A Method for Calculating the Inbreeding Coefficient

AKIO KUDÔ[1]

Department of Human Genetics,
University of Michigan Medical School, Ann Arbor, Mich.

THE COEFFICIENT OF INBREEDING, F, has been defined by Malécot (1948) as the probability that an individual will possess at a given genetic locus two genes identical in their origin. The purpose of this note is to present a new method of calculating this coefficient particularly appropriate to complex biological relationships.

We shall define an *ancestral line* as that stepwise sequence of individuals or links uniting a particular person with an ancestor of his. We shall term this line a *paternal line* if it passes through the father of the individual and a *maternal line* if it passes through the mother. An *ancestral line pair,* or simply a *pair,* consists of a paternal and a maternal line of a given individual. In a pedigree extending over n generations, there are at most 2^{n-2} different paternal and maternal lines; thus in a seven-generation pedigree there are at most $2^5 = 32$ different lines, and hence $32^2 = 1024$ different pairs. It may happen that the paternal and maternal lines of a pair possess one or more common individuals. When this occurs we shall speak of the lines as *joining*. If a line pair contains but one common individual we shall call this individual a *common ancestor;* if more than one common individual exists, the term *common ancestor* will be applied only to that individual nearest, in relationship, to the person whose inbreeding coefficient is being computed. The sequence from the father to the mother through the paternal line, the common ancestor, and the maternal line is called a *loop,* and the *loop length,* denoted by l, is synonymous with the number of individuals in the loop. If the common ancestor is the n_1th ancestor of the father and the n_2th of the mother, we have the relation $l = n_1 + n_2 + 1$. With a particular pair there is associated at most one common ancestor, and hence one loop. There may be, however, a number of pairs with which the same common ancestor is associated. There may also be a number of loops in each of which the same individual is the common ancestor. If parents of a common ancestor have, in turn, a common ancestor, we term this *rejoining*. Thus, triple joining implies an ancestor common to the common ancestor of a common ancestor of an individual. The inbreeding coefficient, F_o, associated with an individual is defined by the following recurrence relation:

$$F_o = \Sigma \frac{1}{2^{n_1 + n_2 + 1}} (1 + F) = \Sigma \frac{1}{2^l} (1 + F)$$

Received May 18, 1962.
Supported in part by U. S. Atomic Energy Commission Contract AT (11-1)-942.

[1]Present address: Mathematical Institute, Faculty of Science, Kyushu University, Fukuoka, Japan.

where summation runs over every loop of length l and F is the inbreeding coefficient of the common ancestor in the loop. To the author's knowledge, this formula was first given by Wright (1950).

PEDIGREE SHEET

For the calculation of the inbreeding coefficient we suggest a form similar to that in Figs. 2 and 4. To use such a form it is essential that each individual in the group under study be identifiable and that his parents be given. Identification may be accomplished by a variety of conventions—e.g., names, letters, numbers—provided the convention gives rise to no relational ambiguities.

In the pedigree sheets the 32 columns represent all possible types of paternal lines, whereas the 32 rows represent all possible types of maternal lines in a pedigree of seven generations. The unshadowed cells represent male ancestors, the shadowed cells females, and each number which is entered is understood to be repeated in each of the blank spaces to the right (or below, for the female line.) The first column represents the paternal line of the type "father-father-father-father-father-father" the second "Fa-Fa-Fa-Fa-Fa-Mother," the third "Fa-Fa-Fa-Fa-Mo-Fa," etc. The lines are so arranged that if zeros and ones are assigned to fathers and mothers, respectively, each line becomes a number in the binary mode. The first, second and third columns, for example, would be (000 000), (000 001), and (000 010). The lines, maternal and paternal,

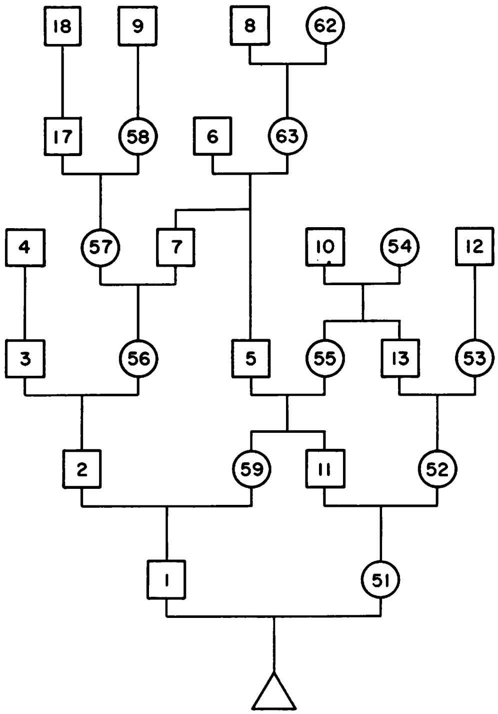

FIG. 1. A human pedigree extracted from a larger pedigree of an isolated community.

should be arranged in ascending binary-number order. For convenience, after the lines are arranged in the sequence just described, the maternal (paternal) lines are numbered 1 through 2^k. When considered jointly, the maternal and paternal lines describe an array of $(2^k)^2$ cells. We shall distinguish among these cells by indicating the appropriate maternal and paternal lines. Thus the i,jth cell is that formed by the ith maternal and jth paternal lines.

We proceed now to the use of the sheet for calculations. In Fig. 2 are set out all of the parental lines associated with the pedigree given in Fig. 1. If the parents of an individual are unknown, parental lines passing through this individual cannot, of necessity, be fully defined. When a line cannot be fully defined, this fact has been indicated by the insertion of a question mark in the location of the first unknown person. It need hardly be mentioned that this reduces the number of unique paternal and/or maternal lines, for all lines passing through the unknown individual are essentially the same. In the present case, for example, the first through fourth paternal lines are, for want of information, identical. The reader may satisfy for himself the construction of the individual lines. We offer as an example the 29th column, a paternal line, which will be recognized as the sequence of individuals numbered 1, 59, 55, 54, ?.

Note, now, that in paternal lines 9 and 10 and maternal lines 1, 2, 3 and

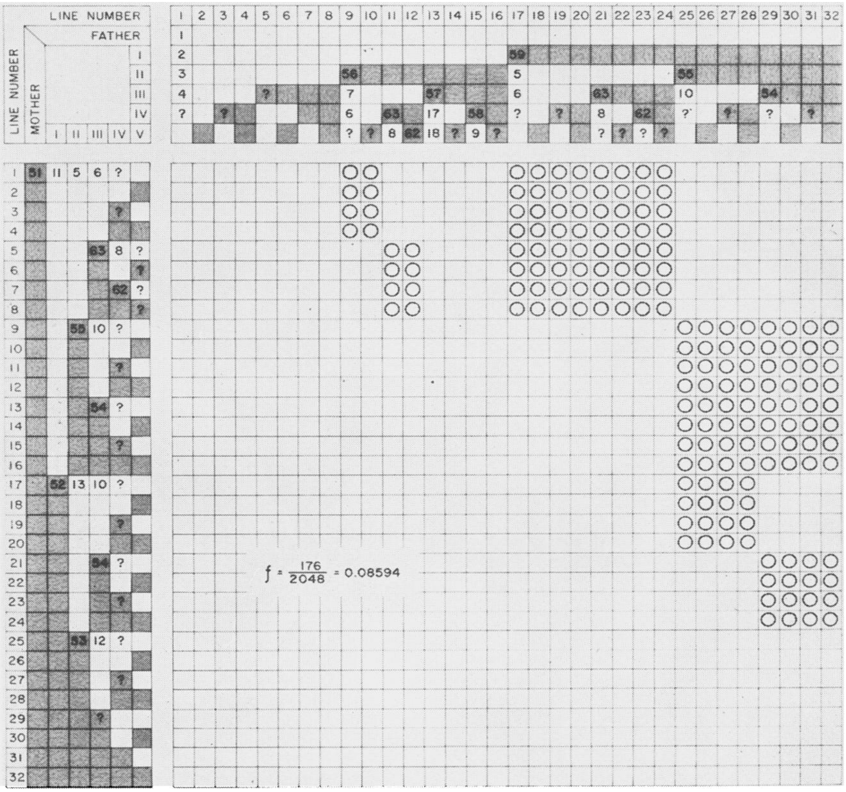

FIG. 2. Analysis of the pedigree in Fig. 1. This example serves to show how simple the analysis of human pedigrees can be when put in the form advocated here.

4 the same individual, 6, occurs. The occurrence of this individual we indicate by inscribing circles in the eight cells, 1,9; 1,10; 2,9; 2,10; 3,9; 3,10; 4,9 and 4,10 formed by the intersections of rows 1-4 with columns 9-10. We also note that in addition to individual 6, individuals 63, 5, 55, 10 and 54 are common ancestors. Again, circles are inscribed in a manner analogous to that for individual 6. There is, it will be observed, no instance of rejoining. The formula for the inbreeding coefficient is, in this case, simply

$$F_0 = \frac{\text{number of circles}}{2(\text{total number of cells})} \qquad \text{(A)}$$

This formula is valid in all instances where rejoining does not occur. This may be proved in the following way: We note that 6 is the 4th ancestor of the father and the 3rd of the mother. Thus $n_1 = 4$ and $n_2 = 3$, and the contribution, f, to F_0 of this loop is

$$\frac{1}{2^{4+3+1}} = \frac{1}{2^8}$$

The number of circles corresponding to this loop is

$$2 \times 2^2 = 2^{5-n_1} \times 2^{5-n_2}$$

and the ratio of these two numbers is

$$\frac{1}{2^{n_1+n_2+1}} \Big/ (2^{5-n_1+5-n_2}) = 1/2^{5+5+1}$$

which is independent of n_1 and n_2 and is equal to the inverse of twice the total number of cells. Thus we have the relation

$$f = \frac{\text{number of circles in the loop}}{2(\text{total number of cells})}$$

and, as F_0 is the sum of the f's, we are led to conclude that (A) is valid.

Let us proceed now to the case of rejoining, i.e., the situation where a common ancestor has a common ancestor. Fig. 3 is an artificial pedigree which, however, illustrates this case. Fig. 4 is a diagrammatic representation of the calcu-

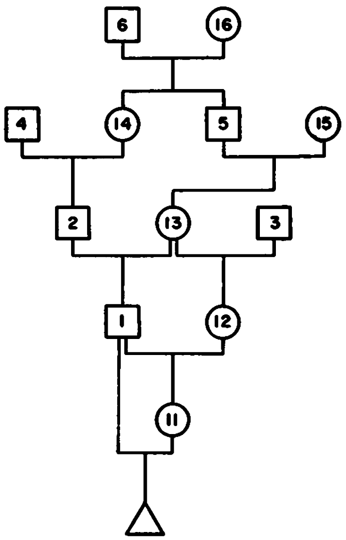

FIG. 3. An artificial pedigree with rejoining and father-daughter marriage.

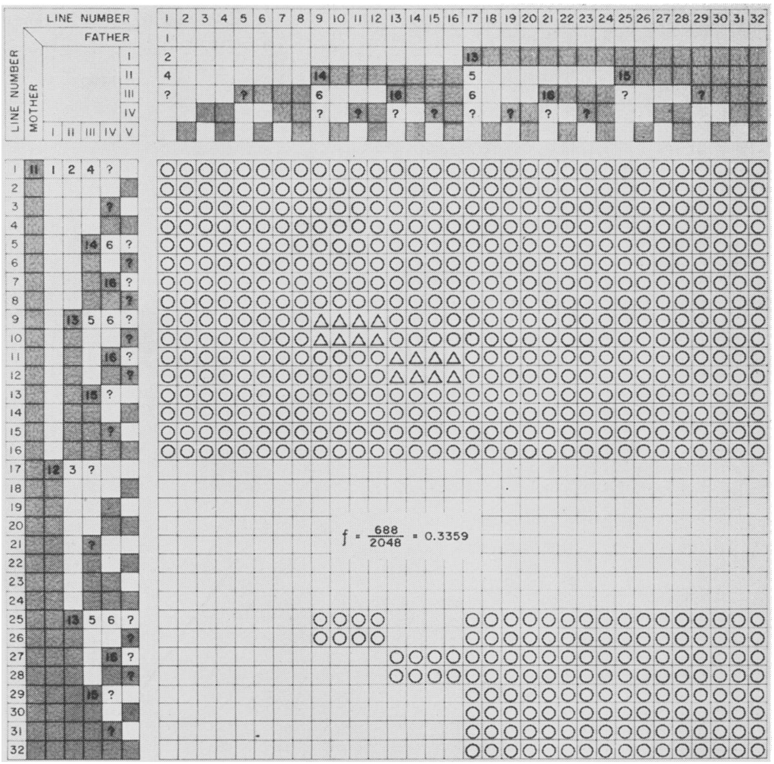

FIG. 4. Analysis of the pedigree in Fig. 3. This example indicates the use of the pedigree sheet in case of rejoining.

lation. We observe four common ancestors, namely, 1, 6, 16, and 13, and as in the previous example we inscribe circles in the cells associated with these common ancestors. To calculate the effect of the rejoining we proceed as follows: We note that the parents of 1, namely 2 and 13, have common ancestors, 6 and 16, and we are obliged, therefore, to calculate the inbreeding coefficient of 1. For this calculation we need not prepare another form. We can use the portion I-V of paternal lines 1-16 and the portion II-V of maternal lines 9-16. Within the rectangle so formed we mark with triangles[1] those cells where 6 occurs simultaneously in the maternal and paternal lines, and similarly for 16. [Note: Maternal and paternal line pairs simultaneously containing 6 and lying outside the rectangle are not scored.] An argument similar to that for (A) leads us to the conclusion that the inbreeding coefficient of 1, F_1, is given by:

$$F_1 = \frac{\text{number of triangles}}{2(\text{number of cells in the rectangle})} \qquad \text{(B)}$$

In the case of the common ancestor, 13, there is no rejoining. This may be verified by examining the pairs corresponding to the lower left portion of the rectangle of circles in the lower right corner of Fig. 5. If 6 and/or 16 had

[1]To avoid ambiguity in the illustrations a triangle inside a circle is shown as a triangle only.

common ancestors, a process similar to that leading to (B) would suffice to calculate the inbreeding coefficients of 6 and/or 16. The cells here which would then contribute to the inbreeding coefficient might be indicated with inscribed small circles.

The general formula for the inbreeding coefficient becomes

$$F_0 = \frac{\#\text{ circles} + 2\#\text{ triangles} + 2^2\#\text{ small circles} + \ldots}{2(\text{total number of cells})} \quad \text{(C)}$$

It should be noted that the circles, triangles, small circles etc. are merely devices to recognize pairs in which loops arise through joining, rejoining, triple joining etc.

SOURCES OF ERROR

As has been indicated, to use the pedigree sheet does not require the drawing of the pedigree chart of the ancestors of the individual. This, however, can lead to incomplete identification which cannot be easily detected and constitutes one of the major sources of error in the technique.

A second possible source of error will be the failure to detect all of the loops of common ancestors. A third lies in the calculation of the inbreeding coefficient after drawing the rectangles of the loop. The latter may not be frequent, but can occur when a large number of inbreeding coefficients is calculated.

In the following we shall propose a check which will eliminate the second kind of error:

Consider Fig. 1. Since the pedigree sheet extends to the 5th generation of ancestors of the parents of the individual whose inbreeding coefficient is to be computed, there can appear at maximum $1+2+2^2+2^3+2^4+2^5 = 63$ individuals in the paternal as well as the maternal lines. Among these individuals there may be some who are unknown and some who appear a number of times, say, p in the paternal and m in the maternal lines. Our formulae use the number of known individuals, number of blocks, and the numbers p and m of individuals involved as defined above. Among these numbers we have the following relations:

p = number of known ancestors in the paternal lines
= 63 minus the number of ? in the paternal lines
m = number of known ancestors in the maternal lines
= 63 minus the number of ? in the maternal lines
pm = number of blocks

where summation runs over all the ancestors.

After the representatives of the blocks have been indicated and checked by the above formulae, we can safely proceed to inscribe circles, triangles etc. on the sheet.

ALTERNATIVE METHODS

One might suggest two alternative methods. One is to use hatching, double hatching and triple hatching instead of a circle, a triangle and a small circle. Another suggestion would be to mark not only the cells in the lower left portion of the rectangle but also those in the upper right portion of the rectangle with triangles or double hatching, in the case of rejoining. In this case the triangle or double hatched cells should appear symmetrically along the diagonal of the rectangle. In this manner we count the pairs of rejoining twice and those of triple joining four times. There is, with this approach, no need for the coefficients of $2, 2^2$ etc. in formula (C).

DISCUSSION

The method here proposed enjoys, it would seem, a number of advantages over the classical method of tracing the pedigree chart (cf., for instance, Neel and Schull). Firstly, one need not draw a pedigree chart, a troublesome and time-consuming job. This has certain disadvantages, as we have seen, and may lead to the omission of some genetic information.

Secondly, this method seems to be less liable to mistakes, and especially when a check is performed during calculation. The method is easy to teach or to learn. With this approach the numerous computations in the evaluation of a complicated pedigree of a large population can be put into the hands of clerical workers, a great advantage.

When the classical method of tracing the pedigree chart is used, and if the pedigree is very complicated, it sometimes happens that two persons will compute different values for the same coefficient, and neither of them can be sure which is correct. There is no effective method of controlling the calculations in the classical method. This difficulty can be overcome by using the pedigree sheet.

When one is obliged to calculate the F of a highly inbred population of considerable size, one is naturally led to the idea of using high speed computers. Any programmer who is acquainted with the use of this pedigree sheet should find it simple to construct the logic to be employed in the programming.

ACKNOWLEDGMENTS

This paper was worked out while the author was at the Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, on leave-of-absence from Kyushu University, Fukuoka, Japan. The author expresses his hearty thanks to Professors J. V. Neel, W. J. Schull and other members of the department for their interest in this inquiry. The author is particularly grateful to Professor Schull, who went through the manuscript and improved the exposition greatly.

REFERENCES

MALÉCOT, G. 1948. *Les Mathématiques de L'hérédité.* Paris: Masson et Cie.

NEEL, J. V., AND SCHULL, W. J. 1954. *Human Heredity.* Chicago: University of Chicago Press.

WRIGHT, S. 1950. The genetical structure of populations. *Ann. Eug.* (Lond.) 15: 323-354.