MIXED SIB AND RANDOM MATING WHEN HOMOZYGOTES ARE AT A DISADVANTAGE

A. R. PAGE and B. I. HAYMAN

Applied Mathematics Laboratory, D.S.I.R., Private Bag, Christchurch, New Zealand

1. INTRODUCTION

AN investigation of several inbreeding systems when homozygotes are at a selective disadvantage compared with heterozygotes was conducted by Hayman and Mather (1953). Hayman (1953) also considered the system of mixed selfing and random mating which occurs with plants which have imperfect selfing mechanisms. The present investigation is of a further mixed system, that of sib and random mating. Since sib mating is the most intensive inbreeding system possible with animals we are performing for animals what Hayman (1953) did for plants.

A further important mixed system is that of parent-offspring and random mating which corresponds fairly closely to the backcrossing system practised in cattle breeding. This would be even more difficult mathematically than the present investigation but Hayman and Mather (1953, figs. 3 and 4) showed that pure parent-offspring mating achieved a similar level of inbreeding to pure sib mating whatever the selective disadvantage of the homozygotes. We may, therefore, expect our present conclusions to be equally valid for mixed parent-offspring and random mating.

We also compare the relative inbreeding effectiveness of selfing and sib mating when some random mating is permitted. In the situation discussed by Hayman and Mather (1953) we can make this comparison with the relative survival rates of homozygotes that produce equal levels of heterozygosity under the two mating systems and we find that selfing is about 60 per cent. more effective as an inbreeding system than sib mating for homozygote survival rates down to 20 per cent., but that as the survival rate is lowered further this difference in inbreeding effectiveness gradually disappears. We will see that with a proportion of random mating the relative inbreeding effectiveness of selfing and sib mating is considerably altered.

2. GENERAL EQUATIONS

One gene with two alleles, a and b, is considered, the population being assumed large enough for random variations in frequency to be neglected. The population can be classified into six distinct mating types

appropriate to sib mating. Let the relative frequencies of the types $aa \times aa$, $bb \times bb$, $aa \times bb$, $ab \times ab$, $aa \times ab$, and $bb \times ab$ be λ, μ, α, β, γ, δ respectively, $(\lambda + \mu + \alpha + \beta + \gamma + \delta = 1)$, and let $x, y, 1$ be the relative viabilities of the genotypes aa, bb and ab.

Starting from a population $(\lambda_0, \mu_0, \alpha_0, \beta_0, \gamma_0, \delta_0)$ the genotypic array in the next generation is

$$l_0 x\ aa + m_0 y\ bb + n_0\ ab$$

where

$$l_0 = \lambda_0 + \tfrac{1}{4}\beta_0 + \tfrac{1}{2}\gamma_0 \quad m_0 = \mu_0 + \tfrac{1}{4}\beta_0 + \tfrac{1}{2}\delta_0 \quad n_0 = \alpha_0 + \tfrac{1}{2}\beta_0 + \tfrac{1}{2}\gamma_0 + \tfrac{1}{2}\delta_0$$

and hence

$$l_0 + m_0 + n_0 = 1.$$

If $l_0 x + m_0 y + n_0 = k_0$ then the frequencies of the genotypes aa, bb, ab in the next generation are p_1, q_1, r_1 respectively $(p + q + r = 1)$ where

$$k_0 p_1 = l_0 x \qquad k_0 q_1 = m_0 y \qquad k_0 r_1 = n_0.$$

Hence the genotypic array may be written

$$k_0 (p_1 aa + q_1 bb + r_1 ab).$$

Thus for random mating with selection the proportions of the mating types in the next generation are

$$k_0^2 (p_1^2,\ q_1^2,\ 2p_1 q_1,\ r_1^2,\ 2r_1 p_1,\ 2q_1 r_1).$$

For sib mating the changes in frequency between the generations are specified by the generation matrix $\mathbf{A}$ (Hayman and Mather, 1953, section 4.2). $\mathbf{A}$ is the matrix.

Offspring	Parents					
	$aa \times aa$	$bb \times bb$	$aa \times bb$	$ab \times ab$	$aa \times ab$	$bb \times ab$
$aa \times aa$	x	.	.	$\tfrac{1}{4}x^2$	$\tfrac{1}{2}x^2$	.
$bb \times bb$	.	y	.	$\tfrac{1}{4}y^2$	.	$\tfrac{1}{2}y^2$
$aa \times bb$	.	.	.	$\tfrac{1}{2}xy$	.	.
$ab \times ab$	.	.	1	1	$\tfrac{1}{2}$	$\tfrac{1}{2}$
$aa \times ab$	.	.	.	x	x	.
$bb \times ab$	.	.	.	y	.	y
Column divisors	1	1	1	$x+y+2$	$x+1$	$y+1$

Thus $\lambda_1 \alpha \mathbf{A}\lambda_0$ where λ is the 6×1 column vector of the mating type frequencies.

Since each column of the generation matrix sums to the proportion of offspring from the corresponding mating type, the constant of proportionality is $1/k_0 (p_1 + q_1 + r_1) = 1/k_0$.

Let s and t $(s+t=1)$ be the frequencies of sib and random mating respectively in the mixed system. The equations governing the frequencies in the next generation are now

$$
\left.\begin{aligned}
k_0(s+k_0t)\lambda_1 &= s\left(x\lambda_0+\frac{x^2}{4(x+y+2)}\beta_0+\frac{x^2}{2(x+1)}\gamma_0\right) &&+tk_0^2p_1^2 \\[2mm]
k_0(s+k_0t)\mu_1 &= s\left(y\mu_0+\frac{y^2}{4(x+y+2)}\beta_0+\frac{y^2}{2(y+1)}\delta_0\right) &&+tk_0^2q_1^2 \\[2mm]
k_0(s+k_0t)\alpha_1 &= s\frac{xy}{2(x+y+2)}\beta_0 &&+2tk_0^2p_1q_1 \\[2mm]
k_0(s+k_0t)\beta_1 &= s\left(\alpha_0+\frac{1}{x+y+2}\beta_0+\frac{1}{2(x+1)}\gamma_0+\frac{1}{2(y+1)}\delta_0\right)+tk_0^2r_1^2 \\[2mm]
k_0(s+k_0t)\gamma_1 &= s\left(\frac{x}{x+y+2}\beta_0+\frac{x}{x+1}\gamma_0\right) &&+2tk_0^2r_1p_1 \\[2mm]
k_0(s+k_0t)\delta_1 &= s\left(\frac{y}{x+y+2}\beta_0+\frac{y}{y+1}\delta_0\right) &&+2tk_0^2q_1r_1
\end{aligned}\right\} \text{(i)}
$$

3. EQUILIBRIUM

The population is in equilibrium if $\lambda_1=\lambda_0$. Let the equilibrium population λ, be $(\lambda,\ \mu,\ \alpha,\ \beta,\ \gamma,\ \delta)$ with genotypic frequencies $(p,\ q,\ r)$, the two frequency distributions being related by the equations

$$p = \lambda+\tfrac{1}{2}\alpha+\tfrac{1}{2}\gamma \qquad q = \mu+\tfrac{1}{2}\alpha+\tfrac{1}{2}\delta \qquad r = \beta+\tfrac{1}{2}\gamma+\tfrac{1}{2}\delta$$

(which also hold in each generation for any population). Hence, $p-q = (\lambda-\mu)+\tfrac{1}{2}(\gamma-\delta) = l-m$ where, as before, $l = \lambda+\tfrac{1}{4}\beta+\tfrac{1}{2}\gamma$, $m = \mu+\tfrac{1}{4}\beta+\tfrac{1}{2}\delta$, $n = \alpha+\tfrac{1}{2}\beta+\tfrac{1}{2}\gamma+\tfrac{1}{2}\delta$.

Since the population is in equilibrium $kp = lx$ and $kq = my$; $k = lx+my+n$. Also $l+m+n = 1$.

Eliminating l, m, n from these equations gives

$$
\begin{aligned}
y(k-x)p-x(k-y)q &= 0 \\
kyp+kxq+kxyr &= xy \\
p+q+r &= 1
\end{aligned}
\qquad\text{(ii)}
$$

Therefore

$$\frac{p}{x(1-k)(k-y)} = \frac{q}{y(1-k)(k-x)} = \frac{r}{2(k-x)(k-y)}$$

$$= \frac{1}{k[(k-x)(1-y)+(k-y)(1-x)]} \qquad\text{(iii)}$$

Substituting these results in equations (i) gives six linear equations in $(\lambda,\ \mu,\ \alpha,\ \beta,\ \gamma,\ \delta)$, their solution being functions of k. Then using the

equation $\lambda+\mu+\alpha+\beta+\gamma+\delta = 1$ an equation is obtained for k in terms of the parameters x, y, s and t. Thus k is known and hence so is the equilibrium distribution. The elimination of λ, μ, α, β, γ, δ is easily performed by matrix methods. Together with $\lambda+\mu+\alpha+\beta+\gamma+\delta = 1$, equations (i) for equilibrium may be written

$$k(s+tk)\begin{pmatrix}\lambda\\1\end{pmatrix} = \begin{pmatrix}s\mathbf{A} & k^2t\mathbf{z}\\k(s+tk)\mathbf{u} & 0\end{pmatrix}\begin{pmatrix}\lambda\\1\end{pmatrix}$$

or $k(s+tk)\begin{pmatrix}\lambda\\1\end{pmatrix} = B\begin{pmatrix}\lambda\\1\end{pmatrix}$ \hfill (iv)

with $\mathbf{u} = (1, 1, 1, 1, 1, 1)$ and $\mathbf{z} = \begin{bmatrix}p^2\\q^2\\2pq\\r^2\\2rp\\2qr\end{bmatrix}$ $\begin{array}{l}p = p(k)\\q = q(k)\\r = r(k)\end{array}$

The equation determining k may be written

$$|\mathbf{B}-k(s+tk)\mathbf{I}| = 0.$$

$\mathbf{I}$ is the 7×7 identity matrix. When $t = 0$, the equation reduces to $|\mathbf{A}-k\mathbf{I}| = 0$, the usual characteristic equation for pure sib mating (Hayman and Mather, 1953, 4.2).

With the appropriate value of k, solving the equations (iv) yields the equilibrium frequencies λ.

The expansion of the determinant is

$$(k-x)(k-y)f(k, x, y, s) = 0,$$

$f(k, x, y, s)$ being an eighth degree polynomial in k.

If $k = x$, $\lambda = (1, 0, 0, 0, 0, 0)$ and the genotype distribution is $(1, 0, 0)$. Thus the population becomes homozygous aa. Similarly $k = y$ corresponds to a homozygous bb population. The roots of $f(k, x, y, s) = 0$ give rise to populations λ containing both heterozygotes and homozygotes, only one of which, however, is a real population.

λ defines a real, stable population equilibrium if it corresponds to the maximum root of the determinantal equation. The maximum root of $f(k, x, y, s) = 0$ lies between the least of x and y, and unity and defines a real population with genotypic frequencies given by equations (iii). This is the stable equilibrium only if $k>x, y$; otherwise one of the homozygous states is the stable solution.

Fig. 1 shows the type of equilibrium population for various values of x, y and s. The areas marked A and B correspond to the homozygous states aa and bb respectively; in C the population, though containing heterozygotes, has more homozygotes than a population

mating purely at random without selection; in D it contains more heterozygotes; on the boundary between C and D the effects of inbreeding and selection exactly counter each other.
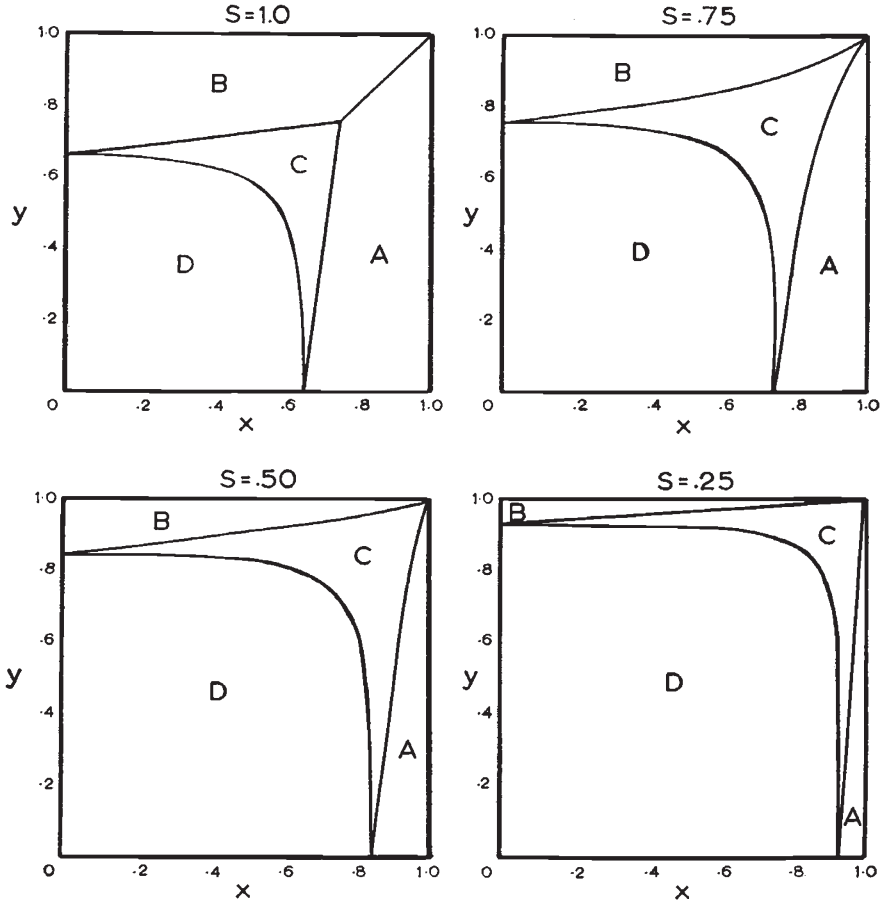

FIG. 1.—Phase diagram showing types of equilibrium with various proportions of sib mating (s) and random mating $(t = 1-s)$. $x =$ survival rate of one homozygote (aa) relative to that of the heterozygote (ab). $y =$ relative survival rate of the other homozygote (bb).

When the point (x, y) is in A, the population becomes homozygous aa, in B, homozygous bb. In C, the population becomes less heterozygous than a population mating at random without selection, and in D, more heterozygous.

The inner boundaries of A and B are where two of the three values of k determining the types of equilibrium coincide. Thus their equations are

$$f(x, x, y, s) = 0 \quad \text{and} \quad f(y, x, y, s) = 0.$$

On the boundary between C and D Hardy's law holds, i.e. $4pq = r^2$. Thus using this equation with equations (iii), the boundary is

$$f\left(\frac{x+y-2xy}{1-xy}, x, y, s\right) = 0.$$

Unless one allele is initially absent, the equilibrium frequencies are independent of the initial state, being determined by x, y and s only.

4. EQUAL VIABILITIES

When $x = y$ some simplification in the equilibrium distribution is obtained. The first of equations (ii) now becomes
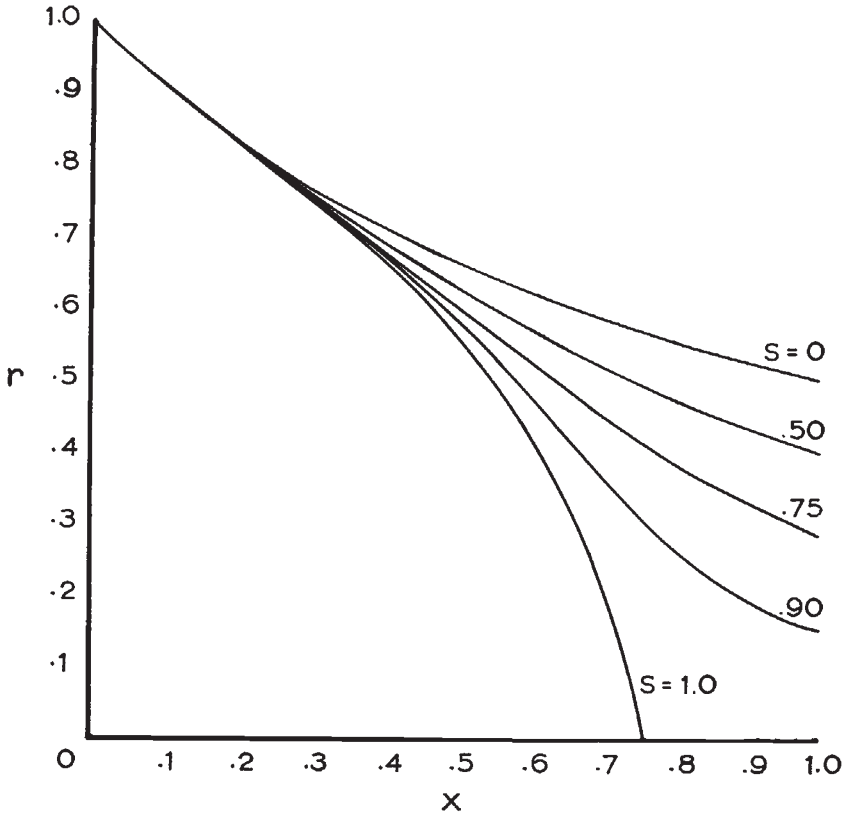
$$(k - x)(p - q) = 0.$$

FIG. 2.—Equilibrium with equal viabilities of the homozygotes. r = proportion of heterozygotes in the population. x = survival rate of both homozygotes relative to the heterozygote. s = proportion of sib mating.

If $k = x$ the equilibrium is homozygous with respect to aa or bb, and these two distributions are unstable, attained only if the initial population has one allele missing. Thus $p = q$ and hence $\lambda = \mu$, and $\gamma = \delta$. Thus in equilibrium the gene frequencies are equal and

$$p = q = \frac{1 - k}{2k(1 - x)}, \qquad r = \frac{k - x}{k(1 - x)} \qquad \text{(v)}$$

The mating type pairs $aa \times aa$, $bb \times bb$; and $aa \times ab$, $bb \times ab$ may be grouped together, and the matrix **B** is

			$aa \times aa$	$aa \times bb$	$ab \times ab$	$aa \times ab$	$tk^2\mathbf{z}$
$aa \times aa$	.	.	sx	.	$\dfrac{sx^2}{4(x+1)}$	$\dfrac{sx^2}{2(x+1)}$	$2tk^2p^2$
$aa \times bb$	.	.	.	.	$\dfrac{sx^2}{4(x+1)}$	.	$2tk^2p^2$
$ab \times ab$	.	.	.	s	$\dfrac{s}{2(x+1)}$	$\dfrac{s}{2(x+1)}$	tk^2r^2
$aa \times ab$	.	.	.	.	$\dfrac{sx}{x+1}$	$\dfrac{sx}{x+1}$	$4tk^2rp$
$k(s+kt)\mathbf{u}$	.	.	$k(s+kt)$	$k(s+kt)$	$k(s+kt)$	$k(s+kt)$	.

The determinantal equation for k is

$$4t^2(1+x)^2k^6 + 2t(1+x)^2\{4s - (1+x)t\}k^5 + 2s\{2(1+x)s - (2+6x+3x^2)t\}k^4$$
$$+ s(1+x)\{-2(1+4x+2x^2)s + x(2+3x)t\}k^3 + x(1+x)(2+3x)s^2k^2$$
$$+ sx^3\{(2+x)s + (1+x)t\}k - s^2x^4 = 0 \qquad \text{(vi)}$$

This equation is also obtained by letting $x = y$ in the general equation $f(k, x, y, s) = 0$. The maximum root lies between x and $\frac{1}{2}(1+x)$, and with this value the genotype frequencies are

$$\left(\tfrac{1}{2}(1-r), \tfrac{1}{2}(1-r), r\right) \text{ with } r = (k-x)/k(1-x).$$

Graphs of r against x for various s are shown in fig. 2.

5. COMPARISON WITH SELFING

Selfing is the more powerful inbreeding system and a higher proportion of random mating is needed to obtain a given frequency of heterozygotes than in a mixed sib and random mating scheme with the same value, x for the viability of the homozygotes. Fig. 3 is a graph of s_1, the proportion of selfing, against s_2, the proportion of sib mating which gives the same frequency r of heterozygotes in the equilibrium population, for different values of x. Hayman (1953) gives the equation determining r under mixed selfing and random mating, with equal homozygote viabilities as $rx(1+s_1 - s_1r) = (1-r)(1-s_1+s_1r)$.

Eliminating r and k between this and equations (v) and (vi) gives the relationship between s_1, s_2 and x.

The ratio $(s_2 - s_1)/s_2$ gives a measure of the relative inbreeding effectiveness of the two systems and this in turn can be measured by

N

the closeness of the (s_1, s_2) curve to the dotted line $s_1 = s_2$. For low homozygote survival rates the curve closely approximates to a straight line whose gradient, s_1/s_2 is constant. For 50 per cent. survival, selfing appears to be about 50 per cent. more effective; for 30 per cent. survival the figure is 80 per cent.; selfing is 90 per cent. more effective
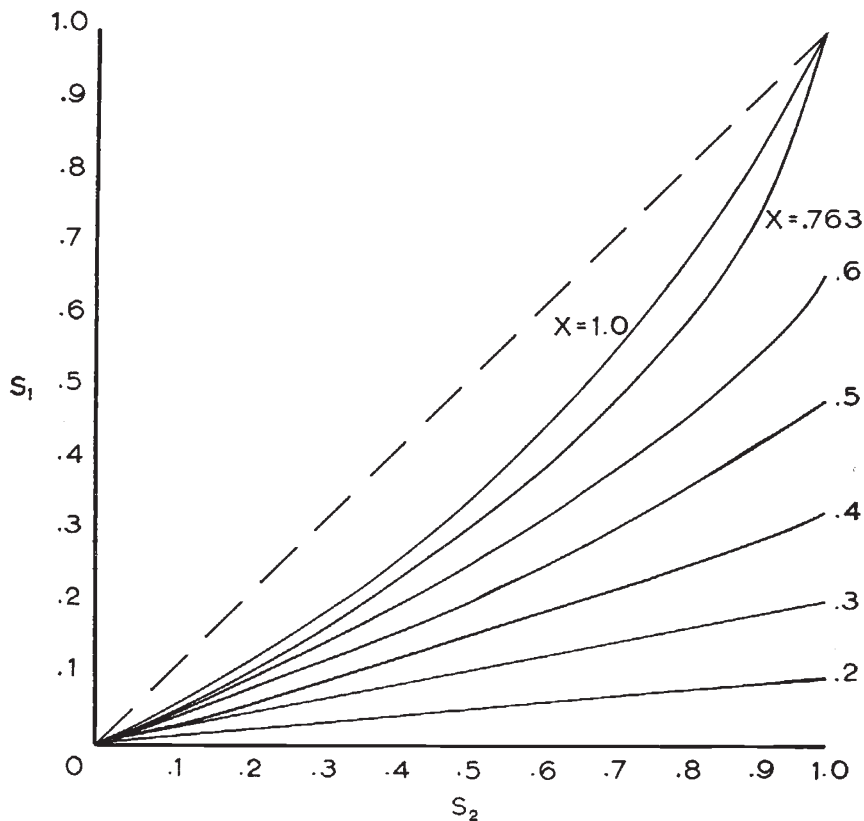

FIG. 3.—Comparison of a mixed sib and random mating system with a mixed selfing and random mating system. $s_1 =$ proportion of selfing. $s_2 =$ proportion of sib mating required to produce the same equilibrium distribution as the mixed selfing system. $x =$ relative survival rate of both homozygotes in either mating system.

Along the broken line, $s_1 = s_2$. The deviation of the (s_1, s_2) curve from this line is a measure of the difference between the inbreeding effectivenesses of the two systems. For small x the curve approximates to a straight line whose gradient measures the effectiveness of the mixed sib mating system relative to that of the mixed selfing system.

For $x \geqslant 0.763$ the curves meet the line $s_1 = s_2$ at the point $(1, 1)$. All other curves deviate entirely from this line.

for 20 per cent. survival and is almost 100 per cent. more effective for less than 10 per cent. survival of the homozygotes.

5. NO SELECTION OPERATING

The method used above breaks down when $x = y = 1$, but in this case the approach to equilibrium can be described. The equations

linking the mating type frequencies between the nth and $(n+1)$th generations are

$$\lambda_{n+1} = s(\lambda_n + \tfrac{1}{16}\beta_n + \tfrac{1}{4}\gamma_n) \qquad + t p_{n+1}^2$$

$$\mu_{n+1} = s(\mu_n + \tfrac{1}{16}\beta_n + \tfrac{1}{4}\delta_n) \qquad + t q_{n+1}^2$$

$$\alpha_{n+1} = s(\tfrac{1}{8}\beta_n) \qquad + 2 t p_{n+1} q_{n+1}$$

$$\beta_{n+1} = s(\alpha_n + \tfrac{1}{4}\beta_n + \tfrac{1}{4}\gamma_n + \tfrac{1}{4}\delta_n) + t r_{n+1}^2$$

$$\gamma_{n+1} = s(\tfrac{1}{4}\beta_n + \tfrac{1}{2}\gamma_n) \qquad + 2 t r_{n+1} p_{n+1}$$

$$\delta_{n+1} = s(\tfrac{1}{4}\beta_n + \tfrac{1}{2}\delta_n) \qquad + 2 t q_{n+1} r_{n+1}$$

where

$$p_{n+1} = \lambda_n + \tfrac{1}{4}\beta_n + \tfrac{1}{2}\gamma_n$$

$$q_{n+1} = \mu_n + \tfrac{1}{4}\beta_n + \tfrac{1}{2}\delta_n$$

$$r_{n+1} = \alpha_n + \tfrac{1}{2}\beta_n + \tfrac{1}{2}\gamma_n + \tfrac{1}{2}\delta_n$$

and are the genotype frequencies in the $(n+1)$th generation.

From these equations it follows that

$$4 p_{n+2} - 2 s p_{n+1} - s p_n = u(s + 4 t u)$$

$$4 q_{n+2} - 2 s q_{n+1} - s q_n = v(s + 4 t v) \qquad \text{(vii)}$$

$$4 r_{n+2} - 2 s r_{n+1} - s r_n = 8 t u v$$

where u and v are the (constant) frequencies of the genes a and b (assuming no mutation). Solving these difference equations gives the distribution at any generation in terms of the initial distribution.

$$p_n = A_1 \epsilon_1{}^n + B_1 \epsilon_2{}^n + u(s + 4 t u)/(1 + 3 t) \rightarrow u^2 + u v s/(1 + 3 t)$$

$$q_n = A_2 \epsilon_1{}^n + B_2 \epsilon_2{}^n + v(s + 4 t u)/(1 + 3 t) \rightarrow v^2 + u v s/(1 + 3 t)$$

$$r_n = A_3 \epsilon_1{}^n + B_3 \epsilon_2{}^n + t u v/(1 + 3 t) \rightarrow 2 u v - 2 u v s/(1 + 3 t).$$

The constants A_i, B_i are determined by the initial mating type frequencies. ϵ_1 and ϵ_2 are the roots of

$$4 z^2 - 2 s z - s = 0 \qquad \text{(viii)}$$

and hence $|\epsilon_i| < 1$ for $|s| \leqslant 1$.

The form of the limiting or equilibrium frequencies indicates the departure from the random mating population $(u^2, v^2, 2uv)$. The equilibrium distribution is dependent only on the initial gene frequencies.

For a given value of s the equilibrium point, as represented in a trilinear diagram based on an equilateral reference triangle, lies on the parabola $\tfrac{1}{2} s r = t(4 p q - r^2)$. Approach to this equilibrium parabola is along a straight line perpendicular to the base of the reference triangle from the initial point. The distance from the equilibrium decreases geometrically each generation.

This approach to equilibrium is very similar to that of mixed selfing and random mating with no selection (Hayman, 1953). We may also expect the approach under selection to be qualitatively similar for the genotypes in the two cases and so the general remarks by Hayman (1953, 4.2) should also apply here.

The mating types follow a much more complicated approach to equilibrium than the genotypes, being determined by a seventh order linear difference equation compared with the second order equation (vii) for the genotypes. Its characteristic equation contains equation (viii) as a factor. The equilibrium frequencies depend only on the gene frequencies u and v.

7. SUMMARY

The equilibrium is investigated of a large population under a system of mixed sib and random mating with selection against the homozygotes. Diagrams are produced giving the types of equilibrium and comparing this system with a system of mixed selfing and random mating.

8. REFERENCES

HAYMAN, B. I.   1953.   Mixed selfing and random mating when homozygotes are at a disadvantage. *Heredity*, 7, 185-192.

HAYMAN, B. I., AND MATHER, K.   1953.   The progress of inbreeding when homozygotes are at a disadvantage. *Heredity*, 7, 165-183.