

THE EFFECTS OF ISOLATES ON THE FREQUENCY OF A RARE HUMAN GENE

By L. C. DUNN

COLUMBIA UNIVERSITY, AND STATE INSTITUTE OF RACE BIOLOGY AND HUMAN GENETICS,
UPPSALA, SWEDEN*

Communicated November 4, 1947.

An important corollary of Mendel's principle of segregation was noted by Weinberg¹ and by Hardy² in 1908. They showed that in a large population undergoing random mating (panmixia), the relative frequencies of a pair of adaptively neutral alleles such as A and a tend to remain constant in the equilibrium state $q^2AA:2q(1-q)Aa:(1-q)^2aa$ in which q and $1-q$ are the proportions of A and a in the population. This simple relationship formed the basis for the development of the population genetics of cross-fertilizing animals and plants and has been particularly fruitful in studies of human genetics. It has been found, however, that the ideal populations assumed in the Hardy equilibrium are seldom if ever found in nature. Random mating or panmixia could hardly extend over a wide-spread animal or plant species. Human populations also are broken into rather small groups known as isolates within which marriages tend to occur at random, while genes are less frequently exchanged between the isolates. The isolating factors here may be of various kinds such as geographic, social, religious and others.

It is obvious that if random mating and free exchange of genes do not occur in large populations, the gene frequencies will be affected by circumstances peculiar to small populations, and by inbreeding or mating among relatives. Inbreeding assumes an unusual importance in human populations which generally contain recessive mutant genes for deleterious characters of various kinds. When such genes are rare, homozygotes are likely to be produced only by marriage between relatives who carry the same recessive gene. If such marriages, as between first cousins, occur at random, they will have a high probability of occurrence only when the circle of possible mates is small, as in an isolate. Therefore it is to be expected that such rare recessives will not be distributed evenly in large populations, but that different rare hereditary defects will tend to appear in different small communities, some in one, others in another.

Wahlund³ (1928) has worked out, in theory, the differences to be expected in the frequencies of the several genotypes as between a whole population under random mating and one which is divided into isolates; while Dahlberg⁴ (1943) has shown how the frequency of cousin marriages may be used to estimate the average size of the isolates, and has calculated the theoretical relationship between gene frequency, frequency of cousin

marriages in the isolate and frequency of cousin marriages among parents of persons homozygous for a given gene.

An actual example of isolates was found by Sjøgren⁵ (1931) in studying the distribution of a gene for a rare and always fatal disease, juvenile amaurotic idiocy, as it occurs in Sweden. The disease begins with failing sight and blindness in young children of 4 to 7 years, and progresses through loss of sensory, mental and physical powers to its terminus in death some 10 or 12 years later. It was possible to detect nearly all cases occurring in Sweden since 1890, when special instruction and registration of the blind became obligatory, and to work out the ancestry and geographic origins of the affected persons by using the excellent and continuous records kept in the Swedish parishes. The heterozygous ancestors responsible for the 115 primary cases studied by Sjøgren were found to group themselves into 59 families coming from 23 rather restricted localities in southern and central Sweden; while the proportion of first cousin marriages among the parents of the amaurotics proved to be 15% which is probably 20 to 30 times the rate for the population at large. Thus the Swedish population was shown to consist, in respect to the marriages responsible for juvenile amaurotic idiots, of restricted marriage circles or isolates within which marriage between relatives occurs with higher frequency than in the population at large.

In the present note it is proposed to test some of the more recent methods by applying them to Sjøgren's data. Specifically it is of interest to determine the frequency of the gene for juvenile amaurotic idiocy in the isolates and to estimate the fraction of the whole population which is contained within the isolates.

We may begin by estimating the frequency of the gene on the assumption of random mating in the whole population of Sweden.

Sjøgren found that for the years 1913-1922, the average number of juvenile amaurotics alive at one time in Sweden was 52.5. Since the average age of onset of the disease (detected by blindness) is 6.7 years and the average age at death is 18.2 years, its average duration is about 11.5 years. Thus about 52.5/11.5, or 4.6 new cases appear each year. For the ten-year period concerned, about 1,200,000 children reached the age of seven, when the disease would be detected, or about 120,000 per year. The frequency of homozygotes can thus be estimated as 4.6/120,000 or 0.0000383, roughly 4 per 100,000 (0.004%). Assuming random mating the proportions of homozygotes and heterozygotes at equilibrium would be $AA = 0.9876$; $Aa = 0.0123$; $aa = 0.00004$. Sjøgren recognized that the heterozygote frequency of 1.2% thus calculated for this gene could not be correct, since it applied to the whole population and must therefore be "für die Populationen in den Herdgebieten (isolates) und angrenzenden Teilen zu niedrig, für die Lakunen dazwischen zu hoch. . . ."

We must therefore calculate the gene frequency within the isolates with the aid of Dahlberg's (1943) formula,⁴

$$k = \frac{c(1 + 15r)}{16r},$$

in which k is the frequency of first cousin marriages among parents of juvenile amaurotics, c is the frequency of cousin marriages arising at random in the isolate population and r is the frequency of the gene. In the present case we may use for k the value of 15% found by Sjøgren, and for c the value 0.45% derived from data supplied by Professor Dahlberg on cousin marriages among parents of 17,016 children in public schools in country districts of southern and central Sweden in 1947. This gives $r = 0.0019$, corresponding to $r^2 = 0.0000361$ or about 4 homozygotes per million as compared with 4 per 100,000 as found by Sjøgren. The assumption of 0.45% cousin marriages thus leads to an estimate of homozygote frequency in the isolates which is $1/10$ of the actual frequency in the whole population in 1930. This evidently is absurd. The frequency of cousin marriages in the isolate population must therefore be higher.

Sjøgren, in considering this problem, assumed a frequency of cousin marriages of 1% and when this value is substituted in equation 1, the gene frequency $r = 0.0044$ is found, corresponding to $r^2 = 0.000019$ or about 2 homozygotes per 100,000. This figure is also too low. In these cases, it is probable that the frequency of cousin marriages in the isolates has been underestimated, for Sjøgren's data were chiefly from idiots diagnosed in the period 1896-1930, whose parents would therefore have been married some 30-40 years earlier when the rate of cousin marriages in restricted country districts was probably much higher. (Unpublished data of Professor Dahlberg indicate that the cousin marriage rate in the Swedish nobility declined sharply in the period 1870 to 1930.) It is probable that a cousin marriage rate of 2% in the isolates for the relevant period is not an overestimate. Using $c = 2\%$ and $k = 15\%$ we get a gene frequency in the isolates of 0.009524 or $r^2 = 0.00009$, about 1 homozygote per 10,000 in the isolates.

If we assume that all cases of juvenile amaurotic idiocy are born in the isolates, where alone the gene frequency and rate of cousin marriage are high enough to produce them, then we may estimate the fraction of the total population contained in the isolates as 42.5%. (Roughly, if the frequency of homozygotes in the whole population is about 1 per 25,000, as found by Sjøgren, and about 1 per 10,000 in the isolates, as estimated from the cousin marriage fraction, then the isolates constitute about 40% of the population.) Of the Swedish population of 6,074,368 in 1930 the isolates containing this gene thus comprised 42.5% or 2,580,000 persons. With a homozygote frequency of 0.00009 we should expect to find about 230

homozygotes in such a population. The frequency can be applied, however, only to that age fraction of the population in which juvenile amaurotic idiocy occurs, namely, in children between 7 years (onset of blindness) and 18 years (age at death). In 1930 there were 1,226,000 persons in these age groups in Sweden, or about 20% of the population. Since the chance of detection is thus 20%, we take 20% of 230, or 46, as the number of homozygotes which we may expect to find alive at one time. For the ten-year period 1912-1921 Sjøgren found an average of 52 juvenile amaurotic idiots alive at one time. The discrepancy between the actual figure and that calculated on the assumption of a cousin marriage rate of 2% and a gene frequency in the isolates of 0.009524 is not so great as to discredit the latter computation.

It is possible to estimate the average size of the isolates from the frequency of cousin marriages using the relation $n - 1 = \frac{2b(b - 1)}{c}$ (Dahlberg, 1943), where n is the size of the isolate population, c is frequency of cousin marriage and b is number of children per family. (In the stationary Swedish population b may be taken as 2.) Where $c = 0.02$ the isolate size is about 200. The minimum frequency of the gene could hardly be less than 2 heterozygotes per isolate, in this case a gene frequency of about 0.01, which is not far from the value 0.009 as estimated above.

There is thus a fair agreement between the frequency of homozygotes predicted by the theory and the numbers actually found by Sjøgren. It is important, however, to emphasize that even this measure of agreement is reached only by assuming a rate of cousin marriage which is much higher than that which occurs today. This is to say that the calculations apply to conditions which no longer exist and illustrates one of the difficulties in testing such theories by data from human populations. A rigid test of the methods used would require not only actual data on the cousin marriage rate but estimates of the numbers of people in some of the isolates within which the gene occurs, for comparison with predicted numbers. An attempt has been made to estimate the numbers of people in the "Heterozygoten-Herden" identified by Sjøgren but the uncertainty of isolate boundaries and the unlikeness in the population distribution in different areas have made this impractical.

Nevertheless the distribution of this rare gene, with a higher frequency in its centers, and a low or zero frequency in the rest of the population illustrates a fact about human populations which is essential to recognize. The methods for dealing with gene frequencies in isolates can be used as rough approximations, and indicate the variables which must be measured in future studies of gene distribution.

* The author acknowledges with gratitude the hospitality of the Institute and the advice and kindness of the Director, Professor Gunnar Dahlberg.

¹ Weinberg, W., *Jahres L. Ver. Vaterl. Naturk. Württemberg*, 64, 368-382 (1908).

² Hardy, G., *Science*, 28, 49-50 (1908).

³ Wahlund, S., *Hereditas*, 11, 65-106 (1928).

⁴ Dahlberg, G., *Acta Med. Scand. Suppl.*, 148, 1-219 (1943).

⁵ Sjögren, T., *Die Juvenile amaurotische Idiotie*, Lund, 1931.

THE NATURE OF GENE ACTION AS SHOWN BY CELL-LIMITED
AND CELL-DIFFUSIBLE GENE PRODUCTS

BY DONALD F. JONES

CONNECTICUT AGRICULTURAL EXPERIMENT STATION, NEW HAVEN

Communicated September 28, 1947

The gene markers previously used to identify chromosome aberrations in the endosperm of the maize seed produce cell areas with well-defined boundaries when dominant genes are removed from the cell allowing their recessive alleles to operate. These tissue alterations are brought about during development by any process that removes chromosomes in whole or in part from the cell. The change occurs in the nucleus but the effect is visible in the cytoplasm. Isolated single cells show these changes. Obviously something passes through the nuclear membrane, either from the nucleus to the cytoplasm or in the reverse direction, but does not go beyond the cell membrane since no cell division has occurred in the single cell alterations to liberate nuclear products into the cytoplasm or the reverse.

Where the effects of the gene products are confined within the cell such genes may be considered as cell-limited. Genes of this type in maize aleurone are *C*, *R*, *Pr* and *I*. For illustration of this type of gene action see figure 1 and Jones^{1, 2} and Clark and Copeland.³

In marked contrast to these cell-limited genes are the cell-diffusible genes where the gene products pass through the cell wall and affect adjoining cells over a considerable area. The *A*₁, *A*₂ and *Y* color genes in maize are of this type. The *A* series of anthocyanin genes are necessary to produce color in all parts of the plant. In the recessive condition the aleurone is colorless, the cob and pericarp are brown and the other parts of the plant are green or brown, depending upon other genes present. The dominant allele produces anthocyanin in the leaves, silks, glumes, anthers, aleurone and scutellum when the complementary genes are present. When *A* is removed from the aleurone by the loss of the locus containing this gene that part of the seed is colorless. However, there is a gradual diminution in color from the pigmented to the unpigmented area extending over an area of several cells so that the border is not distinct as it is in other color changes involving *C*, *R*, *Pr* and *I*. Evidently something diffuses through