ASSOCIATION BETWEEN MENDELIAN FACTORS WITH MIXED SELFING AND RANDOM MATING

J. H. BENNETT
Department of Genetics, Cambridge University
and
F. E. BINET
Department of Statistics, Melbourne University

Received 20.vii.55

In most plant species in which reproduction is effected both by cross- and self-fertilisation, the relative contributions from these two kinds of matings depend to some extent on climatic and other environmental conditions and so vary from one generation to the next. In the present article, however, we shall confine our attention to the simplified but fundamental situation in which these contributions bear a constant ratio to one another throughout all generations.

Consider a large population of plants reproducing in non-overlapping generations in such a way that there is a constant probability s that any plant will be self-fertilised and a probability $1-s$ that it will cross with some plant chosen at random from the population. We shall suppose that all crosses are equally fertile and all genotypes equally viable. With this system of mating, gene frequency clearly remains constant. Suppose A and a are the genes present at some locus and $p(\mathrm{A})$, $p(a)$, $[p(\mathrm{A})+p(a) = 1]$ their frequencies in the population. If $p_n(\mathrm{A}a)$ denotes the frequency of occurrence of the genotype Aa in the nth generation, we have

$$p_{n+1}(\mathrm{A}a) = \tfrac{1}{2}sp_n(\mathrm{A}a)+2(1-s)p(\mathrm{A})p(a)$$

and for the genotype AA,

$$p_{n+1}(\mathrm{AA}) = s[p_n(\mathrm{AA})+\tfrac{1}{4}p_n(\mathrm{A}a)]+(1-s)p(\mathrm{A})^2.$$

Hence, if we put

$$\mathrm{K}_n = p_n(\mathrm{A}a)-\frac{4(1-s)}{2-s}\,p(\mathrm{A})p(a),$$

then

$$\mathrm{K}_n = (\tfrac{1}{2}s)^n\mathrm{K}_0.$$

At equilibrium, K = 0,

$$i.e.\ p_\infty\ (\mathrm{A}a) = \lim_{n\to\infty} p_n(\mathrm{A}a) = \frac{4(1-s)}{2-s}\,p(\mathrm{A})p(a)$$

$$\text{and } p_\infty\ (\mathrm{AA}) = \lim_{n\to\infty} p_n(\mathrm{AA}) = p(\mathrm{A})\left[1-\frac{2(1-s)}{2-s}\,p(a)\right]$$

$$= p(\mathrm{A})^2+\frac{s}{2-s}\,p(\mathrm{A})p(a).$$

These equilibrium genotypic frequencies are those which would exist in a composite population comprising a random mating component supplying the proportion $\frac{2(1-s)}{2-s}$ of the whole population and a selfing component making up the remainder. Genotypic equilibrium is not attained in one generation as with random mating but is approached at a rate measured by $-\log_e(\frac{1}{2}s)$.

If B and b are two genes present at a second locus and y is the frequency of recombination between the two loci, the frequency of occurrence of a given chromosome or gene-combination such as AB in the genotypes of the $(n+1)$-th generation is

$$p_{n+1}(\mathrm{AB}) = p_n(\mathrm{AB}) + \tfrac{1}{2}y[p_n(\mathrm{A}b/a\mathrm{B}) - p_n(\mathrm{AB}/ab)].$$

It can also be shown that

$$\begin{aligned} p_{n+1}(\mathrm{A}b/a\mathrm{B}) - p_{n+1}(\mathrm{AB}/ab) &= \tfrac{1}{2}(s-2y)[p_n(\mathrm{A}b/a\mathrm{B}) - p_n(\mathrm{AB}/ab)] \\ &\quad + 2(1-s)[p(\mathrm{A})p(\mathrm{B}) - p_n(\mathrm{AB})]. \end{aligned}$$

Hence, if we put

$$u_n = p(\mathrm{A})p(\mathrm{B}) - p_n(\mathrm{AB})$$

and

$$v_n = p_n(\mathrm{A}b/a\mathrm{B}) - p_n(\mathrm{AB}/ab),$$

we have

$$\begin{pmatrix} u_n \\ \hline v_n \end{pmatrix} = \mathrm{Z}^n . \begin{pmatrix} u_0 \\ \hline v_0 \end{pmatrix}$$

where Z stands for the matrix

$$\begin{pmatrix} 1 & -\frac{1}{2}y \\ \hline 2(1-s) & \frac{1}{2}(s-2y) \end{pmatrix}$$

The matrix Z has two positive and unequal latent roots λ, μ which satisfy the equation,

$$2z^2 - (2+s-2y)z + s(1-2y) = 0.$$

The latent roots are therefore

$$\frac{\lambda}{\mu} = \tfrac{1}{4}\{2+s-2y \pm \sqrt{[(2-s-2y)^2 + 8sy]}\}.$$

Clearly, if neither $1-s$ nor y is zero, $\lambda(>\mu)$ is less than unity and

$$\lim_{n\to\infty} p_n(\mathrm{AB}) = p(\mathrm{A})p(\mathrm{B})$$

$$\lim_{n\to\infty} [p_n(\mathrm{A}b/a\mathrm{B}) - p_n(\mathrm{AB}/ab)] = 0.$$

A measure of the rate of approach to equilibrium is given by $-\log_e\lambda$ (see table 1).

Corresponding with the two latent roots λ and μ, there exist two principal linear components of frequency,

$$\mathrm{L}_n = 2(1-s)u_n - (1-\lambda)v_n = \lambda^n.\mathrm{L}_0$$

$$\mathrm{M}_n = 2(1-s)u_n - (1-\mu)v_n = \mu^n.\mathrm{M}_0$$

from which the values of u_n and v_n can readily be determined for any positive integer n.

When $s = 1$, *i.e.* with complete selfing, the latent roots are 1 and $\frac{1}{2}-y$, the corresponding principal components of frequency being

$$(1+2y)u_n - yv_n = \text{constant},$$

and

$$v_n = (\tfrac{1}{2}-y)^n.v_0.$$

Hence, in this case,

$$p_\infty(\mathrm{AB}) = p_\infty(\mathrm{AB}/\mathrm{AB}) = p_0(\mathrm{AB}) + \frac{y}{1+2y}[p_0(\mathrm{A}b/a\mathrm{B}) - p_0(\mathrm{AB}/ab)]$$

giving the limiting genotypic frequencies when complete homozygosity has been attained. In general, when $s = 1$,

$$p_\infty(\mathrm{AB}) \neq p(\mathrm{A})p(\mathrm{B}).$$

With complete random mating, *i.e.* $s = 0$, the latent roots are $1-y$ and 0, the corresponding principal components of frequency being

$$\begin{aligned} u_n - \tfrac{1}{2}y\, v_n &= (1-y)^n . \, [u_0 - \tfrac{1}{2}y \, . \, v_0] \\ u_n - \tfrac{1}{2}v_n &= 0 \quad (n > 0). \end{aligned}$$

Hence

$$\begin{aligned} p(\mathrm{A})p(\mathrm{B}) - p_n(\mathrm{AB}) &= \tfrac{1}{2}[p_n(\mathrm{A}b/a\mathrm{B}) - p_n(\mathrm{AB}/ab)] \\ &= (1-y)^{n-1}[p(\mathrm{A})p(\mathrm{B}) - p_0(\mathrm{AB}) - \tfrac{1}{2}p_0(\mathrm{A}b/a\mathrm{B}) + \tfrac{1}{2}p_0(\mathrm{AB}/ab.)] \end{aligned}$$

TABLE 1

Rate of approach to equilibrium at two linked loci with recombination frequency y *in a population with 100* s *per cent. self-fertilisation and 100 (1−*s*) per cent. random mating*

	Recombination frequency y											
Frequency s of self-fertilisation	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50	0·55	0·60
0·00	·0513	·1054	·1625	·2231	·2877	·3567	·4308	·5108	·5978	·6932	·7985	·9163
0·05	·0499	·1022	·1572	·2152	·2764	·3412	·4099	·4831	·5611	·6444	·7337	·8291
0·10	·0484	·0989	·1517	·2069	·2647	·3253	·3888	·4553	·5250	·5978	·6737	·7524
0·20	·0452	·0918	·1398	·1894	·2403	·2925	·3459	·4002	·4553	·5108	·5664	·6213
0·30	·0416	·0839	·1269	·1705	·2144	·2584	·3023	·3459	·3888	·4308	·4715	·5108
0·40	·0376	·0753	·1129	·1502	·1870	·2231	·2584	·2925	·3253	·3567	·3865	·4147
0·50	·0331	·0657	·0975	·1285	·1584	·1870	·2144	·2403	·2647	·2877	·3092	·3292
0·60	·0281	·0551	·0809	·1054	·1285	·1502	·1705	·1894	·2069	·2231	·2382	·2523
0·70	·0224	·0433	·0628	·0809	·0975	·1129	·1270	·1398	·1517	·1625	·1725	·1808
0·80	·0159	·0303	·0433	·0551	·0657	·0753	·0839	·0918	·0989	·1054	·1113	·1165
0·90	·0084	·0159	·0224	·0281	·0332	·0377	·0416	·0453	·0484	·0512	·0538	·0537
0·95	·0044	·0081	·0114	·0142	·0166	·0188	·0207	·0224	·0239	·0253	·0266	·0277

With mixed selfing and random mating ($s \neq 1$), the genotypic frequencies at equilibrium are given by

$$\begin{aligned} p_\infty(\mathrm{AB}/ab) &= p_\infty(\mathrm{A}b/a\mathrm{B}) = \\ &\qquad 4(1-s)p(\mathrm{A})p(a)p(\mathrm{B})p(b)/[2-s+2sy(1-y)] \\ p_\infty(\mathrm{AB}/\mathrm{A}b) &= -p_\infty(\mathrm{AB}/ab) + 4(1-s)p(\mathrm{A})p(a)p(\mathrm{B})p(b)/(2-s) \\ p_\infty(\mathrm{AB}/\mathrm{AB}) &= \\ &\qquad \tfrac{1}{2}p_\infty(\mathrm{AB}/ab) + p(\mathrm{A})p(\mathrm{B}) - 2(1-s)p(\mathrm{A})p(\mathrm{B})[p(a)+p(b)]/(2-s). \end{aligned}$$

The equilibrium frequencies for the other genotypes are readily obtainable from these expressions on making appropriate gene substitutions. We see that when equilibrium is attained, linked factors are associated to an extent that depends not only on the proportion of selfing but also on the magnitude of the recombination fraction. In such a case, it might be possible to determine the recombination fraction between two loci from an analysis of the population frequencies. When $y = \frac{1}{2}$, as for two loci on different chromosomes,

$$p_\infty(\mathrm{A}a\mathrm{B}b) - p_\infty(\mathrm{A}a)p_\infty(\mathrm{B}b) = \frac{16s(1-s)}{(4-s)(2-s)^2}p(\mathrm{A})p(a)p(\mathrm{B})p(b)$$

$$p_\infty(\mathrm{AAB}b) - p_\infty(\mathrm{AA})p_\infty(\mathrm{B}b) = \frac{-8s(1-s)}{(4-s)(2-s)^2}p(\mathrm{A})p(a)p(\mathrm{B})p(b)$$

$$p_\infty(\mathrm{AABB}) - p_\infty(\mathrm{AA})p_\infty(\mathrm{BB}) = \frac{4s(1-s)}{(4-s)(2-s)^2}p(\mathrm{A})p(a)p(\mathrm{B})p(b).$$

Two factors which segregate independently are thus not associated at random when equilibrium is reached. There is, in fact, a positive association between the homozygous (or heterozygous) states at loci on different chromosomes. This association may be represented as in table 2 which gives the excess of the equilibrium frequencies for genotypes at two independently segregating loci over the product of the equilibrium frequencies at the separate loci as a multiple of

TABLE 2

Association between the homozygous and heterozygous conditions at two independent loci in a population experiencing mixed selfing and random mating

	AA	A*a*	*aa*
BB	1	−2	1
B*b*	−2	4	−2
bb	1	−2	1

$$\frac{4s(1-s)}{(4-s)(2-s)^2}p(\mathrm{A})p(a)p(\mathrm{B})p(b).$$

For given gene frequencies, this excess has its greatest value

$$0{\cdot}1506\,p(\mathrm{A})p(a)p(\mathrm{B})p(b)$$

when $s = 0{\cdot}6946$. (See fig. 1.)

There are results analogous to these for randomly mating populations of polysomic organisms where, for our present purposes, double reduction may be considered as equivalent to a small amount of selfing. It has been shown (Bennett, 1954) that at equilibrium in randomly mating populations of tetrasomic or hexasomic organisms the probability of occurrence of any chromosome or gene-combination, such as AB, equals the product of the appropriate gene frequencies, *i.e.* $p_\infty(\mathrm{AB}) = p(\mathrm{A})\,p(\mathrm{B})$. If there is no double reduction, linked loci will therefore be associated at random in the gametes or zygotes at equilibrium. However, when double reduction can occur, linked loci may not be associated at random in the gametes and zygotes at equilibrium though loci on different chromosomes certainly will be.

SUMMARY

In a population that experiences mixed selfing and random mating, genes are associated at random on the chromosomes at equilibrium but there is a positive association between the genotypic state (homozygous or heterozygous) at different loci even when the two loci are on different chromosome, *i.e.* there is a tendency for heterozygosis to become concentrated in some members of the population.

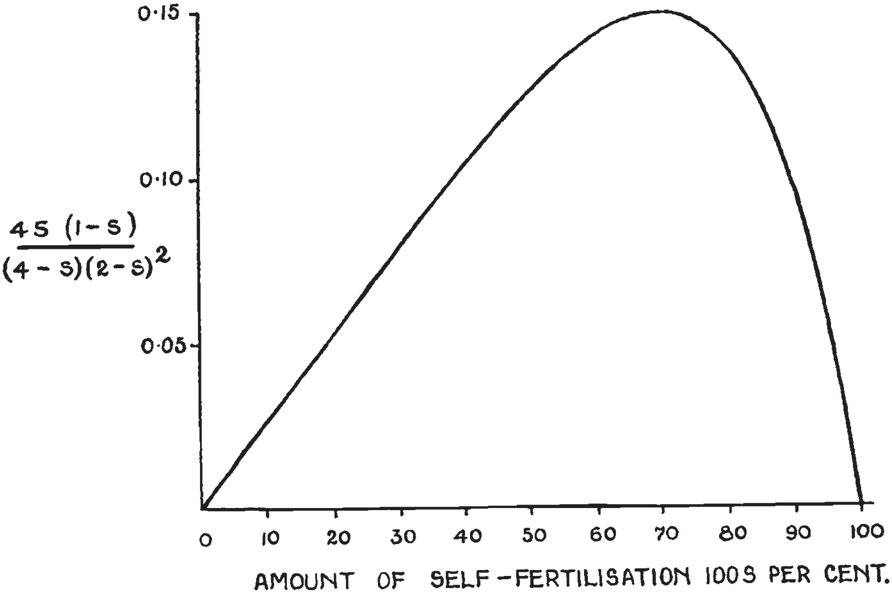

FIG. 1.—Excess of the equilibrium frequency of double homozygotes for two independent loci over the product of the equilibrium frequencies of single homozygotes at these two individual loci expressed as a multiple of the gene frequency product $p(\mathrm{A})p(a)p(\mathrm{B})p(b)$ and as a function of the amount of self-fertilisation.

Acknowledgment.—The authors are grateful to Dr H. T. Clifford, University College of Nigeria and formerly of the University of Melbourne, for some initial suggestions and a discussion of relevant botanical aspects of the problem.

REFERENCE

BENNETT, J. H. 1954. Panmixia with tetrasomic and hexasomic inheritance. *Genetics*, 39, 150.