*Genet. Res., Camb.* (1981), **37**, *pp.* 291–302
*With 3 text-figures*
*Printed in Great Britain*



# A mathematical model of extranuclear genes and the genetic variability maintained in a finite population*

BY NAOYUKI TAKAHATA AND TAKEO MARUYAMA
*National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan*

(*Received* 23 *July* 1980 *and in revised form* 1 *December* 1980)

SUMMARY

A mathematical theory of population genetics accounting for the genes transmitted through mitochondria or chloroplasts has been studied. In the model it is assumed that a population consists of $N_m$ males and $N_f$ females, the genetic contribution from a male is $\beta$ and that from a female $1-\beta$, and each cell line contains $n$ effective copies of a gene in its cytoplasm. Assuming selective neutrality and an infinite alleles model, it is shown that the sum ($H$) of squares of allelic frequencies within an individual and the corresponding sum ($Q$) for the entire population are, at equilibrium, given by

$$\hat{H} \simeq \frac{1}{\left\{1+\rho\,\frac{(1-1/n)^{\lambda}}{1-(1-1/n)^{\lambda}}\,\frac{2N_e v\lambda}{1+2N_e v\lambda}\right\}(1+2nv)}$$

and

$$\hat{Q} \simeq \hat{H}/(1+2N_e v\lambda)$$

where $\rho = 2\beta(1-\beta)$, $N_e = \{\beta^2/N_m+(1-\beta)^2/N_f\}^{-1}$, $\lambda$ is the number of somatic cell divisions in one generation, and $v$ is the mutation rate per cell division. If the genes are transmitted entirely through the female the formulae reduce to $\hat{H} \simeq 1/(1+2nv)$ and $\hat{Q} \simeq 1/\{1+(2N_e\lambda+2n)v\}$. Non-equilibrium behaviours of $H$ and $\hat{Q}$ are also studied in the case of a panmictic population. These results are extended to geographically structured models, and applied to existing experimental data.

## 1. INTRODUCTION

Transmission genetics of mitochondria and chloroplasts in yeast has been extensively studied by Birky and his associates, (Birky, 1975, 1978; Birky & Skavaril, 1976 and Birky *et al.* 1978). Using the molecular method of restriction enzymes, several studies have been carried out recently to investigate the evolutionary rate and the genetic variability maintained in the populations of extranuclear genes of man, rodent species and others (Brown, George & Wilson, 1979;

* Contribution no. 1346 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan.

Avise *et al.* 1979*a*, *b*; Hayashi *et al.* 1979 and others). Among characteristic features of genetical interest, Brown *et al.* have reported that the evolutionary rate of mitochondrial genomes is considerably faster than that of ordinary Mendelian genes. On the other hand, according to Avise *et al.*, phylogenetic relationships among different mitochondrial genomes can be identified and geographically closer individuals tend to share genomes of closer relationship. Since these studies are topical and important in population genetics and molecular evolution, we can expect that in the near future a large amount of data of the similar nature will be reported on various organisms. Therefore we believe it is quite appropriate to formulate mathematical models for such genes contained in extranuclear genomes. In this paper, we attempt to advance population genetics theory for genes existing in indefinite multiple copies in each individual and transmitted mainly through the cytoplasm.

## 2. MODEL

(1) *Mode of inheritance.* We assume that the genes in question reside in the cytoplasm and the number of copies existing in a cell is not necessarily constant, but can vary in time without becoming extinct or too numerous. Prior to a cell division the number of genes is approximately doubled, and reduced to half after the division. The genes are transmitted through the gametes, which contribute different proportions depending on the sex. We denote the average proportion of the male's contribution by $\beta$ and that of the female by $1-\beta$. Therefore, if $\beta = 0$ the genes are transmitted entirely through eggs, and if $\beta = \frac{1}{2}$ both sexes contribute equally. In one generation, every cell in the germ line goes through a certain number of somatic cell divisions, at each of which the total number of genes is doubled and then reduced by random partition. We denote by $\lambda$ the number of cell divisions. As in the infinite neutral allele model of Kimura & Crow (1964), we assume that every mutant is different from the pre-existing ones in the population, and that the mutation rate for every cell division is $v$. We denote by $n$ the harmonic mean of the number of existing gene copies through cell generations. For instance, if the numbers of genes in cell generations 1, 2, ..., $t$ are $n_1$, $n_2$, ..., $n_t$ then $n = t \Big/ \left\{ \sum_{i=1}^{t} \frac{1}{n_i} \right\}$. Another potential factor which can make the actual and effective numbers different is the variation of replicating copies among the genes (see Wright, 1931; Kimura & Crow, 1963). One generation cycle of cytoplasmic genes modelled here is illustrated in Fig. 1.

(2) *Population structure.* We will deal with two situations. A simple one is of course a random mating population consisting of $N_f$ females and $N_m$ males. The other is a one- or two-dimensional stepping-stone structure in which the population consists of circularly arranged $L$ separate colonies or $L_1 \times L_2$ colonies on a surface of a torus-like space. We assume that all the colonies in the population have equal size and that each colony receives immigrants from the two (one-dimensional case) or four (two-dimensional case) neighbouring colonies at a rate $m$ per generation.

### 3. ANALYSES

(1) *Random mating populations.* Let $H$ be the probability that two randomly chosen genes from a single individual are identical by descent, and $Q$ be the analogous probability for two genes taken from separate individuals. To study the stationary values and transient behaviour for $H$ and $Q$, we derive equations de-

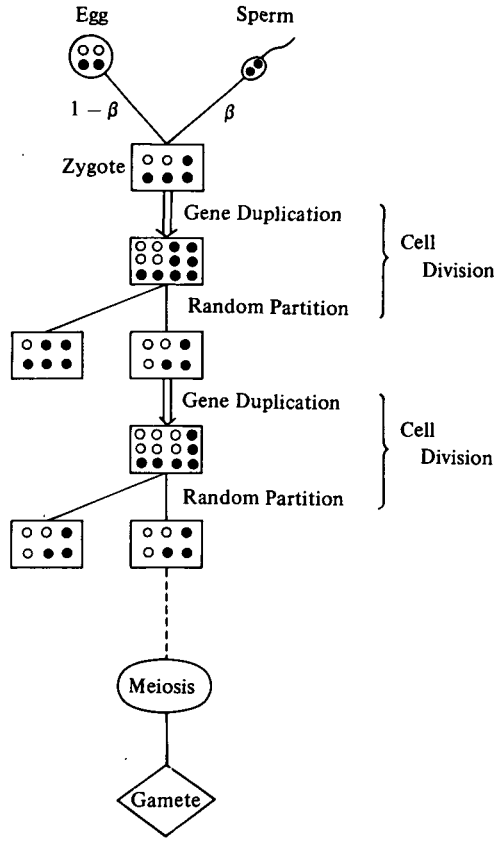


Fig. 1. Diagram illustrating one generation cycle of cytoplasmic genes, assuming $n = 6$, $\beta = \frac{1}{3}$, and every gene duplicates exactly once at each cell division. Symbols ○ and ● indicate two allelic forms of a gene.

scribing the change in these quantities. We will first see the change due to random sampling of gametes and formation of zygotes. If we denote by $H'$ and $Q'$ the quantities after the sampling and fertilization, we have

$$H' = \{1-2\beta(1-\beta)\}H + 2\beta(1-\beta)Q \tag{1}$$

and

$$Q' = \frac{1}{N_e}H + \left(1-\frac{1}{N_e}\right)Q, \tag{2}$$

https://doi.org/10.1017/S0016672300020292 Published online by Cambridge University Press

in which $1/N_e = \beta^2/N_m + (1-\beta)^2/N_f$. The first term in the right-hand side of (1) is the probability that two genes taken from an individual are derived from a single individual in the previous generation, and the second term is the probability that they are derived from two separate individuals. Similarly, the first term in the right-hand side of (2) is the probability that two genes in separate individuals are from a single individual and the second term is the probability that they are derived from two separate individuals of the previous generation.

After the fertilization, every cell goes through somatic cell division $\lambda$ times, at each of which changes in $H$ and $Q$ are as follows:

$$\left.\begin{aligned} H' &= (1-v)^2\left\{\frac{1}{n} + \left(1-\frac{1}{n}\right)H\right\}, \\ Q' &= (1-v)^2 Q, \end{aligned}\right\} \tag{3}$$

where $v$ is the mutation rate per cell division and $n$ is the inbreeding effective number of genes. If the cell division is repeated $\lambda$ times and we denote the resulting quantities by $H'$ and $Q'$, we have

$$H' = (1-v)^{2\lambda}\left(1-\frac{1}{n}\right)^{\lambda} H + b \tag{4}$$

and

$$Q' = (1-v)^{2\lambda} Q, \tag{5}$$

in which

$$b = (1-v)^2\left\{1-(1-v)^{2\lambda}\left(1-\frac{1}{n}\right)^{\lambda}\right\}\Big/\left[n\left\{1-(1-v)^2\left(1-\frac{1}{n}\right)\right\}\right]. \tag{6}$$

Combining these equations (1), (2), (4) and (5), we have the following equations describing the change in one complete generation,

$$\mathbf{c}' = M\mathbf{c} + \mathbf{b}, \tag{7}$$

in which $\mathbf{c}$ is the column vector $(H, Q)$, $M$ is the matrix

$$(1-v)^{2\lambda}\begin{pmatrix} (1-\rho)\left(1-\frac{1}{n}\right)^{\lambda}, & \rho\left(1-\frac{1}{n}\right)^{\lambda} \\ \frac{1}{N_e}, & 1-\frac{1}{N_e} \end{pmatrix} \tag{8}$$

and $\mathbf{b}$ is the column vector $(b, 0)$ where $b$ is given in (6).

The equilibrium values of $\mathbf{c}$ can be obtained by replacing $\mathbf{c}'$ by $\mathbf{c}$ in (7) and solving the resulting equation. Denoting these equilibrium values by $\hat{H}$ and $\hat{Q}$, we have

$$\hat{H} = \frac{1}{\left[1+\rho\,\dfrac{w(1-w)\,aN_e}{\{w+(1-w)N_e\}(1-wa)}\right]\left[n\left\{(1-v)^{-2}-\left(1-\frac{1}{n}\right)\right\}\right]} \tag{9}$$

$$\hat{Q} = \frac{1}{\left[1+\left(\frac{1}{w}-1\right)N_e\right]\left[1+\rho\frac{w(1-w)\,aN_e}{\{w+(1-w)N_e\}\,(1-wa)}\right]\left[n\left\{(1-v)^{-2}-\left(1-\frac{1}{n}\right)\right\}\right]}, \tag{10}$$

where $w = (1-v)^{2\lambda}$ and $a = \left(1-\frac{1}{n}\right)^{\lambda}$. It is important to note that if $\rho \ll 1-\left(1-\frac{1}{n}\right)^{\lambda}$ or equivalently if $\beta \ll \frac{\lambda}{n}$, the genetic contribution from males can be ignored without altering the results significantly. When this is the case, the main

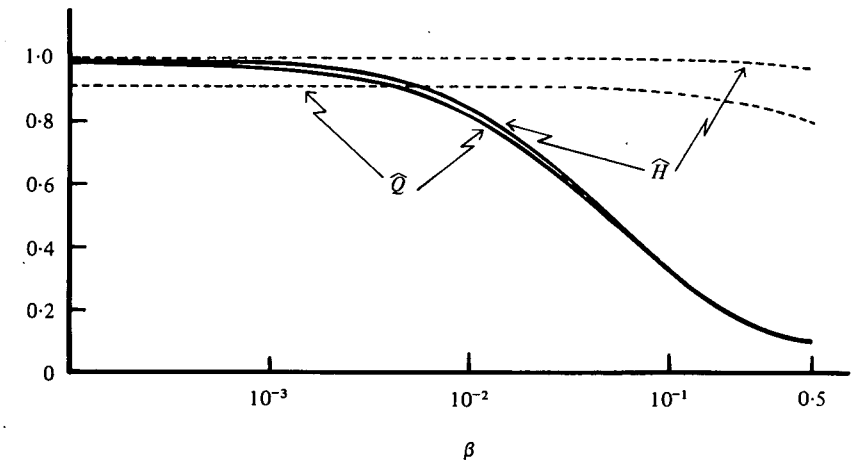


Fig. 2. Genetic identities ($\hat{H}$ and $\hat{Q}$) as functions of $\beta$, for random mating populations, where $N_f = N_m = N/2$. $N = 10^4$, $v = 10^6$; ---, $\lambda = 10$, $n = 10$; —, $\lambda = 1$, $n = 10^3$.

feature of the genetic polymorphisms will be determined by the balance between mutation and the ratio of the number of cell divisions in a generation to the number of genes. It might be worth mentioning, however, that even if the value of $\beta$ alone is small, the contributions of male gametes cannot always be ignored. In fact, if $\beta$ is small, but $\lambda$ is also small and $n$ is large, the male sex can play a significant role in determining the amount of genetic variability. The relationships among these parameters are shown in Fig. 2.

It is interesting to note that either from the recurrence equation (7) or from the explicit formulae (9) and (10), we can get

$$\frac{\hat{Q}}{\hat{H}} = \frac{1}{1+\{(1-v)^{-2\lambda}-1\}N_e} \simeq \frac{1}{1+2N_e v\lambda}.$$

This is a simple formula giving the relationship between $\hat{H}$ and $\hat{Q}$, and interestingly it is independent of the values of $n$, $\hat{H}$ and $\hat{Q}$. Analogous formulae have been worked out for the multigene family model by Ohta (1978), and also for a geo-

graphically structured population model of Mendelian genes by Maruyama & Kimura (1980).

The equilibria given in (9) and (10) are values just before each meiotic division. Modifying (7) slightly, however, we can obtain the equilibrium values of $H$ and $Q$ immediately after fertilization, which can be significantly different, particularly when $\lambda$ is large or $n$ is small. They are

$$\hat{H}^1 \simeq \left(1-\rho\,\frac{2N_e v\lambda}{1+2N_e v\lambda}\right)\hat{H}$$

and

$$\hat{Q}^1 \simeq (1+2v\lambda)\,\hat{Q} \simeq \hat{Q},$$

where the superscript 1 refers to the values immediately after fertilization. The differences between these values at the two different times will be small if values of $\rho$ and $2N_e v\lambda$ are small. However, if $\rho$ and $2N_e v\lambda$ are large, the difference of $\hat{H}^1$ and $\hat{H}$ can be large, while $\hat{Q}^1$ and $\hat{Q}$ are always approximately the same. This is intuitively clear; if $\rho$ and $2N_e v\lambda$ are large the entire population will be highly polymorphic (small $Q$) and the individual level ($H$) of the homozygosity will also be low at the time of fertilization, but with a large number of somatic divisions the value of $H$ will be much increased. Therefore under this circumstance we have $\hat{H}^1 \ll \hat{H}$.

So far we have dealt with the stationary values of $H$ and $Q$. It is also important to see the rate at which the equilibria are attained when these values are perturbed. To do this we need to change the recurrence equation (7) into a linear system. Let $\boldsymbol{\Delta}$ be the vector consisting of the differences of $H$ from its equilibrium value $\hat{H}$ and of $Q$ from $\hat{Q}$, i.e. $\boldsymbol{\Delta}$ is the column vector $(H-\hat{H},\ Q-\hat{Q})$. Through simple algebra, we have then

$$\boldsymbol{\Delta}' = M\boldsymbol{\Delta}, \tag{11}$$

where $\boldsymbol{\Delta}'$ refers to the difference vector one generation later and $M$ the matrix given in (8). Now the difference vector $\boldsymbol{\Delta}$ decreases to zero and its rate is related to the eigenvalues of the $M$ matrix. Since the recurrence relationship is given for discrete time measurement, the actual rate of approach is given by one minus the eigenvalues, and they are

$$1-\lambda_1 \simeq 2v\lambda + \frac{1-\left(1-\frac{1}{n}\right)^{\lambda}}{1-(1-\rho)\left(1-\frac{1}{n}\right)^{\lambda}}\,\frac{1}{N_e} \tag{12}$$

and

$$1-\lambda_2 \simeq 1-\{1-(\rho+2v\lambda)\}\left(1-\frac{1}{n}\right)^{\lambda},$$

where $\lambda_i$s refer to the eigenvalues of $M$. Obviously $1-\lambda_1 < 1-\lambda_2$, and thus asymptotically the difference vector $\boldsymbol{\Delta}$ approaches zero at a rate given by (12). It is interesting to note that when $\rho$ is small, the eigenvector associated with $\lambda_1$ co-

indices nearly with the $Q$ axis in a two-dimensional $H$ and $Q$ plane. This implies that when $\rho$ is small, perturbed values of $H$ and $Q$ return quickly to equilibrium in $H$, and then slowly to $\hat{Q}$. On the other hand, if $\rho$ is significantly large, both $H$ and $Q$ approach their stationary values at a rate given by (12). Equation (11) can be explicitly solved and the differences at generation $t$ are given by

$$H_t - \hat{H} = \frac{1}{\lambda_1 - \lambda_2}\left[\{(a_{11} - \lambda_2)\lambda_1^t - (a_{11} - \lambda_1)\lambda_2^t\} \times (H_0 - \hat{H}) + a_{12}(\lambda_1^t - \lambda_2^t)(Q_0 - \hat{Q})\right]$$

and

$$Q_t - \hat{Q} = \frac{1}{\lambda_1 - \lambda_2}\left[\{(a_{22} - \lambda_2)\lambda_1^t - (a_{22} - \lambda_1)\lambda_2^t\} \times (Q_0 - \hat{Q}) + a_{21}(\lambda_1^t - \lambda_2^t)(H_0 - \hat{H})\right],$$

in which the subscripts in $Q$ and $H$ indicate the time measured in generations, and

$$\left.\begin{aligned} a_{11} &= (1-v)^{2\lambda}(1-\rho)\left(1-\frac{1}{n}\right)^{\lambda}, \quad a_{12} = (1-v)^{2\lambda}\rho\left(1-\frac{1}{n}\right)^{\lambda}, \\ a_{21} &= (1-v)^{2\lambda}\frac{1}{N_e} \quad \text{and} \quad a_{22} = (1-v)^{2\lambda}\left(1-\frac{1}{N_e}\right). \end{aligned}\right\} \tag{13}$$

We should like to note that the theory presented for the random mating population is mathematically analogous to the models of multigene families studied by Ohta (1978, 1980). The fusion of female and male gametes corresponds to the crossing-over between different chromosomes in the multigene family model. The sampling and distribution of genes at each somatic cell division correspond to the intrachromosomal crossing-over. The present model also has a mathematical analogy with a geographical population model for Mendelian genes, in which extinction and replacement of local subpopulations are incorporated (Maruyama & Kimura, 1980). The extinction of a local subpopulation is followed by a replacement in the geographical model. Succession of local populations without their becoming extinct corresponds to the somatic cell divisions.

(2) *Geographically structured populations*. Some species are quite sessile or move very little geographically. For such species, geographically closer individuals are more closely related genetically, and therefore the population structure becomes an important factor in determining the genetic variability. In this section we shall attempt to cope with this problem using the stepping-stone model (Kimura & Weiss, 1964). The mathematics used here is very close to that used in Maruyama (1970), and therefore we shall present key equations and the final formulae.

Following Maruyama (1970), let the population occupy a circular habitat with $L$ colonies and each colony have effective size $N_e$. Let us assume that short-range migration occurs only between adjacent colonies with the rate $m$, i.e. the $i$th colony receives migration from the $i+1$th colony at the rate $m/2$ and it also receives migration from the $i-1$th colony at the same rate. This is a special case of geographically structured populations studied, and for details the reader may refer to Kimura & Weiss (1964) and Maruyama (1970).

Let $Q_i$ ($i = 0, 1, 2, \ldots, L-1$) be the probability that two homologous genes taken

randomly from two individuals belonging to populations $i$ steps apart are identical by descent. And let $H$ be the probability that two homologous genes sampled from a single individual are identical by descent. Then the total changes in $H$ and $Q_0$ due to the random sampling of gametes, mutation and cell division are given by the same formula as (7) of the panmictic case. Now the changes in $Q_i$ for $i \geqslant 1$ during somatic cell division are given by

$$Q_i' = (1-v)^{2\lambda} Q_i. \tag{14}$$

And by migration, $Q_i$ for $i \geqslant 0$ are further altered,

$$Q_i' = \left(\frac{m}{2}\right)^2 Q_{i-2} + m(1-m)\, Q_{i-1} + \left\{(1-m)^2 + \frac{m^2}{2}\right\} Q_i + m(1-m)\, Q_{i+1} + \left(\frac{m}{2}\right)^2 Q_{i+2}, \tag{15}$$

whereas $H$ remains unchanged. Therefore the time-dependent solutions of $H$ and $Q_i'$s can be determined by these recurrence relations. In particular, the equilibrium values can be explicitly obtained by substituting $H'$ for $H$ and $Q_i'$ for $Q_i$:

$$\hat{H} = (1-v)^{2\lambda}\left(1-\frac{1}{n}\right)^{\lambda}\{(1-\rho)\hat{H} + \rho\hat{Q}_0\} + \frac{(1-v)^2\left\{1-(1-v)^{2\lambda}\left(1-\frac{1}{n}\right)^{\lambda}\right\}}{n\left\{1-(1-v)^2\left(1-\frac{1}{n}\right)\right\}} \tag{16}$$

and

$$\hat{Q}_i = \frac{g}{L} \sum_{k=0}^{[L/2]} \frac{\xi_k}{\Delta_k\{1-(1-v)^{2\lambda}\xi_k\}} \cos\frac{2\pi k i}{L}, \tag{17}$$

where $\Delta_k = 1$ if $k = 0$ or $k = L/2$ and $\Delta_k = \frac{1}{2}$ otherwise, $\xi_k = 0$ and $\xi_k = \{1-m(1-\cos[2\pi k/L])\}^2$ for $k > 0$, $g = (1-v)^{2\lambda}(\hat{H}/N_e - \hat{Q}_0/N_e)$. In the above formula, $\hat{H}$ and $\hat{Q}_0$ are implicit, but they can be determined explicitly. Let $a_{11}$, $a_{12}$, $a_{21}$ be the same as given in (13), and let $a_{22} = -a_{21}$. Then the explicit formulae for $\hat{H}$ and $\hat{Q}_0$ are

$$\hat{H} = \frac{b(1-w_1 a_{22})}{1-(a_{11}+w_1 a_{22})+w_1|A|}$$

and

$$\hat{Q}_0 = \frac{w_1 a_{21}}{1-w_1 a_{22}}\hat{H},$$

where $b$ is the non-zero element of $\mathbf{b}$ given in (6), $|A|$ is the determinant of the matrix whose entries are $a_{ij}$s, and

$$w_1 = \frac{1}{L} \sum_{k=0}^{[L/2]} \frac{\xi_k}{\Delta_k\{1-(1-v)^{2\lambda}\xi_k\}}.$$

An interesting relationship between the local and global variabilities follows from (17). To see this let

$$\hat{Q} = \frac{1}{L}\sum_{i=0}^{L-1}\hat{Q}_i.$$

Thus $\hat{Q}$ is the probability of identity for two randomly chosen genes from the whole population. Upon substitution of $\hat{Q}_i$ given by (17) into the above equation and rearranging, we have

$$\hat{Q} = \frac{(1-v)^{2\lambda}(\hat{H}-\hat{Q}_0)}{N_e L\{1-(1-v)^{2\lambda}\}} \simeq \frac{\hat{H}-\hat{Q}_0}{2N_T\lambda v} \tag{18}$$

where $N_T = LN_e$. In deriving (18), we assumed the equal colony size and restricted migration pattern, but the formula can be proved to be true for more general models of population structure. The only requirement is that the population has stable structure which allows an equilibrium state to exist. A similar formula has been derived for Mendelian genes in a very general situation (Maruyama, 1970; Crow & Maruyama, 1971).

Although one- and two-dimensional models are mathematically the same, the dimension of the space is known to have large effects on the amount of genetic variability and other quantities of biological interest (Kimura & Weiss, 1964; Maruyama, 1972). Therefore we feel it necessary to present corresponding formulae for a two-dimensional case. We assume that the population consists of $L_1 \times L_2$ colonies of equal size and that they are located on grid points of a two-dimensional lattice forming the surface of a torus. Let $Q_{kl}$ be the probability analogous to $Q_k$, in which $k$ indicates the number of steps apart along one axis and $l$ the steps along the other axis. Also let $\xi_0^{(i)} = 0$ and

$$\xi_k^{(i)} = \left\{1-\frac{m}{2}\left(1-\cos\frac{2\pi k}{L_i}\right)\right\}^2 \quad (i = 1, 2 \text{ and } k = 1, 2, \ldots, L_i),$$

where $m$ is the migration rate. Further let $\Delta_k^{(i)} = 1$ if $j = 0$ or $= L_i/2$ and $\Delta_k^{(i)} = \frac{1}{2}$ otherwise. Now using these notations, we define

$$w_2 = \frac{1}{L_1 L_2}\sum_{l=0}^{[L_1/2]}\sum_{k=0}^{[L_2/2]}\frac{\xi_l^{(1)}\xi_k^{(2)}}{\{1-(1-v)^{2\lambda}\xi_l^{(1)}\xi_k^{(2)}\}\Delta_l^{(1)}\Delta_k^{(2)}}.$$

Then we have the following formulae for the two-dimensional model:

$$\left.\begin{aligned}\hat{H} &= \frac{b(1-w_2 a_{22})}{1-(a_{11}+w_2 a_{22})+w_2|A|},\\ \hat{Q}_{00} &= \frac{w_2 a_{21}\hat{H}}{1-w_2 a_{22}},\end{aligned}\right\} \tag{19}$$

$$\hat{Q}_{ij} = \frac{(1-v)^{2\lambda}(\hat{H}-\hat{Q}_{00})}{N_e L_1 L_2}\sum_l\sum_k\frac{\xi_l^{(1)}\xi_k^{(2)}\cos 2\pi il/L_1\cos 2\pi jk/L_2}{\{1-(1-v)^{2\lambda}\xi_l^{(1)}\xi_k^{(2)}\}\Delta_l^{(1)}\Delta_k^{(2)}}$$

and finally

$$\hat{Q} = \frac{1}{L_1 L_2}\sum_{i,j}\hat{Q}_{ij} = \frac{(1-v)^{2\lambda}(\hat{H}-\hat{Q}_{00})}{N_T\{1-(1-v)^{2\lambda}\}}, \tag{20}$$

where $N_e$ is the effective size of a single subpopulation and $N_T = L_1 L_2 N_e$. Formula (20) is a two-dimensional version of the relationship given in (18).

https://doi.org/10.1017/S0016672300020292 Published online by Cambridge University Press

## 4. DISCUSSION

We have attempted to provide a mathematical model and analysis for the genetic variability maintained by the extranuclear genes transmitted through mitochondria or chloroplasts. We assumed that all the alleles are selectively neutral. This appears to be consistent with some of the observations made on these genes. For instance, Brown *et al.* (1979) argue that a seemingly rapid evolutionary rate of these genes is probably due to the possible dispensability of mitochondrial genes existing in multiple copies. They also suggest that these genes may have less functional constraints and therefore are relatively more free to be substituted. Furthermore it is known that many of the important mitochondrial proteins are genetically coded by nuclear genes (see Beale & Knowles, 1978). Taking these considerations into account, it may not be too unrealistic to assume that there are a large number of possible alleles to which a mitochondrial gene can mutate without altering its function.

Precise modes of replication and partition of the mitochondrial genes are not known, but we think the following two situations are likely to include the reality. The first is to assume that every gene replicates exactly once prior to a cell division and then the genes are partitioned randomly into two daughter cells. The second model assumes that prior to a cell division, genes replicate independently for a random number of times and theoretically form an infinite gene pool, and then each resulting daughter cell receives approximately $n$ copies randomly chosen from the pool. In the first case, the effective number of the genes is $2n-1$, while in the second case it is equal to $n$. The difference arises for the same reason as for populations in which each individual is assumed to have exactly the same number of offspring and in which the offspring number is a random variable.

We shall examine the effect of some parameter values on $\hat{H}$ and $\hat{Q}$. It has been shown that if the product $n\beta$ is less than $\lambda$, $\hat{H}$ is close to unity. And in general this appears to be the case in most higher organisms, because $\beta$ is likely to be at most a few per cent and $n$ is also not greater than 1000, while $\lambda$ seems to be of the order of 10 or 100. Therefore it is reasonable to take the $\hat{H}$ value as very close to one, as was observed by Avise *et al.* (1979*a*, *b*). On the other hand, the values of $\hat{Q}$ are much influenced by the total population size $N_T$, the mutation rate $v$, the migration rate $m$ and $\lambda$. If $\lambda$ is large and $n$ is small, the value of $\beta$ has small effect on $\hat{H}$ and $\hat{Q}$. On the other hand, when $\lambda$ is small and $n$ is large, the paternal effect on $\hat{H}$ and $\hat{Q}$ becomes large. These relationships are illustrated in Figs. 2 and 3, where considerable decreases in $\hat{H}$ and $\hat{Q}$ are apparent as $\beta$ increases. When $\beta = \frac{1}{2}$ and $N_e m \gg 1$, $\hat{Q}$ is nearly equal to $1/(1+nN_T v)$, which is the mean level of homozygosity expressed by a population consisting of $nN_T/2$ nuclear genes. The requirement that the female gametes determine primarily the level of $\hat{H}$ and $\hat{Q}$ is $\beta \ll \lambda/n$, and if this condition is satisfied the effect of the male gametes can be ignored without altering the results much.

According to Avise *et al.* (1979*a*, *b*), the mitochondrial genes from a single individual or from individuals belonging to a single local population show very

little genetic diversity, while the genes collected from a large geographical area are strongly divergent. This implies the necessity of an analysis in which the geographical structure of the population is considered. Using formulae (19) and (20) for a two-dimensional torus-like space, we have carried out some calculations to examine the effect of the geographical structure, We found that the local differentiation seems to be essentially the same as that expected for the nuclear

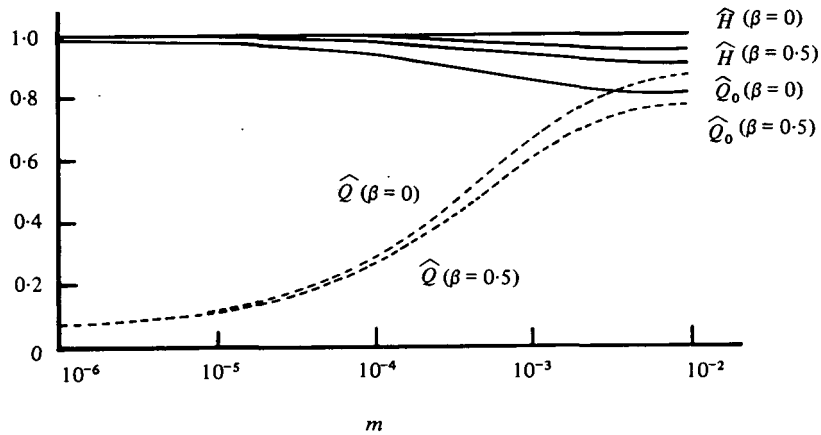


Fig. 3. Genetic identities ($\hat{H}$, $\hat{Q}_0$ and $\hat{Q}$) as functions of migration rate, for one-dimensional geographically structured populations where $N$ denotes the total population number and $N_f = N_m = N/2$. $N = 10^4$; $v = 10^6$; $L = 16$; $\lambda = 10$; $n = 10$.

genes, except that the mitochondrial genes are slightly more polymorphic than the Mendelian genes. Therefore it appears that the results of Avise *et al.* are consistent with the expectation based on our present theory.

We are quite confident that the model dealt with here takes nearly all the biologically known facts into account and therefore is fairly general. However, we are also aware that the biology of these genes is still at a rather early stage, particularly in regard to the ways of replication and partition of the extranuclear genes, and thus when it is fully understood the model may need to be slightly altered. In this context, the effective number of extranuclear genes in a cell used here is rather ambiguous and should be defined more precisely in future like that of nuclear genes. At any rate, because of its generality we believe that the present paper will provide a model useful in advancing the population genetics of cytoplasmic genes.

We are grateful to Drs J. Felsenstein, W. Birky, Jr and T. Ohta for many helpful discussions and constructive criticism. This work was supported partly by grants 439016 and 439017 from the Ministry of Education, Japan.

## REFERENCES

AVISE, J. C., LANSMAN, R. A. & SHADE, R. O. (1979*a*). The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus peromyscus. *Genetics* **92**, 279–295.

AVISE, J. C., GIBLIN-DAVIDSON, C., LAERM, J., PATTON, J. C. & LANSMAN, R. A. (1979*b*.) Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences, U.S.A.* **76**, 6694–6698.

BEALE, G. H. & KNOWLES, J. K. C. (1978). *Extranuclear Genetics*. London: Edward Arnold.

BIRKY, C. W. Jr. (1975). Zygote heterogeneity and uniparental inheritance of mitochondrial genes in yeast. *Molecular and General Genetics* **141**, 41–58.

BIRKY, C. W. Jr. (1978). Transmission genetics of mitochondria and chloroplasts. *Annual Review of Genetics* **12**, 471–512.

BIRKY, C. W. Jr., DEMKO, C. A., PERLMAN, P. S. & STRAUSBERG, R. (1978). Uniparental inheritance of mitochondrial genes in yeast: dependence on input bias of mitochondrial DNA and preliminary investigations of the mechanism. *Genetics* **89**, 615–651.

BIRKY, C. W. Jr. & SKAVARIL, R. V. (1976). Maintenance of genetic homogenity in systems with multiple genomes. *Genetical Research* **27**, 249–265.

BIRKY, C. W. Jr. & SKAVARIL, R. V. (1978). Genetic consequences of random replication of organelle DNA. *Genetics* **88**, s11 (abst.).

BROWN, W. M., GEORGE, Jr., M. & WILSON, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences, U.S.A.* **76**, 1967–1971.

CROW, J. F. & MARUYAMA, T. (1971). The number of neutral alleles maintained in a finite, geographical structured population. *Theoretical Population Biology* **2**, 437–453.

HAYASHI, J., YONEKAWA, H., GOTOH, O., TAGASHIRA, Y., MORIWAKI, K. & YOSIDA, T. (1979). Evolutionary aspects of variant types of rat mitochondrial DNAs. *Biochimica et Biophysica Acta* **564**, 202–211.

KIMURA, M. & CROW, J. F. (1963). The measurement of effective population number. *Evolution* **17**, 279–288.

KIMURA, M. & CROW, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

KIMURA, M. & WEISS, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561–576.

MARUYAMA, T. (1970). Stepping stone models of finite length. *Advances in Applied Probability* **2**, 229–258.

MARUYAMA, T. (1972). Rate of decrease of genetic variability in two-dimensional continuous population of finite size. *Genetics* **70**, 639–651.

MARUYAMA, T. & KIMURA, M. (1980). Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 6710–6714.

OHTA, T. (1978). Theoretical study of genetic variation in multigene families. *Genetical Research* **31**, 13–28.

OHTA, T. (1980). Evolution and variation of multigene families. *Lecture Notes in Biomathematics* **37**. Berlin, Heidelberg, New York: Springer-Verlag.

WRIGHT, S. (1931). Evolution in mendelian populations. *Genetics* **16**, 97–159.