-c)\,r_n + c(1-c)\,s_n. \end{aligned}$$

Whence
$$\begin{vmatrix} -\lambda & c^2 & (1-c)^2 \\ 0 & (1-c)^2-\lambda & c^2 \\ 1 & c(1-c) & c(1-c)-\lambda \end{vmatrix} = 0,$$

$$\text{or } \lambda^3-(1-c)\,\lambda^2-(1-c)\,(1-2c+2c^2)\,\lambda+(1-c)^4-c^4=0.$$

If $q_0=1$, we have the values of x_n given in Table III. It will be seen that the values oscillate considerably. Although ultimately $x_n=a_1\lambda_1^{\,n}$, there is a large negative root responsible for the oscillations. Thus for $c=0{\cdot}1$, $\lambda_3=-0{\cdot}856$. For $c=0$, $x_n=\frac{1}{2}-\frac{1}{2}(-1)^n$. For other values of c, the values of a_1 given in Table III are calculated by considering x_9 and x_{10}, and eliminating the oscillatory effects of λ_3.

TABLE III.

x_n, *in case II* (*d*).

Generations ...	1	2	3	4	5	6	7	8	9	10
$c=0$	1	0	1	0	1	0	1	0	1	0
0·1	1	0·1	0·828	0·163	0·692	0·200	0·584	0·219	0·497	0·226
0·2	1	0·2	0·704	0·264	0·513	0·266	0·384	0·243	0·295	0·211
0·3	1	0·3	0·616	0·321	0·405	0·271	0·280	0·212	0·199	0·160
0·4	1	0·4	0·552	0·352	0·342	0·258	0·225	0·180	0·151	0·123
0·5	1	0·5	0·5	0·375	0·313	0·250	0·203	0·164	0·133	0·107

c ...	0	0·1	0·2	0·3	0·4	0·5
a_1	0·5	0·572	0·667	0·764	0·843	0·894
λ_1	1·0	0·9499	0·9013	0·8581	0·8254	0·8090

For $c = \frac{1}{2}$ we have the usual equation for x_n.

When c is small, then $\lambda_1 = 1 - \frac{1}{2}c$, $\lambda_2 = 1 - 2c$, $\lambda_3 = -1 + \frac{3}{2}c$, and

$$x_n = \tfrac{1}{2}\,(1 - \tfrac{1}{2}c)^n - \tfrac{1}{2}\,(-1)^n\,(1 - \tfrac{3}{2}c)^n.$$

Hence the mean length of chromosome heterozygous on each side of the locus of **Z** is

$$\frac{1}{n+1}\,[1 - \tfrac{1}{3}\,(-1)^n]$$

approximately when n is large.

III. INBREEDING WITH HETEROSTYLISM OR PARTIAL SEX LINKAGE.

In the case of heterostylism, while in certain plants pin × pin or thrum × thrum matings are possible, in other cases they are quite infertile. Thus all matings are **Ss** × **ss**, where **S** is the gene for the thrum condition. Hence, in building up a pure line, genes linked with **S** will remain heterozygous for a long time.

Similarly, where a gene is found both in the X- and Y-chromosomes it may cross-over from one to the other, a state of affairs found in *Lebistes* by Winge (1923) and in *Drosophila* by Philip (1935). In this case the male is clearly analogous to the thrum, the female to the pin.

Let the mating types occur with the following frequencies:

$$\begin{array}{lll}
p_n \begin{cases} \mathbf{sA.sA \times SA.sA} \\ \mathbf{sa.sa \times Sa.sa} \end{cases} & \text{or} & \begin{matrix} X\mathbf{A}.X\mathbf{A} \times X\mathbf{A}.Y\mathbf{A} \\ X\mathbf{a}.X\mathbf{a} \times X\mathbf{a}.Y\mathbf{a} \end{matrix} \\
q_n \begin{cases} \mathbf{sA.sA \times Sa.sa} \\ \mathbf{sa.sa \times SA.sA} \end{cases} & \text{or} & \begin{matrix} X\mathbf{A}.X\mathbf{A} \times X\mathbf{a}.Y\mathbf{a} \\ X\mathbf{a}.X\mathbf{a} \times X\mathbf{A}.Y\mathbf{A} \end{matrix} \\
r_n \begin{cases} \mathbf{sA.sA \times Sa.sA} \\ \mathbf{sa.sa \times SA.sa} \end{cases} & \text{or} & \begin{matrix} X\mathbf{A}.X\mathbf{A} \times X\mathbf{A}.Y\mathbf{a} \\ X\mathbf{a}.X\mathbf{a} \times X\mathbf{a}.Y\mathbf{A} \end{matrix} \\
s_n \begin{cases} \mathbf{sA.sA \times SA.sa} \\ \mathbf{sa.sa \times Sa.sA} \end{cases} & \text{or} & \begin{matrix} X\mathbf{A}.X\mathbf{A} \times X\mathbf{a}.Y\mathbf{A} \\ X\mathbf{a}.X\mathbf{a} \times X\mathbf{A}.Y\mathbf{a} \end{matrix} \\
t_n \begin{cases} \mathbf{sA.sa \times SA.sa} \\ \mathbf{sA.sa \times Sa.sA} \end{cases} & \text{or} & \begin{matrix} X\mathbf{A}.X\mathbf{a} \times X\mathbf{A}.Y\mathbf{A} \\ X\mathbf{A}.X\mathbf{a} \times X\mathbf{a}.Y\mathbf{a} \end{matrix} \\
u_n \begin{cases} \mathbf{sA.sa \times SA.sa} \\ \mathbf{sA.sa \times Sa.sA} \end{cases} & \text{or} & \begin{matrix} X\mathbf{A}.X\mathbf{a} \times X\mathbf{a}.Y\mathbf{a} \\ X\mathbf{A}.X\mathbf{a} \times X\mathbf{A}.Y\mathbf{a} \end{matrix}
\end{array}$$

then the proportion of heterozygous females or pins is $x_n = t_n + u_n$ and that of heterozygous males or thrums is $y_n = r_n + s_n + u_n$.

We find

$$\begin{aligned}
p_{n+1} &= p_n + c\,(1-c)\,r_n + c\,(1-c)\,s_n + \tfrac{1}{4}t_n + \tfrac{1}{2}\,c\,(1-c)\,u_n, \\
q_{n+1} &= \tfrac{1}{4}\,\{c^2 + (1-c)^2\}\,u_n, \\
r_{n+1} &= (1-c)^2\,r_n + c^2 s_n + \tfrac{1}{4}\,\{c^2 + (1-c)^2\}\,u_n, \\
s_{n+1} &= \tfrac{1}{4}t_n + \tfrac{1}{2}\,c\,(1-c)\,u_n, \\
t_{n+1} &= c^2 r_n + (1-c)^2\,s_n + \tfrac{1}{4}t_n + \tfrac{1}{4}\,u_n, \\
u_{n+1} &= q_n + c\,(1-c)\,r_n + c\,(1-c)\,s_n + \tfrac{1}{4}t_n + \tfrac{1}{4}\,u_n.
\end{aligned}$$

Whence

$$\begin{vmatrix} -\lambda & 0 & 0 & 0 & 1-2c\,(1-c) \\ 0 & (1-c)^2-\lambda & c^2 & 0 & 1-2c\,(1-c) \\ 0 & 0 & -\lambda & 1 & 2c\,(1-c) \\ 0 & c^2 & (1-c)^2 & 1-4\lambda & 1 \\ 1 & c\,(1-c) & c\,(1-c) & 1 & 1-4\lambda \end{vmatrix}$$

or if $\mu=2\lambda$, $k=c\,(1-c)$,

$$\mu^5-(3-2c-2k)\,\mu^4-c\mu^3+\{(5-7c-2k)\,(1-2k)+(1-2c)\,4k^2\}\,\mu^2 \\ -(1-2c)\,(1-2k)\,(1-4k)\,\mu-2\,(1-2c)\,(1-2k)^2=0.$$

If $q_0=1$, *i.e.* if we start with the mating

$$\mathsf{sA.sA}\times\mathsf{Sa.sa} \text{ or } X\mathsf{A}.X\mathsf{A}\times X\mathsf{a}.Y\mathsf{a}$$

we have the values of x_n and y_n given in Table IV. When n is large we have $x_n=a_1\lambda_1^n$, $y_n=a_1'\lambda_1^n$ approximately, where the values of a_1, a_1' and λ_1 are given in Table IV. When $c=\frac{1}{2}$ we get the usual equations for unlinked genes. When $c=0$ we have the same values for x_n, for since a_1 vanishes we have $x_n=a_2\lambda_2^n$, and this expression is the same as when $c=\frac{1}{2}$. On the other hand $y_n=\frac{2}{3}\{1-(-\frac{1}{2})^n\}$ and clearly two-thirds of the thrums or males remain heterozygous in the absence of crossing-over.

TABLE IV.

x_n and y_n in case III.

Generations ...		1	2	3	4	5	6	7	8	9	10
$c=0$	x_n	1	0·5	0·500	0·375	0·313	0·25	0·203	0·164	0·133	0·107
	y_n	1	0·5	0·750	0·625	0·687	0·656	0·672	0·664	0·668	0·666
$c=0·1$	x_n	1	0·5	0·516	0·397	0·346	0·292	0·252	0·217	0·189	0·166
	y_n	1	0·5	0·644	0·512	0·501	0·442	0·407	0·366	0·332	0·299
$c=0·3$	x_n	1	0·5	0·512	0·389	0·332	0·271	0·224	0·186	0·154	0·126
	y_n	1	0·5	0·528	0·400	0·344	0·281	0·234	0·193	0·160	0·132
$c=0·5$	x_n	1	0·5	0·5	0·375	0·313	0·25	0·203	0·164	0·133	0·107
	y_n	1	0·5	0·5	0·375	0·313	0·25	0·203	0·164	0·133	0·107

c	0	0·1	0·2	0·3	0·4	0·5
a_1 (for x_n)	0	0·432	—	0·848	—	0·894
a_1' (for y_n)	0·667	0·888	—	0·882	—	0·894
λ_1	1·0	0·8984	0·8553	0·8270	0·8136	0·8090

If, however, $r_0=1$, *i.e.* if we start with the mating $\mathsf{sa.sa}\times\mathsf{SA.sa}$, or $X\mathsf{a}.X\mathsf{a}\times X\mathsf{a}.Y\mathsf{A}$, things are rather different. This is the case if, for example, in *Lebistes* we introduce a new gene in the Y-chromosome, or in *Drosophila* inbreed a strain all of whose X-chromosomes contain the bobbed gene, whilst the Y carries the normal allelomorph.

Here the values of λ, when c is very small, are approximately

$$1-\frac{4c}{3},\ \frac{1+\sqrt{5}}{4},\ \tfrac{1}{2},\ \frac{1-\sqrt{5}}{4},\ \text{and }-\tfrac{1}{2}.$$

So the value of y_n ultimately approximates to $\left(1-\frac{4c}{3}\right)^n$ while that of x_n rises gradually to a value of nearly $4c$, and then very slowly falls off to zero.

In all cases, when n is large, the mean length of heterozygous chromosome in pins or females is very small, decreasing in a geometrical series. The length heterozygous in thrums or males is about $1/2n$ on each side of the locus of the genes for heterostylism or sex if $q_0=1$, and $3/4n$ if $r_0=1$. In *Drosophila melanogaster* Philip (1935) finds crossing-over between the locus of bobbed and the sex genes in less than one gamete per thousand. So over 500 generations of brother-sister mating would be needed to produce homozygosis for this locus in half a population, and even in a "pure line" heterozygosis for this locus would hardly ever be abolished.

IV. BROTHER-SISTER MATING WITH SELF-STERILITY OF THE *NICOTIANA* TYPE.

Here there are in any population three or more allelomorphs such that a pollen grain carrying any one of them cannot enter a zygote carrying the same gene. Hence if these allelomorphs are called S_1, S_2, S_3, S_4, etc., then while $S_1S_2 \times S_3S_4$ gives $S_1S_3+S_1S_4+S_2S_3+S_2S_4$, the mating $S_1S_2 \times S_2S_3$ gives $S_2S_3+S_3S_1$ and the reciprocal mating $S_3S_1+S_1S_2$.

It is clear that as the result of any system of inbreeding, ultimately all but three allelomorphs (say S_1, S_2 and S_3) will disappear from a population. Moreover, a population consisting of equal numbers of the three genotypes S_1S_2, S_2S_3 and S_3S_1, will be in stable equilibrium under any mating system in which there is no selection in favour of any genotype.

If **A** is linked with the S locus, twelve genotypes are possible in such a population, and each can be fertilised by pollen grains from eight others. There are thus ninety-six types of fertile mating. In a symmetrical population the frequency of a mating is unchanged if we permute $S_1\,S_2\,S_3$ or exchange **A** and a. Also reciprocal crosses are equally frequent. Let the frequencies of the mating types be as follows:

p_n $S_3\mathbf{A}.S_1\mathbf{A}\times S_1\mathbf{A}.S_2\mathbf{A}$ (one of 12 types),
q_n $S_3\mathbf{A}.S_1\mathbf{A}\times S_1\mathrm{a}.S_2\mathrm{a}$ (one of 12 types),
r_n $S_3\mathbf{A}.S_1\mathbf{A}\times S_1\mathbf{A}.S_2\mathrm{a}$ and reciprocally (one of 24 types),
s_n $S_3\mathbf{A}.S_1\mathbf{A}\times S_1\mathrm{a}.S_2\mathbf{A}$ and reciprocally (one of 24 types),
t_n $S_3\mathrm{a}.S_1\mathbf{A}\times S_1\mathbf{A}.S_2\mathrm{a}$ or $S_3\mathrm{a}.S_1\mathbf{A}\times S_1\mathrm{a}.S_2\mathbf{A}$ (two of 24 types).

While the two groups of mating types under t_n do not occur with equal frequency (the second type having the frequency q_n) they yield progeny in the same proportions, and may be grouped together. The proportion of **Aa** heterozygotes, $x_n = \frac{1}{2}(r_n + s_n) + t_n$. Then

$$\begin{aligned}
p_{n+1} &= p_n + \tfrac{1}{2}cr_n + \tfrac{1}{2}(1-c)\,s_n + k(1-2k)\,t_n,\\
q_{n+1} &= k(1-2k)\,t_n,\\
r_{n+1} &= \tfrac{1}{2}(1-2k)\,r_n + \tfrac{1}{2}(1-2k)\,s_n + (1-2k)^2\,t_n,\\
s_{n+1} &= kr_n + ks_n + 4k^2t_n,\\
t_{n+1} &= q_n + \tfrac{1}{2}(1-c)\,r_n + \tfrac{1}{2}cs_n + 2k(1-2k)\,t_n.
\end{aligned}$$

So

$$\begin{vmatrix}
-\lambda & 0 & 0 & k(1-2k)\\
0 & 1-2k-2\lambda & 1-2k & (1-2k)^2\\
0 & k & k-\lambda & 2k^2\\
1 & 1-c & c & 2k(1-2k)-\lambda
\end{vmatrix} = 0.$$

Hence $\lambda = 0$, or

$$\begin{aligned}
2\lambda^3 - (1+4k-8k^2)\,\lambda^2 - \{(1-c)\,(1-4k)+4k^2\}\,\lambda \\
-2k\,(1-2k)\,\{1-2k-c\,(1-4k)\} = 0.
\end{aligned}$$

If $q_0 = 1$, *i.e.* the initial matings are $S_3\mathbf{A}.S_1\mathbf{A} \times S_1\mathbf{a}.S_2\mathbf{a}$ and reciprocally, or similar matings, then we have the values of x_n given in Table V. The exact equation for $c=0$ is

$$x_n = \tfrac{2}{3}\left\{1 - \left(\frac{-1}{2}\right)^n\right\},$$

TABLE V.

x_n, *in case IV.*

Generations ...	1	2	3	4	5	6	7	8	9	10
$c=0$	1	0·5	0·75	0·625	0·687	0·656	0·672	0·664	0·668	0·666
0·1	1	0·5	0·628	0·503	0·489	0·435	0·402	0·355	0·335	0·305
0·3	1	0·5	0·516	0·391	0·333	0·270	0·224	0·184	0·151	0·127
0·5	1	0·5	0·5	0·375	0·313	0·250	0·203	0·164	0·133	0·107

c ...	0	0·1	0·2	0·3	0·4	0·5
a_1	0·667	0·753	—	0·882	—	0·894
λ_1	1·0	0·9137	0·8554	0·8221	0·8108	0·8090

and for $c=\frac{1}{2}$ the usual equation for brother-sister mating. For large n $x_n = a_1\lambda_1^n$ approximately, the values of a_1 and λ_1 being given in Table V. When c is small, $x_n = \frac{2}{3}(1-c)^n$ approximately, so after n generations the mean length of chromosome heterozygous on each side of the S locus is $2/3n$.

V. Summary.

When a foreign gene is being bred into a pure line, or when a population with a lethal gene, incomplete sex linkage, heterostylism, or self-sterility is being inbred, the population is kept heterozygous for a particular gene or chromosome segment, but otherwise inbred. In these cases expressions are found for the probability of introducing a gene linked with the given gene, or for finding such a linked gene still heterozygous after inbreeding. Expressions are also found for the mean length of foreign chromosome introduced, or the length of chromosome remaining heterozygous (Table VI). The difficulty of introducing a single gene without any linked genes is emphasised.

TABLE VI.

Mean length of chromosome heterozygous on each side of principal locus after n generations (n large).

I (*a*), (*b*)	Introduction of dominant	$1/n$
I (*c*)	Introduction of autosomal recessive ($c'=c$)	$2/n$
I (*c*)	Introduction of autosomal recessive ($c'=0$)	$4/n$
I (*d*)	Introduction of sex-linked recessive	$4/n$
II (*a*)	Self-fertilisation	$1/2n$
II (*b*)	Brother-sister mating, autosomal lethal selected ($c'=c$)	$1/3n$
II (*c*)	Brother-sister mating, autosomal lethal selected ($c'=0$)	$2/3n$
II (*d*)	Brother-sister mating, sex-linked lethal selected	$1/n\ (1-\frac{1}{3}\ (-1)^n)$
III	Brother-sister mating, partial sex-linkage or heterostylism	$1/2n$
IV	Brother-sister mating, self-sterility	$2/3n$

REFERENCES.

Little, C. C. and McPheters, B. W. (1932). "The incidence of mammary cancer in a cross between two strains of mice." *Amer. Nat.* **66**, 568.

Philip, U. (1935). "Crossing-over between the X- and Y-chromosomes in *Drosophila melanogaster*." *J. Genet.* **31**, 341.

Timoféeff-Ressowsky, N. W. (1933). "Ruckgenmutationen und die Genvariabilität in verschiedenen Richtungen." *Z. indukt. Abstamm.- u. VererbLehre*, **66**, 165.

Winge, Ö. (1923). "Crossing-over between the X- and Y-chromosomes in *Lebistes*." *J. Genet.* **13**, 201.