Genet. Res., Camb. (1985), **45**, *pp.* 179–194
With 3 text-figures
Printed in Great Britain

Introgression of extranuclear genomes in finite populations: nucleo-cytoplasmic incompatibility*

By NAOYUKI TAKAHATA

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

(*Received 9 May 1984 and in revised form 7 January 1985*)

SUMMARY

A 'two locus two allele' model is developed with special reference to the introgression of extranuclear genomes between two species of finite size. The model assumes that one locus, coded by a nuclear genome, causes the reproductive barrier while the other locus, coded by an extranuclear genome, causes nucleo-cytoplasmic incompatibility in particular genotypes. To fully study this model, simulations are conducted, and a diffusion equation is derived when introgression or extranuclear gene flow occurs in one direction. It is shown that although selection against the nuclear genome can reduce the levels of extranuclear gene flow and retard the introgression process, the dynamics are very similar to those without such selection. In contrast, the nucleo-cytoplasmic incompatibility directly affects the dynamics of introgressing extranuclear genomes: in large populations the ability of incompatibility to overcome extranuclear gene flow is conspicuous, but in small populations it is overshadowed by random sampling drift. Paternal leakage of extranuclear genomes, if present, is of evolutionary importance only when the male's migration rate is much larger than the female's. When the sizes of two populations are unequal, the introgression is most likely to occur from the larger population to the smaller one in the absence of mating preferences of backcross progeny. Recent observations on interspecific mitochondrial transfer in various species do not support the ubiquitousness of nucleo-cytoplasmic incompatibility as an efficient reproductive barrier.

1. INTRODUCTION

The evolution of reproductive isolation is of fundamental interest in evolutionary biology. Isolation could develop by fixation of different sets of interacting genes (Dobzhansky, 1937; Muller, 1942), or chromosomal rearrangments (White, 1978), or repetitive DNAs (Dover & Flavell, 1982) in different populations. Although the underlying evolutionary mechanisms of developing reproductive isolation are not fully understood, no doubt most reproductive isolation results from changes in the nuclear genome. For example, Zouros (1981) determined the number and chromosomal locations of genes which control the mating behavior of *D. mojavensis*

* The early version of this paper was prepared while I was on leave at Center for Demographic and Population Genetics, University of Texas, Houston, Texas 77225 U.S.A. This is contribution no. 1589 from the National Institute of Genetics.

and *D. arizonensis*. Genes that cause hybrid sterility are found in the nuclear genomes of rice (Oka, 1978), wheat (Tsunewaki & Kihara, 1962), *D. pseudobscura* (Prakash, 1972), and many other species.

However, the incompatibility of crosses between local populations or species can sometimes be traced to nuclear and cytoplasmic (extranuclear) interaction. Laven (1959) discusses the possibility of speciation by cytoplasmic isolation in the *Culex pipiens*-complex; he claims that the dualism of genetic systems should not be overlooked, and that the existence of a dual, or in some plants a triple, system of genetic determinants must be taken into account in evolutionary theory. A number of cases of male sterility are attributed to incompatibility between nuclear and cytoplasmic genes (Grun & Aurbertin, 1965). An enumeration of cases of cytoplasmic pollen sterility is found in Jones (1951). Caspari, Watson & Smith (1966) model the effect of pollen sterility on gene flow (see also Wright, 1969, pp. 163–168), although they find no evidence that the nuclear and cytoplasmic pollen incompatibility by itself can induce permanent reproductive isolation.

Recent molecular genetics analysis sheds light on the interaction of nuclear and extranuclear genomes from different points of view. Mitochondrial and chloroplast DNAs are 'promiscuous' not only between cellular organelles and nucleus (Stern & Lonsdale, 1982, Ellis, 1982; Farrelly & Butow, 1983) but also between organelles in different populations or species (Powell, 1983; Ferris *et al.* 1983; Yonekawa, Tagashira & Moriwaki, 1983; Palmer, Shields & Cohen, 1983; Weissinger *et al.* 1983; Spolsky & Uzzell, 1984). Thus extranuclear DNAs might move and be integrated into the host nucleus, and moreover interspecific transfer could occur rather frequently in the course of evolution. Extranuclear DNAs are remarkably flexible both in their gene organization and in their introgression between different species.

Takahata & Slatkin (1984) studied a simple model that accounts for mithochondrial gene flow over species boundaries. They considered two populations X and Y and assumed that population X, of infinite size, receives only female immigrants from population Y. As the genotypes in population Y are assumed not to change during the course of evolution, their model differs from Wright's (1931) island model only in that migration is sex-dependent and that the effect of random sampling of gametes is neglected even in population X (the island population). From their deterministic analysis it was found that the rate of replacement of extranuclear genomes is mainly determined by the immigration rate g if selection against the foreign extranuclear genome is very weak. Since g must be very small for different species, we can claim that random sampling drift may have played an important role in the introgression process. A purpose of this note is therefore to incorporate the fact that populations consist of finite numbers of males and females. We also extend their model to investigate the effect of nucleo-cytoplasmic incompatibility [in Wright's (1969) terminology] on the introgression of extranuclear genomes. We will approach this problem by both simulation and analytical studies.

2. THE MODEL

We consider two populations X and Y of a diploid organism which differ in a pair of alleles or haplotypes of completely linked loci, A and a, and in their extranuclear genomes, M and m. We assume that there has been geographic isolation between these populations for a sufficient time to establish a fixed genotype in each population. Denote these genotypes by $AAMM$ in population X and $aamm$ in population Y. The reason why we use double letters to designate extranuclear genotypes will be made clear in the following. Suppose that it is possible for the two populations to exchange individiuals because of recent population expansion or change of their geographic ranges. Suppose also that reproductive isolation between the two populations is still incomplete and therefore the hybrids are not completely sterile (otherwise no gene flow exists even for extranuclear genomes). Our problem is to determine under what conditions the extranuclear genome in one population is replaced by the other and how fast the process proceeds, if it occurs.

We assume that A and a are responsible for reproductive isolation between the two populations by lowering the fitness of heterozygotes (Slatkin, 1982). We designate by $1-s$ the reduced fitness independent of extranuclear genotypes. Under this condition, we shall study the influence of nucleo-cytoplasmic incompatibility on extranuclear gene flow. We model incomplete maternal inheritance for generality and denote by β the proportion of extranuclear genomes paternally transmitted into a single cell zygote. If $\beta = 0$ and no heteroplasmic zygotes are produced, our model of extranuclear genomes is equivalent to that of haploid genomes. When a zygote is heteroplasmic due to paternal extranuclear genomes, we assume complete fixation of an extranuclear genotype within a cell before gametogenesis. This fixation is caused by random drift at mitosis occurring in a generation (within-generation drift). With this assumption, we can treat the problem in a simplified way and ignore the multiplicity of copy genomes in a cell (Takahata, 1984, and references therein for mechanisms of within-generation drift). When $AAMM$ males and $aamm$ females mate, the genotype of their offspring is therefore $AaMm$ although the relative frequency of M in their offspring is dependent on β. By nucleo-cytoplasmic incompatibility, we imply that individuals having a combination of Am, or aM, or both are less fit. To specify the selection regime, we assume that the extent of the fitness reduction of incompatible individuals is in proportion to the amount of incompatible gene products of Am, or aM, or both, relative to the total. Thus the fitness of individuals produced by AM male gametes and Am female gametes is reduced by αt ($\alpha = 1-\beta$), that produced by AM male gametes and am female gametes is reduced by $t/2$ and so on. We note that the incompatibility thus defined is nonreciprocal if $\beta \neq 0{\cdot}5$. For instance, if $\beta = 0$, a cross between Am male gametes and AM female gametes gives normal offspring, but the reciprocal cross gives rise to incompatible offspring. With multiplicative action of underdominant selection and nucleo-cytoplasmic incompatibility, we have the selection regime given in Table 1 where $1-s' = (1-s)(1-t/2)$. This selection regime is applied to both sexes unless

https://doi.org/10.1017/S0016672300022102 Published online by Cambridge University Press

otherwise stated. The effect of unequal selection in both sexes will be discussed later.

The life cycle of the populations is as follows:

$$\text{Adults} \xrightarrow[\text{migration}]{} \text{adults} \xrightarrow[\text{gametogenesis}]{} \text{gametes} \xrightarrow[\text{random sampling}]{} \text{gametes} \xrightarrow[\text{fertilization}]{} \text{zygoes} \xrightarrow[\text{selection}]{} \text{adults}$$

$$P_i(P_i^*) \qquad P_i'(P_i^{*\prime}) \qquad P_i''(P_i^{*\prime\prime}) \qquad (i = 1, 2, 3, 4) \qquad P_i^{\dagger}(P_i^{*\dagger})$$

Here $P_1(P_1^*)$, $P_2(P_2^*)$, $P_3(P_3^*)$, and $P_4(P_4^*)$ are the expected gamete frequencies of AM, Am, aM, and am in the females (males) in either of population X or Y. We will add a subscript X or Y to P_i and P_i^* when it is necessary to distinguish them between the two populations. The primes and dagger denote these frequencies at different stages of life cycle. Generations are discrete and nonoverlapping.

Table 1. *Genotypes and selection regime*

	Female gamete			
Male gamete	AM (ξ_1)	Am (ξ_2)	aM (ξ_3)	am (ξ_4)
AM (η_1)	$AAMM$ 1	$AAmM$ $1-\alpha t$	$AaMM$ $1-s'$	$AamM$ $1-s'$
Am (η_2)	$AAMm$ $1-\beta t$	$AAmm$ $1-t$	$AaMm$ $1-s'$	$Aamm$ $1-s'$
aM (η_3)	$AaMM$ $1-s'$	$AamM$ $1-s'$	$aaMM$ $1-t$	$aamM$ $1-\beta t$
am (η_4)	$AaMm$ $1-s'$	$Aamm$ $1-s'$	$aaMm$ $1-\alpha t$	$aamm$ 1

Incomplete maternal inheritance is assumed for the extranuclear genomes M and m. Selection against the autosomal heterozygotes, and nucleocytoplasmic incompatibility against associations Am and aM in an individual are indicated by the reduced fitnesses $1-s$ and $1-\theta t$ (θ is a function of β), respectively, with their combined action being multiplicative [$1-s' = (1-s)(1-t/2)$]. An individual with genotype $AamM$ produces gametes, AM, aM, Am and am with relative frequencies $\beta/2$, $\beta/2$, $\alpha/2$ and $\alpha/2$, respectively and that with $AAMm$ produces gametes AM and Am with α and β and so on.

Let $N_f^{(X)}$ and $N_m^{(X)}$ be the numbers of breeding females and males in population X, and $N_f^{(Y)}$ and $N_m^{(Y)}$ be those in population Y. These numbers are assumed to be constant with time. Migration exchanges $gN_f^{(X)}$ females and $g^*N_m^{(X)}$ males between the two populations. The proportions of females and males exchanged in population X are g and g^*, respectively while those in population Y are $h = (N_f^{(X)}/N_f^{(Y)})\, g$ and $h^* = (N_m^{(X)}/N_m^{(Y)})\, g^*$. Thus we have

$$\left.\begin{aligned} P'_{Xi} &= (1-g)\, P_{Xi} + gP_{Yi}, \\ P'_{Yi} &= (1-h)\, P_{Yi} + hP_{Xi} \end{aligned}\right\} \quad (i = 1, 2, 3, 4) \qquad (1)$$

for female exchange. Male exchange is given by replacing g and h by g^* and h^*.

Random sampling of gametes following gametogenesis occurs independently of sex and population. In each population, N_f female and N_f male gametes are

https://doi.org/10.1017/S0016672300022102 Published online by Cambridge University Press

sampled from each gamete pool to form N_f female zygotes. Likewise, N_m female and N_m male gametes are sampled forming N_m male zygotes. The resulting gamete frequencies are determined by multinominal distributions with index N_f or N_m and parameters $\{P_i'\}$ or $\{P_i^{*\prime}\}$. Let $\{\xi_i\}$ and $\{\eta_i\}$ be the female and male gamete frequencies to form the female zygotes. When N_f is sufficiently large (in practice, even when it is as small as 10), it is known that a diffusion approximation can give a quite satisfactory agreement with the Wright–Fisher reproduction model. As we have considered a time discrete model up until now, an appropriate representation for these multinominal sampling processes may be given by stochastic difference equations

$$\left.\begin{aligned}\xi_i &= P_i' + \Delta B_i(N_f, P_i'),\\ \eta_i &= P_i^{*\prime} + \Delta B_i^*(N_f, P_i^{*\prime})\end{aligned}\right\} \tag{2}$$

from which we obtain $P_i'' = \frac{1}{2}(\xi_i + \eta_i)$. In the above equations, $\{\Delta B_i\}$ and $\{\Delta B_i^*\}$ are two independent sets of random variables. The expectations of ΔB_i and ΔB_i^*, conditioned on the values of P_i' and $P_i^{*\prime}$, are

$$E\{\Delta B_i \mid P_i' = p_i\} = E\{\Delta B_i^* \mid P_i^{*\prime} = p_i^*\} = 0 \tag{3a}$$

and the variances and covariances are

$$\left.\begin{aligned}E\{\Delta B_i \Delta B_j \mid P_i' = p_i\} &= \frac{1}{N_f} p_i(\delta_{ij} - p_j),\\ E\{\Delta B_i^* \Delta B_j^* \mid P_i^{*\prime} = p_i^*\} &= \frac{1}{N_f} p_i^*(\delta_{ij} - p_j^*)\end{aligned}\right\} \tag{3b}$$

in which δ_{ij} is the Kronecker delta function and higher order terms of N_f^{-1} are neglected. Likewise, the gamete frequencies to form the male zygotes, $\{\xi_i^*\}$ and $\{\eta_i^*\}$, can be expressed in terms of two other independent sets of random variables with index N_m and parameters $\{P_i'\}$ or $\{P_i^{*\prime}\}$. We define the frequencies of A and M in the sampled female gametes as

$$\lambda = \xi_1 + \xi_2 \quad \text{and} \quad \rho = \xi_1 + \xi_3 \tag{4a}$$

and those in the sampled male gametes as

$$\lambda^* = \eta_1 + \eta_2 \quad \text{and} \quad \rho^* = \eta_1 + \eta_3. \tag{4b}$$

Equation (2) are a convenient way to describe the change of the gamete frequencies due to random sampling drift in bisexual populations. We will use them in deriving a diffusion equation, but not in the simulation study. In simulations, we shall faithfully sample the gametes following multinomial distributions.

Let us consider $\{P_i^\dagger\}$ in the females of the population after selection. When fertilization occurs, heteroplasmic zygotes can be formed if $\beta \neq 0$. Then within-generation drift acts during development, fixing an extranuclear genotype in a cell. After multiple rounds of mitotic cell divisions and meiosis a heteroplasmic zygote can produce both types of homoplasmic cells, denoted by MM and mm, with relative frequencies $\alpha(\beta)$ and $\beta(\alpha)$ when type $M(m)$ is transmitted by a female gamete. That the frequency of homoplasmic cells produced by such heteroplasmic zygotes equals α (or β) implies that M and m are equivalent in amplification and degeneration. In other words, we assume no selection occurring within an individual (Takahata, 1984).

Now it is a straightforward matter to calculate $\{P_i^\dagger\}$ in terms of $\{\xi_i\}$ and $\{\eta_i\}$. The mean fitness, W, is given by

$$W = 1-\alpha t\{\xi_2\lambda^*+\xi_3(1-\lambda^*)\}-\beta t\{\eta_2\lambda+\eta_3(1-\lambda)\}-s'(\lambda+\lambda^*-2\lambda\lambda^*) \tag{5}$$

and $\{P_i^\dagger\}$ by

$$WP_1^\dagger = (\tfrac{1}{2}\alpha)[\xi_1+\rho\lambda^*-s'\{\xi_1(1-\lambda^*)+\xi_3\lambda^*\}] + (\tfrac{1}{2}\beta)[\eta_1+\rho^*\lambda-s'\{\eta_1(1-\lambda)+\eta_3\lambda\}]-\alpha\beta t(\xi_1\eta_2+\eta_1\xi_2), \tag{6a}$$

$$WP_2^\dagger = (\tfrac{1}{2}\alpha)[\xi_2+(1-\rho)\lambda^*-2\alpha t\xi_2\lambda^*-s'\{\xi_2(1-\lambda^*)+\xi_4\lambda^*\}] + (\tfrac{1}{2}\beta)[\eta_2+(1-\rho^*)\lambda-2\beta t\eta_2\lambda-s'\{\eta_2(1-\lambda)+\eta_4\lambda\}]-2\alpha\beta t\xi_2\eta_2 \tag{6b}$$

$$WP_3^\dagger = (\tfrac{1}{2}\alpha)[\xi_3+\rho(1-\lambda^*)-2\alpha t\xi_3(1-\lambda^*)-s'\{\xi_1(1-\lambda^*)+\xi_3\lambda^*\}] + (\tfrac{1}{2}\beta)[\eta_3+\rho^*(1-\lambda)-2\beta t\eta_3(1-\lambda)-s'\{\eta_1(1-\lambda)+\eta_3\lambda\}]-2\alpha\beta t\xi_3\eta_3. \tag{6c}$$

The change in the frequencies, Q and R, of A and M in the females is obtained by

$$WQ^\dagger = \tfrac{1}{2}[\lambda+\lambda^*-s'(\lambda+\lambda^*-2\lambda\lambda^*)]-t(\alpha\xi_2\lambda^*+\beta\eta_2\lambda) \tag{7a}$$

$$WR^\dagger = \alpha[\rho-s'\{\xi_1(1-\lambda^*)+\xi_3\lambda^*\}]+\beta[\rho^*-s'\{\eta_1(1-\lambda)+\eta_3\lambda\}] -t[\alpha^2\xi_3(1-\lambda^*)+\beta^2\eta_3(1-\lambda)+\alpha\beta\{\xi_1\eta_2+\eta_1\xi_2+2\xi_3\eta_3\}] \tag{7b}$$

where $Q^\dagger = P_1^\dagger+P_2^\dagger$ and $R^\dagger = P_1^\dagger+P_3^\dagger$. The frequencies $\{P_i^{*\dagger}\}$ in the males of the population are given by replacing $\{\xi_i\}$ and $\{\eta_i\}$ by $\{\xi_i^*\}$ and $\{\eta_i^*\}$ and so are $Q^{*\dagger}$ and $R^{*\dagger}$.

It is perhaps instructive to examine the mean and variance of $Q^\dagger(Q^{*\dagger})$ and $R^\dagger(R^{*\dagger})$ when there are no migration and selection involved. From (7), we have

$$Q^\dagger = \tfrac{1}{2}(\lambda+\lambda^*)$$

and

$$R^\dagger = \alpha\rho+\beta\rho^*.$$

Using (2), (3) and (4), the conditional means become

$$E\{Q^\dagger \mid Q = q, Q^* = q^*\} = \tfrac{1}{2}(q+q^*), \tag{8a}$$

$$E\{R^\dagger \mid R = r, R^* = r^*\} = \alpha r+\beta r^* \tag{8b}$$

and the conditional variances become

$$\mathrm{Var}\,(Q^\dagger) = \frac{1}{4N_f}\{q(1-q)+q^*(1-q^*)\}, \tag{9a}$$

$$\mathrm{Var}\,(R^\dagger) = \frac{1}{N_f}\{\alpha^2 r(1-r)+\beta^2 r^*(1-r^*)\}. \tag{9b}$$

The conditional means of $Q^{*\dagger}$ and $R^{*\dagger}$ in the males of the population are the same as (8), whereas the conditonal variances are given by

$$\mathrm{Var}\,(Q^{*\dagger}) = \frac{1}{4N_m}\{q(1-q)+q^*(1-q^*)\}, \tag{10a}$$

$$\mathrm{Var}\,(R^{*\dagger}) = \frac{1}{N_m}\{\alpha^2 r(1-r)+\beta^2 r^*(1-r^*)\}. \tag{10b}$$

In order to define the variance effective number N_e in a bisexual population, we must seek a quasi-Markovian variable (see Ewens, 1979, pp. 104–112). For a nuclear (Mendelian) gene, $Z^{(n)} = \frac{1}{2}(Q+Q^*)$ has such a property and $N_e^{(n)}$ is given by

$$N_e^{(n)} = \frac{4N_f N_m}{N_f + N_m}. \tag{11}$$

In the case of an extranuclear gene, $Z^{(e)} = \alpha R + \beta R^*$ plays the same role as $Z^{(n)}$. The mean per-generation increment of this quasi-Markovian $Z^{(e)}$ is, of couse, 0 as seen from (8*b*). We define the variance effective number $N_e^{(e)}$ for an extranuclear gene by

$$N_e^{(e)} = \frac{(\alpha^2+\beta^2)\, z(1-z)}{\mathrm{Var}\,(Z^{(e)})} \tag{12}$$

where $z = E(Z^{(e)}) = \alpha r + \beta r^*$. The variance of $Z^{(e)}$ is given by

$$\mathrm{Var}\,(Z^{(e)}) = \left(\frac{\alpha^2}{N_f} + \frac{\beta^2}{N_m}\right)\{\alpha^2 r(1-r) + \beta^2 r^*(1-r^*)\} \tag{13}$$

from (9*b*) and (10*b*). We can show that $\{\alpha^2 r(1-r) + \beta^2 r^*(1-r^*)\}$ is well approximated by $(\alpha^2+\beta^2)\, z(1-z)$, the difference being in proportion to β and $r-r^*$. From (12) and (13), and neglecting higher order terms of $1/N_f$ *and* $1/N_m$, we obtain

$$N_e^{(e)} = \frac{N_f N_m}{\beta^2 N_f + \alpha^2 N_m}, \tag{14}$$

which is identical with the inbreeding effective number (Takahata & Maruyama, 1981). However, there is no guarantee of the validity of the above treatment if gene frequency differs to a large extent in both sexes. This can occur either by sex-dependent migration or by sex-dependent selection. In either case, formula (14) is valid only when β is extremely small.

3. SIMULATION RESULTS

The model described in the previous section was simulated to observe the frequency changes of A and M and the linkage disequilibrium in the two populations. Some of the simulation results were used to check the validity of the analytical treatment in the following section. We treated migration and selection deterministically, using (1), (6) and (7). In order to determine the genetic composition of the female and male gametes which produce $N_f^{(X)}$ female zygotes in population X, we generated two sets of multinomial random numbers with index $N_f^{(X)}$ and parameters $\{P_i'\}$ or $\{P_i^{*\prime}\}$. These random numbers were used to compute $\{\xi_i\}$ and $\{\eta_i\}$ in (2). Two other sets of multinomial random numbers with index $N_m^{(X)}$ and parameters $\{P_i'\}$ or $\{P_i^{*\prime}\}$ were generated for $N_m^{(X)}$ male zygotes in population X. The same procedure was employed for population Y if it was finite.

All simulations started with the initial conditions $P_{X1} = P_{X1}^* = 1$ and $P_{Y4} = P_{Y4}^* = 1$, i.e. population X is fixed for AM and population Y for am. To obtain the expectation of any quantity, 100 independent runs were repeated for each set of parameter values. The parameter values we chose were as follows:

$g = 10^{-3}$, $g^* = 0$ and 10^{-3}, $\beta = 0$ and 0·5, $s' = 0$ and 0·5, $t = 0$, 0·01 and 0·05, and $N_f^{(X)} = N_m^{(X)} = 50$ while $N_f^{(Y)} = N_m^{(Y)} = 50$, 100 and ∞.

Fig. 1 shows the result of control simulations and the effect of selection against heterozygotes. Here no nucleo-cytoplasmic incompatibility was assumed ($t = 0$) and $N_f^{(Y)} = N_m^{(Y)} = \infty$ so that migration is unidirectional from population Y to X with $g = g^* = 10^{-3}$. With this setting and $s' = 0$, the expectation of $Z = \alpha R + \beta R^*$ is given by $1 - e^{-(\alpha g + \beta g^*)T}$ where T is the time elapsed since migration started. The agreement between this expectation and the simulation results were satisfactory, showing that 100 independent repeats are sufficient to obtain the expectation of our quantities and that random sampling of gametes is correctly simulated.

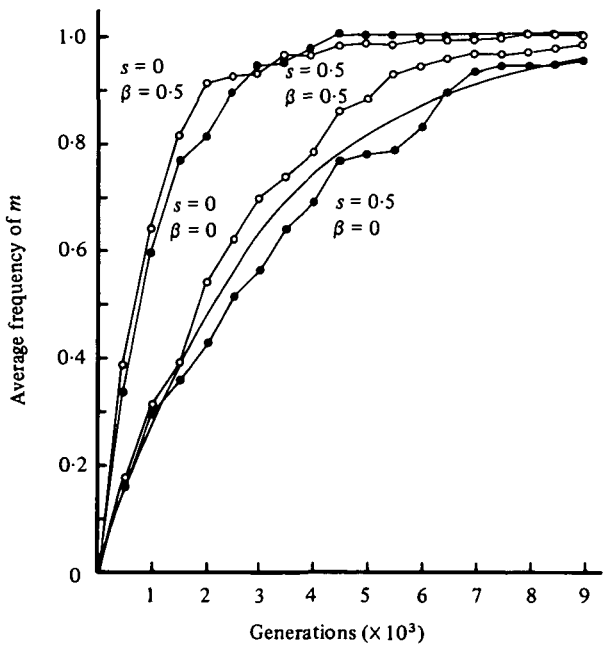

Fig. 1. Average frequency of the invading extranuclear genome m in population X (ordinate) against time elapsed since invasion started (abscissa, in units of 10^3 generations). $N_f^{(X)} = N_m^{(X)} = 50$, $N_f^{(Y)} = N_m^{(Y)} = \infty$, $g = g^* = 10^{-3}$, $t = 0$ and $s = 0$ or 0·5 are assumed. ●, Results for $\beta = 0$; ○, results for $\beta = 0{\cdot}5$ simply as a reference. A control simulation indicated by (●) ($s = 0$, $\beta = 0$) agrees with (18c) with $t = 0$ and $g_e = g$, whereas a simulation indicated by (●) ($s = 0{\cdot}5$, $\beta = 0$) agrees well with the numerical solution (a smooth line) of (21) with $t = 0$ and $g_e = g/3$.

When heterozygotes are semi-lethal ($s' = 0{\cdot}5$), the introgression of foreign extranuclear genome m can occur with a slight retardation while that of the foreign nuclear gene a is nearly completely prevented. We note that the introgression of m is retarded by a factor of about 1/3 compared with the case when $s' = 0$ and that the frequency of A rapidly approaches equilibrium. We will explain these results in the next section.

Fig. 2 shows some simulation results when both populations are of finite size and therefore migration is bidirectional. When t is small, the whole population is eventually fixed for either M or m. The expected frequency of m, however,

https://doi.org/10.1017/S0016672300022102 Published online by Cambridge University Press

approaches an equilibrium value depending on the size of populations X and Y. On the other hand, appreciable incompatibility makes the introgression of extranuclear genomes difficult because the resident nuclear gene is kept high in frequency. In one example in Fig. 2, $N_e^{(e)} t = 2{\cdot}5$ so that the effect of incompatibility is large enough to overcome random sampling drift. The direction of introgression

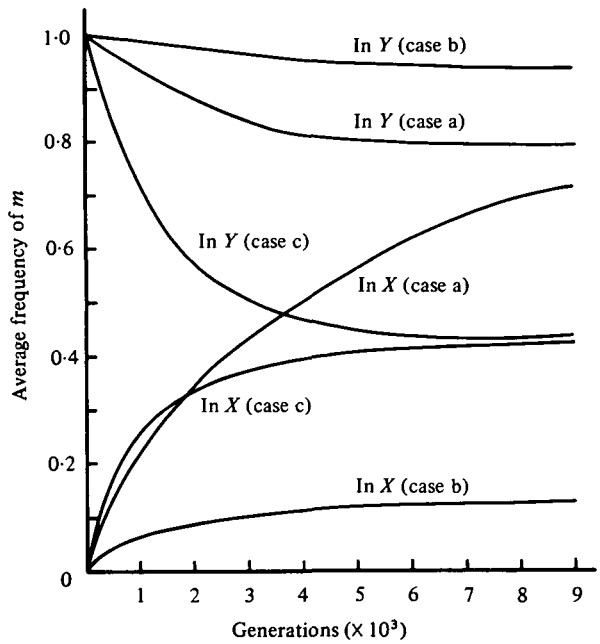

Fig. 2. Average frequencies of the extranuclear genome m in population X and in population Y: Effects of nucleo-cytoplasmic incompatibility and differential population size. The decreasing and increasing lines are regression lines of the simulation results in population Y and population X, respectively. In all simulations, $g = g^* = 10^{-3}, \beta = 0$ and $s' = 0{\cdot}5$: (a) $t = 0{\cdot}01$, $N_f^{(X)} = N_m^{(X)} = 50$ and $N_f^{(Y)} = N_m^{(Y)} = 100$, (b) $t = 0{\cdot}05$, $N_f = N_m = 50$ m in both populations and (c) $t = 0$, $N_f = N_m = 50$ in both populations. The simulation result for the case of $t = 0{\cdot}05$, $N_f^{(X)} = N_m^{(X)} = 50$ and $N_f^{(Y)} = N_m^{(Y)} = 100$, is very similar to (b).

of extranuclear genome is predictable in our model. When two populations differ in size, the introgression tends to occur from the larger population to the smaller population because of the asymmetry of migration rates between the two populations. For example, when $N_f^{(X)}$ $N_m^{(X)} = 50$, $N_f^{(Y)} = N_m^{(Y)} = 100$, $\beta = 0$, $g = g^* = 10^{-3}$, $t = 0{\cdot}01$ and $s' = 0{\cdot}5$, population X was completely invaded by foreign extranuclear genome m at the end of the simulation ($10/g$ generations) in 72 out of 100 repeats. In contrast, population Y still maintained the resident type m in more than 70 repeats with the average frequency of m being about 0·8. Thus, a larger population is more resistant to introgression if hybrids have no mating preference in the backcross generation. As t increases, however, this directionality becomes invisible for the obvious reason. The expected frequencies of m in the populations X and Y for the case when $t = 0{\cdot}05$, $N_f^{(X)} = N_m^{(X)} = 50$ and $N_f^{(Y)} = N_m^{(Y)} = 100$ do not differ much from those for the case (b) of equal population size.

https://doi.org/10.1017/S0016672300022102 Published online by Cambridge University Press

4. ANALYTICAL RESULTS

Here we treat the simulation model analytically, confining ourselves to the case of $\beta = 0$, $g^* = 0$, $N_f^{(Y)} = N_m^{(Y)} = \infty$ and

$$1 \sim s' \gg t, g \tag{15}$$

where the sign $\sim$ is used to indicate that the quantities on both sides are of the same order of magnitude. These parameters are equivalent to those in Takahata & Slatkin (1984) and in the simulation in Fig. 1. An additional assumption is that initially the gamete frequencies were the same in both sexes, i.e. $P_i = P_i^*$ $(i = 1, 2, 3, 4)$. To study this analytical model, we first ignore the effect of random sampling drift.

Recalling that $Q = P_1 + P_2$ and $R = P_1 + P_3$, and defining $U = P_1/R$, we can show that Q and U rapidly approach their equilibrium states and are weakly affected by drift, but in contrast, that the change of R is a very slow process strongly affected by drift. Consider the changes of Q, R and U by migration and selection alone. The lower case letters q, r and u are used to indicate that the processes are treated in the absence of drift. Using (2) and (4), and noting that $\xi_i = P_i'$ and $\eta_i = P_i^{*\prime}$, we have

$$q^\dagger = \frac{1}{2W}[q + q' - 2tq(q' - ur') - s'(q + q' - 2qq')], \tag{16a}$$

$$r^\dagger = \frac{r'}{2W}[1 - t(1-u)(1-q) - s'\{q + u - 2qu\}], \tag{16b}$$

$$u^\dagger = \frac{q + u - s'(q + u - 2qu)}{2\{1 - t(1-u)(1-q) - s'(q + u - 2qu)\}} \tag{16c}$$

where $q' = (1-g)q$, $r' = (1-g)r$ and

$$W = 1 - t\{qq' + (1-q)r' - ur'\} - s'(q + q' - 2qq'). \tag{16d}$$

From (16a) and (16c), we can readily obtain the equilibria of q and u. The stable equilibrium of q is given by

$$\hat{q} \simeq 1 - \frac{g(1-s')}{2s'} \tag{17a}$$

and that of u by

$$\hat{u} \simeq 1 - \frac{g(1-s')^2}{2s'(1+s')} \tag{17b}$$

to a good approximation. If we linearize (16a) and (16c) around $\hat{q}$ and $\hat{u}$, we obtain the rate of return to equilibrium. The larger eigenvalue of the linear equations is $1 - s'$ so that the return occurs quite rapidly for large s'. This fact allows us to substitute $\hat{q}$ and $\hat{u}$ into q and u in (16b) and (16d), and to study the change of r under the condition that q and u have been attained to equilibrium. The assumption (15) yields an approximation for the change of r as

$$r^\dagger = (1 - g_e)r + tr(1-r) \tag{18a}$$

where

$$g_e = \frac{g(1-s')}{1+s'}. \tag{18b}$$

https://doi.org/10.1017/S0016672300022102 Published online by Cambridge University Press

The degree of extranuclear gene flow g_e is a function of s' and equals $(1/3)g$ for $s' = 0{\cdot}5$. This provides a rationale on simulation results for $s' = 0{\cdot}5$ in Fig. 1. Indeed, these simulation results were indistinguishable from the simulation results for the case of $s' = 0$ and $(1/3)g$ instead of g. The time continuous solution of (18*a*) is given by

$$1-r(T) = \frac{g_e[1-\exp\{(g_e-t)T\}]}{t-g_e[\exp\{(g_e-t)T\}]} \tag{18c}$$

under the initial condition $r(0) = 1$. Thus the equilibrium value of r is

$$\begin{aligned} \hat{r} &\simeq 1-\frac{g_e}{t} \quad \text{for} \quad t > g_e \\ &= 0 \qquad\quad \text{otherwise} \end{aligned} \tag{19}$$

and the rate of return to equilibrium is either g_e or t, whichever is larger. To show an adequacy of the present treatment, (18*c*) was compared with the exact solutions of $1-r$, the frequency of m in population X, which were calculated from (16). Fig. 3 shows the exact solutions which are well approximated by (18*c*).

The degree of gene flow g_e in the steady state (i.e. after the nuclear gene frequency reached the equilibrium state) is a key quantity to determine the introgression process of extranuclear genomes. It is interesting to know the levels of gene flow for the nuclear loci which link to the locus being responsible for the reproductive isolation. In this case, recombination plays an important role in the introgression process of nuclear genes. We consider a single neutral locus which has a recombination value c with the selected locus. If we follow the same procedure to derive (18*b*), we obtain

$$g_{e(n)} = \frac{c(1-s')}{s'+c(1-s')}g \tag{20}$$

for the degree of nuclear gene flow in the steady state. Clearly, there is no nuclear gene flow for completely linked loci. However, as c increases, $g_{e(n)}$ increases ($g_{e(n)} = g_e$ when $c = \frac{1}{2}$). Thus the interspecific transfer of nuclear genes is also possible if they are not closely linked to the selected locus (or loci). In fact, Selander, Hunt & Yang (1969) found nuclear introgression at several enzyme loci in *M. musculus* populations in Denmark (see also Lewontin, 1970, pp. 179–183) before Ferris *et al.* (1983) showed clear evidence of mitochondrial introgression in these populations.

Now let us incorporate the effect of random sampling drift into the change of r. When $g_e \sim t \sim 1/N_e^{(e)} \ll 1$, the standard method of diffusion approximations leads to the Kolmogorov backward equation

$$\frac{\partial\phi}{\partial T} = \{-g_e r + tr(1-r)\}\frac{\partial\phi}{\partial r} + \frac{r(1-r)}{2N_e^{(e)}}\frac{\partial^2\phi}{\partial r^2} \tag{21}$$

in which T again stands for the generations elapsed since the two populations started gene exchange and ϕ is the distribution of r at time T or the expectation of any function of r.

Equation (21) is rather familiar in the literature of population genetics (Wright, 1969; Crow & Kimura, 1970; Ewens, 1979). However, the analysis made to date is only for the mean sojourn time and the Green's function (Ewens, 1964, Li & Nei,

1977; Kimura, 1980). If we use Nagylaki's (1974) general formula, we can obtain the distribution of sojourn time (Maruyama & Kimura, 1971). Yet as the mean sojourn time itself is rather complicated (it is possible to express it in an integral form or in terms of Kummer's confluence hypergeometric function), it seems very difficult to obtain the distribution. The spectral decomposition is also unknown.

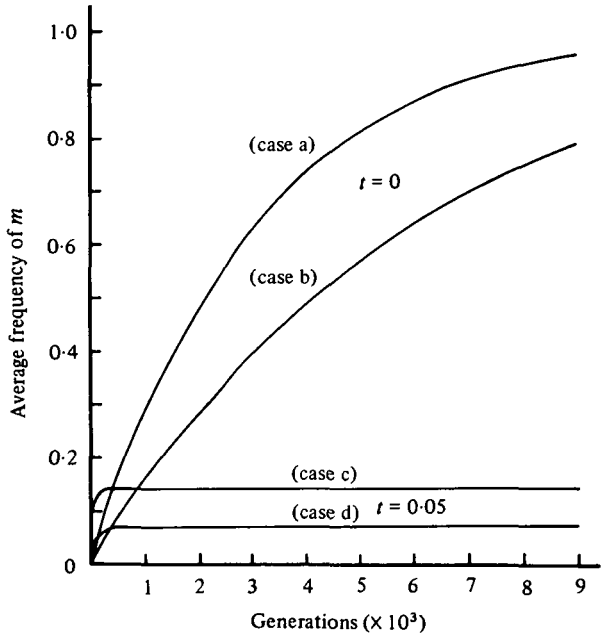

Fig. 3. Frequency of the invading extranuclear genomes m in population X. Both populations are assumed to be of infinite size, but immigration is assumed to occur unidirectionally from Y to X. $s' = 0{\cdot}5$ and $g = 10^{-3}$ in all cases: (a) $g^* = g$ and $\beta = 0{\cdot}5$, (b) $g^* = 0$, $\beta = 0{\cdot}5$ or $g^* = g$ and $\beta = 0$, (c) $g^* = g$, $\beta = 0{\cdot}5$ and (d) $g^* = g$ and $\beta = 0$ or $g^* = 0$ and $\beta = 0{\cdot}5$ [$t = 0$ in (a) and (b), $t = 0{\cdot}05$ in (c) and (d)]. Note that incompatibility nearly prevents the introgression when $t \gg g_e$. The equilibrium value attained is given by g_e/t in (19). The doubled value for case (c) is the result of male immigration and the unrealistic assumption that males transmit as many extranuclear genomes as females.

To obtain the expected frequency of foreign extranuclear genome, we used a method of numerical solution of partial differential equations (e.g. Maruyama, 1977). The results were compared with the simulation results (Fig. 1). The agreement between the two warrants the use of (21) to describe the introgression process under the conditions (15).

When there is no incompatibility, the probability that the foreign extranuclear genome becomes fixed in population X by the Tth generation can be obtained from (8.5.19) in Crow & Kimura (1970, pp. 394). The asymptotic form becomes

$$P \sim 1-(1+G)\exp(-g_e T)+\tfrac{1}{2}G(3+G)\exp[-\{2(1+G)/G\}g_e T] \qquad (22)$$

where $G = 2N_e^{(e)} g_e$. The mean fixation time is given by

$$T_m = \frac{G}{G-1}\{\gamma+\psi(G)\}/g_e \qquad (23)$$

https://doi.org/10.1017/S0016672300022102 Published online by Cambridge University Press

where $\gamma = 0{\cdot}577\ldots$ is Euler's constant and $\psi(\cdot)$ stands for the digamma function (e.g. Kimura, 1980). It is clear from these formulae that the time required for fixation to occur is of the order of the reciprocal of the immigration rate for small and moderate values of G. As G increases, the digamma function increases, implying that the fixation time is prolonged indefinitely. However, since $\gamma + \psi(G)$ increases with $\log G$, the population-size dependence of T_m is rather weak. Within this range of G, no only T_m but also P are determined mainly by g_e.

In contrast to the rather weak effect of population size on the fixation time expected in the *neutral* case, the effect of incompatibility is very strong when coupled with weak random sampling drift. In a large population, only a small amount of incompatibility is sufficient to prevent the establishment of the invading extranuclear genome. To fix the idea, let us assume $t = 10\, g_e$ and $G = 1$. In this case, the mean fixation time is about $10/g_e$ from Table 1 in Kimura (1980) (see also Table 3 in Li & Nei, 1977). If the population is increased by ten times ($G = 10$), the mean fixation time becomes $6 \times 10^9/g_e$, despite the ratio of t to g_e being kept constant. Needless to say, this enormously prolonged fixation time implies that fixation will never occur in a large population.

For a small population, such as $2N_e^{(e)}\, t < 1$, however, fixation occurs with the mean time of $1/g_e$. It is well-known fact that a mutant with selective advantage or disadvantage s behaves as neutral if the effective population number is smaller than $|s|^{-1}$ (Li & Nei, 1977). Thus, nucleo-cytoplasmic incompatibility is not an efficient mechanism to prevent the introgression of extranuclear genomes in a small population unless it is sufficiently strong.

5. DISCUSSION

Our analytical study shows that the immigration rate g_e defined in (18*b*) is a quantity of prime importance in determining the introgression process. The quantity is a function of the intrinsic immigration rate and of the degree of premating isolation between two species. We also assumed that F_1 males are sterile and therefore excluded male migration. In the case of *D. pseudoobscura* and *D. persimilis* in the United States, 3 female hybrids were found out of 30,000 females examined (Powell, 1983). Laboratory studies show the postmating isolation of these species to be such that F_1 females are fertile while F_1 males are sterile (Dobzhansky, 1936). These figures suggest that the reproductive isolation between *D. pseudoobscura* and *D. persimilis* is not complete, which allows gene exchange in nature. Therefore, we set $g = 10^{-4}$ in our previous paper (Takahata & Slatkin, 1984) which is much larger than a typical value of mutation rate per locus (10^{-6} or 10^{-5}). Although F_1 females are fertile, there are no data that indicate the relative fitnesses of F_1 female and their backcross progeny. We can speculate, however, that the relative fitnesses should be much lower than that of normal genotypes, otherwise both species could not have maintained their genetic distinction. If this is the case, g_e is smaller than g because of the reduction in hybrid viability and/or fertility, e.g. $g_e = (\frac{1}{3})g$ when $s' = 0{\cdot}5$ and of course $g_e = 0$ when $s' = 1$. Then the introgression would be slower than expected from the observed value of g.

The effect of male migration can be easily anticipated. If β is extremely small, it increases the frequency of foreign nuclear genomes without increasing the frequency of foreign extranuclear genomes. However, an increased amount would be very small as long as selection against hybrids is strong enough compared with the level of male migration rate. Then the dynamics of introgressing extranuclear genomes are not affected much by male migration. A significant change can happen only in an extreme case where male migration swamps the reproductive barrier.

When the sizes of the two populations are finite and migration is bidirectional, the extranuclear genome m is eventually fixed or lost from the entire population. Yet the expected frequency of m taken over many sample paths may asymptotically take an intermediate value, depending on asymmetric migration rates between populations. In a symmetrical migration model, it should approach $\frac{1}{2}$. The discrepancy between this expectation and the simulation result in Fig. 2(*c*) is due to a sampling error. An asymmetrical migration, on the other hand, can easily result in a bias in the direction of introgression, as shown in Fig. 2(*b*). However, this asymmetry does not seem to be the only determinant of the direction. Some sorts of assortative mating (or Wallace effect) may be more important in this respect (Murray, 1972; Slatkin, 1982).

The effect of 'paternal leakage' of extranuclear genomes was demonstrated by assuming an unrealistic value of β. It is significant only when migration is strongly biased for males. If males migrate more frequently than females do, a small paternal leakage would play an important role in the introgression process of extranuclear genomes. In contrast, more frequent female migration obscures the paternal leakage.

Sex-independent selection against backcross progeny is a conservative assumption for protecting a resident genotype. If we consider selection operating only in one of the sexes, the absence of selection in the other sex would make the introgression of extranuclear and nuclear genomes easier. Reproductive isolation mechanisms must operate in both sexes. When the F_1 males are sterile, the sterility diminishes the level of nuclear gene flow, but does not diminish extranuclear gene flow. As long as the F_1 females are fertile and transmit the majority of extranuclear genomes, male sterility plays a minor role in the introgression of extranuclear genomes.

In conclusion, it is very difficult to prevent the introgression of extranuclear genomes unless the hybrids between two species have very low fitnesses (see also Takahata & Slatkin, 1984). Selection operating on nuclear genomes does not strongly affect the dynamics of extranuclear genomes because of their independent segregation. This conclusion also applies to nuclear genes which do not cause reproductive isolation and which are loosely linked to genes that are responsible for reproductive isolation (see also Barton, 1983). Incompatibility would prevent the introgression of extranuclear genomes only when its effect is strong compared with random sampling drift. Although the intensity of incompatibility is little known, it would not be surprising to find more instances of mitochondrial or chloroplast invasion across species boundaries other than those already reported in Powell (1983), Ferris *et al.* (1983), Yonekawa *et al.* (1983), Palmer *et al.* (1983),

Weissinger *et al.* (1983), and Spolsky & Uzzell (1984). Conversely, these observations suggest that nucleo-cytoplasmic incompatibility has played a very minor role in reproductive isolation between species examined.

I thank M. Nei, M. Slatkin, B. Charlesworth and M. Kimura for their comments and interest on this subject. An anonymous reviewer's comments were very useful in improving the content and presentation of this paper. This research has been supported by NIH and NSF Grants and by Grant from the Ministry of Education, Science and Culture in Japan.

REFERENCES

BARTON, N. H. (1983). Multilocus clines. *Evolution* **37**, 454–471.

CASPARI, E., WATSON, G. S. & SMITH, W. (1966). The influence of cytoplasmic pollen sterility on gene exchange between populations. *Genetics* **53**, 741–746.

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

DOBZHANSKY, T. (1936). Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* **21**, 113–135.

DOBZHANSKY, T. (1937). *Genetics and the Origin of Species*. New York: Columbia University Press.

DOVER, G. A. & FLAVELL, R. B. (ed.) (1982). *Genomic Evolution*. London: Academic Press.

ELLIS, J. (1982). Promiscuous DNA-chloroplast genes inside plant mitochondria. *Nature* **299**, 678–679.

EWENS, W. J. (1964). The pseudo-transient distribution and its use in genetics. *Journal of Applied Probability* **1**, 141–156.

EWENS, W. J. (1979). *Mathematical Population Genetics*. Berlin: Springer-Verlag.

FARRELLY, F. & BUTOW, R. A. (1983). Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* **301**, 296–301.

FERRIS, S. D., SAGE, R. D., HUANG, C. M., NIELSEN, J. T., RITTE, U. & WILSON, A. C. (1983). Flow of mitochondrial DNA across a species boundary. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 2290–2294.

GRUN, P. & AUBERTIN, M. P. (1965). Evolutionary pathways of cytoplasmic male sterility in solanum. *Genetics* **51**, 399–409.

JONES, D. F. (1951). The cytoplasmic separation of species. *Proceedings of the National Academy of Sciences, U.S.A.* **37**, 308–310.

KIMURA, M. (1980). Average time until fixation of a mutant allele in a finite population under continued mutation pressure: Studies by analytical, numerical, and pseudo-sampling methods. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 522–526.

LAVEN, H. (1959). Speciation by cytoplasmic isolation in the *Culex Pipiens*-complex. *Cold Spring Harbor Symposium on Quantitative Biology* **24**, 166–173.

LEWONTIN, R. C. (1970). *The Genetic Basis of Evolutionary Change*. New York and London: Columbia University.

LI, W.-H. & NEI, M. (1977). Persistence of common alleles in two related populations or species. *Genetics* **86**, 901–914.

MARUYAMA, T. (1977). *Stochastic Problems in Population Genetics: Lecture Notes in Biomathematics*, vol. 17. Berlin: Springer-Verlag.

MARUYAMA, T. & KIMURA, M. (1971). Some methods for treating continuous stochastic processes in population genetics. *Japanese Journal of Genetics* **46**, 407–410.

MULLER, H. J. (1942). Isolating mechanisms, evolution and temperature. In *Biological Symposia*, vol. 6 (ed. T. Dobzhansky), pp. 71–125. Lancaster, Pa.: Jacques Cattell Press.

MURRAY, J. (1972). *Genetic Diversity and Natural Selection*. New York: Hafner.

NAGYLAKI, T. (1974). The moments of stochastic integrals and the distribution of sojourn times. *Proceedings of the National Academy of Sciences, U.S.A.* **71**, 746–749.

OKA, H.-I. (1978). Reproductive barriers distributed in plant species and their genetic control. In *The Dynamics of Speciation in Plants and Animals* (ed. H.-I. Oka and O. Kitagawa), pp. 56–61.

PALMER, J. D., SHIELDS, D. R. & COHEN, D. B. (1983). Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. *Theoretical and Applied Genetics* **65**, 181–189.

POWELL, J. R. (1983). Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: Evidence from *Drosophila*. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 492–495.

PRAKASH, S. (1972). Origin of reproductive isolation in the absence of apparent genetic differentiation in a geographic isolate of *Drosophila pseudoobscura*. *Genetics* **72**, 143–155.

SELANDER, R. K., HUNT, W. G. & YANG, S. Y. (1969). Protein polymorphism and genic heterozygosity in two European subspecies of the house mouse. *Evolution* **23**, 379–390.

STERN, D. B. & LONSDALE, D. M. (1982). Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature* **299**, 698–702.

SLATKIN, M. (1982). Pleiotropy and parapatric speciation. *Evolution* **36**, 263–270.

SPOLSKY, C. & UZZELL, T. (1984). Natural interspecies transfer of mitochondrial DNA in amphibians. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 5802–5805.

TAKAHATA, N. (1984). A model of extranuclear genomes and the substitution rate under within-generation selection. *Genetical Research* **44**, 109–116.

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. (1984). Mitochondrial gene flow. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 1764–1767.

TSUNEWAKI, K. & KIHARA, H. (1962). Comparative gene analysis of common wheat and its ancestral species. I. Necrosis. *Japanese Journal of Genetics* **37**, 474–484.

WEISSINGER, A. K., TIMOTHY, D. H., LEVINGS, G. S. III & GOODMAN, M. M. (1983). Patterns of mitochondrial DNA variation in indigenous maize races of Latin America. *Genetics* **104**, 365–379.

WHITE, M. J. D. (1978). *Modes of Speciation*. San Francisco: W. H. Freeman.

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

WRIGHT, S. (1969). *Evolution and the Genetics of Populations*, vol. 2: *The Theory of Gene Frequencies*. Chicago: University of Chicago Press.

YONEKAWA, H., TAGASHIRA, S. & MORIWAKI, K. (1983). Mitochondrial DNA and evolution of mice. *Iden* **37**, 29–37 (in Japanese).

ZOUROS, E. (1981). The chromosomal basis of sexual isolation in two sibling species of *Drosophila: D. arizonensis* and *D. mojavensis*. *Genetics* **97**, 703–718.