Genet. Res., Camb. (1984), **44**, *pp.* 109–116
Printed in Great Britain

A model of extranuclear genomes and the substitution rate under within-generation selection

By NAOYUKI TAKAHATA

Department of Zoology, NJ-15, University of Washington, Seattle, WA 98195 U.S.A.

(*Received 4 January 1984 and in revised form 13 March 1984*)

SUMMARY

A single locus model of extranuclear genomes is developed under the assumption of the complete action of within-generation drift which is caused by random transmission of multiple copy genomes during cell division in a generation. Within-generation drift segregates different copy genomes in a cell into different cells, resulting in homoplasmic cells. Under some conditions, the present model reduces to that for haploid nuclear genomes. A point overlooked in previous models is that the multiplicity also admits of the possibility of selection occurring within a cell or between cells in an individual (within-generation selection). If there is selection mediated by, for instance, differential proliferation of genomes, then a haploid model no longer explains the dynamics of extranuclear genomes. Rather a model analogous to biased gene conversion at a single locus (Nagylaki, 1983; Walsh, 1983) is more appropriate. An application of this model to either the fixation probability or substitution rate of new mutations shows that strictly maternal inheritance does not allow the fullest use of mutations, as it obscures the effect of within-generation selection. But if there is appreciable paternal contribution, within-generation selection could be a strong evolutionary force to which nuclear genomes are never exposed.

In the early 1960s organelle or extranuclear DNA was discovered in chicken mitochondria (Nass & Nass, 1963) and *Chlamydomonas* chloroplasts (Ris & Plaut, 1962). Molecular biological analysis of extranuclear genomes have since revealed unique and unexpected features in their gene organization, genetic code, and gene expression (Rabinowitz & Swift, 1970; Anderson *et al.* 1981; Bibb *et al.* 1981), rekindling interest in their origin and evolution (Margulis, 1981; Wallace, 1982). Sequence comparison of homologous DNA between species has shown also that extranuclear genomes have been evolving with their own tempo and mode (Brown *et al.* 1982). A high substitution rate of mitochondrial genomes in mammals is one of such characteristics.

To understand the evolutionary pattern and to understand the evolutionary significance of their multiplicy in a cell (Beale & Knowles, 1978; Gillham, 1978), unequal cytoplasmic contribution in both sexes (Ursprung & Schabtach, 1965; Lansman *et al.* 1983), and mode of transmission at cell division (Tilney-Basset,

1970; Gillham *et al.* 1974; Birky, 1978, 1983; Thrailkill *et al.* 1980), a theory of extranuclear population genetics combined with more traditional population genetics based on Mendelism is necessary. The cytoplasmic heredity and drift occurring within a generation can cause unique evolutionary forces.

Takahata & Slatkin (1983) have shown that extranuclear genomes differ in their potential exposure to 'individual' selection, demonstrating that the multiplicity of extranuclear genomes strongly affects the evolution of individual mutations if the multiplicity does not result in the dilution of the selection experienced by each copy of the mutation. The conclusion was drawn based on their simulation model in which they ignored selection that may occur within generations. Here I propose a simple model of extranuclear genomes and use it to discuss the substitution rate when selection mediated by differential proliferation and/or functional differences among genomes in a cell or between cells is present. Although little is empirically known about this type of selection and its molecular mechanism, it is suggested that even a small change in extranuclear genomes could have a considerable effect on the cell phenotype (Kearsey & Craig, 1981; Blanc *et al.* 1981). I will show that selection occurring within an individual (referred to as within-generation selection), coupled with the transmission mode, can change the substitution rate to a large extent.

The model developed here assumes two states at a particular site in the extranuclear genome, one of which is henceforth called normal (M) and the other mutant (M'), and random transmission of extranuclear copy genomes at mitosis. One important consequence of random transmission is to homogenize a cell cytogenetically and to produce homoplasmic cells. We call this mechanism within-generation drift (Takahata & Slatkin, 1983) and assume that the cells pass through a sufficient number of mitotic divisions before gametogenesis to insure that complete fixation occurs within each cell. Within-generation drift acts in an analogous manner to random sampling drift in a population (Birky, 1983), fixing a genotype in a cell by segregating different genotypes into different cells. Thus we assume that all gametes produced by a heteroplasmic zygote are exclusively homoplasmic, although they may differ to each other.

Suppose that a population is dioecious, being composed of N_f females and N_m males (total breeding number $N_T = N_f + N_m$). The life cycle of this population is as follows:

$$\underset{}{\text{Zygotes}} \xrightarrow[\text{rounds of mitosis}]{\text{Multiple}} \underset{x_t(x_t^*)}{\text{Adults}} \xrightarrow{\text{Gametogenesis}} \text{Gametes} \xrightarrow[\text{sampling of gametes}]{\text{Random}} \underset{x_t'(x_t^{*\prime})}{\text{Gametes}} \xrightarrow{\text{Fertilization}} \text{Zygotes}$$

Here, x_t (x_t^*) is the frequency of M' in the adult female (male) population in the tth generation, and the frequency of M' in the randomly sampled gametes is denoted by x_t' ($x_t^{*\prime}$). Generations are discrete and nonoverlapping and are measured by successive rounds of meiosis (gametogenesis). Within a single generation, many rounds of mitotic division occur. There are n copies of extranuclear genome in a cell. The number n may vary drastically during development. If n varies, an appropriate number of copy genomes relevant to within-generation drift is the effective number n_e; the harmonic mean of n taken over all cell divisions in a

generation (Wright, 1931). The effective number n_e depends also on the mode of replication and transmission. If some molecules are replicated more often than others, and if there is no mechanism allowing the perfect transmission (Chapman *et al.* 1982), n_e will be smaller (e.g. Crow & Kimura, 1970 pp. 109–110).

For simplicity, we assume that mutation occurs only at the single-cell zygote stage. We denote by p the initial frequency of M' in that zygote and start to number generations when a new mutation appeared. We assume that the fixation probability of M' within a generation is given by $u(p)$. For the moment, we assume merely $u(0) = 0$ and $u(1) = 1$. The zygote may produce both types of gametes M and M', in proportion to $1-u(p)$ and $u(p)$. If selection operates among these different homoplasmic cells within a single individual, their frequencies may change, but at first we exclude this kind of selection. We also assume that the viability and fertility of individuals that carry mutation do not differ from those of normal, i.e. the *neutrality at the individual level*. Thus, when the zygote carrying a new mutation is female (male), the frequency of M' becomes $u(p)/N_f$ ($u(p)/N_m$) in females (males). We let $x_0 = u(p)/N_f$ ($x_0 = 0$) and $x_0^* = 0$ ($x_0^* = u(p)/N_m$) for female (male) mutations where the suscript 0 indicates the initial generation. With these frequencies, random sampling of gametes takes place independently from each gamete pool. Since x_0' ($x_0^{*\prime}$) is the frequency of M' in females (males) following random sampling of gametes, $N_f x_0'$ ($N_m x_0^{*\prime}$) is a random variable, following a binomial distribution with parameters $u(p)/N_f$ ($u(p)/N_m$) and N_f(N_m).

Let x_t (x_t^*) be the frequency of M' in the female (male) adult population in the tth generation. We first consider the change of x_t and x_t^* due to random sampling of gametes. This process is usually described by a Markov chain which may be approximated by a diffusion process under an appropriate scaling of parameters. A caution in our model is that we need take into account the fact that sampling takes place independently of sex. Here we use two independent random variables to express this situation and make use of a diffusion approximation. The random variables, B_1 and B_2, introduced are such that $E(B_1) = E(B_2) = E(B_1 B_2) = 0$, $E(B_1^2) = x(1-x)/N_f$ and $E(B_2^2) = x(1-x)/N_m$ where $E(\ )$ stands for the expectation conditioned on $x_t = x_t^* = x$. By using B_1 and B_2, we have

$$x_t' = x_t + B_1 \text{ and } x_t^{*\prime} = x_t^* + B_2. \tag{1}$$

In general, $x_t' = x_t^{*\prime}$ does not hold true.

Table 1 summarizes the possible combinations of different gametes and the gamete frequencies as a result of random mating. We define the fraction of copy genomes transmitted by an egg by α ($\beta = 1-\alpha$), to take account of the unequality in inheritance. The heteroplasmic zygote fertilized by an egg with M' genome and a sperm with M genome forms both M and M' homoplasmic cells with relative frequencies $u(\alpha)$ and $1-u(\alpha)$. This follows directly since α is the fraction of M' copy genomes in the resulting single-cell zygote. Subsequent within-generation drift fixes M' in $u(\alpha)$ of the cells, while M is fixed in the other cells. Thus the frequency of M' in the adult population becomes

$$x_{t+1} = x_t' x_t^{*\prime} + u(\beta)\, x_t^{*\prime}(1-x_t') + u(\alpha)\, x_t'(1-x_t^{*\prime}) \tag{2}$$

and $x_{t+1} = x_{t+1}^*$ because of no preferential assignment of individuals to sex.

https://doi.org/10.1017/S0016672300026288 Published online by Cambridge University Press

Table 1. *Genotypes and their frequencies in a random mating population as a result of within-generation drift*

		Female Gametes	
		M' x	M $1-x$
Male Gametes	M' x^*	M' xx^*	M' $u(\beta)x^*(1-x)$
			M $(1-u(\beta))x^*(1-x)$
	M $1-x^*$	M' $u(\alpha)x(1-x^*)$	M $(1-x)(1-x^*)$
		M $(1-u(\alpha))x(1-x^*)$	

$u(\alpha)$ ($u(\beta)$) is the fixation probablity of M' type when transmitted from a female (male) gamete with the fraction α (β) relative to a male (female) gamete. $\alpha+\beta=1$ and the prime indicating the frequencies after sampling is suppressed.

We define

$$\delta = u(\alpha)+u(\beta)-1 \tag{3}$$

and assume that δ is so small that we can ignore the higher order terms. Substituting (1) for (2) and taking the conditional expectation, we obtain

$$E(x_{t+1}) = x+\delta x(1-x) \tag{4}$$

$$E(x_{t+1}^2) = x^2+(u(\alpha)^2/N_f+u(\beta)^2/N_m)\,x(1-x) \tag{5}$$

where the terms of δ^2, δ/N_f, δ/N_m and $1/(N_fN_m)$ are all neglected. It is natural from (5) to define the effective population size as

$$N_e = (u(\alpha)^2/N_f+u(\beta)^2/N_m)^{-1}. \tag{6}$$

We note that if $u(x)=x$, i.e. no preferential fixation within a generation, $\delta=0$ and N_e is the same as that defined for neutral mutations (Takahata & Maruyama, 1981).

From (4) and (5), we obtain the diffusion operator

$$L = \delta x(1-x)\frac{\partial}{\partial x}+\frac{1}{2N_e}x(1-x)\frac{\partial^2}{\partial x^2} \tag{7}$$

which is essentially the same as that for a nuclear locus with genic selection (Crow & Kimura, 1970 pp. 396) or biased gene conversion in a finite population (Nagylaki, 1983; Walsh, 1983). In particular, a model of biased gene conversion is more analogous to ours since it postulates the nonreciprocal transfer of genetic information. From (7) and the standard diffusion analysis, it follows that the fixation probability is given by

$$F(x) = \frac{1-\exp(-2N_e\delta x)}{1-\exp(-2N_e\delta)}.$$

https://doi.org/10.1017/S0016672300026288 Published online by Cambridge University Press

When inheritance is completely maternal ($\alpha = 1$), $u(\alpha)+u(\beta) = u(1)+u(0) = 1$ and therefore $\delta = 0$ from (3) and $N_e = N_f$ from (6). In this case, there is no advantage or disadvantage of mutation M' in its substitutional process, i.e. $F(x) = x$, the same as for a neutral mutation.

The above analysis computes the fixation probability for a constant value x in the first generation ($t = 1$). To be complete we need recall that random sampling of gametes following the initial zygotic mutation gives a distribution of starting values. More precisely, $x_1 = u(\alpha)\, x_0'$ ($x_1 = u(\beta)\, x_0^{*\prime}$) for female (male) mutations and $N_f x_0'$($N_m x_0^{*\prime}$) follows a binomial distribution with parameters $u(\alpha)/N_f$($u(\beta)/N_m$) and N_f(N_m) as mentioned earlier. Finally, assuming equal mutation rate in both sexes, N_f/N_T and N_m/N_T are probabilities that a new mutation occurs in a female or male, respectively. Thus to calculate the fixation probability, we need compute

$$E_b(F(x)) = \{N_f\, E_b(F(u(\alpha)x_0')) + N_m\, E_b(F(u(\beta)\, x_0^{*\prime}))\}/N_T$$

where x is the frequency of M' in the adult population at $t = 1$ and E_b stands for taking the expectation with respect to the binomial distribution. The above computation yields a linear combination of binomial generating functions. The approximate formula is

$$E_b(F(x)) \approx u(p)\Big\{1-\frac{N_f}{N_T}\exp(-2N_e\,\delta u(\alpha)/N_f) - \frac{N_m}{N_T}\exp(-2N_e\,\delta u(\beta)/N_m)\Big\}/(1-\exp(-2N_e\delta)). \quad (8)$$

For small values of $2N_e\delta u(\alpha)/N_f$ and $2N_e\,\delta u(\beta)/N_m$, it is further approximated by

$$E_b(F(x)) \approx 2N_e\,\delta u(p)/(N_T(1-\exp(-2N_e\,\delta)).$$

The substitution rate is obtained by multiplying the total number of mutations occurring in a generation by $E_b(F(x))$ with $p = 1/n$

$$\begin{aligned} k &\approx nu(1/n)v2N_e\,\delta/(1-\exp(-2N_e\,\delta)) \\ &= nvu(1/n) \quad \text{if} \quad \delta = 0 \end{aligned} \quad (9)$$

in which v is the mutation rate per site of a copy genome per generation.

Now, let us specify the fixation probability within generations. If genomes M and M' have equivalent replication rates, $u(p) = p$ and from (3) and (9), $\delta = 0$ and $k = v$. An interesting case occurs when $u(p) \neq p$ (for whatever reason). A simple model for differential proliferation may be that M' replicates by a factor r more often than M. Then $u(p) = (1-\exp(-2n_e rp))/(1-\exp(-2n_e r))$, and for the case of $\delta = 0$ (or $\alpha = 1$), $k/v \approx 2n_e r/(1-\exp(-2n_e r))$. This implies that within-generation selection contributes mainly to the establishment of mutation within an individual, but not much to the establishment in the entire population. In other words, within-generation selection increases the fixation probability $u(p)$ only in the first generation, and without recombination between sexes there is no opportunity that this type of selection can operate in later generations. In order for the effect of selection to be significant, there must be genetic variation within a cell.

A more profound effect of differential proliferation on the substitution rate can be found when paternal contribution is significant. In this case, $\delta \neq 0$ because of $r \neq 0$ and $\alpha \neq 1$, and therefore the factor $2N_e\delta/(1-\exp(-2N_e\delta))$ in (9) can be much larger or smaller than 1. In particular, in a large population $N_e|\delta|$ can be large even for small $n_e r$ ($\delta \approx 2n_e r\alpha\beta$). This is because with recombination between sexes, heteroplasmic cells are recurrently produced until a mutant fixes in the entire population. Thus, it is possible that the coupled effects of within-generation drift, differential proliferation and mode of inheritance are profound, altering to a large extent the mutant substitution rate in extranuclear genomes. In this context, strictly maternal inheritance does not allow the fullest use of mutations arising in a population and can overshadow evolutionary forces occurring within an individual.

The model and analysis studied here are critically dependent on the assumption of complete homogenization of a cell within a generation. This situation is ensured by a relatively small value of n_e and/or more multiple rounds of cell divisions λ. In addition, weak paternal leakage is favourable. Under these circumstances, the time required until cell homogenization is very short compared with the time during which a significant genetic change in the whole population can occur. Kimura (1957) is the first who formulated within-generation drift, but in a different context (random assortment of chromosome subunits in cell division in which each chromosome consists of n subunits and the duplicated subunits are randomly sorted into two groups of n subunits to form the daughter chromosomes). He obtained the exact solution for the probability with which a chromosome containing only mutant subunits first appears in a given generation, and found the time to maximize this probability. It is about n_e when $n_e = 2n-1$ and $p = 1/n$. On the other hand, the average time required for mutation to become fixed or extinct in a cell is also about n_e when $p = 0{\cdot}5$ and only several mitotic divisions when $p = 1/n$ or $1-1/n$ (cf. Crow & Kimura, 1970 pp. 431–432). Takahata & Maruyama (1981) show that the condition for the absence of heteroplasmic gametes in an equilibrium population is $\beta n_e \ll \lambda$. For the implication of this condition on linkage disequilibrium and for a simple model of nuclear and extranuclear gene interaction counting on this condition, readers may refer to Takahata (1983*a*, *b*) and Takahata & Slatkin (1984).

An advantage of the present model is to be able to easily incorporate evolutionary forces other than considered so far. For instance, if there is some functional differences among gametes produced by a heteroplasmic zygote, there could be competition among these gametes before gamete pool formation. This changes the relative frequency of M' in each individual, u to u'. When M' is by a factor s more functional than M, u' would be given by $(1+s)u/(1+su)$ as in haploid or genic selection. Then we can see that this within-generation selection plays a role in the substitutional process similar to differential proliferation in replication.

In conclusion, there are many potential evolutionary forces to which extranuclear genomes are uniquely exposed. Differential replication and functional differences occurring within an individual are such candidates, changing the evolutionary dynamics of extranuclear genomes considerably with the help of within-generation drift and recombination between sexes, if any. Although very little is empirically

known about such mechanisms, the model proposed here is simple and potentially useful in analysing the complicated forces concerning extranuclear genomic evolution.

Thanks are due to the University of Washington for giving me unforgettable experiences. I am especially grateful to M. Slatkin and J. Felsenstein for their hospitality and interest on this subject. I am much indebted to J. B. Walsh whose comments were most useful to remove several obscurities in the early version of this paper. This work is supported in part by NSF grant no. DEB-8120580. Offprint requests to: National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan.

ERRATUM

In Takahata & Slatkin (*Genetical Research* 1983, **42**, 257–265), two sentences should be read as:

but, in contrast, the configuration of the genome and the mode of transmission both make a large difference (on page 257, lines 11–12 up)

Contribution no. 1528 from the National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan (on page 264, above REFERENCES).

REFERENCES

ANDERSON, S., BANKIER, A. T., BARRELL, B. G., DE BRUIJIN, 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.

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

BIBB, M. J., VAN ETTEN, R. A., WRIGHT, C. T., WABERG, M. W. & CLAYTON, D. A. (1981). Sequence and gene organization of mouse mitochondial DNA. *Cell* **26**, 167–180.

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

BIRKY, C. W., JR. (1983). Relaxed cellular controls and organelle heredity. *Science* **222**, 468–475.

BLANC, H., WRIGHT, C. T., BIBB, M. J., WALLACE, D. C. & CLAYTON, D. A. (1981). Mitochondrial DNA of chloramphenicol-resistant mouse cells contains a single nucleotide change in the region encoding the 3′ end of the large ribosomal RNA. *Proceedings of the National Academy of Science, U.S.A.* **78**, 3789–3793.

BROWN, W. M., PRAGER, E. M., WANG, A. & WILSON, A. C. (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. *Journal of Molecular Evolution* **18**, 225–239.

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.

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory*. New York, Evanston and London: Harper & Row.

GILLHAM, N. W. (1978). *Organelle Heredity*. New York: Raven Press.

GILLHAM, N. W., BOYNTON, J. E. & LEE, R. W. (1974). Segregation and recombination of non-mendelian genes in *Chlamydomonas*. *Genetics* **78**, 439–457.

KEARSEY, S. E. & CRAIG, I. W. (1981). Altered ribosomal RNA genes in mitochondria from mammalian cells with chloramphenicol resistance. *Nature* **290**, 607–608.

KIMURA, M. (1957). Some problems of stochastic processes in genetics. *Annual Mathematical Statistics* **28**, 882–901.

LANSMAN, R. A., AVISE, J. C. & HUETTEL, M. D. (1983). Critical experimental test of the possibility of 'paternal leakage' of mitochondial DNA. *Proceedings of the National Academy of Science, U.S.A.* **80**, 1969–1971.

MARGULIS, L. (1981). *Symbiosis in Cell Evolution*. San Francisco: W. H. Freeman.

NAGYLAKI, T. (1983). Evolution of a finite population under gene conversion. *Proceedings of the National Academy of Science U.S.A.* **80**, 6278–6281.

NASS, M. M. K. & NASS, S. (1963). Intramitochondrial fibers with DNA characteristics. I. Fixation and electron staining reactions. *Journal of Cell Biology* **19**, 593–611.

RABINOWITZ, M. & SWIFT, H. (1970). Mitochondrial nucleic acids and their relation to the biogenesis of mitochondria. *Physiological Reviews* **50**, 376–427.

RIS, H. & PLAUT, W. (1962). Ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas*. *Journal of Cell Biology* **13**, 383–391.

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. (1983*a*). Linkage disequilibrium of extranuclear genes under neutral mutations and random genetic drift. *Theoretical Population Biology* **24**, 1–24.

TAKAHATA, N. (1983*b*). Population genetics of extranuclear genomes under the neutral mutation hypothesis. *Genetical Research* **42**, 235–255.

TAKAHATA, N. & SLATKIN, M. (1983). Evolutionary dynamics of extranulcear genes. *Genetical Research* **42**, 257–265.

TAKAHATA, N. & SLATKIN, M. (1984). Mitochondrial gene flow. *Proceedings of the National Academy of Science U.S.A.* (In the Press.)

TILNEY-BASSET, R. A. E. (1970). The control of plastid inheritance in pelargonium. *Genetical Research* **16**, 46–61.

THRAILKILL, K. M., BIRKY, C. W. JR., LUCKERMANN, G. & WOLF, K. (1980). Intracellular population genetics: evidence for random drift of mitochondrial allele frequencies in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Genetics* **96**, 237–262.

URSPRUNG, H. & SCHABTACH, E. (1965). Fertilization in tunicate: loss of the paternal mitochondrion prior to sperm entry. *Journal of Experimental Zoology* **159**, 379–384.

WALLACE, D. C. (1982). Structure and evolution of organelle genomes. *Microbiological Reviews* **46**, 208–240.

WALSH, J. B. (1983). Role of biased gene conversion in one-locus neutral theory and genome evolution. *Genetics* **105**, 461–468.

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