

Population genetics of extranuclear genomes under the neutral mutation hypothesis*

BY NAOYUKI TAKAHATA

National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan

(Received 4 February 1982 and in revised form 13 June 1983)

SUMMARY

Population genetics of extranuclear genomes is further developed under the neutral-mutation random-drift hypothesis, and the characteristic evolutionary aspects are summarized. Several formulae derived here are concerned with the variances of genetic variability (gene identity) at a single extranuclear locus and the evolutionary distance between two isolated populations which is estimated from a comparison of homologous linked nucleotide sites. Two types of variance are considered; one is the variance in the entire population (V_Q) and the other is the variance within a single germ cell (V_H). When compared with a Mendelian genetic system in a panmictic population, an extranuclear genetic system has the following equilibrium properties: (1) the mean genetic variability is low if, despite the high multiplicity of the genome in a cell, the proportion of the cytoplasmic contribution from the male's gamete is small, (2) the effect of recombination is small and a large amount of variance of linkage disequilibrium tends to be maintained, (3) the overall relationship between the mean and variance of genetic variability does not much differ but V_Q (V_H) is expected to be small if the paternal contribution is small, and (4) the evolutionary distance estimated depends on the extent of intrapopulation variation in a common ancestor population which in turn depends on within-cell variation. I argue that there is an analogy between the model of extranuclear genomes in a finite population and that of nuclear genes in a subdivided population. The analogy helps our understanding of some properties in an extranuclear genetic system.

1. INTRODUCTION

There have been growing observations to reveal transmission genetics of extranuclear (non-Mendelian) genomes and their gene organization by means of modern techniques of molecular biology. In particular, study of the mitochondrial genomes in a variety of organisms has attracted much attention (Rabinowitz & Swift, 1970; Bogenhagen & Clayton, 1974; Birky, 1978; Giles *et al.* 1980; Anderson *et al.* 1981; Bibb *et al.* 1981; Wallace, 1982). Extranuclear genomes, differing from Mendelian genomes, undergo several unique evolutionary pressures due to the semi-autonomous replicating mechanism and lack of spindle apparatus in both mitosis and meiosis. Different contribution of cytoplasm from male and female

* Contribution no. 1448 from the National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan.

gametes in fertilization is also a characteristic in higher organisms (Lansman, Avise & Huettel, 1983 and references therein). In addition, the evidence that more than 80% enzymes and proteins in a mitochondrion are produced by nuclear genes suggests the coherent evolution with the host nuclear genome (for review see Beale & Knowles, 1978; Gillham, 1978). However, these characteristics alone do not seem to be responsible for the development of the extreme economy of gene organization found in mammalian mitochondrial genomes (Anderson *et al.* 1981; Bibb *et al.* 1981). There may be some unknown responding mechanisms; for instance, physical constraints against the size of genome. Otherwise the economy might be merely the vestige of the genome carried by an ancient organism which symbiosed to the present host cell.

Although the evolutionary process of extracellular genomes is complicated and uncertain, recent accumulation of data, particularly on mitochondrial DNA molecules, has stimulated some theoretical study under simplifying assumptions (Dujon, Slonimski & Weill, 1974; Upholt, 1977; Nei & Li, 1979; Ohta, 1980; Takahata & Maruyama, 1981; Engels, 1981; Chapman *et al.* 1982; Birky, Maruyama & Fuerst, 1983; Takahata, 1983*a*; Takahata & Slatkin, 1983). In this note, assuming that mutations are selectively neutral and replicated copies of extranuclear genome are transmitted randomly to the daughter cells in mitosis, I derive several formulae concerning the variance of genetic variability at a single locus and the evolutionary distance estimated from a comparison of homologous nucleotide sequences between two isolated populations. The model and assumptions presented here are plausible but have not yet been confirmed by observations. Nevertheless, this study surely provides a reference point and could even be the basis of future development in theory. Combining the results of variances of linkage disequilibrium between two extranuclear loci (Takahata, 1983*a*), evolutionary aspects of extranuclear genes are discussed in relation to those of Mendelian genes.

2. MODEL AND ANALYSIS

Our model of the transmission genetics of extranuclear genomes is essentially the same as those described by Takahata & Maruyama (1981), Chapman *et al.* (1982), Birky *et al.* (1983) and Takahata (1983*a*). But they are not exactly the same, so that we first explain our model. We consider a population with separate sexes. We assume that in a germ cell in each sex there exist multiple copies of extranuclear DNA molecules and that these molecules are transmitted randomly to the daughter cells in mitosis and therefore are subject to within-generation drift (Takahata & Slatkin, 1983). We call this model 'random transmission' instead of 'imperfect transmission' in Chapman *et al.* (1982). In reality, the unit of transmission is not a single DNA molecule but a cellular organelle within which several DNA molecules are contained. We ignore this fact, and each DNA molecule within a cell is transmitted independently. From the standpoint of population genetics, however, the theory will be constructed so as to incorporate this fact without a large modification (Ohta, 1980 for the possibility).

Although we assume that a germ cell in each sex contains the same number of extranuclear DNA molecules, we assume neither complete maternal inheritance nor that the two sexes necessarily make the same contribution to later generations.

Maternal inheritance is ubiquitous in higher eukaryotes (Lansman *et al.* 1983 and references therein) but may not be complete (Hauswirth & Laipis, 1982).

We further assume that an individual produces cytogenetically the same sperms or eggs. In other words, we ignore the possibility that a heteroplasmic individual produces different types of gametes. Obviously, this is a simplification, but it does not oversimplify the real situation unless the germ cell lines in an individual are highly heterogeneous. Under this assumption, we can use the terms 'germ cell line' and 'individual' interchangeably and simplify the mathematical treatment. For a model taking into account the above possibility, readers may refer to Birky *et al.* (1983).

It is convenient to list the variables that will be used consistently.

β = average proportion of extranuclear DNA molecules transmitted from the male's gamete so that $(1 - \beta)$ is that from the female's gamete.

n = effective number of extranuclear DNA molecules in a germ cell line, assumed to be the same in both sexes.

$N_m(N_f)$ = number of breeding males (females) in a population. If we consider two isolated populations, each population is assumed to consist of the same number of males and females.

λ = average number of somatic cell divisions per generation. If the number is different in sexes, we regard λ as $(\lambda_m N_m + \lambda_f N_f)/(N_m + N_f)$, where λ_m and λ_f are the number in male and female, respectively.

K = number of possible states per site where the term 'site' may be referred to as a nucleotide site, locus and a whole DNA molecule.

v = mutation rate per site per cell division. The total mutation rate per site per generation is denoted by $\mu = v\lambda$.

$N_p^{-1} = (1 - \beta)^p/N_f + \beta^p/N_m$ for $p = 0, 1, 2, \dots, N_p$ is related to the effective number of individuals. In particular, $N_p = N_f$ for $\beta = 0$ and $N_2 = 4N_m N_f/(N_m + N_f)$ for $\beta = 0.5$. The former is the case of completely maternal inheritance and indicates that the effective numbers are equal to the number of breeding females. The latter is equivalent to the effective number of individuals in Mendelian population with separate sexes.

In the following, we specify a possible state of extranuclear DNA molecule by using a vector i . When we are concerned with a single site, i is a scalar, but in general, when we want to treat r linked sites, i is a vector of $i = (i_1, i_2, \dots, i_r)$. We designate the DNA molecule with a vector i by A_i and the frequency of A_i in the l th individual by $x_i(l)$. We assume that the frequencies of A_i are the same in both sexes. We also assume that all mutations are selectively neutral or equivalent and the mutation scheme at each site obeys the Kimura's K allele model (Kimura, 1968). An extension of the K allele model to the case of multiple nuclear loci is given in Griffiths (1981), Takahata (1982) and Golding & Strobeck (1982).

Now, let us consider the change of frequency of A_i in an individual due to random sampling of gametes followed by fertilization. Suppose that the parents of, say, the l th individual are the m th male ($m = 1, 2, \dots, N_m$) and the f th female ($f = 1, 2, \dots, N_f$). Then taking into account a different contribution of DNA molecules from each parent, we can obtain

$$x_i^*(l) = (1 - \beta)x_i(f) + \beta x_i(m), \quad (1)$$

in which the superscript * stands for the frequency after fertilization, the first term in the right-hand side is the proportion of A_i transmitted from the f th female and the second is that from the m th male. Note here that there are $N_m \times N_f$ combinations in total for each new individual. Therefore when we consider the expectation of a certain quantity, denoted by $E_s\{ \}$, it is taken over all these combinations. Also, note that (1) is a consequence of random sampling of gametes in each sex so that the probability that two molecules in different individuals come from one male (female) is $1/N_m$ ($1/N_f$). Furthermore, since we will be concerned with identity probabilities of extranuclear DNA molecules randomly chosen from the same or different individuals, it is necessary to show which molecule resides in which individual. To indicate this situation and also the expectation of $E_s\{ \}$, we introduce a notation of using angle brackets $\langle \rangle$. A single bracket indicates the expectation of identity probability that all molecules in question are chosen from one individual and the product of two or more brackets shows the expectation of identity probability that the molecules are chosen from two or more different individuals. Thus $\langle x_i^2 \rangle$ is the expectation of identity probability that two A_i are chosen from an individual and $\langle x_i \rangle^2 = \langle x_i \rangle \langle x_i \rangle$ is the expectation of identity probability that two A_i are chosen randomly from two different individuals. More generally, $\langle x_i x_j \rangle \langle x_k \rangle \langle x_l \rangle$ is the expectation of A_i, A_j, A_k and A_l when the first two are chosen from an individual and the third and the fourth are chosen randomly from a second and a third individual, respectively.

To be complete, we give the procedure of calculating $\langle x_i^{*2} \rangle$ and $\langle x_i^* \rangle^2$ as follows:

$$\begin{aligned} \langle x_i^{*2} \rangle &= E_s\{x_i^*(l)^2\} \\ &= E_s\{[(1-\beta)x_i(f) + x_i(m)]^2\} \\ &= (1-\beta)^2 E_s\{x_i^2(f)\} + \beta^2 E_s\{x_i^2(m)\} + 2\beta(1-\beta) E_s\{x_i(f)x_i(m)\} \\ &= (1-\rho)\langle x_i^2 \rangle + \rho\langle x_i \rangle^2, \end{aligned} \tag{2}$$

where

$$\rho = 2\beta(1-\beta), \quad \langle x_i^2 \rangle = \frac{1}{N_f} \sum_f x_i^2(f) = \frac{1}{N_m} \sum_m x_i^2(m),$$

$$\begin{aligned} \langle x_i \rangle^2 &= \frac{1}{N_f(N_f-1)} \sum_{f \neq f'} x_i(f)x_i(f') = \frac{1}{N_m(N_m-1)} \sum_{m \neq m'} x_i(m)x_i(m') \\ &= \frac{1}{N_f N_m} \sum_{f,m} x_i(f)x_i(m), \end{aligned}$$

and

$$\begin{aligned} \langle x_i^* \rangle^2 &= E_s\{x_i^*(l)x_i^*(l')\} \\ &= E_s\{[(1-\beta)x_i(f) + \beta x_i(m)][(1-\beta)x_i(f') + \beta x_i(m')]\} \\ &= (1-\beta)^2 E_s\{x_i(f)x_i(f')\} + \beta^2 E_s\{x_i(m)x_i(m')\} \\ &\quad + \beta(1-\beta) E_s\{x_i(f)x_i(m') + x_i(f')x_i(m)\} \\ &= (1-\beta)^2 \left\{ \frac{1}{N_f} \langle x_i^2 \rangle + \left(1 - \frac{1}{N_f}\right) \langle x_i \rangle^2 \right\} \\ &\quad + \beta^2 \left\{ \frac{1}{N_m} \langle x_i^2 \rangle + \left(1 - \frac{1}{N_m}\right) \langle x_i \rangle^2 \right\} + 2\beta(1-\beta) \langle x_i \rangle^2 \\ &= \frac{1}{N_2} \langle x_i^2 \rangle + \left(1 - \frac{1}{N_2}\right) \langle x_i \rangle^2, \end{aligned} \tag{3}$$

where the prime denotes a different indicator for individuals and $l = l'$ only if $f = f'$ and $m = m'$. Formulae (2) and (3) are identical to (1) and (2) in Takahata & Maruyama (1981).

An advantage of the present formulation for identity probability is to make easier the calculation of higher moments. To derive the formulae of the variances of identity probabilities, the twelve variables up to the fourth moments are required. The third moments are related to the expectation of identity probabilities when we choose three molecules of the same kind randomly from one, two and three individuals. Likewise, the fourth moments are related to the expectation of identity probabilities when we sample four molecules of two different kinds randomly from one, two, three and four different individuals. Thus the expectations of identity probabilities of the third order are denoted by $\langle x_i^3 \rangle$, $\langle x_i^2 \rangle \langle x_i \rangle$ and $\langle x_i \rangle^3$ and those of the fourth moments by $\langle x_i^2 x_j^2 \rangle$, $\langle x_i^2 x_j \rangle \langle x_j \rangle$, $\langle x_i^2 \rangle \langle x_j^2 \rangle$, $\langle x_i^2 \rangle \langle x_j \rangle^2$, $\langle x_i x_j \rangle \langle x_i \rangle \langle x_j \rangle$ and $\langle x_i \rangle^2 \langle x_j \rangle^2$. The calculation of those quantities can be made straightforwardly though it is rather tedious so that we demonstrate only the calculation of $\langle x_i^{*2} \rangle \langle x_i^* \rangle$ here.

$$\begin{aligned}
 \langle x_i^{*2} \rangle \langle x_i^* \rangle &= E_s\{[(1-\beta)x_i(f) + \beta x_i(m)]^2 [(1-\beta)x_i(f') + \beta x_i(m')]\} \\
 &= (1-\beta)^3 E_s\{x_i^2(f)x_i(f')\} + \beta^3 E_s\{x_i^2(m)x_i(m')\} \\
 &\quad + \beta(1-\beta)^2 E_s\{x_i^2(f)x_i(m')\} + \beta^2(1-\beta) E_s\{x_i^2(m)x_i(f')\} \\
 &\quad + 2\beta(1-\beta)^2 E_s\{x_i(f)x_i(f')x_i(m)\} + 2\beta^2(1-\beta) E_s\{x_i(f)x_i(m)x_i(m')\} \\
 &= (1-\beta)^3 \left\{ \frac{1}{N_f} \langle x_i^3 \rangle + \left(1 - \frac{1}{N_f}\right) \langle x_i^2 \rangle \langle x_i \rangle \right\} \\
 &\quad + \beta^3 \left\{ \frac{1}{N_m} \langle x_i^3 \rangle + \left(1 - \frac{1}{N_m}\right) \langle x_i^2 \rangle \langle x_i \rangle \right\} \\
 &\quad + \beta(1-\beta) \langle x_i^2 \rangle \langle x_i \rangle + 2\beta(1-\beta)^2 \left\{ \frac{1}{N_f} \langle x_i^2 \rangle + \left(1 - \frac{1}{N_f}\right) \langle x_i \rangle^2 \right\} \langle x_i \rangle \\
 &\quad + 2\beta^2(1-\beta) \left\{ \frac{1}{N_m} \langle x_i^2 \rangle + \left(1 - \frac{1}{N_m}\right) \langle x_i \rangle^2 \right\} \langle x_i \rangle \\
 &= \frac{1}{N_3} \langle x_i^3 \rangle + \left(1 - \rho + \frac{1}{N_3} + \frac{\rho}{N_1}\right) \langle x_i^2 \rangle \langle x_i \rangle + \rho \left(1 - \frac{1}{N_1}\right) \langle x_i \rangle^3, \tag{4}
 \end{aligned}$$

where the relationship $1/N_p = (1-\beta)^p/N_f + \beta^p/N_m$ for $p = 1$ and 3 is used. For the other identity probabilities, readers may refer to the derivation in Takahata (1983a).

We represent the above twelve variables in vector form $X = (X_1, X_2, \dots, X_{12})^t$ in which the superscript t denotes the transpose of a vector:

$$\begin{aligned}
 X_1 &= \langle x_i^2 \rangle, & X_2 &= \langle x_i \rangle^2, & X_3 &= \langle x_i^3 \rangle, & X_4 &= \langle x_i^2 \rangle \langle x_i \rangle, & X_5 &= \langle x_i \rangle^3, \\
 X_6 &= \langle x_i^2 x_j^2 \rangle, & X_7 &= \langle x_i x_j \rangle^2, & X_8 &= \langle x_i^2 x_j \rangle \langle x_j \rangle, & X_9 &= \langle x_i^2 \rangle \langle x_j^2 \rangle, \\
 X_{10} &= \langle x_i^2 \rangle \langle x_j \rangle^2, & X_{11} &= \langle x_i x_j \rangle \langle x_i \rangle \langle x_j \rangle, & X_{12} &= \langle x_i \rangle^2 \langle x_j \rangle^2.
 \end{aligned}$$

Then, we have $X^* = RX$, (5)

where

$$R = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & R_3 \end{pmatrix}, \quad (6) \quad R_1 = \begin{pmatrix} 1-\rho & \rho \\ \frac{1}{N_2} & 1-\frac{\rho}{N_2} \end{pmatrix}, \quad (7) \quad R_2 = \begin{pmatrix} 1-\frac{3\rho}{2} & 0 \\ \frac{1}{N_3} & 1-\rho-\frac{\rho}{N_3}+\frac{\rho}{N_1} \\ 0 & \frac{3}{N_2} \end{pmatrix} \quad (8)$$

and $R_3 =$

$$\begin{pmatrix} 1-2\rho+\frac{\rho^2}{2} & \rho^2 & 2\rho(1-\rho) & \frac{\rho^2}{2} & 0 & 0 \\ \frac{1}{N_4} & (1-\rho)^2-\frac{1}{N_4} & \frac{2\rho}{N_2} & 0 & \frac{\rho^2}{2N_0} & \rho^2\left(1-\frac{1}{N_0}\right) \\ \frac{1}{N_4} & \frac{\rho^2}{2N_0} & 1-\frac{3\rho}{2}-\frac{1}{N_4}+\frac{3\rho}{2N_2} & \frac{\rho^2}{4N_0} & \frac{\rho}{2}\left(1-\frac{1}{N_1}\right) & 0 \\ \frac{1}{N_4} & 0 & \frac{2\rho}{N_2} & (1-\rho)^2-\frac{1}{N_4} & 2\rho\left(1-\rho-\frac{1}{N_2}+\frac{\rho}{2N_0}\right) & \rho^2\left(1-\frac{1}{N_0}\right) \\ 0 & 0 & \frac{2}{N_3} & \frac{1}{N_4}+\frac{\rho^2}{4N_0} & 1-\rho-\frac{1-\rho}{N_2}-\frac{2}{N_4} & \rho\left(1-\frac{3}{N_2}-\frac{\rho}{N_0}\right) \\ 0 & \frac{1}{N_4}+\frac{\rho^2}{4N_0} & \frac{2}{N_3} & \frac{\rho}{N_1} & 1-\rho+\frac{\rho}{N_1}-\frac{1-\rho}{N_2}-\frac{2}{N_4} & \rho\left(1-\frac{3}{N_2}-\frac{\rho}{N_0}\right) \\ 0 & 0 & 0 & 0 & \frac{4}{N_2} & 1-\frac{6}{N_2} \end{pmatrix} \quad (9)$$

The above formulae are derived for particular types of DNA molecules i and j , but they are homogeneous equations so that the same formulae must hold true for the variables summed over all the possible states of i and j . Then we interpret the variables of the vector X as the expectation of identity probabilities irrespective of the type of DNA molecules.

Next let us consider the change of $x_i^*(t)$ in a germ cell line in an individual which is caused by mutation and within-generation drift. To derive the formulae for the changes of expected identity probabilities, we make use of a diffusion approximation method. We specify the i th DNA molecule by r multiple sites like $\mathbf{i} = (i_1, i_2, \dots, i_r)$ and assume the complete linkage between sites. For convenience, we introduce a function of $x_i(t)$, and drop an indicator of individuals, noting that the process of mutation and within-generation drift takes place independently from individual to individual. Let $x_i(t)$ be the frequency of A_i in an individual after t somatic cell divisions, and $x_{i,p}(t)$ be the frequency of DNA molecules that are specified by the vector $(i_1, i_2, \dots, i_{p-1}, i_{p+1}, \dots, i_r)$, i.e.

$$x_{i,p}(t) = \sum_{i_{p-1}}^K x_{(i_1, i_2, \dots, i_r)}(t). \tag{10}$$

If DNA molecules are specified by only one site, then (10) should be equal to 1. Using the model of mutation mentioned earlier and a diffusion approximation for the change of $x_i(t)$ due to random distribution of replicated molecules in somatic cell division, we have

$$x_i(t+1) = (1-rv)x_i(t) + \frac{v}{K-1} \sum_{p=1}^r \sum_{\substack{i_{p-1} \\ i'_p \neq i_p}}^K x_{(i_1, \dots, i'_p, \dots, i_r)}(t) + \xi_i \tag{11}$$

in which the first term of the right-hand side corresponds to the probability of no change, the second to the contribution coming from all molecules that can produce A_i by a single-step mutation and the third is a random variable with mean 0 and covariance

$$E\{\xi_i \xi_j\} = \frac{1}{n} x_i(t) \{\delta_{ij} - x_j(t)\}. \tag{12}$$

In (12), $E\{\}$ stands for taking the expectation with respect to within-generation drift and $\delta_{ij} = 1$ if $i = j$ and if otherwise 0. If we use (10), (11) becomes

$$x_i(t+1) = \left(1 - \frac{Krv}{K-1}\right)x_i(t) + \frac{v}{K-1} \sum_{p=1}^r x_{i,p}(t) + \xi_i \tag{13}$$

for $r > 1$ and

$$x_i(t+1) = \left(1 - \frac{Kv}{K-1}\right)x_i(t) + \frac{v}{K-1} + \xi_i \tag{13'}$$

for $r = 1$.

(i) Variance of genetic variability

Let us first formulate the variances of identity probabilities at a single site, keeping the number of states K finite. We redefine the vector X in terms of the expectations of identity probabilities:

$$\begin{aligned} X_1 &= \sum_i E\{\langle x_i^2 \rangle\}, & X_2 &= \sum_i E\{\langle x_i \rangle^2\}, & X_3 &= \sum_i E\{\langle x_i^3 \rangle\}, & X_4 &= \sum_i E\{\langle x_i^2 \rangle \langle x_i \rangle\}, \\ X_5 &= \sum_i E\{\langle x_i \rangle^3\}, & X_6 &= \sum_{ij} E\{\langle x_i^2 x_j^2 \rangle\}, & X_7 &= \sum_{ij} E\{\langle x_i x_j \rangle^2\}, & X_8 &= \sum_{ij} E\{\langle x_i^2 x_j \rangle \langle x_j \rangle\}, \\ X_9 &= \sum_{ij} E\{\langle x_i^2 \rangle \langle x_j^2 \rangle\}, & X_{10} &= \sum_{ij} E\{\langle x_i^2 \rangle \langle x_j \rangle^2\}, & X_{11} &= \sum_{ij} E\{\langle x_i x_j \rangle \langle x_i \rangle \langle x_j \rangle\} \end{aligned}$$

and

$$X_{12} = \sum_{ij} E\{\langle x_i \rangle^2 \langle x_j \rangle^2\}.$$

Noting that $h = \sum_i \langle x_i^2 \rangle$ and $q = \sum_i \langle x_i \rangle^2$ are the identity probabilities within an individual and in a population, X_6 and X_{12} are the expectations of $\langle (\sum_i x_i^2) \rangle$ and q^2 , respectively (in our previous paper we denoted the expectations of h and q as H and Q , which are X_1 and X_2 in the present notation). Thus the variance of genetic variability within an individual is given by

$$V_H = X_6 - X_1^2 \tag{15}$$

and that in a population is

$$V_Q = X_{12} - X_2^2. \tag{16}$$

As mentioned before, the transition matrix of X due to random sampling of gametes and fertilization is the same as that in (5). Accordingly, we have

$$X^* = RX. \tag{17}$$

On the other hand, the transition matrix, M , for the change of x due to mutation and within-generation drift in a somatic cell division is readily calculated by using (13'), which is given in the Appendix. Defining $v' = Kv/(K-1)$ and a column vector b with the elements $b_1 = 1/n + 2v'/K$, $b_2 = 2v'/K$ and $b_p = 0$ for $p = 3, 4, \dots, 12$, we can describe the change of X in one cell division as

$$X(t+1) = MX(t) + b. \tag{18}$$

Assuming λ cell divisions in one generation, we have

$$X(\lambda) = M^\lambda X^* + (M^{\lambda-1} + M^{\lambda-2} + \dots + I) b, \tag{19}$$

where I is the identity matrix and $X(\lambda)$ denotes the value of X before random sampling of gametes for the next generation. Formulae (17) and (19) give a complete set of equations to get the variances. In particular, the equilibrium value of X after λ cell divisions is given by

$$\tilde{X} = (I - M^\lambda R)^{-1} (M^{\lambda-1} + M^{\lambda-2} + \dots + I) b. \tag{20}$$

When we want to determine the value immediately after fertilization, $\tilde{\tilde{X}}$, we multiply \tilde{X} by R and obtain

$$\tilde{\tilde{X}} = R\tilde{X}. \tag{21}$$

(ii) *Evolutionary distance*

In this section, we derive a formula for estimating the evolutionary distance taking proper account of intrapopulational and within-cell heterogeneities. Let us consider completely linked r sites of DNA molecules in two isolated populations which were split T generations ago. The evolutionary distance is defined by

$$K_{nuc} = 2\mu T = 2v\lambda T \tag{22}$$

as the average number of mutations per site that have occurred since the divergence of the two populations. This definition includes not only the contribution of nucleotide substitutions but also that of segregating mutations. The simple quantities for estimating the distance are the expectation of identity probabilities within and between populations. Denoting by $y_i(l)$ the frequency of A_i in the l th

individual in the second population we define the expected identity probabilities for the r linked sites within an individual, within a population and between populations as

$$\begin{aligned}
 H_r &= \sum_i E\langle x_i^2 \rangle = \sum_i E\langle y_i^2 \rangle, \\
 Q_r &= \sum_i E\langle x_i \rangle^2 = \sum_i E\langle y_i \rangle^2, \\
 J_r &= \sum_i E\langle x_i \rangle \langle y_i \rangle.
 \end{aligned}
 \tag{23}$$

By using (13) and the corresponding equation for $y_i(l)$, we obtain

$$\Delta H_r = -\left(\frac{1}{n} + \frac{2Kvr}{K-1}\right) H_r + \frac{2vr}{K-1} H_{r-1} + \frac{1}{n}, \quad \Delta Q_r = \frac{2vr}{K-1} (Q_{r-1} - KQ_r)
 \tag{24}$$

and

$$\Delta J_r = \frac{2vr}{K-1} (J_{r-1} - KJ_r)$$

for mutation and within-generation drift. In the above equations, Δ denotes the difference in the time interval between two consecutive cell divisions and a symbol of indicating time is suppressed. As time is measured in units of cell division and λ cell divisions are assumed to occur in one generation, we need the time-dependent solutions of (24). They can be solved as:

$$\begin{aligned}
 H_r^{**} &= \frac{1}{K^r} \sum_{p=0}^r {}_r C_p \left[e^{-(\lambda/n + 2K\mu p/(K-1))} \sum_{q=0}^p {}_p C_q (-1)^q K^{p-q} H_{p-q}^* \right. \\
 &\quad \left. + \{1 - e^{-(\lambda/n + 2K\mu p/(K-1))}\} \sum_{q=0}^p {}_p C_q (-1)^q K^{p-q} \hat{H}_{p-q} \right], \\
 Q_r^{**} &= \frac{1}{K^r} \sum_{p=0}^r {}_r C_p e^{-2K\mu p/(K-1)} \sum_{q=0}^p {}_p C_q (-1)^q K^{p-q} Q_{p-q}^*
 \end{aligned}
 \tag{25}$$

and

$$J_r^{**} = \frac{1}{K^r} \sum_{p=0}^r {}_r C_p e^{-2K\mu p/(K-1)} \sum_{q=0}^p {}_p C_q (-1)^q K^{p-q} J_{p-q}^*.$$

In the above equations, ${}_r C_p$ is the binomial coefficient, \hat{H}_r is the solution of $\Delta H_r = 0$ in (24), and the single and double asterisks denote the values immediately after fertilization and meiosis, respectively.

Recalling that H_r^* and Q_r^* are the outcome of random sampling of gametes and fertilization, we get

$$H_r^* = (1 - \rho) H_r - \rho Q_r, \quad Q_r^* = \frac{1}{N_2} H_r + \left(1 - \frac{1}{N_2}\right) Q_r$$

from (7).

On the other hand, J_r does not change in this process, i.e.

$$J_r^* = J_r,$$

since random sampling of gametes is assumed to take place independently between two populations. Thus we can obtain the time-continuous solution of $J_r(T)$ by integrating (24) directly over λT cell divisions

$$J_r(T) = \frac{1}{K^r} \sum_{p=0}^r {}_r C_p e^{-2K\mu pT/(K-1)} \sum_{q=0}^p {}_p C_q (-1)^q K^{p-q} J_{p-q}(0).
 \tag{26}$$

Differential equations for $H_r(T)$ and $Q_r(T)$ are given by

$$\begin{aligned} \frac{dH_r}{dT} &= -H_r + \frac{1}{K^r} \sum_{p=0}^r {}_rC_p \left[e^{-(\lambda/n+2K\mu p/(K-1))} \sum_{q=0}^p {}_pC_q (-1)^q K^{p-q} \right. \\ &\quad \left. \times \{(1-\rho)H_{p-q} + \rho Q_{p-q}\} + \{1 - e^{-(\lambda/n+2K\mu p/(K-1))}\} \sum_{q=0}^p {}_pC_q (-1)^q K^{p-q} \hat{H}_{p-q} \right], \\ \frac{dQ_r}{dT} &= -Q_r + \frac{1}{K^r} \sum_{p=0}^r {}_rC_p \left[e^{-2K\mu p/(K-1)} \sum_{q=0}^p {}_pC_q (-1)^q K^{p-q} \left\{ \frac{1}{N_2} H_{p-q} + \left(1 - \frac{1}{N_2}\right) Q_{p-q} \right\} \right] \end{aligned} \tag{27}$$

from substitution of H_r^* and Q_r^* for the right-hand side in (25).

If we assume that $J_r(0) = \tilde{Q}_r$, and that $J_r(T)$ and \tilde{Q}_r are estimated from observation, then we can estimate K_{nuc} from (26) and the equilibrium solution of Q_r in (27). Fortunately, \hat{H}_r and \tilde{Q}_r can be determined from the lower order successively under the condition of $\hat{H}_0 = 1$ and $\tilde{Q}_0 = 1$. It should be noted in (26) that the functional relationship between $J_r(T)$ and K_{nuc} does not include \hat{H}_r , so that when $J_r(T)$ and \tilde{Q}_r are directly calculated from data, we can estimate K_{nuc} without any knowledge about \hat{H}_r .

The formulae for a single site are rather simple, as given below:

$$J_1(T) = \frac{1}{K} + \left(J_1(0) - \frac{1}{K} \right) e^{-2K\mu T/(K-1)}$$

or substituting (22) and solving the above equation for K_{nuc} ,

$$K_{nuc} = -\frac{K-1}{K} \log \left\{ \frac{KJ_1(T) - 1}{KJ_1(0) - 1} \right\}, \tag{28}$$

$$\tilde{Q}_1 = \left\{ \hat{H}_1 + \frac{N_2}{K} (e^{2K\mu/(K-1)} - 1) \right\} / \left\{ 1 + N_2(e^{2K\mu/(K-1)} - 1) \right\} \tag{29}$$

and

$$\hat{H}_1 = \frac{\hat{H}_1 + \frac{\rho}{K} \frac{2N_2\mu}{1+2N_2\mu} \left\{ e^{\lambda/n+2K\mu/(K-1)} - 1 \right\}}{1 + \rho \frac{2N_2\mu}{1+2N_2\mu} \left\{ e^{\lambda/n+2K\mu/(K-1)} - 1 \right\}}, \tag{30}$$

where

$$\hat{H}_1 = \frac{K-1+2nv}{K-1+2Knv} \approx 1.$$

The formulae (29) and (30) are the time-continuous K -allele model version of (9) and (10) in Takahata & Maruyama (1981). From the comparison we can check the validity of the diffusion approximation, and we found that it is quite satisfactory unless n is very small, say smaller than 10.

3. DISCUSSION

To understand the evolution and variation of extranuclear genomes, there are several quantities to be confirmed experimentally. The studies on DNA content per cell reviewed by Nass (1969) indicate a yield of 250 mitochondria per mouse fibroblast cell grown in culture (L cell) each of which contains about 6 DNA

molecules (see also Bogenhagen & Clayton, 1974). Thus, a single L cell contains roughly 1500 mitochondrial DNA molecules. More importantly, however, as she pointed out, the DNA content differs in different cell types and the highest content occurs in tumour cells. The number in a germ cell, n , that is relevant to the present problem is not well known and may be much smaller than 10^3 or 10^4 .

The number of nuclear DNA replications in a germ cell line per generation is estimated as 10–20, depending upon the species examined (Kondo, 1977). But it is still unclear if extranuclear DNA replication in the germ line occurs at the same rate. There is evidence that turnover of mitochondrial DNA is very rapid in some tissues, and hence λ may be much larger than 10 or 20 (Gross, Getz & Rabinowitz, 1969; Rabinowitz & Swift, 1970).

On the other hand, there are many conclusive indications of maternal inheritance of mitochondrial DNA. They come from the study of *Drosophila* (Reilly & Thomas, 1980), *Xenopus* (Dawid & Blackler, 1972), mouse (Avisé *et al.* 1979) and human (Giles *et al.* 1980). It seems important, however, to note that the resolution of those experiments is at most 4%, i.e. the methods can only detect heterogeneous DNA molecules, if present, at a level of more than 4%. Therefore, although β is small in most higher organisms, observations do not prove the completely maternal inheritance. From the theoretical standpoint, it is critical to distinguish the case of $\beta = 0$ from that of β close to 0. The significance of the distinction depends on λ as well as n . Recently, Lansman *et al.* (1983) improved the method for detecting the 'paternal leakage' of mitochondrial DNA and demonstrated that β is less than 0.004% in the tobacco budworm. If this is the case in other organisms, the evolutionary implication of paternal leakage is extremely small. On the other hand, an unusual pattern of mitochondrial DNA polymorphism observed in a maternal lineage of Holstein cows (Hauswirth & Laipis, 1982) may suggest the small but non-negligible contribution of paternal leakage.

We have as yet little precise knowledge about such key quantities. The parameter values used in the following should be regarded as tentative.

(i) Mean genetic variability

Assuming that $K = \infty$, $\lambda = 20$, $n = 10^3$ and $v = 10^{-6}$, and using (29) and (30), we plotted \hat{H}_1 and \hat{Q}_1 in Fig. 1 as functions of $N(N_f = N_m)$ and β . When $\beta = 0$, within-cell variation, $1 - \hat{H}_1$, is very small and independent of N whereas the intrapopulation variation, $1 - \hat{Q}_1$, increases as N increases. Both values are, however, smaller than the expected heterozygosity at a nuclear locus with the same mutation rate μ (the homozygosity is indicated by a thick line in the figure). This is because the effective population size N_e and the ploidy are different in the two genetic systems (Takahata & Maruyama, 1981; Engels, 1981). When $\beta = 0$, $N_e = N_2 = N_f$ in the extranuclear system but, in contrast, $N_e = 4N_m N_f / (N_m + N_f)$ in the nuclear system. In addition, if the organism is diploid, each individual contributes exactly two nuclear genomes to the population. Thus there could be a difference of four times in the 'effective number of genomes' between both genetic systems.

However, as the cytoplasmic contribution from the male gamete becomes large,

a different feature appears. The critical value of β is the order of magnitude of λ/n even though β itself is very small. In the present example, if we assume $\beta = 0.02$, then \bar{Q}_1 takes roughly the same value as that expected at a nuclear locus (note that λ/n is set as 0.02). As β further increases, not only \bar{Q}_1 but also \bar{H}_1 rapidly decrease, and a large amount of genetic variability can be maintained at an extranuclear locus. Thus whether 4% paternal contribution can be neglected entirely depends on the values of λ and n .

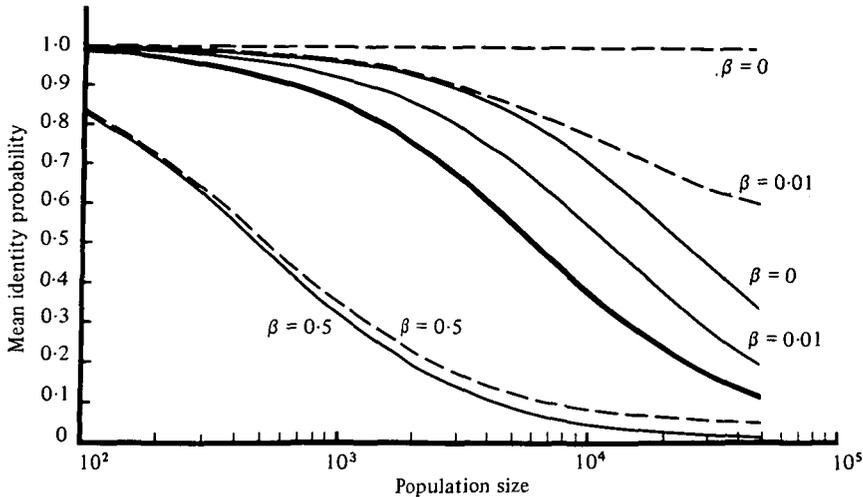


Fig. 1. Equilibrium mean genetic variability at an extranuclear locus as functions of β and N ($N_f = N_m$). The broken and solid lines represent within-cell and intrapopulational variations (\bar{H}_1 and \bar{Q}_1), respectively. The thick line indicates the average homozygosity at a nuclear locus with population size $2N$ and mutation rate μ . Here $n = 10^3$, $\lambda = 20$, $v = 10^{-6}$ ($\mu = v\lambda = 2 \times 10^{-5}$) and the infinite allele model are assumed.

(ii) Linkage disequilibrium

The amount of variance of linkage disequilibrium for the entire population, \bar{D}_b^2 (Takahata, 1983a) is depicted in Fig. 2 as a function of recombination fraction c , where we assumed that $K = \infty$, $n = 10$, $\lambda = 20$ and $2N_2\mu = 1$. When β is small, two genes linked on an extranuclear genome tend to be non-randomly associated, since \bar{H}_1 is close to 1 and therefore recombination has no apparent effect. The reduction of variance of linkage disequilibrium for the case of nuclear loci is similar to the line of \bar{D}_b^2 for $\beta = 0.5$. Under this circumstance, within-cell variation is high, so that the effect of recombination between two loci within a cell becomes efficient.

To see the n -dependence of \bar{D}_b^2 , the case of $n = 10^4$ was also computed for a small value of β ($= 10^{-3}$). Under these conditions, $\bar{D}_b^2 = 0.061$ for $c = 0$ and 0.0067 for $c = 0.001$. Comparison with the line indicated by $\beta = 10^{-3}$ in Fig. 2 shows that the effect of recombination is very strong in this case due to the high level of polymorphism within a cell maintained by relatively large n and β . Thus, roughly speaking, we can conclude that so long as \bar{H}_1 is larger than the mean homozygosity at a nuclear locus, the effect of recombination between extranuclear genes is less

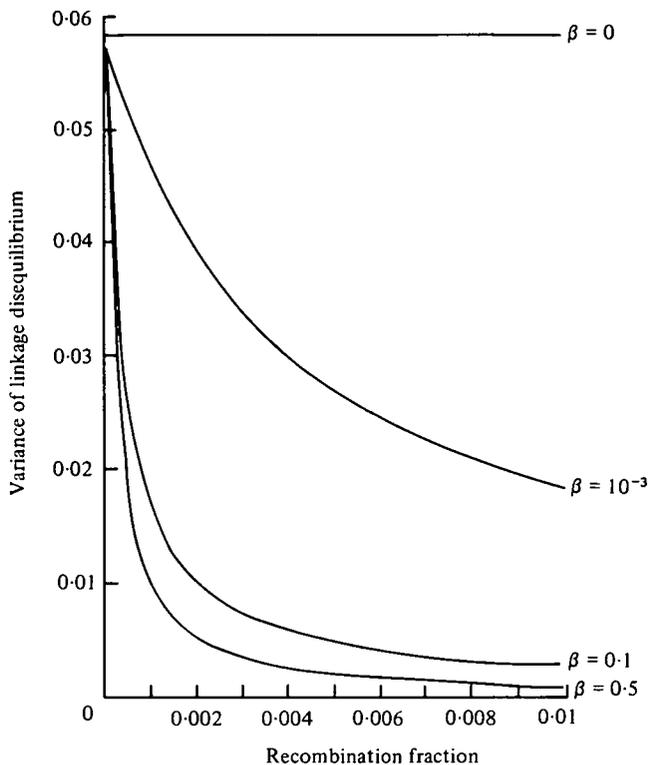


Fig. 2. Equilibrium variance of linkage disequilibrium in a population is plotted as a function of recombination fraction between two extranuclear loci with K infinite. Here mutation rate v at both loci, n and λ are assumed to be 10^{-6} , 10 and 20, respectively. For the mathematical formula, see Takahata (1983*a*).

efficient than that between nuclear genes with the same fraction c , and that a large amount of \hat{D}_b^2 is expected in an extranuclear genetic system even when recombination is frequent.

(iii) Variances

Variance of genetic variability measured by \hat{V}_H and \hat{V}_Q at equilibrium is plotted in Fig. 3(*a*) by using (20), while coefficient of variation is given in Fig. 3(*b*). The solid curves represent the expected relationships for the case of neutral mutations at a nuclear locus (Stewart, 1976). It is seen from Figs. 3(*a*, *b*) that the overall relationship between the mean and variance of genetic variability at an extranuclear locus is rather similar to that at a nuclear locus (Birky *et al.* 1983). The close relationship is, however, very complicated and the variances \hat{V}_H and \hat{V}_Q deviate either upward or downward from the variance, \hat{V}_n , at a nuclear locus with the same mean genetic variability.

To understand the reason, it may be helpful to note that the model of extranuclear genes treated here is analogous to that of nuclear genes in a subdivided population. An extranuclear genome in a cell corresponds to an individual in a subpopulation, an individual or cell to a single colony consisting

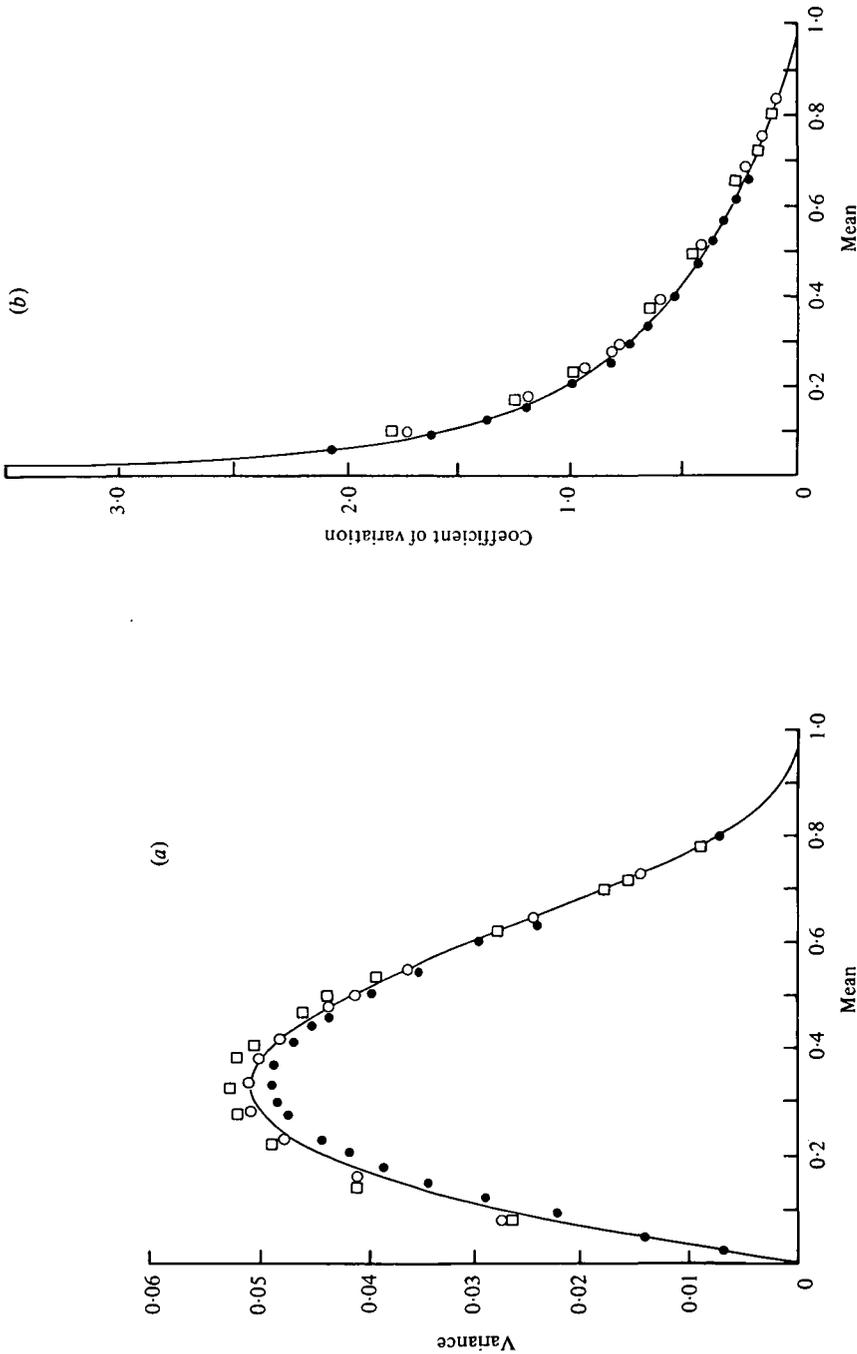


Fig. 3. (a) Relationship between variance and mean of genetic variability ($1 - \bar{Q}_1$ or $1 - \bar{H}_1$) at equilibrium. The solid line shows the relationship at a nuclear locus. We assumed $n = 10^4$, $\lambda = 20$, $v = 10^{-6}$ and $K = \infty$. Variances \bar{V}_Q and \bar{V}_H in the case of $\beta = 0.5$ are indicated by the open squares and circles, respectively, which are obtained by changes in population size N . The solid circles indicate \bar{V}_Q for $\beta = 0$, and the value of \bar{V}_H does not change and equals 0.0063. (b) Coefficient of variation as a function of mean genetic variability ($1 - \bar{Q}_1$ or $1 - \bar{H}_1$). The symbols and parameters are the same as in Fig. 3a.

of many individuals, and the cytoplasmic contribution from a male gamete to gene flow between subpopulations. Furthermore, random partition of replicated extranuclear genomes following cell division corresponds to random sampling of gametes within a subpopulation, and random sampling of gametes to a sort of group selection among subpopulations resulting in extinction and recolonization of subpopulations. The mathematical formulae for determining the variance of homozygosity in an incompletely isolated subdivided population have recently been developed by Golding & Strobeck (1983) and Takahata (1983*b*), independently. Their formulae allow us to study the deviation of variances of homozygosity within and between subpopulations as well as that for the entire population. Using these formulae, we can show that the deviation results entirely from that of between-subpopulation variance. In a nearly completely isolated population, the deviation occurs downward, since any pair of subpopulations are genetically dissimilar and therefore the distribution of homozygosity between subpopulations tends to be L-shaped. This indicates the small variance of homozygosity between subpopulations. The same situation occurs in the case of extranuclear genes, where the maternal effect is strong. The dissimilarity between individuals causes the reduction in \bar{V}_H and \bar{V}_Q . However, for the genetic variability to be intermediate, the population size should be small. A small population size counterbalances the effect of maternal inheritance and makes the distribution of identity probability between individuals U-shaped. Thus we can expect only a slight reduction in the variances of extranuclear identity probability in a small population.

On the other hand, when the extent of gene flow is intermediate there is an increase in variance of homozygosity in a subdivided population. This increase is due to intermediate between-subpopulation similarity. Moderate gene flow makes subpopulations either similar or dissimilar to each other in time and space and therefore the distribution of homozygosity tends to be U-shaped. A U-shaped distribution has a large variance, and it is expected in the case of extranuclear genes when β is large. In order for individuals to be genetically similar, the genetic contribution from a male gamete must be substantial or the population size must be small, or both.

The above tendencies are shown in Fig. 3, where \bar{V}_H and \bar{V}_Q are obtained by changing N and keeping the other parameters constant. As in the case of a subdivided population, a larger number of individuals in a population make the above deviations less significant. The larger the number, the smaller the relative probability of genetically identical individuals in a population. Thus the deviations in variance become invisible.

(iv) *Evolutionary distance*

The formulae (26) and (27) enable us to estimate the evolutionary distance of (22), taking into account not only intrapopulation variation but also within-cell variation. These formulae are depicted in Fig. 4 according to the type of data available. When data on nucleotide sequences of extranuclear gene or genome can be used, the formulae (28)–(30) allow us to estimate the distance (Fig. 4*a*). When observation is made by means of restriction enzymes, estimation of K_{nuc} is based

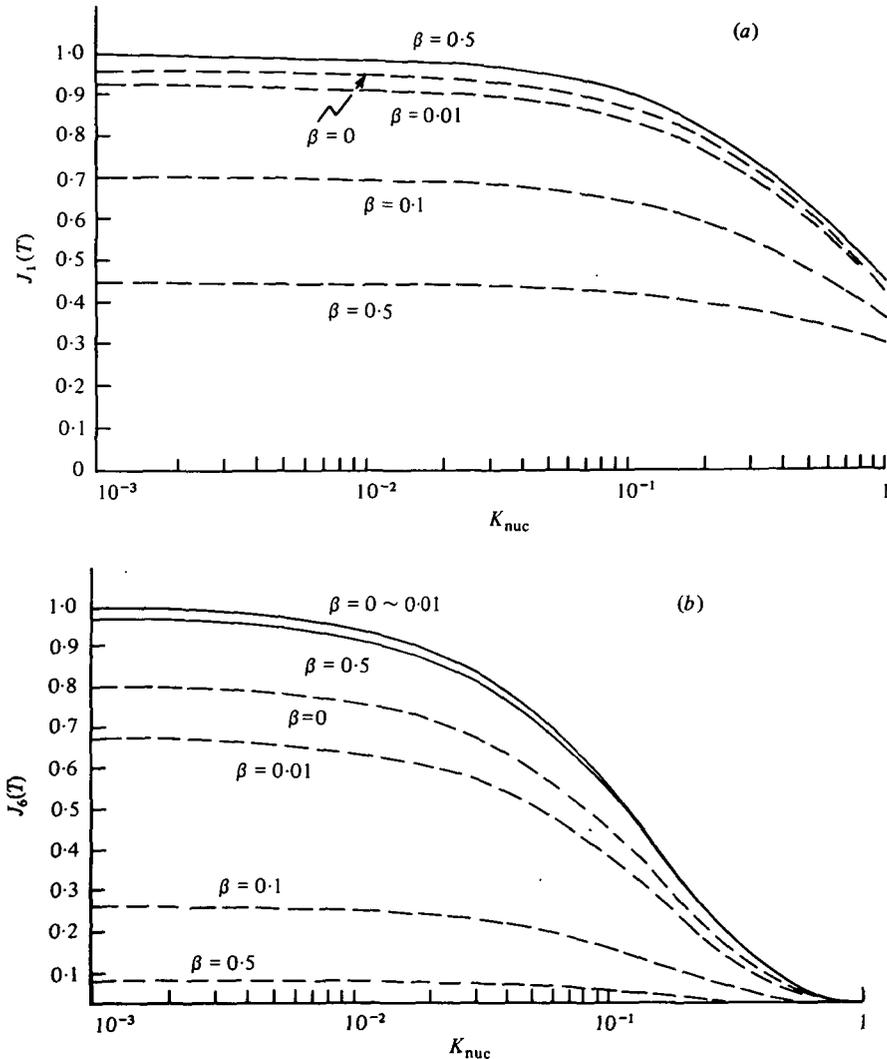


Fig. 4. (a) Relationship between $J_1(T)$ and $K_{\text{nuc}} = 2\mu T$ between two isolated populations with the divergence time of T generations. The values of parameters are taken as $\lambda = 20$, $n = 10^3$, $K = 4$ and $v = 10^{-9}$, and $J_1(0)$ are assumed to be equal to \bar{Q}_1 . The solid line represents the case of $N_f = N_m = 2500$, though the value of β varies from 0 to 0.5. Under the conditions the β dependence of $J_1(T)$ is not conspicuous. The broken lines represent the case of $N_f = N_m = 10^6$ and show the remarkable effect of β on $J_1(T)$. (b) Relationship between $J_6(T)$ and K_{nuc} . The implication of lines and values of parameters are the same as in Fig. 4a. Note that the declination of the lines begins at a relatively small value of K_{nuc} as compared with $J_1(T)$.

on the cleavage pattern of completely linked r (4–6) nucleotide sequences in the genome, so that the relationship among K_{nuc} , $J_r(t)$ and \bar{Q}_r is required (Fig. 4b). In both figures, the uppermost line approximately represents the relationship between K_{nuc} and $J_r(t)$ when intrapopulational variation is absent. Neglect of intrapopulational variation is usually made in estimating the distance from

nucleotide sequence data (Jukes & Cantor, 1969; Kimura & Ohta, 1972). This validity is guaranteed by the small mutation rate and population size, which are assumed here to be 2×10^{-8} per generation and 2500, respectively.

As N increases, intrapopulational variation becomes higher, and this trend is much enhanced when completely maternal inheritance is not warranted. The broken lines in Fig. 4 represent the β dependence. The left-most value of each line corresponds to \bar{Q}_r . If n is not small, the effect of β on \bar{Q}_r is remarkable and the population with $\beta = 0.5$ is much more polymorphic even than that with the doubled population size $2N$ but with $\beta = 0$. When the product of β , n , N_2 and v is not very small, not only intrapopulational variation but also within-cell variation is significantly high. Under the circumstances, disregard of the variation in estimating the distance provides a considerable overestimate of K_{nuc} , particularly when the two populations compared are closely related (Fig. 4; see also Nei & Li, 1979). It should be also noted that when there is high polymorphism the estimation of K_{nuc} becomes difficult because the K_{nuc} -dependence of $J_r(T)$ becomes weak. The steep declination of $J_r(T)$ occurs after K_{nuc} exceeds about $(10r\bar{Q}_r)^{-1}$, and thus an enormous amount of polymorphism entirely prevents an estimate of the distance. There may exist few organisms, however, that have such extensive polymorphism at the nucleotide level (Engels, 1981 and references therein). If the amount of polymorphism measured by $1 - \bar{Q}_1$ is a few per cent at most, the problem of polymorphism is not important in estimating the distance over 0.1.

The usual method for estimating K_{nuc} from comparison of homologous DNA sequences treats each nucleotide in each sequence independently and does not take into account the linkage relationship between nucleotides. Although the method is appropriate when estimating intermediate values of K_{nuc} , say 0.1–1 or a little greater, it does not give a reliable estimate of K_{nuc} less than 0.1. To estimate a small value of K_{nuc} accurately, it is desirable to use an alternative method, as suggested by a comparison of Figs. 4(a and b). This method takes into account the information about linkage among nucleotides instead of treating each independently. The more linked sites are analysed as a unit, the earlier the steep declination of $J_r(T)$ should be observed. This is simply because r -linked nucleotide sites can accumulate more mutations than a single site can. In this regard, the use of hexanucleotide restriction enzymes gives a more accurate estimate of small K_{nuc} than the use of tetranucleotide restriction enzymes, all else being equal. When nucleotide sequences are available, this method for arbitrary number r is applicable and can estimate a relatively small distance. Thus, the formulae (26) and (27) can be used for such an analysis and are potentially useful even though any degree of polymorphism is present.

The population genetics theory of extranuclear genomes was developed based on the neutral mutation hypothesis. Several evolutionary characteristics were summarized concerning the mean and variance of genetic variability, the variance of linkage disequilibrium and the evolutionary distance that would help us further study the evolution of extranuclear genomes.

I wish to express my gratitude to Dr Motoo Kimura for his interest and encouragement. Major revision of this paper was made while I was on leave at the University of Washington. I am

especially grateful to Drs Monty Slatkin, Joe Felsenstein, Steve Palumbi, Alan Harper, Mark Kirkpatrick, Bruce Walsh and Betty and Gordon Orians for their help and encouragement. Thanks are due to anonymous referees for their useful suggestions and to the editor for his patience. This work is supported in part by Grant-in-Aid 57120009 from the Ministry of Education, Science and Culture of Japan.

APPENDIX

We give the transition matrix, M , of identity probabilities at a single site due to mutation and within-generation drift. We note that this process takes place in each individual independently, i.e.

$$E\{\xi_i(l) \xi_j(l)\} = \frac{1}{n} x_i (\delta_{ij} - x_j) \delta_{il}$$

in which δ_{ij} and δ_{il} are the delta functions and δ_{il} indicates that the above expectation is 0 for two different individuals ($l \neq l'$). The formula (13') is equivalent to that for the gene frequency change in the finite island model with no migration. As the derivation of M is the same as that in the island model (Takahata, 1983*b*), we give only the result opposite.

REFERENCES

- ANDERSON, S., BANKIER, A. T., BARRELL, B. G., DE BRUIJN, M. H. L., COULSON, A. R., DROUIN, J., EPERON, I. C., NIERLICH, D. P., ROE, B. A., SANGER, F., SCHREIER, P. H., SMITH, A. J. H., STADEN, R. & YOUNG, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
- AVISE, J. C., LANSMAN, R. A. & SHADE, R. O. (1979). 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.
- BEALE, G. H. & KNOWLES, J. K. C. (1978). *Extranuclear Genetics*. London: Edward Arnold.
- BIBB, M. J., VAN ETTEN, R. A., WRIGHT, C. T., WALBERG, M. W. & CLAYTON, D. A. (1981). Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26**, 167–180.
- BIRKY, C. W., JR. (1978). Transmission genetics of mitochondria and chloroplasts. *Annual Review of Genetics* **12**, 417–512.
- BIRKY, C. W., JR, MARUYAMA, T. & FUERST, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**, 513–527.
- BOGENHAGEN, D. & CLAYTON, D. A. (1974). The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. *Journal of Biological Chemistry* **249**, 7991–7995.
- CHAPMAN, R. W., STEPHENS, J. C., LANSMAN, R. A. & AVISE, J. C. (1982). Models of mitochondrial DNA transmission genetics and evolution in higher eucaryotes. *Genetical Research* **40**, 41–57.
- DAWID, I. B. & BLACKLER, A. W. (1972). Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. *Developmental Biology* **29**, 152–161.
- DUJON, B., SLONIMSKI, P. P. & WEILL, L. (1974). Mitochondrial genetics. IX. A model for recombination and segregation of mitochondrial genomes in *Saccharomyces cerevisiae*. *Genetics* **78**, 415–437.
- Engels, W. R. (1981). Estimating genetic divergence and genetic variability with restriction endonucleases. *Proceedings of the National Academy of Sciences, U.S.A.* **78**, 6329–6333.
- GILES, R. E., BLANC, H., CANN, H. M. & WALLACE, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 6715–6719.
- GILLHAM, N. W. (1978). *Organelle Heredity*. New York: Raven Press.
- GOLDING, G. B. & STROBECK, C. (1982). The distribution of nucleotide site differences between two finite sequences. *Theoretical Population Biology* **22**, 96–107.
- GOLDING, G. B. & STROBECK, C. (1983). Variance and covariance of homozygosity in a structured population. *Genetics* **104**, 513–529.
- GRIFFITHS, R. C. (1981). Transient distribution of the number of segregating sites in a neutral infinite-sites model with no recombination. *Journal of Applied Probability* **18**, 42–51.
- GROSS, N. J., GETZ, G. S. & RABINOWITZ, M. (1969). Apparent turnover of mitochondrial deoxyribonucleic acid and mitochondrial phospholipids in the tissues of the rat. *Journal of Biological Chemistry* **244**, 1552–1562.
- HAUSWIRTH, W. W. & LAIPIS, P. J. (1982). Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 4686–4690.
- JUKES, T. H. & CANTOR, C. H. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism* (ed. H. N. Munro), pp. 21–123. New York: Academic Press.
- KIMURA, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research* **11**, 247–269.
- KIMURA, M. & OHTA, T. (1972). On the stochastic model for estimation of mutational distance between homologous proteins. *Journal of Molecular Evolution* **2**, 87–90.
- KONDO, S. (1977). Evolutionary considerations on DNA repair and mutagenesis. In *Molecular Evolution and Polymorphism (Proceedings of the Second Taniguchi International Symposium on Biophysics)*, (ed. Motoo Kimura) pp. 313–331.
- LANSMAN, R. A., AVISE, J. C. & HUETTEL, M. D. (1983). Critical experimental test of the possibility of 'paternal leakage' of mitochondrial DNA. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 1969–1971.
- NASS, M. K. (1969). Mitochondrial DNA: advances, problems, and goals. *Science* **165**, 25–35.

- NEI, M. & LI, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, U.S.A.* **76**, 5269–5273.
- OHTA, T. (1980). Two-locus problems in transmission genetics of mitochondria and chloroplasts. *Genetics* **96**, 543–555.
- RABINOWITZ, M. & SWIFT, H. (1970). Mitochondrial nucleic acids and their relation to the biogenesis of mitochondria. *Physiological Reviews* **50**, 376–427.
- REILLY, J. G. & THOMAS, C. A., JR (1980). Length polymorphisms, restriction site variation, and maternal inheritance of mitochondrial DNA of *Drosophila melanogaster*. *Plasmid* **3**, 109–115.
- STEWART, F. M. (1976). Variability in the amount of heterozygosity maintained by neutral mutations. *Theoretical Population Biology* **9**, 188–201.
- TAKAHATA, N. (1982). Linkage disequilibrium, genetic distance and evolutionary distance under a general model of linked genes or a part of the genome. *Genetical Research* **39**, 63–77.
- TAKAHATA, N. (1983*a*). Linkage disequilibrium of extranuclear genes under neutral mutations and random genetic drift. *Theoretical Population Biology* **24**, 1–21.
- TAKAHATA, N. (1983*b*). Gene identity and genetic differentiation of populations in the finite island model. *Genetics* **104**, 497–512.
- TAKAHATA, N. & MARUYAMA, T. (1981). A mathematical model of extranuclear genes and the genetic variability maintained in a finite population. *Genetical Research* **37**, 291–302.
- TAKAHATA, N. & SLATKIN, M. (1983). Evolutionary dynamics of extranuclear genes. *Genetical Research* **42**, 257–262.
- UPHOLT, W. B. (1977). Estimation of DNA sequence divergence from comparison of restriction endonuclease digests. *Nucleic Acids Research* **4**, 1257–1265.
- WALLACE, D. C. (1982). Structure and evolution of organelle genomes. *Microbiological Reviews* **46**, 208–240.