RANDOM MATING WITH THE EXCEPTION OF SISTER BY BROTHER MATING

RAINARD B. ROBBINS

University of Michigan, Ann Arbor, Michigan

[Received March 23, 1918]

In tracing the fate of a Mendelian character, the results for random mating are well known. The results for brother and sister mating have been published. It is at once seen to be of interest to know what will be the results for mating at random excepting that brothers shall not mate with sisters. This paper deals with the problem just stated. For the reader who does not care to follow the details of the discussion, the principal results are included in a summary.

We assume that all families have the same number of individuals, equally divided between males and females and that the division as to dominant (A) and recessive (a) characters is independent of the division as to sex.

Let each family consist of $4k$ male and $4k$ female children, $4k$ being an integer. It is necessary to consider the types of families which arise from different crosses. These are tabulated below.

Type of cross	Composition of resulting family (AA , Aa , aa)	Letter indicating type of family and number of families of this type
$AA \times AA$	$(8k, 0, 0)$	p
$AA \times Aa$	$(4k, 4k, 0)$	q
$AA \times aa$	$(0, 8k, 0)$	r
$Aa \times Aa$	$(2k, 4k, 2k)$	s
$Aa \times aa$	$(0, 4k, 4k)$	t
$aa \times aa$	$(0, 0, 8k)$	u

It seems essential to express the numbers of families of the various types in one generation in terms of the corresponding numbers for the preceding generation. For this purpose we will use capital letters when

referring to one generation and small letters when referring to the preceding generation. If we let

$$f = p + q + r + s + t + u,$$

a study of the possible crosses enables us to write down the following recurrence relations:

$$1) \left\{ \begin{array}{l} P = \frac{k}{f-1} [4pq + 2ps + 4p(p-1) + q(q-1) + qs + \frac{1}{4}s(s-1)] \\ Q = \frac{k}{f-1} [4pq + 8pr + 4ps + 4pt + 2q(q-1) + 4qr + 3qs + 2qt + 2rs + s(s-1) + st] \\ R = \frac{k}{f-1} [2ps + 4pt + 8pu + qs + 2qt + 4qu + \frac{1}{2}s(s-1) + st + 2us] \\ S = \frac{k}{f-1} [q(q-1) + 4qr + 2qs + 2qt + 4r(r-1) + 4rt + 4rs + s(s-1) + 2st + t(t-1)] \\ T = \frac{k}{f-1} [qs + 2qt + 4qu + 2rs + 4rt + s(s-1) + 3st + 4us + 2t(t-1) + 8ru + 4ut] \\ U = \frac{k}{f-1} [\frac{1}{4}s(s-1) + st + 2us + t(t-1) + 4ut + 4u(u-1)] \end{array} \right.$$

It is from these equations that we must derive whatever information we get regarding the outcome of the type of mating under consideration. The complicated nature of the equations makes a direct solution of them seem hopeless. However, certain combinations of them lead to hopeful simplifications and finally yield much desired information. If we add the six equations just as they stand and use the notation

$$F = P + Q + R + S + T + U,$$

we have

$$F = 4kf$$

A moment's thought shows that this must be the case, i.e., if each family has $8k$ children, the number of families in one generation will be $4k$ times the number in the preceding generation, it being assumed that one male mates with one female. If f_n is the number of families in the n th generation we have at once

$$2) \quad f_n = f_0(4k)^n.$$

Let a, b, c be proportional respectively to the number of AA, Aa and aa individuals in a generation. The table giving the composition of the various types of families enables us to choose a, b, c as follows:

$$3) \quad \begin{cases} a = 2p + q + s/2, \\ b = q + 2r + s + t, \\ c = s/2 + t + 2u. \end{cases}$$

We note that $a + b + c = 2f$. By using equations 3) we are able to put equations 1) in a simpler form:

$$4) \quad \begin{cases} P = \frac{k}{f-1} [a^2 - a - 2p + s/4], \\ Q = \frac{k}{f-1} [2ab - 2a + 4p], \\ R = \frac{k}{f-1} [2ac - s/2], \\ S = \frac{k}{f-1} [b^2 - b - 2r], \\ T = \frac{k}{f-1} [2bc - 2c + 4u], \\ U = \frac{k}{f-1} [c^2 - c - 2u + s/4]. \end{cases}$$

If one wishes to use recurrence relations for the purpose of calculating the numbers for one generation from those for the preceding generation, equations 4) are much more easily used than equations 1). By thinking of equations 3) as written in capital letters, we can use equations 4) to obtain A, B, C in terms of the small letters thus:

$$5) \quad \begin{cases} A = \frac{k}{2(f-1)} [(b+2a)^2 - 8a - b - 2r + s], \\ B = \frac{k}{f-1} [(b+2a)(b+2c) - 3b + 2r - s], \\ C = \frac{k}{2(f-1)} [(b+2c)^2 - 8c - b - 2r + s]. \end{cases}$$

Equations 5) are more compact than equations 4), but give us less information; in fact we could not use equations 5) in calculating the numbers for succeeding generations because they do not give us R and S and these would be needed in the next step. However, if we combine with equations 5) the following pair from equations 4),

$$R = \frac{k}{f-1} [2ac - s/2],$$

$$S = \frac{k}{f-1} [b^2 - b - 2r],$$

we have a set better adapted for calculating the numbers for successive generations than is set 4).

Combining equations 4) we find that

$$6) \begin{cases} 2P + Q + R = 4ka, \\ 2U + T + R = 4kc, \\ 2S + Q + T = 4kb. \end{cases}$$

Here we have a set of recurrence relations solved for the variables which, if we were using subscript notation, would have lower subscripts. The equations are useless for direct computation but they will soon become important in an indirect way.

From equations 5) we find that

$$7) \begin{cases} B + 2A = 4k(b + 2a), \\ B + 2C = 4k(b + 2c), \end{cases}$$

Whence,

$$8) \begin{cases} \frac{B + 2A}{B + 2C} = \frac{b + 2a}{b + 2c}. \end{cases}$$

The ratio of dominant to recessive gametes in a particular generation is $(b + 2a)/(b + 2c)$. Equation 8) shows that this ratio is the same in one generation as in the preceding generation; from this we deduce immediately that the ratio of dominant to recessive gametes combining to form the individuals of the various generations is constant. This is a significant fact in that it is also true for complete random mating and for brother and sister mating. Using subscript notation, the solutions of system 7) are

$$9) \begin{cases} b_n + 2a_n = (b_0 + 2a_0)(4k)^n, \\ b_n + 2c_n = (b_0 + 2c_0)(4k)^n. \end{cases}$$

By means of equations 3), equations 6) may be rewritten with subscripts thus:

$$10) \begin{cases} r_n - s_n/2 = 4k a_{n-1} - a_n, \\ r_n - s_n/2 = 4k c_{n-1} - c_n, \\ -2r_n + s_n = 4k b_{n-1} - b_n. \end{cases}$$

In the light of 9) and 10), equations 5) may be written with subscripts as follows:

$$\text{II)} \left\{ \begin{aligned} a_n &= \frac{k}{f_0(4k)^{n-1}-1} \left[\frac{1}{2}(b_0 + 2a_0)^2(4k)^{2n-2} - \frac{1}{2}(b_0 + 2a_0)(4k)^{n-1} \right. \\ &\quad \left. - 2a_{n-1} - 4ka_{n-2} \right], \\ b_n &= \frac{k}{f_0(4k)^{n-1}-1} \left[(b_0 + 2a_0)(b_0 + 2c_0)(4k)^{2n-2} - 2b_{n-1} \right. \\ &\quad \left. - 4kb_{n-2} \right], \\ c_n &= \frac{k}{f_0(4k)^{n-1}-1} \left[\frac{1}{2}(b_0 + 2c_0)^2(4k)^{2n-2} - \frac{1}{2}(b_0 + 2c_0)(4k)^{n-1} \right. \\ &\quad \left. - 2c_{n-1} - 4kc_{n-2} \right]. \end{aligned} \right.$$

The solutions of this set of equations would give us a_n , b_n , c_n in terms of the original data.

It is well now to distinguish two cases. Consider first the case in which there are just two children in each family; in this case, $4k = 1$. Then equations II) simplify considerably, giving us,

$$\text{I2)} \left\{ \begin{aligned} 8(f_0-1)a_n + 4a_{n-1} + 2a_{n-2} &= (b_0 + 2a_0)(b_0 + 2a_0-1), \\ 4(f_0-1)b_n + 2b_{n-1} + b_{n-2} &= (b_0 + 2a_0)(b_0 + 2c_0), \\ 8(f_0-1)c_n + 4c_{n-1} + 2c_{n-2} &= (b_0 + 2c_0)(b_0 + 2c_0-1). \end{aligned} \right.$$

The solutions are

$$\text{I3)} \left\{ \begin{aligned} a_n &= \rho^n [k_1 \cos n\theta + k_2 \sin n\theta] + \frac{(b_0 + 2a_0)(b_0 + 2a_0-1)}{8f_0-2}, \\ b_n &= -2\rho^n [k_1 \cos n\theta + k_2 \sin n\theta] + \frac{2(b_0 + 2a_0)(b_0 + 2c_0)}{8f_0-2}, \\ c_n &= \rho^n [k_1 \cos n\theta + k_2 \sin n\theta] + \frac{(b_0 + 2c_0)(b_0 + 2c_0-1)}{8f_0-2}, \end{aligned} \right.$$

in which $1/\rho = 2\sqrt{f_0-1}$ and $\tan \theta = -\sqrt{4f_0-5}$.

Aside from the value of these equations for purposes of computation, they give us the limiting values of a_n , b_n , c_n as n increases indefinitely. In other words we get the final results for this type of mating if indefinitely continued. If we denote the limiting values of a_n , b_n , c_n , as n increases indefinitely, by a' , b' , c' we have

$$\begin{aligned} a' &= \frac{(b_0 + 2a_0)(b_0 + 2a_0-1)}{8f_0-2} & b' &= \frac{2(b_0 + 2a_0)(b_0 + 2c_0)}{8f_0-2}, \\ c' &= \frac{(b_0 + 2c_0)(b_0 + 2c_0-1)}{8f_0-2} \end{aligned}$$

For completely random mating, the proportions of the three types are fixed after the first random mating and are¹

$$a' : b' : c' :: (b_0 + 2a_0)^2 : 2(b_0 + 2a_0)(b_0 + 2c_0) : (b_0 + 2c_0)^2.$$

In brother and sister mating, the heterozygous type tends to disappear and for a limiting population we have (ROBBINS 1917)

$$a' : c' :: (b_0 + 2a_0) : (b_0 + 2c_0).$$

By comparing these three sets of results we note that omitting the brother and sister mating has had the effect of increasing the proportion of heterozygous individuals over what it is in completely random mating, *when there are only two offspring in a family*.

Consider now the case in which $4k > 1$, i.e., the case in which there are more than two offspring to a family. Returning to equations 11), it would be desirable to solve these equations without restrictions on k if possible. However, the nature of the solutions of the simple case in which $4k = 1$ seems to indicate that probably the solution of the general case would be of little use for purposes of computation even if we had it. We can get all desirable information about the limiting population without solving equations 11). We can calculate immediately from equations 11) that if $4k > 1$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{a_n}{(4k)^n} &= \frac{(b_0 + 2a_0)^2}{8f_0}, \\ \lim_{n \rightarrow \infty} \frac{b_n}{(4k)^n} &= \frac{2(b_0 + 2c_0)(b_0 + 2a_0)}{8f_0}, \\ \lim_{n \rightarrow \infty} \frac{c_n}{(4k)^n} &= \frac{(b_0 + 2c_0)^2}{8f_0}. \end{aligned}$$

Then using a' , b' , c' for the limiting values as above,

$$a' : b' : c' :: (b_0 + 2a_0)^2 : 2(b_0 + 2a_0)(b_0 + 2c_0) : (b_0 + 2c_0)^2.$$

This is the same result as in random mating. Thus in case a typical family consists of more than two offspring, the effect of omitting brother and sister mating from otherwise random mating is only temporary, the final results being the same as for completely random mating.

SUMMARY

This paper is a discussion of the problem of random mating with the exception that brothers shall not mate with their sisters, a single pair of Mendelian factors being considered. Three principal results are accomplished. First, formulae are developed from which we can calculate

¹ See for instance WENTWORTH and REMICK (1916).

the distribution of the offspring for succeeding generations (a) as to type of family,—equations 3) and 4)—and (b) as to type of individual—equations 5). Second, in case a family consists of two offspring the problem is solved completely in equations 13). Third, the distribution of the limiting population is given in case there are two offspring in a family and also in case there are more than two offspring in a family. The most important conclusions may be stated thus:

When brother and sister mating is omitted from otherwise random mating, the progeny in succeeding generations approaches a fixed distribution as the number of generations increases, in which pure dominants, heterozygotes and recessives all appear. (a) In case there are but two offspring in a typical family, the proportion of heterozygotes in the limiting population is greater than for completely random mating. (b) In case there are more than two offspring in a typical family, the limiting population will be the same as if the mating had been completely random.

LITERATURE CITED

- WENTWORTH, E. N., and REMICK, B. L., 1916 Some breeding properties of the generalized Mendelian population. *Genetics* **1**: 608-616.
ROBBINS, R. B., 1917 Some applications of mathematics to breeding problems. *Genetics* **2**: 489-504.