THE NATURE OF HUMAN GENETIC LOADS

By J. B. S. HALDANE and S. D. JAYAKAR

Genetics and Biometry Laboratory, Government of Orissa, Bhubaneswar-3, Orissa, India

The first attempt to estimate genetic loads was probably that of Haldane (1937). His estimates were undoubtedly crude, though of the correct order of magnitude. Recently two types of work have led to quantitative estimates. On the one hand large volumes of work on cytologically polymorphic Drosophila populations have revealed loads of the order of 10% or much more due to lowered viability or fertility of homozygotes both of which are less fit than their heterozygote (e.g. Dobzhansky, 1951). On the other hand Morton, Crow and Muller (1956), from an examination of the effects of human inbreeding, conclude that over-dominance (i.e. higher fitness of heterozygotes than either homozygote) is unimportant as compared with mutation as a cause lowering human fitness.

Those who doubt or deny this can point to apparently stable human polymorphism at a number of loci. At three of these loci, two concerned in haemoglobin synthesis, and one in glucose-6-phosphate dehydrogenase deficiency, the heterozygotes appear only to be fitter when malaria, and particularly malaria caused by *Plasmodium falciparum*, is common. At other loci, for example those responsible for the synthesis of the ABO, MN, and Rh groups of antigens, polymorphism is found in all human groups; and no explanation, other than the increased fitness of heterozygotes, is available. The total loss of fitness at the β haemoglobin locus, associated with sickle-cell disease, is often over 5%, and can reach nearly 20%. That at the universally polymorphic loci is probably much less, for at least at the ABO locus ancestral gene frequencies can persist for many generations when a human group migrates. However the number of such loci is considerable, and their total contribution might be expected to be of the order of 10%.

Let us now examine Morton, Crow and Muller's argument in detail. They consider a mutant **a** at an autosomal locus, whose frequency among gametes is a small quantity q. The fitness of $+/\mathbf{a}$ is $1-sh$, that of **aa** is $1-s$, relative to a unit fitness of $+/+$, s and h are taken to be positive, h being usually of the order of ·05 or less. If h is negative we have balanced polymorphism with q fairly large. The coefficient of inbreeding in the group of the population considered is F. They then show that the loss of fitness due to **a** is

$$q[q+2h\ (1-q)]\ s+Fq(1-q)\ (1-2h)s. \qquad (1)$$

On the other hand if the heterozygote $\mathbf{a}_1\mathbf{a}_2$ is fitter than either homozygote, and the relative fitness of $\mathbf{a}_1\mathbf{a}_1$, $\mathbf{a}_1\mathbf{a}_2$, and $\mathbf{a}_2\mathbf{a}_2$ are $1-s_1 : 1 : 1-s_2$, while the mean coefficient of inbreeding is f, it can easily be shown that at equilibrium the frequencies of $\mathbf{a}_1$ and $\mathbf{a}_2$ are

$$q_1=\frac{s_2-fs_1}{(1-f)\ (s_1+s_2)}, \text{ and } q_2=\frac{s_1-fs_2}{(1-f)\ (s_1+s_2)}.$$

Morton, Crow, and Muller assumed $f=0$, which is quite adequate for most loci, since f rarely exceeds ·01. The load, or loss of fitness, is $\frac{(1+f)\,s_1\,s_2}{s_1+s_2}$, and in a group of the population with $F=0$, it is $\frac{s_1\,s_2}{s_1+s_2} + \frac{f^2(s_1-s_2)}{(1-f)^2\,(s_1+s_2)}$, whilst in a group with coefficient F it is

$$\frac{s_1\,s_2}{s_1+s_2} + \frac{f^2(s_1-s_2)}{(1-f)^2\,(s_1+s_2)} + F\left[\frac{s_1\,s_2}{s_1+s_2} - \frac{f\,(s_1-s_2)^2}{(1-f)^2\,(s_1+s_2)}\right],$$

or nearly $\frac{(1+F)s_1s_2}{s_1+s_2}$, the value given by Morton Crow, and Muller. Morton, Crow, and Muller now assume that there are a number of nongenetical causes of premature death or sterility, and that the loss of fitness at any one locus is small. If so the negative natural logarithm of the mean fitness in a group with mean coefficient of inbreeding F is $A+FB$, where

$$A = \Sigma x + \sum_i \left[q_i\,(q_i+2h_i-2h_iq_i)\,s_i\right] + \sum_j \left[(s_{1j}+s_{2j})^{-1}\,\{s_{1j}\,s_{2j}+f^2\,(1-f)^{-2}\,(s_{1j}-s_{2j})^2\}\right]$$

$$= \Sigma x + \sum_i \left[q_i\,(q_i+2h_i)\,s_i\right] + \sum_j s_{1j}\,s_{2j}\,(s_{1j}+s_{2j})^{-2} \text{ nearly,} \qquad (2)$$

$$B = \sum_i \left[(1-2h_i)\,q_i\,(1-q_i)\,s_i\right] + \Sigma \left[(s_{1j}+s_{2j})^{-1}\,\{s_{1j}\,s_{2j}-f\,(1-f)^{-2}\,(s_{1j}-s_{2j})^2\}\right]$$

$$= \sum_i \left[(1-2h_i)\,q_i\,s_i\right] + \sum_j \left[s_{1j}\,s_{2j}\,(s_{1j}+s_{2j})^{-2}\right] \text{ nearly.} \qquad (3)$$

Here Σx refers to non-genetical causes of death, q_i, h_i, and s_i, refer to the i-th locus where recessives are kept in being by mutation, s_{1j} and s_{2j} to the j-th locus where heterosis, or overdominance, preserves polymorphism.

To estimate A and B they used six sets of data. Four were due to Sutter and Tabah (1953) and one each to Arner (1908) and Bemiss (1858). Sutter and Tabah computed the mortalities of children whose parents were "unrelated", second cousins, first cousins once removed, and first cousins, in two French departments. The mortalities were separated into recorded miscarriages plus still births plus neonatal deaths, and infantile plus juvenile deaths. The pooled mortalities (correcting the obvious misprint of 25 for 75 in Table 2) are ·111 when parents were unrelated, ·151 when they were second cousins, ·148 when first cousins once removed, and ·222 when first cousins. These give $A=$·118, and, on the basis of the first cousin marriages, $B=2{\cdot}13=18{\cdot}0\,A$. The earlier data give similar results, the six values of BA^{-1} ranging from 24·41 to 7·94, with a median about 16. Now the first term in (2) contributes to A only, if we neglect the fact that some recessive genotypes increase liability to infection, and the last terms give $BA^{-1}=1$, nearly. Hence, it is argued with considerable force, the mortality in each group is largely due to lethal and sublethal autosomal recessives kept in being by

mutation. It may perhaps be added that dominant and sex-linked recessive sublethal mutants will contribute to A, and hardly at all to B, so they cannot be an important cause of death.

Since Morton, Crow, and Muller's paper more data have accumulated. Neel (1963) lists 16 results on the effects of inbreeding on early human mortality. In five of them B was negative, though never significantly. The median value of BA^{-1} is 3·9, or if we include Bemiss' data, 6·3. This may perhaps be a more plausible figure than 16. And it may be that in a few populations the load due to mutation is unimportant. But even with the lower median value Morton, Crow, and Muller have a very strong case provided their theory is unassailable.

Let us however examine the matter more closely. The "genetic deaths" or extinctions of genes, fall into three groups. A fertilized ovum or early embryo may die without record. To judge from what is known as to Drosophila and the mouse, such deaths may account for a very large fraction of the genetic deaths. Secondly individuals may die in the later stage of pregnancy, at birth, or before maturity. These are the deaths recorded. Finally there are a few deaths during the reproductive age, and a great deal of total or partial sterility. It will be very hard to determine the importance of the latter fact in man, since one cannot separate the effects of sexual continence and birth control from those of genetically determined sterility. But this last cause was probably much more important some centuries ago, when the existing frequencies of common genes were determined, than it is today. It should, however, be much easier to obtain data on genetically determined sterility in countries such as India, where there is as yet very little birth control, than in those where this practice is prevalent. Since birth control is likely to spread, the matter should be investigated as soon as possible.

If we consider two allelomorphs at the same locus, with a heterozygote fitter than either, it is unlikely that they will bring about death from the same cause. Thus in the case of corpuscular sickling the deaths are due to malaria or to anaemia. In many cases I suggest that one homozygote shortens life, while the other lowers fertility. Juvenile mortality is higher in large than small human families. This is no doubt partly, but perhaps not wholly, due to greater poverty. However, in the absence of birth control, one may argue as follows. Genes conferring high fertility exist. But they are not universal. It is hard to suppose that all genetically determined sterility is due to a balance between selection and mutation. It may be balanced by high early mortality of the potentially fertile genotypes.

Matthew (1926) pointed out that natural selection tended to increase tooth length and diminish the lateral toes in the evolving Equidae. Contemporaneous species were advanced in one respect or the other. Kermack (1954) found the same for morphological characters in a single species of *Micraster*. Those which were more marked in descendants were negatively correlated in their ancestors. Pearson, Lee, and Bramley-Moore (1899) found a negative correlation of longevity and fertility in man. This is to be expected, as they pointed out, on *a priori* grounds. They postulated a conflict between "natural selection" or selection for survival, and "reproductive selection",

or selection for fertility. Their work demands repetition with various modifications which are obvious after sixty years. In particular it seems likely that the same genotype might have a higher fertility than the average in one sex, and a lower one in the other. The net effect might be slight, which, as will be seen, supports the theory developed later.

Where a heterozygote is over-dominant as regards fitness, it is reasonable to suppose that in many cases $\mathbf{a}_1\mathbf{a}_1$ is more fertile than the average, but has a higher infantile and juvenile mortality, $\mathbf{a}_2\mathbf{a}_2$ is less fertile but more viable, while $\mathbf{a}_1\mathbf{a}_2$ may be intermediate in both respects, but achieves a combination with a higher degree of fitness. In such a case $\mathbf{a}_1\mathbf{a}_2$ will usually have a viability nearer that of $\mathbf{a}_2\mathbf{a}_2$ than that of $\mathbf{a}_1\mathbf{a}_1$. From the point of view of viability, $\mathbf{a}_1$ will be more or less completely recessive. Let us examine this situation. I suppose the frequencies of the genes $\mathbf{a}_1$ and $\mathbf{a}_2$ to be q_n and p_n ($p_n+q_n=1$) in the nth generation. The fertility and viability of $\mathbf{a}_1\mathbf{a}_2$ are taken as unity, the viability of $\mathbf{a}_1\mathbf{a}_2$ being $(1-l)$ while its fertility is unity. The viability of $\mathbf{a}_2\mathbf{a}_2$ is unity, its fertility $(1-k)$. In a study of infantile and juvenile mortality only the lesser viability of $\mathbf{a}_1\mathbf{a}_1$ will be noted. Let f be the coefficient of inbreeding of the whole population, F that of a section of it (e.g. children of first cousins) specially studied. The effective breeding population in generation n is

$$(1-k)\ (p_n^2+fp_n\,q_n)\ \mathbf{a}_2\mathbf{a}_2 : 2(1-f)\ p_n\,q_n\ \mathbf{a}_1\mathbf{a}_2 : (1-l)\ (q_n^2+fp_n\,q_n)\ \mathbf{a}_1\mathbf{a}_1.$$

Hence $$q_{n+1} = \frac{q_n - lq_n\,(fp_n+q_n)}{1-kp_n\,(p_n+fq_n)-lq_n\,(fp_n+q_n)},$$

$$\Delta\, q_n = \frac{p_nq_n\,[k\,(p_n+fq_n)-l\,(fp_n+q_n)]}{1-kp_n\,(p_n+fq_n)-lq_n\,(fp_n+q_n)}.$$

At equilibrium this is zero, hence $p_n=0$, $q_n=0$, or

$$q_n = Q = \frac{k-fl}{(1-f)\,(k+l)},\quad p_n = P = \frac{l-fk}{(1-f)\,(k+l)}$$

The third equilibrium is stable if $k>fl$, $l>fk$.

The death rate in the whole population is $lQ\,(fP+Q)$, or

$$D_f = lQ\,(fP+Q) = \frac{(1+f)\,kl\,(k-fl)}{(1-f)\,(k+l)^2}.$$

The deathrate among those which are not inbred is

$$D_0 = lQ^2 \qquad = \frac{l\,(k-fl)^2}{(1-f)^2\,(k+l)^2}.$$

The deathrate of persons with a coefficient of inbreeding F is

$$D_F = lQ\,(FP+Q) = \frac{l\,(k-fl)^2 + Fl\,(l-fk)\,(k-fl)}{(1-f)^2\,(k+l)^2}.$$

Or, if this were the only source of increased mortality from inbreeding, in Morton, Crow, and Muller's terminology,

$$A = -\ln\ (1-D_0)$$

$$A^{-1}B = F^{-1}\ [\ln\ (1-D_0) - \ln\ (1-D_F)]$$

$$= F^{-1} \ln \left(1 - \frac{D_F - D_0}{1 - D_0}\right).$$

$$\text{Thus } A = \frac{l\ (k-fl)^2}{(1-f)^2\ (k+l)^2} + \tfrac{1}{2}\left[\frac{l\ (k-fl)^2}{(1-f)^2\ (k+l)^2}\right]^2 + \tfrac{1}{3}\left[\frac{l\ (k-fl)^2}{(1-f)^2\ (k+l^2)}\right]^3 + - -$$

$$= \frac{k^2 l}{(k+l)^2}, \text{ nearly,}$$

if k or l is sufficiently small.

$$BA^{-1} = \frac{l-fk}{k-fl}, \text{ nearly,}$$

$$= \frac{l}{k}, \text{ more roughly.}$$

Hence BA^{-1} has the fairly large value about 10, found by Morton, Crow, and Muller if l is about ten times as large as k, for example if the viability of $\mathbf{a}_1\mathbf{a}_1$ were 90% of normal, and the fertility of $\mathbf{a}_2\mathbf{a}_2$ 99%. In this case the locus would make a contribution of about ·00083 to A, and about ·0083 to B. Such a situation may not be common. It is certainly not impossible. Nor is a balance situation where one homozygote has a low vitality in early embryonic stages, the other at ages when deaths can be detected.

There is however another possibility. Morton, Crow, and Muller assume that the populations investigated were in genetic equilibrium. This seems to me highly improbable. The French peasants studied by Sutter and Tabah were exposed to more infectious diseases than modern Americans, but to far fewer than the French royal family a century earlier. Haldane has repeatedly stated his opinion that the main agent of natural selection in the human species during the last five thousand years has been infectious disease. This process has not often reached equilibrium, because the pathogenic organisms die out or evolve. Sutter and Tabah's data show that one or more recessive phenotypes are specially liable to tuberculosis. Clearly genes making for resistance to the principal diseases killing before the end of the reproductive period must tend to spread. They may spread right through a population, so that the gene or genes for abnormal susceptibility to tuberculosis found in homozygous condition in inbred French children are reduced to moderately rare recessive. Had tuberculosis continued as a major killer of European youth for another thousand years, they might

have been reduced to a frequency of ·001 or so. We suggest that they are still fairly common in France because they have not been completely eliminated. If the genes conferring disease resistance give rise to a homozygote which is of subnormal fitness for some other reason, as with sickle cell haemoglobin, an equilibrium will be reached. But this will give low values of BA^{-1}, unless the commoner homozygote acts by depressing fertility or very early viability, rather than viability at readily observable ages.

However exposure to disease has another effect. When it causes a gene for resistance which was formerly rare to spread through a population, the gene in question was presumably rare because, in the absence of disease, it lowered fitness. Thus, any disease which kills or sterilizes many people over many generations will leave a genetical scar, a gene which is somewhat harmful in the absence of the disease, and is much commoner than it would have been had the disease never occurred. We owe this notion to L. S. Penrose. We suggest that modern European and American populations are now riddled with genes in the course of slow elimination by natural selection which in the past conferred at least a measure of resistance to bubonic plague, cholera, typhus, leprosy, and other nearly extinct diseases, as well as to infections such as typhoid fever and tuberculosis, which are now infrequent in Europe. Some of these relic genes are now fairly rare. If the effect of such a gene on infantile and juvenile viability is fully recessive, a frequency $q=·0625$ will give $BA^{-1}=16$; if the lowering of the viability of heterozygotes is 5% of that of homozygotes ($h=·05$), $BA^{-1}>9$. But genes with h as high as ·05 must be eliminated much more rapidly than fully recessive genes.

The argument may be extended. Many genes which made for fitness in palaeolithic men must have an opposite effect today, and may well have lowered fitness for the last 5,000 years. If they were even moderately dominant they were very quickly selected out. It is the probably small fraction which are nearly recessive which persist, and appear as the result of inbreeding. It is rarely safe to overlook the facts that mankind has evolved and is evolving.

We suggest, then, that the fairly rare recessive or nearly recessive genes which are responsible for the higher mortality of human beings of consanguineous parentage fall mainly into three classes, namely:

(1) Genes kept in being by mutation.

(2) Genes whose allelomorphs, when homozygous, have an effect less than that of the gene considered, in lowering fertility, or causing very early embryonic death.

(3) Relic genes, which were advantageous in the past, by conferring resistance to infection, or otherwise.

If this is so, Morton, Crow, and Muller's estimate of our load of recessive mutations remains correct. But their estimate of the rate of mutation needed to keep it in being is somewhat too high. It might conceivably be two or three times the true value. It might be only ten per cent or less above it. We may hope to solve the problem in at least two ways. One is by discovering the intensity in Haldane's (1954) sense of the selection needed to produce observed polymorphisms, such as that at the MN locus. Another is by studying the genetics of disease resistance in countries where lethal

infectious diseases are still common. The latter is the more urgent task, as such diseases are being rapidly eliminated all over our planet, and human genetics are little studied in the countries where death from infection is still common.

SUMMARY

The rare recessive human genes which increase the mortality of inbred human beings may not all be kept in being by mutation. They may include relic genes which were useful in the past but are harmful today. They may also include genes kept in being by greater fitness of the heterozygote at loci where the other homozygote has a lowered fertility rather than a lowered viability.

REFERENCES

ARNER, G. B. L. (1908)*. *Colombia Univ. studies in history, economics, and public law*, **31**, No. 3, pp. 1–99.

BEMISS, S. M. (1858)*. *Trans. Am. Med. Ass.*, **11**, 319–425.

DOBZHANSKY, TH. (1951). *Genetics and the origin of species.* Columbia University Press: New York.

HALDANE, J. B. S. (1937). The effect of variation on fitness. *Amer. Nat.*, **71**, 337–349.

HALDANE, J. B. S. (1954). The measurement of natural selection. *Caryologia*, Vol. Suppl., 480–487.

KERMACK, K. A. (1954). A biometrical study of *Micraster coranguinum* and *M.* (*Isomicraster*) *senonensis*. *Phil. Trans.*, **B 237**, 375–428.

MATTHEW, W. D. (1926). The evolution of the horse: a record and its interpretation. *Quart. Rev. Biol.*, **1**, 139.

MORTON, N. E., CROW, J. F. AND MULLER, H. J. (1956). An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Nat. Ac. Sci.*, **42**, 855–863.

NEEL, J. V. (1963). In *Genetic Selection in man* (p. 82). Ann Arbor: Univ. Mich.

PEARSON, K., LEE, A. AND BRAMLEY-MOORE, L. (1899). Mathematical contributions to the theory of evolution, VI. *Phil. Trans. Roy. Soc.*, **A 192**, 257–330.

SUTTER, J. AND TABAH, L. (1952). Effects de la consanguinité et de l'endogamie. *Population*, **7**, 249–266.

SUTTER, J. AND TABAH, L. (1953). Structure de la mortalité dans les familles consanguines. *Population*, **8**, 511–526.

* not read by the authors, references taken from Morton, Crow, and Muller.