数理解析研究所講究録
第195巻 1973年 87-106



# A Markov process with stochastic time parameter: the gene frequency change in a geographically-structured population of finite and fixed total number

TAKEO MARUYAMA*

National Institute of Genetics, Mishima

Abstract

A Markov process (chain) of gene frequency change was derived for a geographically-structured model of a population of fixed total size. The population consists of colonies, which are connected by migration, and mating and selection are done in each colony independently. It was shown that, if the sum of the heterozygosity appeared in the population is used as the (stochastic) time parameter, the process becomes a random walk which is independent of the geographical structure of the population. This time parameter is not exact if the mutants are under selection, but is a close approximation. As a limit of large population size, a diffusion process was derived. The transition probabilities of the Markov chain and of the diffusion process were obtained explicitly. Certain quantities of biological interest are shown to be independent of the population structure. The quantities are the fixation probability of a mutant, the sum of the number of heterozygotes for a mutant, and the heterozygosity summed over the generations in which the gene frequency in the whole population assumes a specified value.

* 国立遺伝学研究所　丸山毅夫

## 1. Introduction

The stochastic model of gene frequency change in a population of fixed finite size put forward by Fisher (1930) and Wright (1931 and later) has become increasingly important in population genetics and evolutionary theory. For reviews, see Moran (1962); Karlin and McGregor (1964); Kimura (1964); Ewens (1969); Wright (1969); Crow and Kimura (1970); Cavalli-Sforza and Bodmer (1971); Kimura and Ohta (1971).

Most of the theory, however, have dealt with random mating populations, while hardly any natural population meets this assumption for various reasons. One of them would be the geographical structure of the population which prevents the random mating, and when we take this into consideration the mathematics becomes very difficult. Therefore the purposes of the present study is to develop a Markov process of gene frequency change in a geographically structured case and investigate its properties. Throughout this paper, we consider one locus where two alleles, A and a, are segregating. The fitnesses of A and a are $1 + s$ and 1 respectively, and the fitnesses are constant in time and throughout the whole population. No mutation or migration from outside world occurs during the time under consideration.

## 2. Derivation of the Markov process.

The population consists of colonies. Migration and/or reformation of the population occur and they are the factors which connect colonies genetically. Each individual has the negative exponential life time distribution with unit mean.

When an individual dies, it is immediately replaced by a new individual born in the same colony. This assumption implies that the total population size is constant. We denote the size of the total population by N. Each pre-existing individual in the colony has a chance of being chosen to reproduce, which is proportional to that individual's fitness. This is a geographically-structured version of Moran's model, (Moran 1958). This stochastic process is not a branching type process. The most important biological limitation of a branching process is that the fate of each individual (particle) is independent from others.

Let $N_{t,k}$ be the colony size at time t and let $M_{t,k}$ be the number of A in colony k at time t. Let $M_t = \sum_k M_{t,k}$. Then random variable $M_t$ is a stochastic process whose state space is [0, 1, 2, $\cdots$, N - 1, N], where N is a fixed integer. This process with time parameter t is very complicated because of the geographical structure of the population. Therefore, in below, we will introduce a new time measure which makes the process into a simple Markov process.

First, note that, in any short time interval, $\Delta t$, two or more death-birth events occur with probability of order $(\Delta t)^2$. Therefore random variable $M_t$ changes by one at any moment. Now consider a particular moment when a birth-death occurs, and ask what is the probability that the value of $M_t$ changes from i to i + 1 through this event? Let $i_k$ be the value of $M_{t,k}$ at this moment, i. e., $i = \sum_k i_k$.

The probability that the death-birth occurs in colony k whose size is $N_{t,k}$ is $N_{t,k}/N$, and given this, the probability,

$q_{i_k,i_k+1}$, that one a in the colony dies and it is replaced by one A, is

$$q_{i_k,i_k+1} = \frac{i_k(N_{t,k} - i_k)(1 + s)}{N_{t,k}(N_{t,k} + si_k)} .$$

The probability, $q_{i_k,i_k-1}$, that one A dies and replaced by one a, is

$$q_{i_k,i_k-1} = \frac{i_k(N_{t,k} - i_k)}{N_{t,k}(N_{t,k} + si_k)} .$$

The probability, $q_{i_k,i_k}$, that no change in the number of A occurs, is

$$q_{i_k,i_k} = 1 - q_{i_k,i_k+1} - q_{i_k,i_k-1} .$$

Therefore the probability, H(t), that the value of $M_t$ changes through the death-birth event is

$$H(t) = \frac{1}{N} \sum_k N_{t,k}[q_{i_k,i_k+1} + q_{i_k,i_k-1}]$$

$$= \frac{1}{N} \sum_k N_{t,k} \left[\frac{i_k(N_{t,k} - i_k)}{N_{t,k}^2}\right] \left\{1 + \frac{s\left(\frac{i_k}{N_{t,k}} - \frac{1}{2}\right)}{1 + \frac{s}{2}}\right\} . \qquad (2\text{-}1)$$

The H(t) can be defined for all t and independently of death-birth events. We assume that migration takes place instantaneously and independently of death-birth events. The value of H(t) changes discontinuously as migration or a death-birth event occurs, and otherwise the value of H(t) remains constant. For example, assume

that the population consists of two colonies, say 1 and 2, and that each of the two colonies has two A genes and two a genes and s = 0. Under this situation, H(t) = (2/8)(4 x 2 x 2/16 + 4 x 2 x 2/16) = 1/2. Now suppose that one A gene moves from colony 1 to colony 2, then H(t) = (2/8)(3 x 1 x 2/9 + 5 x 3 x 2/25) = 7/15. It is important that the effects of migration and of the population structure are incorporated in H(t). Note that the quantity in the square blacket is the probability that two randomly chosen gametes from a colony with replacement are different type. We call this quantity the average local genetic variation. Therefore, H(t) is equal to the average local genetic variation if the alleles are selectively neutral, and otherwise they is approximately equal.

Now ask, if $M_t$ is changed, what is the probability that $M_t$ is changed from i to i + 1? This conditional probability is equal to

$$\frac{1}{H(t)} \frac{1}{N} \sum_k N_{t,k} q_{i_k, i_k+1} .$$

Through simple algebraic manuplations, this becomes

$$\frac{1 + s}{2 + s} .$$

Similarly, the conditional probability that $M_t$ is changed from i to i - 1 is

$$\frac{1}{2 + s} .$$

We should note two properties here; (1) these conditional probabilities are independent of state i, (2) they are independent of the geographical structure of the population.

Consider a short time interval $\Delta t$. Then the probability of occurrence of one death-birth event is

$$N\Delta t + o(\Delta t) \quad .$$

Let $R_t$ be the probability that $M_t$ is unchanged for time interval t. Then

$$R_{t+\Delta t} = R_t(1 - N\Delta t) + R_t N\Delta t \left(1 - \frac{1}{\Delta t}\int_t^{t+\Delta t} H(\xi)d\xi\right) + o(\Delta t) \quad .$$

The first term on the right side of the above equation is the probability that $M_t$ remains unchanged until time t and no death-birth occurs in $(t, t + \Delta t)$. The second term is the probability that $M_t$ remains unchanged until t and one death-birth occurs in $(t, t + \Delta t)$ but $M_t$ remains unchanged. From this equation we have

$$\frac{1}{R_t}\frac{dR_t}{dt} = -NH(t) \quad ,$$

and therefore

$$R_t = e^{-\int_0^t NH(\xi)d\xi} \quad . \qquad (2\text{-}2)$$

Consider the stochastic process as a collection of sample paths $\{\omega\}$, and let

$$\tau \equiv \tau(\omega, t) \equiv \int_0^t H(\omega, \xi)d\xi \qquad (2\text{-}3)$$

where $\omega$ indicates a particular sample path and $H(\omega, \xi)$ is the same quantity as $H(\xi)$ of (2-1) associated with sample path $\omega$. Consider this $\tau$ as a new time measure. Let $M_\tau$ be the same stochastic process as $M_t$, except the time is measured by $\tau$. Let $R_\tau$ be the probability that $M_\tau$ is unchanged for time interval $\tau$. Then from (2-2) and (2-3), we have

$$R_\tau = e^{-N\tau}$$

This process $M_\tau$ is characterized as follows: (1) the state space is $[0, 1, 2, \cdots, N - 1, N]$; (2) the sojourn time at each state at each visit is exponentially distributed with mean $1/N$; (3) the conditional probability that $M_\tau$ is changed from i to i + 1 is $(1 + s)/(2 + s)$, and the conditional probability that $M_\tau$ is changed from i to i - 1 is $1/(2 + s)$. These conditional probabilities are independent of state and of geographical structure of the population. Therefore $M_\tau$ is a Poisson type process.

Let $q_{\tau,i,j}$ be the probability that $M_\tau$ is i at time 0 and it is j at time $\tau$. Then we have

$$q_{\tau-\Delta\tau,i,j} = e^{-N\Delta\tau} q_{\tau,i,j} + (1 - e^{-N\Delta\tau})[\lambda q_{\tau,i-1,j} + \mu q_{\tau,i+1,j}] + o(\Delta\tau)$$

where $\lambda = (1 + s)/(2 + s)$ and $\mu = 1/(2 + s)$. From this equation,

we derive

$$\frac{dq_{\tau,i,j}}{d\tau} = -N[-q_{\tau,i,j} + \lambda q_{\tau,i-1,j} + \mu q_{\tau,i+1,j}] \quad , \qquad (2\text{-}4)$$

Kolmogorov backward equation. Using the time measure H(t) of (2-3), equation (2-4) can be derived from Dynkin's formula, (cf. Dynkin 1965, p.133, formula 5-10). We can derive the Kolmogorov forward equation similarly:

$$\frac{dq_{\tau,i,j}}{d\tau} = N[-q_{\tau,i,j} + \lambda q_{\tau,i,j-1} + \mu q_{\tau,i,j+1}] \quad . \qquad (2\text{-}5)$$

These differential equations characterize the Markov process, and all the information concerning the process can be obtained from these equations.

It is remarkable and fortunate that the stochastic time $\tau$, that transforms the originally very complicated process into a simple random walk, happens to be the sum of heterozygosity which is a very familiar quantity in genetics. As far as I know, this is the first example of a stochastic time parameter used in biology.

It is often convenient to express the system of the equation in matrix notation. We define the following matrices:

$$Q \equiv [q_{\tau,i,j}]_{(N-1)x(N-1)}$$

$$\frac{dQ_\tau}{d\tau} \equiv \left[\frac{dq_{\tau,i,j}}{d\tau}\right]_{(N-1)x(N-1)}$$

$$A \equiv \begin{bmatrix} -1 & \mu & 0 & 0 & \cdots & 0 & 0 & 0 \\ \lambda & -1 & \mu & 0 & \cdots & 0 & 0 & 0 \\ 0 & \lambda & -1 & \mu & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda & -1 & \mu \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda & -1 \end{bmatrix}_{(N-1)x(N-1)}$$

where $\lambda = (1 + s)/(2 + s)$ and $\mu = 1/(2 + s)$. Then the systems of the differential equations in (2-4) become

$$\frac{dQ_\tau}{d\tau} = - NAQ_\tau \tag{2-7}$$

and (2-5) becomes

$$\frac{dQ}{d\tau} = NA^*Q \tag{2-8}$$

where $A^*$ is the transpose of A.

We shall next derive a diffusion process as a limit of large N. The backward equations (2-4) can be rewritten as

$$-\frac{dq_{\tau,i,j}}{d\tau} = \frac{1}{2} N[q_{\tau,i-1,j} - 2q_{\tau,i,j} + q_{\tau,i+1,j}]$$

$$+ \frac{Ns}{2(2 + s)} [q_{\tau,i-1,j} - q_{\tau,i+1,j}] \quad . \tag{2-9}$$

Let T be a new time measure such that $NT = \tau$, and $Ns = S$

a constant. Let $q_N(T, \frac{i}{N}, \frac{j}{N}) \equiv q_{T,i,j}$. Then (2-9) can be written as

$$-\frac{dq_N(T, \frac{i}{N}, \frac{j}{N})}{dT} = \frac{1}{2}\frac{1}{(\frac{1}{N})^2}\left[q_N(T, \frac{i-1}{N}, \frac{j}{N}) - 2q_N(T, \frac{i}{N}, \frac{j}{N}) + q_N(T, \frac{i+1}{N}, \frac{j}{N})\right] + \frac{Ns}{(2+s)}\frac{1}{(\frac{2}{N})}\left[q_N(T, \frac{i-1}{N}, \frac{j}{N}) - q_N(T, \frac{i+1}{N}, \frac{j}{N})\right] .$$

Therefore as $N \to \infty$, the above difference equation formally converges to

$$-\frac{\partial q_\infty(T, x, y)}{\partial T} = \frac{1}{2}\frac{\partial^2 q_\infty(T, x, y)}{\partial x^2} + \frac{S}{2}\frac{\partial q_\infty(T, x, y)}{\partial x} . \qquad (2\text{-}10)$$

The derivation of this equation is not mathematically rigorous, but this can be justified, (cf. Kac 1947). The Markov process governed by (2-10) is a diffusion process and it is a Brownian motion. If we let $q'_\infty = \exp(-Sx/2)q_\infty(T, x, y)$, the above equation becomes

$$-\frac{\partial q'_\infty(T, x, y)}{\partial T} = \frac{1}{2}\frac{\partial^2 q'_\infty(T, x, y)}{\partial x^2} .$$

I believe that this is the first time in population genetics that a diffusion process is rigorously derived for a selective case, regardless the population structure. Karlin and McGregor (1962) has derived several diffusion processes for a non-selective case

of a random mating population.

3. Applications

We shall investigate biologically interesting quantities. A standard procedure of dealing with equation (2-7) is to obtain a complete spectral analysis of matrix A. The right eigen vectors of matrix A are the column vectors

$$e_k = \sqrt{\frac{2}{N}}\left( \left(\frac{\lambda}{\mu}\right)^{\frac{1}{2}} \sin\frac{\pi k}{N} , \left(\frac{\lambda}{\mu}\right)^{\frac{2}{2}} \sin\frac{2\pi k}{N} , \left(\frac{\lambda}{\mu}\right)^{\frac{3}{2}} \sin\frac{3\pi k}{N} , \cdots, \left(\frac{\lambda}{\mu}\right)^{\frac{N-1}{N}} \sin\frac{(N-1)\pi k}{N}\right)^{*} , \quad k = 1, 2, \cdots, N-1. \tag{3-1}$$

where * indicates transpose of a vector. It can be easily shown that

$$Ae_k = \lambda_k e_k$$

where

$$\lambda_k = -1 + 2\sqrt{\lambda\mu}\cos\frac{\pi k}{N} \tag{3-2}$$

The left eigenvectors of A are

$$d_k = \sqrt{\frac{2}{N}}\left( \left(\frac{\mu}{\lambda}\right)^{\frac{1}{2}} \sin\frac{\pi k}{N} , \left(\frac{\mu}{\lambda}\right)^{\frac{2}{2}} \sin\frac{2\pi k}{N} , \left(\frac{\mu}{\lambda}\right)^{\frac{3}{2}} \sin\frac{3\pi k}{N} , \cdots, \left(\frac{\mu}{\lambda}\right)^{\frac{N-1}{N}} \sin\frac{(N-1)\pi k}{N}\right) , \quad k = 1, 2, \cdots, N-1.$$

and it is easy to show

$$d_k A = \lambda_k d_k$$

where the eigenvalues are the same as (3-2). Note that $\{e_k\}$, or $\{d_k\}$, themselves are not orthogonal system, but $\{e_k\}$ and $\{d_k\}$ are biorthogonal, i. e.,

$$(e_k, d_\ell) = \frac{2}{N} \sum_{i=1}^{N-1} \left(\frac{\lambda}{\mu}\right)^i \sin \frac{i\pi k}{N} \left(\frac{\mu}{\lambda}\right)^i \sin \frac{i\pi \ell}{N} = \delta_{k\ell} \quad .$$

where $\delta_{kk} = 1$ and $\delta_{k\ell} = 0$ if $k \neq \ell$.

With these eigenvectors and eigenvalues, we can easily obtain the fundamental solution of the process of equation (2-7), that is the transition probability:

$$Q_\tau = \frac{2}{N} \sum_{k=1}^{N-1} e^{-\lambda_k \tau} e_k \otimes d_k$$

where $\otimes$ indicates the direct product of two vectors. This can be written as

$$q_{\tau,i,j} = \frac{2}{N} \sum_{k=1}^{N-1} \exp\left\{ - N(1 - 2\sqrt{\lambda\mu} \cos \frac{\pi k}{N} )\tau \right\} \left(\frac{\tau}{\mu}\right)^{\frac{i-j}{2}} \sin \frac{i\pi k}{N} \sin \frac{j\pi k}{N} \quad .$$

$$q_\infty(T, x, y) = 2e^{\frac{Sx}{2}} \sum_{k=1}^{\infty} e^{- \frac{1}{2}(k^2\pi^2 - \frac{S^2}{4})T} \sin k\pi x \sin k\pi y \quad .$$

The Markov process governed by $q_\infty(T, x, y)$ is independent of the geographical structure and it is a Brownian motion.

The ultimate fixation probability, $u_i$, of allele A in the whole population is given by the solution of equation

$$Au = 0$$

where $u = (u_1, u_2, \cdots, u_{N-1})^*$:

$$u_i = \frac{1 - \left(\frac{1}{1+s}\right)^i}{1 - \left(\frac{1}{1+s}\right)^N} \approx \frac{1 - e^{-si}}{1 - e^{-sN}} \qquad (3\text{-}3)$$

where i is the initial number of A in the whole population. This is the same as Moran (1959) in which the formula was derived for a panmictic population. Therefore the fixation probability is independent of the geographical structure. If we replace si and sN of (3-1) by Sx and S respectively, the right side in (3-3) is the solution of the differential equation

$$\frac{1}{2}\frac{d^2f}{dx^2} + \frac{S}{2}\frac{df}{dx} = 0$$

with boundary conditions $f(0) = 0$ and $f(1) = 1$. The independence of the fixation probability here was derived under the assumption of constant size of the total population and local random mating, but importantly the structure of the population was not assumed to be fixed. Even if population structure depends on the genetic constitution, such as genetically similar individuals tend to live closer, or the structure depends on the gene frequency, the fixation probability is unaltered. This independence of the fixation probability was suggested by Maruyama (1970a), based on

an approximation. However it was not clear whether the independence is exact or approximation.

The first exit time of the process from $[1, 2, \cdots, N-1]$ is the solution of equation

$$Af + 1 = 0$$

where $1 = (1, 1, \cdots, 1)^*$ and $0 = (0, 0, \cdots, 0)^*$. The i-th entry, $f_i$, of f is the duration of time, measured by $\tau$, until allele A is fixed in the population or A is lost from the population:

$$f_i = \frac{1}{N^2} \sum_{j=1}^{N-1} \sum_{k=1}^{N-1} \frac{\left(\frac{\lambda}{\mu}\right)^{\frac{i-j}{2}} \sin \frac{i\pi k}{N} \sin \frac{j\pi k}{N}}{\left(1 - 2\sqrt{\lambda\mu} \cos \frac{\pi k}{N}\right)}$$

The biological meaning of $f_i$ is the sum of the number of heterozygote that appears in the population while the two alleles are segregating, provided that initially there are i A genes.

The second moment, $f_i^{(2)}$, of the first exit time measured by $\tau$ is the solution of

$$Af^{(2)} + 2f = 0$$

where f is the mean obtained above and $f^{(2)} = (f_1^{(2)}, f_2^{(2)}, \cdots, f_{N-1}^{(2)})$. This is the second moment of the sum of the number of heterozygotes. The higher moments are similarly obtained:

$$Af^{(n)} + nf^{(n-1)} = 0 \quad .$$

In the above calculations, we included both kinds of sample paths in which A is established in the whole population and in which A is lost. We can make distinction between the two kinds, and calculate the quantities for those paths in which A is fixed, (or in which A is lost). Let $f_i'$ be the mean exit time in the paths in which A is fixed. Then $g \equiv (u_1 f_1', u_2 f_2', \cdots, u_{N-1} f_{N-1}')^*$, is the solution of equation

$$Ag + u = 0$$

where u is the vector of the fixation probabilities, $u_i$:

$$f_i' = \frac{1}{u_i N^2} \sum_{j=1}^{N-1} \sum_{k=1}^{N-1} \frac{\left(\frac{\lambda}{\mu}\right)^{\frac{i-j}{2}} u_j \sin\frac{i\pi k}{N} \sin\frac{j\pi k}{N}}{(1 - 2\sqrt{\lambda\mu}\cos\frac{\pi k}{N})}$$

where $u_i$ is the fixation probability given in (3-3). Biologically, $f_i'$ is the expectation of the amount of heterozygosity given that A is established in the population. A special case of this result was first given in Maruyama (1971a).

Finally we shall obtain the sojourn time at a given state, or in a given range of gene frequency. Let $\phi_{ij}$ be the sojourn time, measured by $\tau$, at state j while A and a are segregating in the population, provided the process starts from state i. Let $\Phi = [\phi_{ij}]$. Then

$$\Phi = -A^{-1},$$

or

$$\phi_{ij} = \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{\left(\frac{\lambda}{\mu}\right)^{\frac{i-j}{2}} \sin\frac{i\pi k}{N} \sin\frac{j\pi k}{N}}{\left(1 - 2\sqrt{\lambda\mu}\cos\frac{\pi k}{N}\right)}$$

As N becomes large,

$$\left.\begin{aligned} \phi_{ij} &\approx \frac{(1 - e^{-2s(N-j)})(1 - e^{-2si})}{s(1 - e^{-2Ns})} \qquad \text{if } j > i \\ \phi_{ij} &\approx \frac{1}{s}\left\{\frac{(1 - e^{-2s(N-j)})(1 - e^{-2si})}{(1 - e^{-2Ns})} - (1 - e^{-2s(i-j)})\right\} \\ &\qquad\qquad \text{if } j < i \end{aligned}\right\} \quad (3\text{-}4)$$

If $s = 0$

$$\left.\begin{aligned} \phi_{ij} &\approx \frac{2i(N - j)}{N} \qquad \text{if} \quad j > i \\ \phi_{ij} &\approx \frac{2j(N - i)}{N} \qquad \text{if} \quad j < i \end{aligned}\right\} \quad (3\text{-}5)$$

If we let $\phi(x, y) = \phi\left(\frac{i}{N}, \frac{j}{N}\right) = \phi_{ij}$, this satisfy the following differential equation

$$\frac{1}{2}\frac{d^2\phi(x, y)}{dx^2} + \frac{S}{2}\frac{d^2\phi(x, y)}{dx} + \delta(x - y) = 0$$

where $\delta(\cdot)$ is the Dirac delta function, (cf. Maruyama 1972a).

Approximation formulas (3-4) and (3-5) were obtained in Maruyama (1972b) by an entirely different method, and were used in Yamazaki and Maruyama (1972) in which an attempt is made to answer the important question as to whether or not the naturally

occurring polymorphisms are mainly due to selectively neutral mutants. See also Crow (1973). These formulas were also used in Maruyama (1972c) in which the consistency between the neutral hypotheses of molecular evolution and of naturally occurring polymorphism was shown to hold. It should be emphasized that the sojourn time measured by the number of generation, the time until fixation of an allele, the distribution of the gene frequencies and the rate of decay of genetic variability are influenced by the population structure.

If the population is at steady situation in which input of mutants at some loci (or nucleotide sites) is balanced by the extinction of genetic variation at other loci, the sum of heterozygosities at loci having a specified gene frequency, say j/N, is proportional to $\phi_{1j}$. Therefore the distribution of heterozygosities is independent of the geographical structure. This is an extension of Wright (1938) in which the distribution of gene frequencies was obtained. It is interesting to note that the quantity independent of the geographical structure is not the gene frequency distribution but it is the distribution of heterozygosities. Extensive investigations on these matters were given in Maruyama (1969, 1970b, 1970c, 1971b, 1971c, 1971d, 1972d, 1972e, 1972f, 1972g).

## References

Cavalli-Sforza, L. L., and W. F. Bodmer, (1971), The Genetics of Human Population. Freeman, San Francisco.

Crow, J. F., (1973), The dilemma of nearly neutral mutations:

How important are they for evolution and human welfare? Jour. of Heredity. (in press).

Crow, J. F., and M. Kimura, (1970), An Introduction to Population Genetics Theory. Harper and Row, New York.

Dynkin, E. B., (1965), Markov Processes vol. 1. Academic Press, New York.

Ewens, W. J., (1969), Population Genetics. Methuen, London.

Fisher, R. A., (1930), The genetical theory of Natural Selection. Clarendon Press, Oxford. Rev. Ed., 1958, New York, Dover Press.

Kac, M., (1947), Random walk and the theory of Brownian motion, Am. Math. Monthly 54, 369-391.

Karlin, S., and J. McGregor, (1962), On a genetic model of Moran. Proc. Camb. Philos. Soc. 58, 299-311.

Karlin, S., and J. McGregor, (1964), On some stochastic models in genetics. Stochastic Models in Medicine and Biology, edited by J. Gurland. The Univ. of Wisconsin Press.

Kimura, M., (1964), Diffusion models in population genetics. Journal of Applied Probability, 1, 177-232.

Kimura, M., and T. Ohta, (1971), Theoretical Aspects of Population Genetics. Princeton University Press, Princeton, New Jersey.

Maruyama, T., (1969), Genetic correlation in the stepping stone model with non-symmetrical migration rates. J. Appl. Prob. 6, 463-477.

Maruyama, T., (1970a), On the fixation probability of mutant genes in a subdivided population. Genet. Res. Camb. 15, 221-225.

Maruyama, T., (1970b), Stepping stone models of finite length. Advances in Applied Probability 2, 229-258.

Maruyama, T., (1970c), Analysis of population structure I. One-

dimensional stepping-stone models of finite length. Ann. Human Genet. Lond. 34, 201-219.

Maruyama, T., (1971a), An invariant property of a structured population. Genet. Res. Camb. 18, 81-84.

Maruyama, T., (1971b), The rate of decrease of heterozygosity in a population occupying a circular or a linear habitat. Genetics 67, 437-454.

Maruyama, T., (1971c), Speed of gene substitution in a geographically structured population. American Naturalist 105, 253-265.

Maruyama, T., (1971d), Analysis of population structure II. Two-dimensional stepping stone models of finite length and other geographically structured populations. Ann. Human Genet. Lond. 35, 179-196.

Maruyama, T., (1972a), The average number and the variance of generation at particular gene frequency in the course of fixation of a mutant gene in a finite population. Genet. Res. Camb. 19, 109-114.

Maruyama, T., (1972b), Some invariant properties of a geographically structured finite population: distribution of heterozygotes under irreversible mutation. Genet. Res. Camb. 20, 141-149.

Maruyama, T., (1972c), A note on the hypothesis: Protein polymorphism as a phase of molecular evolution. J. Molecular Evolution 1, 368-370.

Maruyama, T., (1972d), Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. Genetics 70, 639-651.

Maruyama, T., (1972e), Distribution of gene frequencies in a

geographically structured finite population I. Distribution of neutral genes and of genes with small effect. Ann. Human Genet. Lond. 35, 411-423.

Maruyama, T., (1972f), Distribution of gene frequencies in a geographically structured population II. Distribution of deleterious genes and of lethal genes. Ann. Human Genet. Lond. 35, 425-432.

Maruyama, T., (1972g), Distribution of gene frequencies in a geographically structured population III. Distribution of deleterious genes and genetic correlation between different localities. Ann. Human Genet. Lond. 36, 99-108.

Moran, P. A. P., (1958), Random processes in genetics. Proc. Camb. Phil. Soc. 54, 60-71.

Moran, P. A. P., (1959), The survival of a mutant gene under selection. Jour. Aust. Math. Soc. 1, 121-126.

Moran, P. A. P., (1962), The Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

Wright, S., (1931), Evolution in Mendelian populations. Genetics 16, 97-159.

Wright, S., (1938), The distribution of gene frequencies under irreversible mutation. Proc. Nat. Acad. Sci. 24, 253-259.

Wright, S., (1969), Evolution and the Genetics of Populations. Vol. 2. The Theory of Gene Frequencies. University of Chicago Press, Chicago.

Yamazaki, T., and T. Maruyama, (1972), Evidence for the neutral hypothesis of protein polymorphism. Science 178, 56-58.