PRODUCTION OF PURE LINES IN BEES

By H. KALMUS and C. A. B. SMITH

(With Three Text-figures)

It is little known that for many years Mendel tried to produce pure lines in bees but did not succeed (Iltis, 1924). Indeed, the production of pure lines in bees (*Apis mellifica* L.) remained impracticable because of the impossibility of controlling the mating of bees, until it was made possible by the perfection of a technique of artificial insemination (Nolan, 1937). It is now also possible to rear up to five generations of bees per year in small nuclear colonies. It is therefore most important to work out what method of mating will give us the quickest approach to the pure line, and in doing so, we must take into account the following considerations:

(1) In bees the male is haploid, i.e. has only one set of chromosomes, so *all* genes must follow the 'sex-linked' type of inheritance (see Whiting, 1945).

(2) It is difficult to produce males and females at the same time from the same mother. Consequently the standard method of inbreeding by 'brother-sister' mating may be difficult, and it will be necessary to consider other mating systems, such as continued 'mother-son' or 'aunt-nephew' matings.

(3) Besides male haploidy there might also be some system of genes determining sex, similar to those in *Habrobracon*. Since in *Habrobracon* the homozygotes for these genes are sterile males, it may be difficult to obtain fertile homozygotes for genes situated near these sex genes, and on the same chromosome. Genes situated far away or on other chromosomes will, however, behave normally. The situation would be similar if there exist genes which are lethal when homozygous. Such genes, according to Mackensen reported by Whiting (1947), appear to occur in the honey-bee. On the other hand, it is difficult to see how the nuptial flight and mating habits of the honey-bee can result in consistent outbreeding. Thus it seems reasonable to assume that honey-bees are fairly inbred, and that it would be possible to achieve a high degree of homozygosity in the species. We shall now consider the possible mating systems in more detail.

A. Mother-son mating

In this system we start with an unfertilized female f_0 which is allowed to lay eggs, which will accordingly develop into males. One of these males, m_0, is now mated with f_0, and the resulting fertilized eggs will become females. One of these females, f_1, is taken, and the whole process repeated, getting a son m_1 of f_1, and a daughter f_2 by the mating $m_1 \times f_1$, and so on.

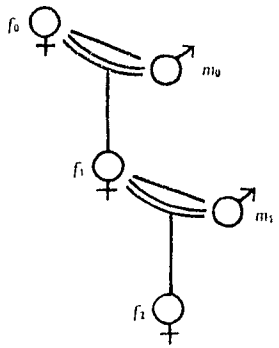

Fig. 1.

Each step from f_r to f_{r+1} may conveniently be referred to as a *double generation*.

Now consider any particular locus in any chromosome. The same gene **A**, say, may occur at this locus in both chromosomes of f_0 (i.e. f_0 is homozygous for **A**). In this case, neglecting the unlikely possibility of mutation, every descendant of f_0 will have the gene **A** at this locus on both her chromosomes of that kind,

if female, or on the one chromosome of that kind, if male. Thus we shall have a pure breeding stock for this gene. Alternatively f_0 may be heterozygous, having one gene **A** at the locus in one chromosome, and an allelomorph of **A**, say **A′**, at the same locus in the other chromosome. In this case m_0 will receive one of the genes **A** and **A′** from f_0, and will in turn pass it on to f_1. The daughter f_1 will also receive one of the genes **A**, **A′** directly from her mother f_0, independently of the gene derived from m_0. Thus f_1 can have four genetic constitutions with equal probability $\frac{1}{4}$ (writing the gene derived from f_0 first), **AA**, **AA′**, **A′A**, **A′A′**; but **AA′** and **A′A** are genetically indistinguishable. Hence, provided that these four possible types are equally viable, there is a probability $\frac{1}{2}$ that f_1 is homozygous, of type **AA** or **A′A′**, and a probability $\frac{1}{2}$ that f_1 is heterozygous **AA′**. If, however, **A** is closely linked to a gene **G**, say, which causes death or sterility in homozygous form, then most individuals **AA** would also be of type **GG**, and would not be chosen to act as f_1, i.e. there would not be such a high probability of homozygosity in the first double generation. Neglecting this complication we see, however, that, since homozygotes produce homozygotes, the probability that the bee is still heterozygous is halved at each double-generation, and so after n double generations is only 2^{-n}.

Mathematically this case is exactly equivalent to self-fertilization in plants (Mendel, 1865; Jennings, 1916), one double-generation of mother-son mating corresponding to a single generation of self-fertilization.

If, instead of one pair of alleles, we consider m pairs of unlinked genes, on different chromosomes, then since after n double generations the chance that any one pair will be homozygous is $(1-2^{-n})$, the chance for m pairs will be $(1-2^{-n})^m$, or, nearly enough, $1-m.2^{-n}$, when n is fairly large.

A similar result will hold provided that the genes are not situated too close together on the chromosome; again, the probability of heterozygosity will be halved at each double generation, except for the first few generations. Haldane (1936) has shown by a rough argument that if we start with a pair of chromosomes of genetical length 100 centimorgans,* and heterozygous for numerous gene pairs, then after n double generations (equivalent to the n generations of self-fertilization considered by Haldane) there is approximately a probability $(1-2n.2^{-n})$ that we shall have complete homozygosity for all the gene pairs. If, therefore, we suppose that there were $2N$ chromosomes altogether in the female (N in the male), each of genetical length 100 centimorgans, we see that after n generations the probability of complete homozygosity will be approximately $(1-2n.2^{-n})^N$, or nearly enough $(1-2nN.2^{-n})$. Since, to a first rough approximation, the two ends of a chromosome of 200 centimorgans length will behave independently, we may consider it as behaving very like two separate chromosomes of 100 centimorgans each. Thus we may say that the probability of complete homozygosity after n generations for any arrangement of chromosomes will be approximately $(1-Kn.2^{-n})$, where K is a constant of the same order of magnitude as the total length of chromosome in the female, measured in units of 1 morgan.

As an example, suppose that we breed for 20 double generations; and suppose we assume a value of 10 for K. Then in any one gene pair the chance of heterozygosity remaining is only $2^{-20}=9{\cdot}5\times10^{-7}$; while the chance of any heterozygosity remaining anywhere,

* Or 'cross-over units of chromosome map distance', defined in such a way that, for small x, two genes on a chromosome between which there is x% recombination are taken to be x centimorgans apart.

neglecting mutation, is $Kn.2^{-n}=10.20.2^{-20}=1{\cdot}9\times10^{-4}$. Clearly, then, this system of inbreeding produces the desired result quickly.

B. BROTHER-SISTER MATING

In this system, as is implied by its name, a female f_0 is mated to one of her brothers m_0. In the next generation one of her daughters f_1 is mated to one of her sons m_1, and so on.

We shall first consider a particular locus in any chromosome, again making the simplifying assumptions that it is not near any lethal genes, or genes for sterility, and that there is no selection for heterozygosity. Clearly we may distinguish four different possible types of mating, as regards genes at this locus. In type P only one gene occurs both in the brother m_r and the sister f_r, say the gene **A**, so that the brother m_r is genotypically **A** and the sister f_r is **AA**. In type Q the sister f_r is homozygous for one gene, while the brother has a different gene; for example, f_r might be **AA** and m_r **A′**.

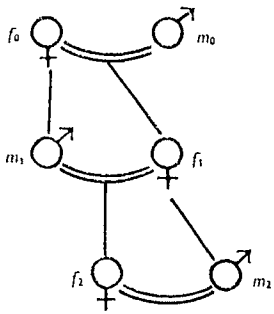

Fig. 2.

In type R the sister is heterozygous, say **AA′**, but has one gene in common with the brother, who is **A**. Finally, in type S all three genes are different; for example, the sister might be **AA′** and the brother **A″**. Each of these types of mating will have a certain probability of occurring in the nth generation. We shall suppose that type P has a probability p_n, type Q a probability q_n, R a probability r_n, and S, s_n. We then readily find, using Haldane's (1937) method for sex-linked genes, that the corresponding probabilities in the $(n+1)$th generation are:

$$\left.\begin{aligned} p_{n+1}&=p_n+\tfrac{1}{4}r_n,\\ q_{n+1}&=\tfrac{1}{4}r_n,\\ r_{n+1}&=q_n+\tfrac{1}{4}r_n+\tfrac{1}{2}s_n,\\ s_{n+1}&=\tfrac{1}{2}s_n. \end{aligned}\right\} \qquad (1)$$

These equations can be solved to give p_n, q_n, r_n and s_n (Haldane, 1937; the same method is used for aunt-nephew mating below). It is clear that once we have reached a mating of type P, such as **AA** × **A**, the same type of mating will persist in all succeeding generations, except for rare mutations, and we have established a pure line.

Thus the quantity we are most interested in is p_n, the probability of a mating of type P, and which may be named the 'purity'. If we take the worst possible case, a mating of type S, to start with, then

$$\begin{aligned} p_n&=1-\frac{(15+7\sqrt{5})}{20}\left(\frac{1+\sqrt{5}}{4}\right)^n-\frac{(15-7\sqrt{5})}{20}\left(\frac{1-\sqrt{5}}{4}\right)^n+2^{-n-1}\\ &=1-1{\cdot}533\ (0{\cdot}8090)^n \text{ for moderately large } n. \end{aligned} \qquad (2)$$

Successive values of p_n are $p_0=0$, $p_1=0$, $p_2=\frac{1}{8}$, $p_3=\frac{1}{4}$, $p_4=\frac{3}{8}$, ..., obeying the recurrence relation

$$p_n=p_{n-1}+\tfrac{1}{8}(1-p_{n-3}),$$

from which further values may be readily calculated. Haldane also shows that the probability h_n that the female f_n is homozygous, after n generations, is

$$h_n=1-\frac{5+3\sqrt{5}}{10}\left(\frac{1+\sqrt{5}}{4}\right)^n-\frac{5-3\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{4}\right)^n.$$

Successive values of h_n are $h_0=0$, $h_1=0$, $h_2=\frac{1}{4}$, $h_3=\frac{3}{8}$, $h_4=\frac{1}{2}$, ..., obeying the same relation as p_n

$$h_n=h_{n-1}+\tfrac{1}{8}(1-h_{n-3}).$$

This is the result for a single gene locus. We may consider linked genes using the rough argument of Haldane (1936), who shows that starting with a segment of 100 centimorgans length, and a mating of type Q, the probability of complete homozygosity in the female after n generations is approximately $1-2n(0{\cdot}8090)^n$. A similar argument in our case might be expected to show that the purity after n generations, i.e. the chance of matings at all loci being of the type P, would be $1-Ln(0{\cdot}8090)^n$, where L is a constant of the same order of magnitude as the total length of chromosome in the female, measured in morgans. (The exact evaluation of L would be difficult and tedious, and the error introduced by this approximation is probably much smaller than that introduced by other assumptions.)

As an example of this we may take $n=40$, $L=10$ (for the sake of argument). Thus after 40 generations we see that the probability of impurity (i.e. mating type Q, R or S) at any one locus is only $1{\cdot}533(0{\cdot}8090)^{40}=0{\cdot}0003$, while the chance of any residual impurity remaining is $10\times 40\times(0{\cdot}8090)^{40}=0{\cdot}08$.

C. Aunt-nephew matings

In this system we start with a mating between a male m_0 and a female f_0, and breed from it two females, say f_1* and f_1. From f_1* we breed a male m_1 which is then mated to f_1, and the whole process is repeated. Each step from one mating to the next may be called a 'double generation'.

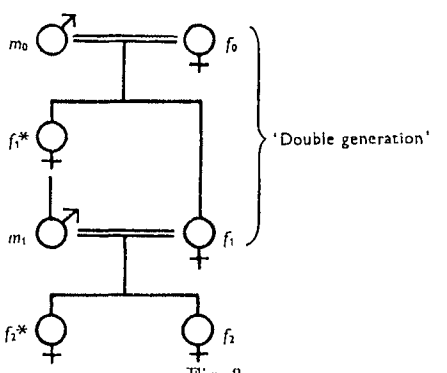

Fig. 3.

Now consider any one locus, at which we may have any one of a series of allelomorphic genes **A**, **A**′, **A**″, etc. With respect to this locus we may have the same mating type as before, viz. the 'pure-breeding' type P, in which only one gene occurs throughout, as, for example, in the mating **A** × **AA**; the type Q, in which the female is homozygous for a gene different from that of the male, as, for example, in **A**′ × **AA**; the type R, in which the male and female have one gene in common, as, for example, in **A** × **AA**′; and finally, the type S, in which all three genes are different, as **A** × **A**′**A**″. We shall again let p_n denote the probability that the mating $m_n \times f_n$ after n double generations is of type P, and similarly q_n, r_n and s_n the probabilities of mating of types Q, R and S respectively. We may then show that after a further double generation these probabilities are given by

$$\left.\begin{aligned} p_{n+1}&=p_n+\tfrac{3}{8}r_n,\\ q_{n+1}&=\tfrac{1}{8}r_n,\\ r_{n+1}&=q_n+\tfrac{1}{2}r_n+\tfrac{3}{4}s_n,\\ s_{n+1}&=\tfrac{1}{4}s_n. \end{aligned}\right\} \qquad (3)$$

For example, it may easily be seen that a mating of type Q can only arise if the immediately preceding mating was of type R, **A** × **AA**′, say, and then only if the two daughters f^*_{n+1} and f_{n+1} of this mating have genotypes **AA**′, and **AA** respectively, and m_{n+1}, the son of f^*_{n+1}, has genotype **A**′. But the probability of getting this series of

genotypes is $\frac{1}{8}$, by the usual Mendelian ratios, and the probability that the previous mating was in fact of type R is r_n; hence the probability of getting type Q is $q_{n+1}=\frac{1}{8}r_n$. Similarly, type P can only arise when the previous mating was of type P or type R, and then with probabilities 1 and $\frac{3}{8}$ respectively. Hence $p_{n+1}=p_n+\frac{3}{8}r_n$.

We now require the general solution of equations (3). We first notice that we may obtain particular solutions of the form $p_n=\lambda^n p_0$, $q_n=\lambda^n q_0$, $r_n=\lambda^n r_0$, $s_n=\lambda^n s_0$, by substituting these values in equations (3), and cancelling out the common factor λ^n.

In this way we obtain four equations, similar in type to equations (3) above, except that on the left-hand sides we have λp_0, λq_0, λr_0, λs_0 in place of p_{n+1}, q_{n+1}, r_{n+1}, s_{n+1}, respectively, and on the right-hand sides p_0, q_0, r_0 and s_0 are to be written for p_n, q_n, r_n and s_n. Eliminating p_0, q_0, r_0 and s_0 we obtain the equation

$$(\lambda-1)\,(32\lambda^3-24\lambda^2+1)=0,$$

which has four roots, say λ_1, λ_2, λ_3, λ_4. Substituting back any one of these roots λ_u, we obtain a possible set of values of p_0, q_0, r_0, s_0 which we shall call p_{0u}, q_{0u}, r_{0u}, s_{0u}. These four particular solutions are given in Table 1 below—they are uniquely determined except that in each of them p_{0u}, q_{0u}, r_{0u}, s_{0u} may be multiplied throughout by an arbitrary constant.

Table 1. *Values of* λ_u, p_{0u}, q_{0u}, r_{0u}, s_{0u}

u	λ_u	p_{0u}	q_{0u}	r_{0u}	s_{0u}
1	1	1	0	0	0
2	$\frac{1}{4}(1+\sqrt{3})$	$-\frac{1}{4}(3+2\sqrt{3})$	$\frac{1}{4}$	$\frac{1}{2}(1+\sqrt{3})$	0
3	$\frac{1}{4}(1-\sqrt{3})$	$-\frac{1}{4}(3-2\sqrt{3})$	$\frac{1}{4}$	$\frac{1}{2}(1-\sqrt{3})$	0
4	$\frac{1}{4}$	$\frac{1}{2}$	$-\frac{1}{2}$	-1	1

Now any linear combination of these solutions will also be a solution of equations (3); i.e. if c_1, c_2, c_3, c_4 are any four constants, then

$$\begin{aligned}
p_n &= c_1 p_{n1}+c_2 p_{n2}+c_3 p_{n3}+c_4 p_{n4}\\
&= c_1 p_{01}\lambda_1^n+c_2 p_{02}\lambda_2^n+c_3 p_{03}\lambda_3^n+c_4 p_{04}\lambda_4^n,\\
q_n &= c_1 q_{01}\lambda_1^n+c_2 q_{02}\lambda_2^n+c_3 q_{03}\lambda_3^n+c_4 q_{04}\lambda_4^n,\\
r_n &= c_1 r_{01}\lambda_1^n+c_2 r_{02}\lambda_2^n+c_3 r_{03}\lambda_3^n+c_4 r_{04}\lambda_4^n,\\
s_n &= c_1 s_{01}\lambda_1^n+c_2 s_{02}\lambda_2^n+c_3 s_{03}\lambda_3^n+c_4 s_{04}\lambda_4^n,
\end{aligned}$$

is a solution. This can be shown to be the general solution, as it is only necessary to choose the constants c_1, c_2, c_3, c_4 in such a way as to fit the initial values of p_0, q_0, r_0, s_0 when $n=0$. Thus if we start off with the worst possible initial mating, one of the type S, we have $p_0=q_0=r_0=0$, $s_0=1$; and this is fitted by taking $c_1=c_2=c_3=c_4=1$. Hence p_n, the probability of a 'pure' mating, is given by

$$\begin{aligned}
p_n &= p_{01}\lambda_1^n+p_{02}\lambda_2^n+p_{03}\lambda_3^n+p_{04}\lambda_4^n\\
&= 1-\tfrac{1}{4}(3+2\sqrt{3})\,[\tfrac{1}{4}(1+\sqrt{3})]^n-\tfrac{1}{4}(3-2\sqrt{3})\,[\tfrac{1}{4}(1-\sqrt{3})]^n+\tfrac{1}{2}(\tfrac{1}{4})^n \text{ exactly}\\
&= 1-1{\cdot}616(0{\cdot}6830)^n \text{ for moderately large } n. \qquad (4)
\end{aligned}$$

Similarly, the probability of homozygosity in the female bee f_n after n double generations is $h_n=p_n+q_n$, since by definition only in mating types P and Q is f_n homozygous. This is

$$h_n=1-\tfrac{1}{2}(1+\sqrt{3})\,[\tfrac{1}{4}(1+\sqrt{3})]^n+\text{further terms}, \qquad (5)$$

or approximately $1-1{\cdot}366(0{\cdot}6830)^n$ for moderately large n. Successive values of p_n and h_n are, $p_0=0$, $p_1=0$, $p_2=\frac{9}{32}$, $p_3=\frac{63}{128}$, ..., and $h_0=0$, $h_1=0$, $h_2=\frac{3}{8}$, $h_3=\frac{9}{16}$, ...; further

values may readily be calculated from the recurrence formulae [which may be verified from equations (4) and (5)]

$$h_n = \tfrac{3}{4}h_{n-1} + \tfrac{1}{32}(9 - h_{n-3}), \quad p_n = \tfrac{3}{4}p_{n-1} + \tfrac{1}{32}(9 - p_{n-3}).$$

By analogy with the previous cases we may also argue that the probability of complete purity after n double generations will be of the form $1 - Mn(0{\cdot}6830)^n$, where M is a constant of the same order of magnitude as the total length of chromosome in the female. Thus taking $n = 20$, and supposing for the sake of argument that $M = 10$, we have after twenty double generations

$$\text{Probability of impurity at one locus} = 1{\cdot}62 \times 0{\cdot}6830^{20} = 0{\cdot}000017,$$
$$\text{Probability of impurity anywhere} = 20 \times 10 \times 0{\cdot}6830^{20} = 0{\cdot}002.$$

Comparison of the three different mating systems

To a first approximation we may summarize our formulae by saying that in mother-son mating the impurity is reduced by a factor $0{\cdot}5^n$, where n is the number of double generations; similarly, in aunt-nephew mating by a factor $0{\cdot}6830^n$, and in n single generations of brother-sister mating by a factor $0{\cdot}8090^n$. Since

$$0{\cdot}5^6 = 0{\cdot}016, \; 0{\cdot}6830^{11} = 0{\cdot}015, \; 0{\cdot}8090^{20} = 0{\cdot}015,$$

we may say that six double generations of mother-son mating are about as effective as eleven double generations of aunt-nephew mating, or twenty single generations of brother-sister mating.

Alternatively, we may use Haldane's index, which is equal to the number of generations or double generations required on the average to reduce impurity to 10% of its former value. Using common logarithms, this is for mother-son mating $-1/\log(0{\cdot}5) = 3{\cdot}32$ double generations, for aunt-nephew $-1/\log(0{\cdot}6830) = 6{\cdot}04$ double generations, and for brother-sister $-1/\log(0{\cdot}8090) = 10{\cdot}86$ generations.

Summary

We discuss various systems of inbreeding, by artificial insemination, with the object of producing a pure line of honey-bees. It is shown that mother-son mating, if practicable, is easily the most efficient system. Brother-sister and aunt-nephew mating systems are also possible, but much slower. They may, however, be easier to work in practice. There is yet not much information as to what extent it is possible to obtain pure lines of bees by inbreeding.

REFERENCES

Haldane, J. B. S. (1936). The amount of heterozygosis to be expected in an approximately pure line. *J. Genet.* **32**, 375–91.

Haldane, J. B. S. (1937). Some theoretical aspects of continued brother-sister mating. *J. Genet.* **34**, 265–74.

Iltis, H. (1924). *Gregor Johann Mendel, Leben, Werk und Wirkung.* 426 pp. Berlin.

Jennings, H. S. (1916). Numerical results of diverse systems of breeding. *Genetics*, **1**, 53.

Mendel, G. (1865). Versuche über Pflanzenhybriden. *Verh. natf. Ver. Brünn.* **4**, 3.

Nolan, W. J. (1937). Bee Breeding. *Yearb. U.S. Dep. Agric.* Separate 1604.

Whiting, P. W. (1945). The evolution of male haploidy. *Quart. Rev. Biol.* **20**, 231–60.

Whiting, P. W. (1947). Some experiments with *Mellitobia* and other wasps. *J. Hered.* **38**, 11–20.