VOLUME XXXI OCTOBER, 1935 No. 3

THE RATE OF SPONTANEOUS MUTATION OF A HUMAN GENE.

BY J. B. S. HALDANE.

SATISFACTORY data on rates of spontaneous mutation exist for *Zea* and *Drosophila*, but not for vertebrates. However, such data may be determined for man by indirect methods. Clearly the rate at which new autosomal genes recessive to the normal type appear can only be accurately estimated where either inbreeding or very extensive back-crossing to recessives is possible. But under ordinary conditions new dominants or sex-linked recessives can be detected more readily.

The sex-linked recessive condition haemophilia has been known for over a century. Since only a small minority of haemophilics live long enough to breed, and (as will be seen) over one-third of all haemophilia genes in new-born babies are in the *X*-chromosome of males, the condition would rapidly disappear unless new haemophilia genes arose by mutation. The only alternatives would be that heterozygous females were more fertile than normal, or that in their meiosis the normal allelomorph of haemophilia was preferentially extruded into a polar body. Neither of these alternatives seems likely.

It is, moreover, certain that new haemophilia genes sometimes arise by mutation. In the Gross family investigated by C. V. Green (Davenport, 1930) a woman (III, 7) had two haemophilic sons with normal colour vision, and two colour-blind sons with normal blood, besides a daughter who transmitted haemophilia to two sons. The mother therefore carried the gene for colour blindness in one *X*-chromosome and that for haemophilia in the other. As she had a colour-blind brother, maternal uncle and maternal aunt, but no haemophilic relatives, she must have received the *X*-chromosome carrying colour blindness from her mother, and that carrying haemophilia from her father. But her father was tested for haemophilia, and found to be normal. (Through error in Davenport's paper he is referred to as II, 3 in the pedigree, as II, 6 in the text.) Hence a mutation must have occurred in the *X*-chromosome received from the father. Davenport suggests that it occurred in the ovary of III, 7. But as all her sons were colour blind or haemophilic it seems more likely that it occurred in the testis of her father, or during her early embryonic life.

318 *The Rate of Spontaneous Mutation of a Human Gene*

Gettings' case (Fig. 603 in Bulloch and Fildes, 1912) is also good evidence for mutation. A woman had one normal and five haemophilic sons. Her father and maternal grandfather were normal. She had three normal brothers and three normal sisters who bore normal sons. Her mother's brothers and the male descendants of her mother's sisters were also normal. Hence her haemophilia gene probably, although not certainly, arose by mutation. There are a number of similar cases in the literature.

We now assume, and will later attempt to show, that most large human populations are in approximate equilibrium as regards haemophilia, selection being balanced by mutation.

If x be the proportion of haemophilic males in the population, and f their effective fertility, that is to say their chance, compared with a normal male, of producing offspring, then in a large population of $2N$, $(1-f)\,xN$ haemophilia genes are effectively wiped out per generation. The same number must be replaced by mutation. But as each of the N females has two X-chromosomes per cell, and each of the N males one, the mean mutation rate per X-chromosome per generation is $\frac{1}{3}(1-f)\,x$, or if f is small, a little less than $\frac{1}{3}x$. Hence we have only to determine the frequency of haemophilia in males to arrive at the approximate mutation rate.

This simple calculation is confirmed, and some other data arrived at, by a more formal treatment. Let **h** represent the gene for haemophilia, **H** its normal allelomorph, so that men are **H** or **h**, women **HH**, **Hh** or possibly **hh**. Homozygous haemophilic women are certainly very rare, and may not exist.

Let eggs be formed in the ratio 1 **H** : x**h**, and spermatozoa in the ratio 1 **H** : y**h**. Then if mating is at random the zygotes are formed in the proportions 1 **HH** : $(x+y)$ **Hh** : xy **hh** ♀, and 1 **H** : x **h** ♂. Inbreeding on the scale found in human populations would only serve to increase the proportion of **hh** females, which is in any case negligible. x and y are clearly small numbers.

Let f be the relative effective fertility of haemophilic males, f' that of haemophilic females. f means the number of children ultimately begotten by 1000 haemophilic males, divided by the number begotten by 1000 normal males. As many haemophilics die young and others do not marry, this number is a small fraction. Pearson pointed out to Bulloch and Fildes (1912) that the marriage rate of male bleeders is 9·6 per cent., that of females in the same families being 36·8 per cent. Further, a larger proportion of such males than of normal males presumably die while their wives are still potentially fertile.

So we may take f as 0·25 or less, though probably exceeding 0·1. It is likely that several allelomorphs of haemophilia exist. In Hay's case (408 of Bulloch and Fildes' collection) only four out of twenty-one affected males died of the disease, and four begot children. This family has now apparently been lost sight of, or haemophilia has died out in it. However, another family in the United States, the Molyneux of Pennsylvania, exhibits the same phenomenon. Green (Davenport, 1930) investigated branches of it, including nineteen haemophilics, besides other haemophilic families. Dr Green writes of one of them: "His does not seem to be a very severe case of haemophilia, but it is by far the most severe of any of the Molyneux." In the case of others haemophilia was not discovered until the age of five or ten years. Several of the boys play football, though suffering abnormally from bruises. There is no doubt that the character is a sex-linked recessive. As it has preserved its distinctive character of mildness in a number of different families, it cannot be due to the ordinary haemophilia gene along with modifiers in other chromosomes. There is strong reason to suspect that we are dealing with a distinct gene. If it is not allelomorphic with the ordinary gene we have the surprising, but not impossible, situation of two genes with very similar effects in the same one out of twenty-four chromosomes. It seems more likely that we are dealing with a relatively harmless allelomorph.

Whereas the normal allelomorph is at best very dangerous, there is no recorded death from haemophilia among Green's nineteen cases, and many have married and begotten children. Clearly the effective fertility of these "minor haemophilics" is not greatly depressed, and f probably exceeds $\frac{1}{2}$. If the two allelomorphs arose by mutation with equal frequency, the milder form would clearly be much more frequent. We therefore conclude that mutation to the major form is more frequent. This is quite in keeping with what is known elsewhere. In *Drosophila* mutations of the eye-colour to white are more frequent than to all its allelomorphs (eosin, apricot, etc.) together.

The value of f' for hh females is conjectural. Perhaps no such female has ever been observed. It may be that $f'=0$, the condition being lethal, as Davenport suggests, or that $f'=1$, the disease being sex-limited as well as sex-linked. However, the value of f' does not affect our calculations.

In their total of reliable families containing haemophilics, Bulloch and Fildes found a large excess of males over females and of haemophilic over normal males. In the forty pedigrees collected in the first part of their Table II, which they regard as the most satisfactory, there are 189 sibships containing 406 male bleeders, two doubtful males, 236 normal

320 *The Rate of Spontaneous Mutation of a Human Gene*

males, 464 females, and fifty-six whose sex is not stated. Eleven of the females were said to be bleeders. Now no family is included unless it contains at least one bleeder. The above total therefore does not give a fair sample of the progeny of heterozygous mothers, but it does give us the sibs of all the haemophilic males in these pedigrees, provided that we subtract one male haemophilic from each family, that is to say 189 from the total. We then get 217 bleeders, two doubtful, and 236 normal among the males, and 455 males, fifty-six doubtful, and 464 females for the sex distribution. Both are very satisfactory approximations to equality, and there is no reason whatever to doubt that segregation, both as regards sex and haemophilia, is perfectly normal. Hogben (1931), using different data and methods, found 28·5 per cent. of haemophilics in the progeny of haemophilic females, or quite a small deviation from the expected 25·5 per cent.

Bulloch and Fildes also find a rather high fertility among the mothers of haemophilics. In a group of forty pedigrees which they regard as their most reliable, they find that the average size of a family containing at least one haemophilic male is 6·1. This large size is partly an expression of bias in selection. For suppose we have in a population n heterozygous women who have had families of s children. The chance that any particular child is haemophilic is approximately $\frac{1}{4}$. So the chance that a family of size s will contain no haemophilic member is $1-(\frac{3}{4})^s$. Hence the number of such families actually recorded will be $n' = n\,[1-(\frac{3}{4})^s]$. Hence when in a collection of families we find n' of size s containing $n's$ children, these represent $\frac{n'}{1-(\frac{3}{4})^s}$ families from **Hh** mothers, containing $\frac{n's}{1-(\frac{3}{4})^s}$ children. For example Bulloch and Fildes' collection (in Table II, part 1) includes seven families of one child and twenty-two of two children. These figures should be raised to twenty-eight and 50·386 respectively. Applying the correction we find that the average fertility of **Hh** women in Bulloch and Fildes' data is reduced from their figure of 6·1 to 5·0. This figure is still rather high, but a glance at the Tenna and Mampel pedigrees shows that the homozygous women in them were also very fertile.

In a sample of the collected haemophilia pedigrees much more than half of those sisters of haemophilics who are recorded as having had sons had at least one haemophilic son. I do not regard this as significant. For a pedigree generally begins with a fairly young male. It is then discovered that he had haemophilic uncles and cousins. Hence the uncles'

in the pedigree are assured of at least one heterozygous sister. Also the progeny of haemophilics' sisters are likely to be included in the pedigree if they include haemophilic sons, but not otherwise. Until further data are available it seems best to assume that haemophilia obeys the same laws as the large majority of sex-linked characters in other animals.

Let μ be the frequency of mutation of the normal allelomorph **H** to the abnormal **h** per generation in women, ν in men. That is to say μ is the probability of finding, in any particular gamete of a woman, an X-chromosome now carrying **h**, although it carried **H** when she was conceived, and similarly with ν.

Then the effective breeding population is in the proportions

$$1\ \mathbf{HH} : (x+y)\ \mathbf{Hh} : f'xy\ \mathbf{hh}\ ♀, \text{ and } 1\ \mathbf{H} : fx\ \mathbf{h}.$$

So apart from mutation they would produce eggs in the ratio $1\ \mathbf{H} : x'\ \mathbf{h}$, and spermatozoa in the ratio $1\ \mathbf{H} : y'\ \mathbf{h}$, where $x' = \frac{x+y+2f'xy}{2+x+y} = \frac{1}{2}(x+y)$ very nearly, and $y' = fx$.

As a result of mutation these numbers must be equal to x and y respectively. Hence $x = \frac{1}{2}(x+y)+\mu$, $y = fx+\nu$. So

$$x = \frac{2\mu+\nu}{1-f},\quad y = \frac{2f\mu+\nu}{1-f},$$

and the mean mutation rate is $\frac{1}{3}(2\mu-\nu) = \frac{1}{3}(1-f)\,x$ as found above.

Before attempting to assess x, a few other calculations may be made. The ratio of heterozygous females to haemophilic males is $1+\frac{2f\mu+\nu}{2\mu+\nu}$, a quantity lying between $1+f$ and 2. This is the ratio at birth. In the actual population, owing to the high death-rate of the haemophilics, there may well be over twice as many heterozygotes as haemophilics.

In a population of N females and N males at fertilisation there are $3N$ X-chromosomes, of which a total $\frac{(4\mu+2f\mu+3\nu)\,N}{1-f}$ carry **h**. The number of new **h** genes arising by mutation is $(2\mu+\nu)\,N$, and the same number must die out, in a stationary population. Hence in such a population the mean life of a haemophilic gene in generations is

$$\frac{(4+2f)\,\mu+3\nu}{(1-f)\,(2\mu+\nu)},\ \text{or}\ \frac{3}{1-f}-\frac{2\mu}{2\mu+\nu}.$$

The first term certainly exceeds 3 and is probably less than 4; the second lies between 0 and 1, and is equal to $\frac{2}{3}$ if $\mu=\nu$. Hence the mean life of a haemophilia gene in a stationary population, *i.e.* the number of

individuals in whom it appears, lies between two and four generations, and is probably about three. In these calculations we neglect genes which *only affect a single cell or a group of cells, and do not determine the* genotype of an individual. In a phase of rapid population growth, such as is now closing in Western Europe and North America, the mean life is much longer. Thus if a population increases by 50 per cent. in each generation, the fact that one-third of the h genes are destroyed in each generation does not diminish the number of "old" h genes.

If the above analysis is correct the number of sporadic cases of haemophilia must be far larger than a study of pedigrees would suggest. This may well be the case, since it is unlikely that pedigrees containing single isolated cases would be published. Of all cases of haemophilia a fraction $\frac{(1-f)\,\mu}{2\mu+\nu}$ should be sons of homozygous mothers and wholly isolated. A further fraction $\frac{(1-f)\,(\mu+\nu)}{2\mu+\nu}$ should be sons of heterozygous mothers who had arisen by mutation, and should therefore have no haemophilic relatives except brothers, and other descendants of their mothers. An adequate survey of a whole population for haemophilia would allow of a verification or disproof of these predictions, and an estimate of the difference, if any, of mutation rate in the two sexes. The figures quoted above are all derived from a study of large pedigrees, so the proportion of haemophilic males was *only slightly below a half.*

We must now attempt to determine x, the fraction of all males born who develop haemophilia. No really satisfactory data exist. Bulloch and Fildes found only two haemophilics out of 137,676 consecutive admissions to the London Hospital (sex ratio not given). On the other hand Manson (1928) found three cases among 25,500 consecutive examined English recruits during the war.

Dr Julia Bell has records of thirty-four certain and seven doubtful cases at present living in Greater London, supplied by seven London hospitals. I find it difficult to suppose that this number represents as much as half the living cases in Greater London, or less than one-tenth. The number of cases among the four million or so males of London would thus be 70–350. If the expectation of life in haemophilics were normal, these numbers would probably be at least doubled. So a rough estimate of the proportion of haemophilics among male births in London is 35–175 per million. Thus x quite certainly exceeds 10^{-6}, and probably lies between 0·00004 and 0·00017. The population of London is drawn from *many different areas*, and this figure is likely to give a fairer idea of the

frequency in England as a whole than would a figure based on a similar population in northern England.

Now $1-f$ is at least 0·75 and probably larger. The mutation rate, $\frac{1}{3}(1-f)x$, is therefore certainly greater than 1 in 400,000, probably lying between 1 in 100,000 and 1 in 20,000. We may take 1 in 50,000 as a plausible figure.

This estimate is very little affected by the value of f, provided this lies between 0 and $\frac{1}{2}$, which is undoubtedly true. It has been assumed that the population is in approximate equilibrium. Let us suppose that this is not the case, but that f, μ and ν are constant. Let x_n be the value of x in the nth generation, and let $x_n = z_n + \frac{2\mu+\nu}{1-f}$. Then

$$x_{n+2} - \tfrac{1}{2}x_{n+1} - \tfrac{1}{2}fx_n = \mu + \tfrac{1}{2}\nu.$$

Whence $$z_{n+2} - \tfrac{1}{2}z_{n+1} - \tfrac{1}{2}fz_n = 0,\ z_n = a\lambda^n + b\kappa^n,$$

where $$\lambda = \tfrac{1}{4}(1+\sqrt{8f+1}),\ \ \kappa = \tfrac{1}{4}(1-\sqrt{8f+1}),$$

and a and b are constants determined by the earlier state of the population. After a few generations the second term becomes negligible, and

$$x_n = \frac{2\mu+\nu}{1-f} + a\lambda^n.$$

λ lies between 0·5 for $f=0$ and 0·683 for $f=0{\cdot}25$. So any departures from equilibrium diminish in geometric progression. How rapidly they do so may be seen by examining the hypothesis that all existing haemophilics are derived from haemophilics or heterozygotes in an earlier population, and that mutation does not occur. In that case the frequency of haemophilia thirty generations ago, if $f=0{\cdot}25$, the least favourable hypothesis for the mutation theory, would have been 100,000 times as great as at present, in other words all Englishmen at the time of the Norman conquest would have been haemophilics!

Hence the existing population must be close to equilibrium unless the mutation rate has recently changed. If this is so our estimate lies between the present value and the value a few generations ago. The only alternatives to accepting a mutation rate of the order calculated seem to be the postulation of a fertility in heterozygous females of $\frac{2}{1+f}$, or over 1·6 times the normal, or a correspondingly great abnormality in their segregation. But in the male sex segregation is entirely normal. So if the females are more likely to receive a haemophilic than a normal gene, while males are not, we must assume selective fertilisation. Unless

324 *The Rate of Spontaneous Mutation of a Human Gene*

some such unjustifiable assumption is made, it must be accepted that the mutation rate is of the order of magnitude here given.

DISCUSSION.

The figure arrived at, of about one spontaneous mutation in about 50,000 life cycles, is within the limits found for other organisms. Thus in *Zea* Stadler (1932) found rates from just over 1 per 1000 to less than 1 per 1,000,000 for seven colour genes. In *Drosophila melanogaster* Muller (1928) found a rate of nearly 1 in 100 for the appearance of lethal genes at a particular locus in the second chromosome. These, however, may have been deficiencies, as they occurred near the end of an inversion in a balanced lethal stock. On the other hand Patterson and Muller (1930) estimate the rate of mutation at the white locus as 1 in 400,000 to white, and 1 in 600,000 to other allelomorphs. A few other loci showed a rate of the same order. These rates can be increased about 100 times by X-rays.

Thus expressed in frequency per life cycle the rate of mutation at the haemophilia locus is higher than any well-authenticated mutation rate in *Drosophila*. The same is probably true if we consider frequency per cell generation, since the number per life cycle in man is only about twice that in *Drosophila*. On the other hand the rate per year is very much larger in *Drosophila* than in man. Taking the mean lengths of a generation as 14 days and 30 years, the human mutation rate corresponds to a rate of 1 in 40,000,000 per *Drosophila* generation, and less if spontaneous mutation rate is a function of temperature. A gene appearing with this rate, if obvious in its manifestation, would probably have been detected once or twice if sex-linked, and not at all if autosomal.

Penrose's (1934) data on epiloia, a rare human autosomal dominant which does not often survive for more than two or three generations before it is extinguished by selection, suggest a mutation rate of the same order of magnitude as that here found for haemophilia.

Bulloch and Fildes, from a survey of the literature, think that haemophilia is a good deal commoner in northern than southern Europe. The data collected by Komai (1934) show that it is not very rare in Japan. If differences exist they are almost certainly due to differences in the mutation rate rather than the effective fertility. These in turn may be due to genetical or environmental differences between the peoples concerned. It is clear that this question, like that of the exact value of the mutation rate, can only be solved when an adequate biological survey of several human populations of a million or more have been made.

Similar calculations may be made for other rare lethal or sublethal human abnormalities. Haldane (1927) showed that for a rare dominant or an autosomal recessive of frequency x the mutation rate required to preserve equilibrium in a random mating population is $(1-f)\,x$. This is equally true for any other mating system. For in a population of N, $2N\,(1-f)\,x$ abnormal recessive genes are eliminated by selection in each generation, and the same number must be furnished anew by mutation. This argument is unassailable for dominants or sex-linked recessives, but must be applied with great care to autosomal recessives. For if a recessive character has a frequency of 1 per 1,000,000, then in a random mating population there are 2000 heterozygotes per 1,000,000. If the gene is lethal, an increased effective fertility of 0·05 per cent. in the heterozygotes would be enough to counteract the selection of the homozygotes. It is quite possible that heterozygosis for amaurotic idiocy may decrease the intelligence quotient by 1 per cent. This would probably cause an increased fertility under existing social conditions.

Moreover, in the case of autosomal recessives, after any disturbance, such as might be produced by a change in the amount of inbreeding or by several other causes, the equilibrium between selection and mutation is only re-established with extreme slowness, whereas in other cases the process is very rapid. Hence the frequency of occurrence of lethal and sublethal recessive abnormalities does not do more than suggest the order of magnitude of the mutation rates of the genes concerned.

Besides haemophilia and epiloia a number of other sublethal dominant or sex-linked conditions are known. Thus cleidocranial dysostosis, neuro-fibromatosis, and blue sclerotics associated with bone fragility are dominants, while anidrotic ectodermal dysplasia can be a sex-linked recessive, or sex-linked with occasional dominance in females. These diseases are not very uncommon, and Cockayne (1933) points out that neurofibromatosis often arises by mutation, while his data show that the effective fertility is quite low. In these cases we must assume a mutation rate of over 1 per 1,000,000, and the same is probably true for lethal and sublethal recessives such as the amaurotic idiocies and xeroderma pigmentosum. If the above arguments are correct it would seem that, taking the generation as the unit of time, man is a rather more mutable species than *Drosophila*.

SUMMARY.

The rate of mutation at which the gene for haemophilia appears in the population of London is estimated at about once in 50,000 human

life cycles. There are probably two distinct allelomorphs at the same locus, the milder type arising less frequently by mutation than the severe type.

I have to thank Dr Julia Bell and Dr C. V. Green for most generously placing at my disposal data collected on behalf of the Medical Research Council and the Research Committee of the American Medical Association.

REFERENCES.

BULLOCH, W. and FILDES, P. (1912). *Treasury of Human Inheritance*, **1**, 169.
COCKAYNE, E. A. (1933). *Inherited Abnormalities of the Skin and Appendages*. Oxford.
DAVENPORT, C. B. (1930). "Sex linkage in man." *Genetics*, **15**, 401.
HALDANE, J. B. S. (1927). "A mathematical theory of natural and artificial selection. Part V." *Proc. Camb. phil. Soc.* **23**, 838.
HOGBEN, L. T. (1931). *Genetic Principles in Medicine and Social Science*. London.
KOMAI, T. (1934). *Pedigrees of Hereditary Diseases and Abnormalities found in the Japanese Race*. Kyoto.
MANSON, J. S. (1928). *Observations on Human Heredity*. London.
MULLER, H. J. (1928). "The measurement of gene mutation rate in *Drosophila*, its high variability, and its dependence upon temperature." *Genetics*, **13**, 279.
PATTERSON, J. T. and MULLER, H. J. (1930). "Are 'progressive' mutations produced by X-rays?" *Ibid.* **15**, 495.
PENROSE, L. S. (1934). *The Influence of Heredity on Disease*. London.
STADLER, L. J. (1932). "On the genetic nature of induced mutations in plants." *Proc. 6th Int. Cong. Genet.* Menasha, Wis.