

- ²⁸ See, for example, C. W. Beck, R. G. Stevenson, Jr., and L. LaPaz, *Popular Astronomy*, 59, No. 2 (February, 1951), 93 ff.
- ²⁹ W. Wahl, *Geochim. et cosmochim. acta*, 2, 91, 1952.
- ³⁰ See H. C. Urey, *Geochim. et cosmochim. acta*, 1, 209-277, 1951; see, particularly, pp. 233 ff. This paper was reproduced in *The Planets*, pp. 124 and 141, particularly. Later G. P. Kuiper published the same idea in *Atmospheres of the Earth and Planets* (Chicago: University of Chicago Press, 1952), pp. 330, 314 (ref. 18), and 421.
- ³¹ See H. Jeffreys, *The Earth* (4th ed.; Cambridge: At the University Press, 1952), chap iv, p. 159, and papers there cited; H. C. Urey, *The Planets*, pp. 20 ff.
- ³² See H. C. Urey, *The Planets*, p. 124, and *Astrophys. J., Suppl.*, 1, 155, 1954.
- ³³ H. C. Urey, *The Planets*, pp. 49 ff.
- ³⁴ See B. Gutenberg, *Internal Constitution of the Earth* (New York: Dover Publications, Inc., 1951), pp. 150 ff.
- ³⁵ H. C. Urey, *Proc. Roy. Soc. London, A*, 219, 281, 1953.
- ³⁶ H. C. Urey, *The Planets*, p. 176.
- ³⁷ See H. C. Urey, *Phys. Rev.*, 80, 295, 1950; *The Planets*, pp. 64 ff.
- ³⁸ D. B. McLaughlin, *Observatory*, 74, 166, 1954; *Pub. Astron. Soc. Pac.*, 66, 161, 221, 1954.
- ³⁹ M. Davidson, *J. Brit. Astron. Assoc.*, 29, 194, 1919; see H. C. Urey, *The Planets*, pp. 43 ff.
- ⁴⁰ H. C. Urey, *Astrophys. J., Suppl.*, 1, 167-172, 1954.
- ⁴¹ A. Blaauw, *Bull. Astron. Inst. Neth.*, 11, 405, 414, 1952; A. Blaauw and M. P. Savendoff, *Bull. Astron. Inst. Neth.*, 12, 69, 1953.
- ⁴² J. R. Simonton, R. A. Rightmire, A. L. Long, and T. P. Kohman, unclassified report No. NYO-66.25; *Phys. Rev.* (in press).

SOLUTION OF A PROCESS OF RANDOM GENETIC DRIFT WITH A CONTINUOUS MODEL*

BY MOTOO KIMURA

DEPARTMENT OF GENETICS, † UNIVERSITY OF WISCONSIN

Communicated by Sewall Wright, November 15, 1954

The problem of random genetic drift in finite populations due to random sampling of gametes in reproduction was first treated mathematically by R. A. Fisher,¹ using a differential equation. Fisher's general method was appropriate, but owing to the omission of a term in the equation, his result for the rate of decay of variance was only half large enough. The correct solution for the state of steady decay was first supplied by S. Wright,² using the method of path coefficients and an integral equation.

Later Fisher³ corrected his results and also elaborated the terminal part of the distribution in the statistical equilibrium by his method of functional equations.

In all these works, however, it was assumed that a state of steady decay had been attained, but nothing was known about the complete solution which might show how the process finally leads to the state of steady decay. The present writer,⁴ by calculating the moments of the distribution and with the help of the Fokker-Planck equation, obtained a solution which assumed an infinite series under the continuous model, showing that the process approaches asymptotically the state of steady decay. At that time, however, only the first few coefficients in the terms of the series could be determined. Pursuing the problem further, he arrived at the complete solution, which will be reported here. After obtaining these results, the writer recently discovered the work of S. Goldberg.⁵ In his unpublished thesis Goldberg

solved the diffusion equation for the gene-frequency distribution in a finite population when recurrent mutations occur. His solutions have a direct connection with the frequency distribution of unfixed classes in the case of pure random genetic drift, as will be seen below.

Consider a random mating population of N diploid parents. Let A and A' be a pair of alleles with frequencies x and $1 - x$, respectively. In order to single out the effect of random drift, we shall assume an idealized situation in which selection, migration, and mutation are absent and generations do not overlap. The process of the change in gene frequencies is most adequately described by giving the frequency distribution $f(x, t)$ at the t th generation, where x takes on a series of discrete values: $0, 1/2N, 2/2N, \dots, 1 - 1/2N, 1$. For fairly large N , however, x can be treated as a continuous variable without serious error.

First, we shall derive the moment formula which is useful heuristically. Let x_t be the gene frequency in the t th generation, and let δx_t be the amount of change due to random sampling of gametes per generation, such that

$$x_{t+1} = x_t + \delta x_t. \tag{1}$$

Let $\mu_n^{(t+1)} = E(x_{t+1}^n)$ be the n th moment of the distribution about 0 in the $(t + 1)$ th generation. We write the expectation of x_{t+1}^n in terms of $(x_t + \delta x_t)^n$ in two steps: (1) taking expectation for the random change δx_t , which will be denoted by E_δ , and (2) taking the expectation for the existing distribution, which will be denoted by E_ϕ .

Noting that $E_\delta(\delta x_t) = 0, E_\delta(\delta x_t)^2 = x_t(1 - x_t)/2N$, etc.,

$$\begin{aligned} \mu_n^{(t+1)} &= E(x_t + \delta x_t)^n \\ &= E_\phi \left\{ x_t^n + \binom{n}{1} x_t^{n-1} E_\delta(\delta x_t) + \binom{n}{2} x_t^{n-2} E_\delta(\delta x_t)^2 + \dots \right\} \\ &= E_\phi \left\{ x_t^n + \frac{n(n-1)}{2} x_t^{n-2} \frac{x_t(1-x_t)}{2N} + O\left(\frac{1}{N^2}\right) \right\}. \end{aligned} \tag{2}$$

The intrinsic assumption in the continuous model is that the effective size N is sufficiently large so that terms of order $1/N^2$ and higher can be omitted without serious error.

Thus

$$\mu_n^{(t+1)} = \left\{ 1 - \frac{n(n-1)}{4N} \right\} \mu_n^{(t)} + \frac{n(n-1)}{4N} \mu_{n-1}^{(t)}.$$

For large N , the moments change very slowly per generation, and we can replace the above equation by the system of differential equations:

$$\frac{d\mu_n^{(t)}}{dt} = -\frac{n(n-1)}{4N} \{ \mu_n^{(t)} - \mu_{n-1}^{(t)} \} \quad (n = 1, 2, 3, \dots). \tag{3}$$

If the population starts from the gene frequency p ($0 < p < 1$), $\mu_n^{(0)} = p^n$, and we can obtain the n th moment as the solution of (3),

$$\begin{aligned} \mu_n^{(t)} &= p + \sum_{i=1}^{\infty} (2i+1)pq(-1)^i F(1-i, i+2, 2, p) \times \\ &\quad \frac{(n-1) \dots (n-i)}{(n+1) \dots (n+i)} e^{-[i(i+1)/4N]t}, \end{aligned} \tag{4}$$

where $F(1 - i, i + 2, 2, p)$ is the hypergeometric function and $q = 1 - p$. For finite n the series is finite. Putting $n = 1, 2, 3$, and 4 , it will be seen that, for large N , the resulting formulas give a very good approximation to the exact moment formulas obtained by A. Robertson⁶ (p. 205).

The probability $f(1, t)$ of the gene A being fixed in the population by the t th generation can be obtained by using the relation

$$f(1, t) = \lim_{n \rightarrow \infty} \sum_{x=0}^1 x^n f(x, t) = \lim_{n \rightarrow \infty} \mu_n^{(t)}.$$

The resulting series,

$$f(1, t) = p + \sum_{i=1}^{\infty} (2i + 1)pq(-1)^i F(1 - i, i + 2, 2, p)e^{-[i(i+1)/4N]t}, \quad (5)$$

is now an infinite series whose convergence must be examined. It is convenient here to introduce the Gegenbauer polynomial $T_{i-1}^1(z)$ which is related to the hypergeometric function by

$$T_{i-1}^1(z) = \frac{i(i + 1)}{2} F\left(i + 2, 1 - i, 2, \frac{1 - z}{2}\right).$$

The properties of this function have been thoroughly studied (see Morse and Feshbach⁷ [pp. 782-783]). Using this relation and putting $p = (1 - r)/2$, ($-1 < r < 1$), we obtain

$$f(1, t) = p + \sum_{i=1}^{\infty} (-1)^i \frac{(2i + 1)}{2i(i + 1)} (1 - r^2) T_{i-1}^1(r) e^{-[i(i+1)/4N]t}, \quad (6)$$

where $T_0^1(r) = 1$, $T_1^1(r) = 3r$, $T_2^1(r) = \frac{3}{2}(5r^2 - 1)$, $T_3^1(r) = \frac{5}{2}(7r^3 - 3r)$, etc. Here, if we use the recurrence relation: $(2i + 1)(1 - r^2)T_{i-1}^1(r) = i(i + 1)P_{i-1}(r) - i(i + 1)P_{i+1}(r)$, the above formula becomes

$$f(1, t) = p + \sum_{i=1}^{\infty} \frac{(-1)^i}{2} \{P_{i-1}(r) - P_{i+1}(r)\} e^{-[i(i + 1)/4N]t}, \quad (7)$$

where $P_n(r)$ represents a Legendre polynomial: $P_0 = 1$, $P_1 = r$, $P_2 = \frac{1}{2}(3r^2 - 1)$, $P_3 = \frac{1}{2}(5r^3 - 3r)$, etc. For $t = 0$, the partial sum of the first n terms of equation (7) is $(-1)^{n-1}(P_{n-1} - P_n)/2$ ($n \geq 3$). By using a proper integral expression (see later part), it can easily be shown that, if $-1 < r < 1$, the partial sum tends to zero as n goes to infinity. For $t > 0$, the series (7) is uniformly convergent and tends to p as $t \rightarrow \infty$. The probability of the gene A' being fixed in the population by the t th generation $f(0, t)$ can be obtained by replacing p with q and r with $-r$. If we note that $P_n(-r) = (-1)^n P_n(r)$, the frequency of the fixed classes is seen to be

$$f(1, t) + f(0, t) = 1 - \sum_{j=0}^{\infty} \{P_{2j}(r) - P_{2j+2}(r)\} e^{-[(2j+1)(2j+2)/4N]t}, \quad (8)$$

which is 0 when $t = 0$ and tends to 1 as $t \rightarrow \infty$.

Let us now consider the probability distribution of unfixed classes. Let $\phi(x, t)$ be the probability density that the gene frequency in the t th generation is between x and $x + dx$ ($0 < x < 1$). It has been shown that, under the assumption of the continuous model, $\phi(x, t)$ satisfies the following partial differential equation:

$$\frac{\partial \phi}{\partial t} = \frac{1}{4N} \frac{\partial^2}{\partial X^2} \{x(1-x)\phi\}, \tag{9}$$

which is a Fokker-Planck equation for the case of the random drift.^{4, 8} This equation has singularities at the boundaries, and no arbitrary conditions can be imposed there.⁹ But the moment formula which can be obtained from equations (4) and (5) by calculating,

$$\mu_n^{(t)} - 1^n f(1, t) = \int_0^1 x^n \phi(x, t) dx, \tag{10}$$

suggests that equation (9) must have the solution of the form

$$\sum_{i=1}^{\infty} C_i X_i(x) e^{-[i(i+1)/4N]t},$$

where C_i are constants and $X_i(x)$ are functions of x only. In order to solve equation (9), if we put $\phi \propto X_i(x) \exp \{-i(i+1)t/4N\}$ ($i = 1, 2, 3, \dots$), we obtain the hypergeometric equation

$$x(1-x) \frac{d^2 X_i}{dx^2} + 2(1-2x) \frac{dX_i}{dx} - (1-i)(i+2)X_i = 0,$$

or, putting $x = (1-z)/2$ such that $z = 1 - 2x$ ($-1 < z < 1$), we obtain the Gegenbauer equation:

$$(z^2 - 1) \frac{d^2 X_i}{dz^2} + 4z \frac{dX_i}{dz} - (i-1)(i+2)X_i = 0. \tag{11}$$

From the comparison of the results obtained from equation (10), it can be found that a Gegenbauer polynomial $X_i = T_{i-1}^1(z)$ is a pertinent solution.

Thus

$$\phi(x, t) = \sum_{i=1}^{\infty} C_i T_{i-1}^1(z) e^{-[i(i+1)/4N]t}. \tag{12}$$

The coefficients C_i can be determined from the initial condition that the population starts from the gene frequency p . Mathematically,

$$\delta(x - p) = \sum_{i=1}^{\infty} C_i T_{i-1}^1(z), \tag{13}$$

where $\delta(x)$ represents the delta function. Multiplying $(1-z^2)T_{i-1}^1(z)$ on both sides of equation (13) and using the orthogonal property,

$$\int_{-1}^1 (1-z^2) T_m^1(z) T_{i-1}^1(z) dz = \delta_{m, i-1} \frac{2(i+1)i}{(2i+1)}, \tag{14}$$

where m in Kronecker's notation represents zero or a positive integer, we obtain

$$2\{1 - (1-2p)^2\} T_{i-1}^1(1-2p) = C_i \frac{2(i+1)i}{(2i+1)}$$

or

$$C_i = 4pq \frac{(2i+1)}{i(i+1)} T_{i-1}^1(1-2p).$$

Thus the formal solution is

$$\phi(x, t) = \sum_{i=1}^{\infty} \frac{(2i+1)(1-r^2)}{i(i+1)} T_{i-1}^1(r) T_{i-1}^1(z) e^{-[i(i+1)/4N]t} \tag{15}$$

or, in terms of the hypergeometric function,

$$\phi(x, t) = \sum_{i=1}^{\infty} pqi(i+1)(2i+1)F(1-i, i+2, 2, p) \times F(1-i, i+2, 2, x) e^{-[i(i+1)/4N]t}. \tag{15}'$$

For $t > 0$, the series is uniformly convergent for x and p , since the exponential term approaches zero rapidly. It is interesting that this solution agrees with the "absorbing barrier solution" of Goldberg⁵ by putting $\alpha = \beta = 0$ in his formula.

The probability that both A and A' coexist in the population in the t th generation (Ω_t) is easily obtained from equation (15), by noting that $dP_i(z)/dz = T_{i-1}^1(z)$ and $P_n(1) = 1$:

$$\begin{aligned} \Omega_t &= \int_0^1 \phi(x, t) dx = \int_{-1}^1 \phi(x, t) \frac{dz}{2} \\ &= \sum_{m=1}^{\infty} \frac{(4m-1)(1-r^2)}{(2m-1)2m} T_{2m-2}^1(r) e^{-[(2m-1)2m/4N]t}. \end{aligned} \tag{16}$$

For $t > 0$, the series is easily seen to be convergent, and, as $t \rightarrow \infty$, Ω_t goes to zero. For $t = 0$, we must show that this series converges to 1. Let $\Omega_{0, n}$ be a partial sum of the first n terms; then, by the recurrence relation $(4m-1)(1-r^2)T_{2m-2}^1(r)/(2m-1)2m = P_{2m-2}(r) - P_{2m}(r)$, we have $\Omega_{0, n} = 1 - P_{2n}(r)$. By using an integral expression of P_n , i.e., $P_n(z) = (1/\pi) \int_0^\pi \{z + \sqrt{z^2 - 1} \cos t\}^n dt$, we can show that, for $|r| < 1$ $P_{2n}(r) \rightarrow 0$ as $n \rightarrow \infty$. For

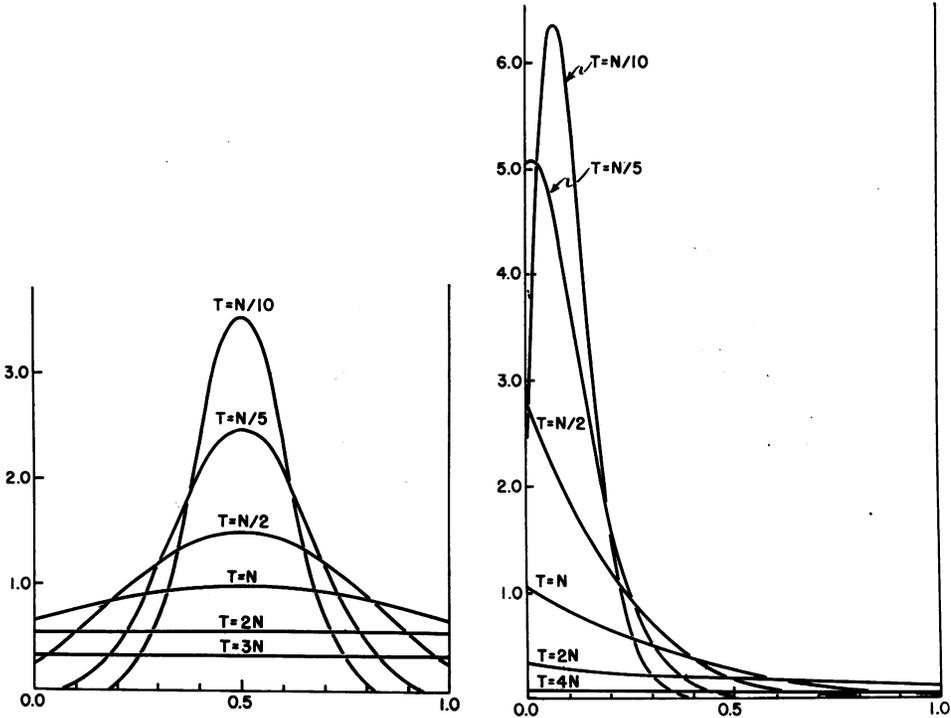
$$\begin{aligned} |P_{2n}(r)| &\leq \frac{1}{\pi} \int_0^\pi |r + \sqrt{r^2 - 1} \cos t|^{2n} dt = \\ &\frac{1}{\pi} \int_0^\pi \{r^2 + (1-r^2) \cos^2 t\}^n dt \rightarrow 0 \quad (n \rightarrow \infty). \end{aligned}$$

Furthermore, from (16),

$$\Omega_t = \sum_{j=0}^{\infty} \{P_{2j}(r) - P_{2j+2}(r)\} e^{-[(2j+1)(2j+2)/4N]t}. \tag{17}$$

Consequently, it can be seen that, from equations (8) and (17), $f(1, t) + \Omega_t + f(0, t) = 1$, as it should be.

The processes of the change in the distribution of the unfixed classes when the population starts from $p = 0.5$ and $p = 0.1$ are illustrated in Figures 1 and 2, respectively. In Figure 1 it will be seen that after $2N$ generations the distribution curve becomes almost flat, and the genes are still unfixed in about 50 per cent of the cases. In Figure 2, the initial gene frequency is assumed to be 10 per cent, and it takes $4N$ or $5N$ generations before the distribution curve becomes practically flat. By that time, however, the genes are fixed in more than 90 per cent of the cases, and the simplest asymptotic formula $Ce^{-(1/2N)t}$ may not be useful as in the case of $p = 0.5$.



FIGS. 1-2.—The processes of the change in the probability distribution of heterallelic classes, due to random sampling of gametes in reproduction. It is assumed that the population starts from the gene frequency 0.5 in Fig. 1 (left) and 0.1 in Fig. 2 (right). t = time in generation; N = effective size of the population; abscissa is gene frequency; ordinate is probability density.

The probability of heterozygosis is calculated by equation (15):

$$H_t = \int_0^1 2x(1-x)\phi(x, t) dx = \sum_{i=1}^{\infty} pq \frac{(2i+1)}{i(i+1)} T_{i-1}^1 (1-2p) \times \int_{-1}^1 (1-z^2) T_{i-1}^1(z) e^{-[i(i+1)/4N]t} dz.$$

By virtue of equation (14) (put $m = 0$), the last integral is 0 except for $i = 1$. Hence

$$H_t = pq \cdot \frac{3}{2} \cdot 1 \cdot \frac{4}{3} \cdot e^{-(1/2N)t} = 2pqe^{-(1/2N)t} = H_0e^{-(1/2N)t}, \tag{18}$$

showing that the heterozygosis decreases exactly at the rate of $1/(2N)$ per generation. This is readily confirmed by a simple calculation: Let p be the frequency of A in the population, where the frequency of the heterozygotes is $2p(1-p)$. The amount of heterozygosis to be expected after one generation of random sampling of the gametes is

$$E\{2(p + \delta p)(1 - p - \delta p)\} = 2p(1-p) - 2E(\delta p)^2 = 2p(1-p) - 2 \frac{p(1-p)}{2N} = \left(1 - \frac{1}{2N}\right) 2p(1-p),$$

as was to be shown.

The author expresses his appreciation to Dr. James F. Crow for valuable help during the course of this work. Thanks are also due to Dr. E. R. Immel for his helpful suggestions.

* Contribution No. 84 of the National Institute of Genetics, Mishima-shi, Japan.

† Contribution No. 570. This work was supported by a grant from the University Research Committee from funds supplied by the Wisconsin Alumni Research Foundation.

¹ R. A. Fisher, *Proc. Roy. Soc. Edinburgh*, **42**, 321-341, 1922.

² S. Wright, *Genetics*, **16**, 97-159, 1931.

³ R. A. Fisher, *The Genetical Theory of Natural Selection* (Oxford, 1930).

⁴ M. Kimura, *Genetics*, **39**, 280-295, 1954.

⁵ S. Goldberg, Ph.D. thesis, Cornell University, 1950.

⁶ A. Robertson, *Genetics*, **37**, 189-207, 1952.

⁷ M. Morse and H. Feshbach, *Methods of Theoretical Physics* (New York, 1953).

⁸ S. Wright, these PROCEEDINGS, **31**, 382-389, 1945.

⁹ W. Feller, *Proc. Second Berkeley Symposium on Math. Stat. and Prob., Univ. of California*, pp. 227-246, 1951.

AN X-RAY ANALYSIS OF CHROMOSOME DUPLICATION*

BY KARL SAX AND EDWARD D. KING

ARNOLD ARBORETUM AND BUSSEY INSTITUTION, HARVARD UNIVERSITY, JAMAICA PLAIN,
MASSACHUSETTS

Communicated January 5, 1955

The induction of half-chromatid aberrations by X-rays provides evidence regarding the multiple nature of the chromosome and may be of considerable significance in the analysis of chromosome duplication. The occurrence of half-chromatid exchanges was first described by Swanson¹ in 1943. These aberrations, found in pollen-tube mitoses of *Tradescantia*, were of sporadic occurrence. Recently Crouse² has found that practically all the X-ray-induced aberrations induced at the first meiotic metaphase stage in *Lilium* involve half-chromatid breaks and exchanges. Breaks in one of the half-chromatids of each of two chromatids are followed by reciprocal translocation to produce a chromatid bridge at anaphase. The close association of the coiled half-chromatids prevents the terminal separation of the two anaphase chromosomes (cf. Fig. 1).

Occasional half-chromatid aberrations have been found at anaphase in the division of the microspore nucleus of *Tradescantia* following X-irradiation. The dosages commonly used (50-150 r) produced considerable stickiness of the chromosomes, so that accurate analysis of chromosomal aberrations could not be made until about 6 hours after raying. By reducing the dosage to 25 r and keeping the inflorescences at 3° C. during irradiation, it was possible to obtain clear figures of mitoses and induced aberrations as early as 3 hours after raying. Irradiation at 3° C. was done to compensate for the low dosage, since it has been shown that the chromosome aberration frequency can be increased by raying at low temperatures.³

Chromosomal aberrations induced at 4-6 hours before anaphase consisted almost entirely of half-chromatid bridges at anaphase. Presumably, these chromosomes were irradiated at prometaphase or metaphase. At these stages the two sister