

Evolutionary dynamics of a multigroup fluctuating-population system

D. P. Bhatia, D. Arora, and M. A. Prasad

Division of Radiological Protection, Bhabha Atomic Research Centre, Trombay, Bombay 400 085, India

(Received 1 July 1992)

We studied the evolutionary dynamics of a population undergoing asexual reproduction in a flat-fitness landscape. The quantity of interest is the distribution of the overlap function q which is a measure of the similarity in the genome structure between two individuals. We obtain analytical expressions for $\langle q \rangle$, $\langle q^2 \rangle$, and $p(q)$ in a model with the following features: continuous time, fluctuating population divided into many compartments, and a finite number of genes per genome. A few special cases of interest are also discussed.

PACS number(s): 05.40.+j, 02.50.-r, 87.10.+e

Recently, evolutionary dynamics of populations has aroused some interest. Derrida and Peliti [1], Serva and Peliti [2], Higgs and Derrida [3], and Maruyama and Kimura [4] have studied the dynamics of evolving populations under both sexual and asexual reproduction in a flat-fitness landscape (i.e., survival and reproduction are independent of the genome structure) [5]. In analogy with spin-glass systems, they consider a two-state (+1, -1) model for the genes and compute an overlap function q and its distribution. The overlap function q , which is a measure of the similarity in gene structure between two individuals, is given by

$$q_{\alpha,\beta} = \sum_{i=1}^N \sigma_i^\alpha \sigma_i^\beta / N. \quad (1)$$

σ_i^α is the state of the i th gene of individual α and N is the number of genes per genome. Serva and Peliti [2] obtained expressions for the process average of q ($\langle q \rangle$). They showed that $\langle q \rangle$ is the same for both sexual and asexual reproduction. Serva and Peliti further showed that the variance of $\langle q \rangle$ is almost zero for sexual reproduction but is of the same order as $\langle q \rangle$ for asexual reproduction. Similar results were obtained by Higgs and Derrida [3] who obtained the distribution $p(q)$ by direct simulation and showed that this quantity is self-averaging for sexual reproduction and not so for asexual reproduction.

These results were obtained by them with $N \rightarrow \infty$, and for a discrete time and constant population model. In this paper, we obtain a finite- N , fluctuating-population (which may also grow), and continuous-time model with asexual reproduction and obtain $\langle q \rangle$, $\langle q^2 \rangle$, and $\langle p(q) \rangle$. We consider the population to be distributed into compartments with a certain rate of nearest-neighbor migration to incorporate loosely some kind of geographical separation.

FORMULATION

Let N be the number of genes, each of which can be in either of the states +1 or -1. Thus there are 2^N possible states of the genome. The number of individuals in any genome state and population compartment can change

due to any one of the following processes: (1) Birth: an individual can divide into two exactly identical individuals with probability $\alpha \Delta t$ in time Δt ; (2) death: an individual can die with probability $\beta \Delta t$ in time Δt ; (3) mutation: a gene can change its state with probability $\mu \Delta t$ in time Δt ; and (4) migration: an individual can migrate from compartment J to $J+1$ or $J-1$ with probability $\eta \Delta t$ in time Δt .

Let us consider two individuals, one each in genome states i_1 and i_2 . Out of the N genes in each of the states i_1 and i_2 , suppose i is the number of genes which are in different states, i.e.,

$$i = 0.5 \left[N - \sum_{k=1}^N \sigma_k(i_1) \sigma_k(i_2) \right], \quad (2)$$

where $\sigma_k(i_1)$ is the state (+1 or -1) of the k th gene in genome state i_1 . This pair will contribute $(1-2i/N)$ to the overlap function q . We define the quantities of interest $\langle q \rangle$ and $\langle q^2 \rangle$ as the process average weighted over the number of pairs of individuals, i.e.,

$$\langle q_j \rangle = \sum_{i_1=1}^{2^N} \sum_{i_2=1}^{2^N} \frac{\langle m_{i_1,J} (m_{i_2,J} - \delta_{i_1,i_2}) \rangle}{\langle M_J (M_J - 1) \rangle} (1 - 2i/N), \quad (3)$$

$$\langle q_j^2 \rangle = \sum_{i_1=1}^{2^N} \sum_{i_2=1}^{2^N} \frac{\langle m_{i_1,J} (m_{i_2,J} - \delta_{i_1,i_2}) \rangle^2}{\langle M_J (M_J - 1) \rangle} (1 - 2i/N)^2, \quad (4)$$

where $m_{i_1,J}$ and $m_{i_2,J}$ denote the number of individuals in the genome states i_1 and i_2 , respectively, in compartment J ; M_J is the total number of individuals in compartment J ; and i as defined in Eq. (2) is the number of gene locations at which states i_1 and i_2 differ. The q 's, m 's, and M 's are, in general, dependent on time. We have, however, not explicitly indicated this dependence for reasons of simplicity. $\langle q_j' \rangle$, the unbiased process average of q , will be given by

$$\langle q_j' \rangle = \sum_{i_1=1}^{2^N} \sum_{i_2=1}^{2^N} \left\langle \frac{m_{i_1,J}(m_{i_2,J} - \delta_{i_1,i_2})}{M_J(M_J - 1)} \right\rangle (1 - 2i/N). \quad (5)$$

We have used the former definitions, given by Eqs. (3) and (4), for $\langle q \rangle$ and $\langle q^2 \rangle$, respectively, because of two reasons. First, it is not easy to compute these quantities

using the latter definition [given by Eq. (5)] because of the term in the denominator. Second, in this definition the right-hand side is not defined for $M_J = 0, 1$. However, the two definitions become identical for a constant population (as used in Ref. [2]).

Let $P(\{m_{i,J}\}, t)$ be the probability that there are exactly $m_{i,J}$ (for $i = 1$ to 2^N and $J = 1$ to L) individuals in genome state i and compartment J . Then the master equation for $P(\{m_{i,J}\}, t)$ is given by

$$\begin{aligned} \frac{dP(\{m_{i,J}\}, t)}{dt} = & \alpha \sum_{i,J} (\bar{A}_{i,J} - 1) m_{i,J} P(\{m_{i,J}\}, t) + \beta \sum_{i,J} (A_{i,J} - 1) m_{i,J} P(\{m_{i,J}\}, t) \\ & + \mu \sum_{i,J,k} (A_{i,J} \bar{A}_{i_k,J} - 1) m_{i,J} P(\{m_{i,J}\}, t) \\ & + 0.5\eta \sum_{i,J} [(A_{i,J} A_{i,J+1} - 1) + (A_{i,J} A_{i,J-1} - 1)] m_{i,J} P(\{m_{i,J}\}, t). \end{aligned} \quad (6)$$

i_k ($k = 1, 2, \dots, N$) refers to a state which differs from state i at gene location k only. The summations are from $i = 1$ to 2^N , $J = 1$ to L , and $K = 1$ to N . Here, $A_{i,J}$ and $\bar{A}_{i,J}$ are Van Kampen's [6] raising and lowering operators defined by

$$\begin{aligned} A_{i,J} f(\{m_{i',J'}\}, t) &= f(\{m_{i',J'} + \delta_{i,i'} \delta_{J,J'}\}, t), \\ \bar{A}_{i,J} f(\{m_{i',J'}\}, t) &= f(\{m_{i',J'} - \delta_{i,i'} \delta_{J,J'}\}, t). \end{aligned} \quad (7)$$

For $\langle q_j \rangle$ and $\langle q_j^2 \rangle$ the quantities to be computed are the correlated moments $\langle m_{i_1,J_1} m_{i_2,J_2} \rangle$. We assume uniform initial conditions, i.e., at $t = 0$ the probability of an individual being in any one of the $L 2^N$ states is constant ($= 1/L 2^N$), and periodic boundary conditions in J , i.e., $m_{i,J+L} = m_{i,J}$. Because of this, the quantities $\langle q_j \rangle$ and $\langle q_j^2 \rangle$ depend only on the difference of i_1 and i_2 [as given by Eq. (2)] and $J = J_1 - J_2$. Multiplying both sides of Eq. (6) with $m_{i_1,J_1} m_{i_2,J_2}$ and summing over all values of m_{i_1,J_1} and m_{i_2,J_2} , we get, after some algebra,

$$\begin{aligned} \frac{dP_{i,J}}{dt} = & 2(\alpha - \beta) P_{i,J} + 2\mu [i P_{i-1,J} - N P_{i,J} + (N - i) P_{i+1,J}] + 0.5\eta [P_{i,J+1} - 2P_{i,J} + P_{i,J-1}] \\ & + C_0 \delta_{i,0} \delta_{J,0} + C_1 \delta_{i,1} \delta_{J,0} + C_2 \delta_{i,0} \delta_{J,1} + C_2 \delta_{i,0} \delta_{J,-1}, \end{aligned} \quad (8)$$

where

$$C_0 = [\alpha + \beta + 2\mu N + 2\eta] m_0 \exp(\alpha - \beta) t,$$

$$C_1 = -2\mu m_0 \exp(\alpha - \beta) t,$$

and

$$C_2 = -\eta m_0 \exp(\alpha - \beta) t,$$

where m_0 is the average number of individuals at $t = 0$, in any of the $L 2^N$ states. The quantity $P_{i,J}$ in the above expression stands for $\langle m_{i_1,J_1} m_{i_2,J_2} \rangle$.

Defining

$$Q_{i,J} = \binom{N}{i} P_{i,J},$$

we get

$$\begin{aligned} \frac{dQ_{i,J}}{dt} = & 2(\alpha - \beta) Q_{i,J} + 2\mu [(N - i + 1) Q_{i-1,J} - N Q_{i,J} + (i + 1) Q_{i+1,J}] \\ & + 0.5\eta [Q_{i,J+1} - 2Q_{i,J} + Q_{i,J-1}] + C_0 \delta_{i,0} \delta_{J,0} + N C_1 \delta_{i,1} \delta_{J,0} + C_2 \delta_{i,0} \delta_{J,1} + C_2 \delta_{i,0} \delta_{J,-1}. \end{aligned} \quad (9)$$

Defining $\tilde{Q}_{i,k}$, the Fourier transform of $Q_{i,J}$,

$$\tilde{Q}_{i,k} = \sum_{J=0}^{L-1} Q_{i,J} \exp(i 2\pi k J / L)$$

and

$$Q_{i,J} = (1/L) \sum_{k=0}^{L-1} \tilde{Q}_{i,k} \exp[-i2\pi kJ/L],$$

we get

$$\begin{aligned} \frac{d\tilde{Q}_{i,k}}{dt} = & 2(\alpha - \beta)\tilde{Q}_{i,k} + 2\mu[(N - i + 1)\tilde{Q}_{j-1,k} - N\tilde{Q}_{i,k} + (i + 1)\tilde{Q}_{j+1,k}] \\ & + 2\eta[\cos(2\pi k/L) - 1]Q_{i,k} + C_0\delta_{i,0} + C_1\delta_{i,1} + 2C_2\cos(2\pi k/L)\delta_{i,0}. \end{aligned} \quad (10)$$

Mean overlap $\langle q \rangle$

To obtain an expression for $\langle q \rangle$, we define a quantity $R_{1,J}$ as

$$R_{1,J} = \sum_{i=0}^N Q_{i,J}(1 - 2i/N),$$

so that $\bar{R}_{1,k}$, the Fourier transform of $R_{1,J}$, is

$$\begin{aligned} \bar{R}_{1,k} &= \sum_{J=0}^{L-1} R_{1,J} \exp[i2\pi kJ/L] \\ &= \sum_{i=0}^N Q_{i,k}(1 - 2i/N). \end{aligned}$$

Multiplying Eq. (10) by $(1 - 2i/N)$ and summing over $i = 0$ to N , we get

$$\begin{aligned} \frac{dR_{1,k}}{dt} &= \{ [2(\alpha - \beta) - 4\mu] + 2\eta[\cos(2\pi k/L) - 1] \} R_{1,k} \\ &+ C_0 + C_1(N - 2) + 2C_2\cos(2\pi k/L). \end{aligned} \quad (11)$$

The solution of this equation, subject to the initial condition $R_{1,k}(0) = m_0$, is

$$\begin{aligned} R_{1,k}(t) &= m_0 \exp(\gamma t) \\ &+ [S/(\alpha - \beta - \gamma)] [\exp(\alpha - \beta)t - \exp(\gamma t)], \end{aligned}$$

where

$$\gamma = 2(\alpha - \beta) - 4\mu - 2\eta[1 - \cos(2\pi k/L)]$$

and

$$\begin{aligned} S &= [C_0 + (N - 2)C_1 + 2C_2\cos(2\pi k/L)] \exp[-(\alpha - \beta)t] \\ &= \{ \alpha + \beta + 4\mu + 2\eta[1 - \cos(2\pi k/L)] \} m_0. \end{aligned}$$

The equation can be inverted for any value of t , and $\langle q \rangle$ as a function of time can be obtained. To simplify things, we take a large value of t and get

$$\begin{aligned} R_{1,0}(t) &= (m_0/L) \sum_{k=0}^{L-1} \exp(\{2(\alpha - \beta) - 4\mu - 2\eta[1 - \cos(2\pi k/L)]\}t) \\ &+ (m_0/L) \sum_{k=0}^{L-1} \frac{\alpha + \beta + 4\mu + 2\eta[1 - \cos(2\pi k/L)]}{-\alpha + \beta + 4\mu + 2\eta[1 - \cos(2\pi k/L)]} \\ &\quad \times (\exp(\alpha - \beta)t - \exp\{2(\alpha - \beta) - 4\mu - 2\eta[1 - \cos(2\pi k/L)]\}t). \end{aligned} \quad (12)$$

Using this expression for $R_{1,k}$ we get $\langle q \rangle$ as

$$\langle q \rangle = 2^N [R_{1,0}(t) - m_0 \exp(\alpha - \beta)t] / \langle M^2 - M \rangle. \quad (13)$$

To obtain the quantity $\langle M^2 - M \rangle$, we write the master equation for $P_1(\{M_J\}, t)$, which is defined as the probability of getting M_J individuals in compartment J ($J = 1, 2, \dots, L$) at time t .

$$\begin{aligned} \frac{dP_1(\{M_J\}, t)}{dt} &= \alpha \sum_{J=1}^L (\bar{A}_J - 1) M_J P_1(\{M_J\}, t) + \beta \sum_{J=1}^L (A_J - 1) M_J P_1(\{M_J\}, t) \\ &+ 0.5\eta \sum_{J=1}^L [(A_J A_{J-1} - 1) + (A_J \bar{A}_{J+1} - 1)] M_J P_1(\{M_J\}, t). \end{aligned} \quad (14)$$

Multiplying by $M_J(M_J - 1)$, summing, and using the initial conditions of uniform distribution, we get after some algebra, the expression

$$\langle M^2 - M \rangle = (M_0^2 - M_0/L) \exp[2(\alpha - \beta)t] + \left[\frac{2\alpha M_0}{L} \right] \sum_{k=0}^{L-1} [\exp\{2(\alpha - \beta) - 2\eta[1 - \cos(2\pi k/L)]\}t] \\ - \exp[(\alpha - \beta)t] / \{\alpha - \beta - 2\eta[1 - \cos(2\pi k/L)]\} . \quad (15)$$

Variance $\langle q^2 \rangle - \langle q \rangle^2$

To obtain $\langle q^2 \rangle$, we define a quantity $\bar{R}_{2,k}$ on the lines of $\bar{R}_{1,k}$ as

$$R_{2,k}(t) = \sum_{i=0}^N \bar{Q}_{i,k}(t) (1 - 2i/N)^2 .$$

Using this in Eq. (10), we see that $R_{2,k}$ satisfies the differential equation

$$\frac{dR_{2,k}}{dt} = [2(\alpha - \beta) - 8\mu] \\ - 2\eta[1 - \cos(2\pi k/L)]R_{2,k} + (8\mu/N)R_{0,k} \\ + C_0 + NC_1(1 - 2/N)^2 \\ + 2C_2 \cos(2\pi k/L) , \quad (16)$$

where

$$R_{0,k}(t) = \sum_{j=0}^{L-1} \sum_{i=0}^N Q_{i,j}(t) \exp(i2\pi kJ/L)$$

is the solution of the differential equation

$$\frac{dR_{0,k}}{dt} = 2\{(\alpha - \beta) - \eta[1 - \cos(2\pi k/L)]\}R_{0,k} \\ + C_0 + NC_1 + 2C_2 \cos(2\pi k/L) . \quad (17)$$

This coupled set of differential equations is to be solved subject to the initial conditions

$$R_{2,0}(t) = m_0 \exp(\alpha - \beta)t + \langle M^2 - M \rangle / N + (1 - 1/N)(2\alpha m_0 / L) \\ \times \sum_{k=0}^{L-1} \frac{\exp\{2(\alpha - \beta) - 8\mu - 2\eta[1 - \cos(2\pi k/L)]\}t - \exp[(\alpha - \beta)t]}{\alpha - \beta - 8\mu - 2\eta[1 - \cos(2\pi k/L)]} . \quad (20)$$

Using this value of $R_{2,0}(t)$, we can get $\langle q^2 \rangle$ as

$$\langle q^2 \rangle = 2^N [R_{2,0}(t) - m_0 \exp(\alpha - \beta)t] / \langle M^2 - M \rangle . \quad (21)$$

The term $m_0 \exp(\alpha - \beta)t$ in the above expression arises out of the fact that the number of pairs in the same state is

$$m_{0,0}(m_{0,0} - 1) = m_{0,0}^2 - m_{0,0}$$

whereas in $R_{2,0}$, $m_{0,0}^2$ is used. The expectation value of $m_{0,0}$ is $m_0 \exp(\alpha - \beta)t$. Let us now consider the special cases $\alpha = \beta$, $L = 1$, etc.

Case I: ($\alpha = \beta$). When $\alpha = \beta$, we get, from Eq. (15),

$$\langle M^2 - M \rangle = (M_0^2 - M_0/L) + 2\alpha t M_0 / L \\ + (2\alpha M_0 / L) \sum_{k=1}^{L-1} (1 - \exp\{-2\eta[1 - \cos(2\pi k/L)]\}t) / \{2\eta[1 - \cos(2\pi k/L)]\}t . \quad (22)$$

Using Eqs. (12), (13), and (15), we get for $\langle q \rangle$ the expression

$$R_{2,k}(0) = L2^N m_0 (1 - 1/Lm_0) \delta_{k,0} / N + m_0 , \quad (18)$$

$$R_{0,k}(0) = L2^N m_0 (1 - 1/Lm_0) \delta_{k,0} + m_0 .$$

After some algebra, we get for $R_{2,k}$ the equation

$$R_{2,k}(t) = Y_1 \exp(\gamma_2 t) \\ + (8/N) Y_2 [\exp(\gamma_1 t) - \exp(\gamma_2 t)] / (\gamma_1 - \gamma_2) \\ + Y_3 [\exp(\alpha - \beta)t - \exp(\gamma_2 t)] / (\alpha - \beta - \gamma_2) , \quad (19)$$

where

$$Y_1 = m_0 + L2^N (m_0^2 - m_0 / L2^N) \delta_{k,0} / N ,$$

$$Y_2 = L2^N (m_0^2 - m_0 / L2^N) \delta_{k,0} \\ + 2\alpha m_0 / \{\alpha - \beta - 2\eta[1 - \cos(2\pi k/L)]\} ,$$

$$Y_3 = S_1 / (\alpha - \beta - \gamma_1) + S_2 ,$$

$$S_1 = [C_0 + NC_1 + 2C_2 \cos(2\pi k/L)] \exp[-(\alpha - \beta)t] \\ = \{\alpha + \beta + 2\eta[1 - \cos(2\pi k/L)]\} m_0 ,$$

$$S_2 = [C_0 + C_1(N - 2)^2 / N \\ + 2C_2 \cos(2\pi k/L)] \exp[-(\alpha - \beta)t] \\ = \{\alpha + \beta + 8\mu(1 - 1/N) + 2\eta[1 - \cos(2\pi k/L)]\} m_0 ,$$

$$\gamma_1 = 2(\alpha - \beta) - 2\eta[1 - \cos(2\pi k/L)] ,$$

$$\gamma_2 = 2(\alpha - \beta) - 8\mu - 2\eta[1 - \cos(2\pi k/L)] .$$

This equation can be inverted for any t to obtain the time dependence of $\langle q^2 \rangle$. We then get

$$\langle q \rangle = \frac{(M_0/L) \sum_{K=0}^{L-1} \frac{\alpha(1 - \exp\{-4\mu - 2\eta[1 - \cos(2\pi K/L)]\}t)}{2\mu + \eta[1 - \cos(2\pi K/L)]}}{(M_0^2 - M_0/L) + 2\alpha t M_0/L + (2\alpha M_0/L) \sum_{K=1}^{L-1} (1 - \exp\{-2\eta[1 - \cos(2\pi K/L)]\}t) / \{2\eta[1 - \cos(2\pi K/L)]\}} \quad (23)$$

For very large values of t ,

$$\langle q \rangle = \frac{(M_0/L) \sum_{K=0}^{L-1} \frac{\alpha}{2\mu + \eta[1 - \cos(2\pi K/L)]}}{(M_0^2 - M_0/L) + 2\alpha t M_0/L + (2\alpha M_0/L) \sum_{K=1}^{L-1} (1 - \exp\{-2\eta[1 - \cos(2\pi K/L)]\}t) / \{2\eta[1 - \cos(2\pi K/L)]\}} \quad (24)$$

Note that the expression for $\langle q \rangle$ does not contain N , the number of genes per genome. This is to be expected since all locations are identical (because of the uniform initial conditions), and the overlap function can be related to the probability that genes in the corresponding locations in two different individuals are the same. It may also be noted that Eq. (23) gives the overlap function for a single compartment. Similar expressions can be written for the overlap function averaged over the whole population and for the average overlap function between two individuals coming from two different compartments [the index J in Eq. (8) corresponds to the distance between compartments]. We have, however, not pursued this in the present work.

Equation (23) is for a general expression for $\langle q \rangle$ for any arbitrary L . We now specifically obtain the asymptotic behavior for $L=1$, $L=2$, and $L \rightarrow \infty$.

$L=1$ case. Equation (23) for $\langle q \rangle$ simplifies to

$$\langle q \rangle = \frac{\alpha M_0 [1 - \exp(-4\mu t)]}{2\mu(M_0^2 + 2\alpha M_0 t - M_0)} \quad (25)$$

$$\langle q \rangle = \frac{(\alpha M_0/2)([1 - \exp(-4\mu t)]/(2\mu) + \{1 - \exp[-4(\mu + \eta)t]\}/(2\mu + 2\eta))}{M_0^2 - M_0/2 + \alpha t M_0 + \alpha M_0 [1 - \exp(-4\eta t)]/4\eta} \sim (\alpha/4M_0\mu)[1 + 1/(1 + \eta/\mu)] \quad (\text{for } M_0 \gg 2\alpha t \text{ and } t \gg 1/4\mu). \quad (27)$$

$\langle q \rangle$ now depends on the value of η/μ , the relative importance of migration with respect to mutation, also. When η/μ is small, we get back the result of the single compartment [Eq. (26a)]. This is because in this case the two compartments become almost insulated from each other. When $\eta/\mu \gg 1$, the second term in the above expression drops out and we get $\langle q \rangle \sim \alpha/2\mu(2M_0)$. A large value of η/μ means that the two compartments act as a single compartment.

$L \rightarrow \infty$ case. Here we consider the case $\mu t \gg 1$ and $M_0 \gg 2\alpha t$. Equation (23), in this case, can be written as

$$\langle q \rangle \sim (1/M_0 L) \sum_{K=0}^{L-1} \alpha / \{2\mu + \eta[1 - \cos(2\pi K/L)]\} \quad (28a)$$

$$\langle q \rangle \sim \begin{cases} \alpha/2\mu M_0 & (\text{for } M_0 \gg 2\alpha t \text{ and } t \gg 1/4\mu) \\ 1/4\mu t & (\text{for } 2\alpha t \gg M_0 \text{ and } t \gg 1/4\mu). \end{cases} \quad (26a, 26b)$$

The value given by Eq. (26a) is very similar to that obtained by Serva and Peliti [2] if we interchange their μ (mutations per generation) with μ/α in Eq. (26a). The average time taken for an individual to die or to duplicate itself is $1/(\alpha + \beta) = 1/2\alpha$ and therefore corresponds to the generation time. This agreement between our result and that of Serva and Peliti is expected since in this range of time, the fluctuations in the population ($\sim \sqrt{2\alpha t M_0}$) are small compared to the initial size of the population, and hence the constant-population model applies.

At much larger times, Eq. (26b) applies. Here, the fluctuations are larger than the average population size. In fact as $t \rightarrow \infty$, the populations tend to become extinct with probability approaching 1. The few realizations which survive have a large m , and the net value of $\langle q \rangle$ (tending to 0) given by Eq. (26b) is due to these realizations. Note that the averaging procedure [see Eq. (3)] gives a higher weight to the larger populations.

$L=2$ case. When there are two compartments, Eq. (23) may be written as

$$\sim (\alpha/M_0 L) \int_0^L dK / [2\mu + 2\eta\pi^2 K^2/L^2] \quad (28b)$$

$$= [\alpha/(2M_0\pi\sqrt{\eta\mu})] \tan^{-1}[\pi\sqrt{(\eta/\mu)}]. \quad (28c)$$

When $\eta \rightarrow 0$, Eq. (28c) reduces to Eq. (26a). As

$$\tan^{-1}[\pi\sqrt{(\eta/\mu)}] \leq \pi\sqrt{(\eta/\mu)},$$

$\langle q \rangle$ given by Eq. (28c) (for $\eta \neq 0$) is less than that given for a single compartment. The other extreme limit of

complete mixing is obtained for $\eta/L^2 \gg \mu$. In this case, only the first term in the sum in Eq. (28a) contributes and $\langle q \rangle \sim \alpha/2\mu M_0 L$. This result can be easily understood by noting that L^2/η is the order of time for migration to the

L th compartment and $1/\mu$ is the average time for a mutation.

For $\alpha = \beta$, the expression for $\langle q^2 \rangle$ given by Eq. (21) reduces to

$$\langle q^2 \rangle = \frac{1}{N} + \frac{2^N (2\alpha m_0/L)(1-1/N) \sum_{K=0}^{L-1} (1 - \exp\{8\mu - 2\eta[1 - \cos(2\pi K/L)]\}t) / \{8\mu + 2\eta[1 - \cos(2\pi K/L)]\}}{(M_0^2 - M_0/L) + 2\alpha t M_0/L + (2\alpha M_0/L) \sum_{K=1}^{L-1} (1 - \exp\{-2\eta[1 - \cos(2\pi K/L)]\}t) / \{2\eta[1 - \cos(2\pi K/L)]\}} . \quad (29)$$

For $L = 1$, the above expression simplifies to

$$\langle q^2 \rangle \sim \begin{cases} 1/N + \alpha(1-1/N)/(4\mu M_0) & (\text{for } M_0 \gg 2\alpha t \text{ and } t \gg 1/4\mu) \\ \alpha/(4\mu M_0) & (\text{for } N \rightarrow \infty) . \end{cases} \quad (30a)$$

Expression (30b) for $\langle q^2 \rangle$ is similar to that obtained by Serva and Peliti in the limit $N \rightarrow \infty$. For finite N , $\langle q^2 \rangle$ [Eq. (30a)] shows a $1/N$ dependence which may even dominate if $N < \alpha/4\mu M_0$.

Case II: ($\alpha > \beta$). For a growing population ($\alpha > \beta$), we get from Eqs. (12), (13), and (15), after some algebra,

$$\langle q \rangle = [2\alpha/(\alpha - \beta - 4\mu)] \{ \exp(-4\mu t) - \exp[-(\alpha - \beta)t] \} / \{ [M_0 + (\alpha + \beta)/(\alpha - \beta)] - [2\alpha/(\alpha - \beta)] \exp[-(\alpha - \beta)t] \} . \quad (31)$$

This equation gives the relative interplay of the population growth parameters (α and β) and the gene mutation characteristics (μ). We see that the large- t behavior of $\langle q \rangle$ is exponential. The leading term is $\exp[-(\alpha - \beta)t]$ or $\exp(-4\mu t)$ according to whether $\alpha - \beta < 4\mu$ or $\alpha - \beta > 4\mu$. When $\alpha - \beta = 4\mu$, the variation is $t \exp(-4\mu t)$. That is, for large t , we get

$$\langle q \rangle \sim \begin{cases} [2\alpha/(\alpha - \beta - 4\mu)] \exp(-4\mu t) / [M_0 + (\alpha + \beta)/(\alpha - \beta)] & (\text{for } \alpha - \beta > 4\mu) \\ [2\alpha/(4\mu - \alpha + \beta)] \exp[-(\alpha - \beta)t] / [M_0 + (\alpha + \beta)/(\alpha - \beta)] & (\text{for } \alpha - \beta < 4\mu) \\ 2\alpha t [\exp(-4\mu t)] / [M_0 + (\alpha + \beta)/(\alpha - \beta)] & (\text{for } \alpha - \beta = 4\mu) . \end{cases}$$

Further, it is easy to see that the first term on the right-hand side of Eq. (30a) dominates and therefore

$$\langle q^2 \rangle \sim 1/N .$$

We find that for finite N , as $t \rightarrow \infty$, $\langle q \rangle \rightarrow 0$, whereas $\langle q^2 \rangle \rightarrow 1/N$ is nonzero. This is because the value of q varies from -1 to $+1$, whereas q^2 is always non-negative, varying only between 0 and 1.

Probability distribution of q

Once again, we define $\langle p(q) \rangle$ as the weighted average over the number of pairs in the realization, i.e.,

$$\langle p(q_i) \rangle = [Q_{i,0} - \delta_{i,0} \langle m \rangle] / \langle M^2 - M \rangle , \quad (32)$$

where $q_i = (N - 2i)/N$ for $i = 0, 1, 2, \dots, N$.

To make things simple, we consider only the case $\alpha = \beta$ and $L = 1$. Equation (9) then reduces to

$$\begin{aligned} \frac{dQ_i(t)}{dt} &= 2\mu[(N - i + 1)Q_{i-1} + (i + 1)Q_{i+1} - (N - i)Q_i] \\ &+ C_0 \delta_{i,0} + NC_1 \delta_{i,1} . \end{aligned} \quad (33)$$

(As we are considering only one compartment, so the suffix J of Q is dropped.)

Defining the generation function Q_z as

$$\tilde{Q}_z = \sum_{i=0}^N Q_i z^i .$$

We have

$$\frac{d\tilde{Q}_z(t)}{dt} = 2\mu(1-z)^2 \partial \tilde{Q}_z(t) / \partial z - 2\mu N(1-z) \tilde{Q}_z(t) + S , \quad (34)$$

where $S = C_0 + NC_1 z$. This partial differential equation has to be solved subject to the initial condition

$$\tilde{Q}_z(0) = (m_0^2 - m_0/2^N)(1+z)^N + m_0 . \quad (35)$$

Using the method of characteristics, one obtains the solution of the equation, after some algebra, as

$$\begin{aligned} \tilde{Q}_z(t) &= (m_0^2 - m_0/2^N)(1+z)^N \\ &+ (m_0/2^N)[(1+z) + (1-z)\exp(-4\mu t)]^N \\ &+ [f(z') - f(z)](1+z)^N , \end{aligned} \quad (36)$$

where

$$z' = \frac{[(1+z) - (1-z)\exp(-4\mu t)]}{[(1+z) + (1-z)\exp(-4\mu t)]}$$

and

$$f(z) = \int \{(C_0 + NC_1 z) / [(1+z)^{N+1}(1-z)]\} dz .$$

On inverting this expression, we get for $Q_i(t)$ the relation

$$Q_i(t) = m_0 \delta_{i,0} + (m_0^2 - m_0 / 2^N) \binom{N}{i} + T_i(N) , \quad (37)$$

where $T_i(N)$ satisfies the recursion relations

$$\begin{aligned} T_i(N) = & (m_0 \alpha / 2\mu N) \left[\delta_{i,0} - \binom{N}{i} \{ [1 + \exp(-4\mu t)] / 2 \}^N \right] \\ & \times \{ [1 - \exp(-4\mu t)] / [1 + \exp(-4\mu t)] \}^i \\ & + 0.5 [T_i(N-1) + T_{i-1}(N-1)] , \end{aligned} \quad (38)$$

with

$$\begin{aligned} T_0(1) &= (m_0 \alpha / 4\mu) [1 - \exp(-4\mu t)] + m_0 \alpha t , \\ T_1(1) &= -(m_0 \alpha / 4\mu) [1 - \exp(-4\mu t)] + m_0 \alpha t . \end{aligned} \quad (39)$$

RESULTS AND CONCLUSIONS

Using the recursion relations given by Eqs. (37) and (38) in Eq. (32), we obtained $\langle p(q) \rangle$ for a number of values of N . Some simulation studies were also carried out for $N \leq 11$. It was seen that $\langle p(q) \rangle$, $\langle q \rangle$, and $\langle q^2 \rangle$ obtained by the two methods agreed very well.

Figure 1 shows the plot of $\langle p(q) \rangle$ versus q for a few representative values of N [computed using Eq. (37)]. It is seen that initially (at low N) the distribution is almost symmetrical about $q = 0$. As N increases, the distribution

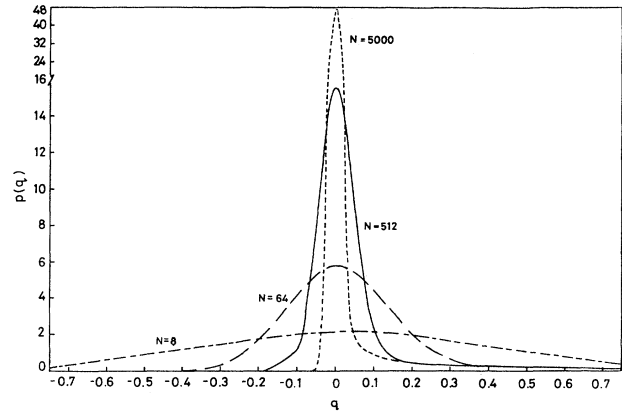


FIG. 1. Probability distribution of q for $N = 8, 64, 512$, and 5000 .

becomes skewed and tends to fall off more sharply for values of $q < 0$. Derrida and Peliti [1] obtained an expression for $\langle p(q) \rangle$ in the infinite- N limit which has nonzero values only for $q > 0$.

For constant population ($\alpha = \beta$), single compartment ($L = 1$), and $t \ll M_0$, we get for $\langle q \rangle$ and $\langle q^2 \rangle$ the same expressions as Serva and Peliti [2] for $N \rightarrow \infty$. For finite N , $\langle q^2 \rangle$ contains terms of order $1/N$, whereas $\langle q \rangle$ is independent of N . For $t \gg M_0$, $\langle q \rangle = 0$, but $\langle q^2 \rangle$ still is nonzero. For $\alpha > \beta$, $\langle q \rangle \rightarrow 0$, whereas $\langle q^2 \rangle$ is not zero. For small N the probability distribution obtained by us differs considerably from the distribution (for $N \rightarrow \infty$) obtained by Derrida and Peliti [1]. According to our expression, $\langle p(q) \rangle$ is nonzero for q ranging from -1 to $+1$, whereas in their case, $p(q)$ is nonzero only for $q > 0$.

- [1] B. Derrida and L. Peliti, *Bull. Math. Biol.* **53**, 355 (1991).
 [2] M. Serva and L. Peliti, *J. Phys. A* **24**, L705 (1991).
 [3] P. G. Higgs and B. Derrida, *J. Phys. A* **24**, L985 (1991).
 [4] T. Maruyama and M. Kimura, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6710 (1980).

- [5] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, England, 1983).
 [6] N. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).