Finite population size effects in quasispecies models with single-peak fitness landscape

2012 EPL 98 18001

(http://iopscience.iop.org/0295-5075/98/1/18001)

A Letters Journal Exploring the Frontiers of Physics

April 2012

EPL, **98** (2012) 18001
doi: 10.1209/0295-5075/98/18001

www.epljournal.org

Finite population size effects in quasispecies models with single-peak fitness landscape

David B. Saakian[1,2,3(a)], Michael W. Deem[4] and Chin-Kun Hu[1(b)]

[1] *Institute of Physics, Academia Sinica - Nankang, Taipei 11529, Taiwan*
[2] *A.I. Alikhanyan National Science Laboratory (Yerevan Physics Institute) Foundation - 2 Alikhanian Brothers St., Yerevan 375036, Armenia*
[3] *National Center for Theoretical Sciences: Physics Division, National Taiwan University - Taipei 106, Taiwan*
[4] *Department of Physics & Astronomy, Rice University - Houston, TX 77005-1892, USA*

received 5 December 2011; accepted in final form 2 March 2012
published online 10 April 2012

PACS 87.23.Kg – Dynamics of evolution
PACS 87.10.Mn – Stochastic modeling

Abstract – We consider finite population size effects for Crow-Kimura and Eigen quasispecies models with single-peak fitness landscape. We formulate accurately the iteration procedure for the finite population models, then derive the Hamilton-Jacobi equation (HJE) to describe the dynamic of the probability distribution. The steady-state solution of HJE gives the variance of the mean fitness. Our results are useful for understanding the population sizes of viruses in which the infinite population models can give reliable results for biological evolution problems.

Introduction. – Investigation of biological evolution models [1–6], such as the Eigen model [3,4] and the Crow-Kimura model [5], by methods of statistical or theoretical physics is highly fruitful in evolution research. The methods used include quantum-statistical mechanics [7–11]; quantum field theory [10–14]; Hamilton-Jacobi equation (HJE) [15–18], etc. Such approach has given many exact results for evolution models [3–19], solved a paradox of the origin of life [20], and produced exact finite genome length corrections for the mean fitness and gene probabilities in some evolution models [21].

In the original formulation of the Eigen and Crow-Kimura models, the configurations of the genome of length L are represented by $M \equiv 2^L$ spin configurations $(s_1, s_2, \ldots, s_L)$, where s_k for $1 \leqslant k \leqslant L$ take $+1$ or -1. Such representation was used by Peng *et al.* to study the long-range correlation in the nucleotide sequences [22]. The M configurations S_i are labelled by $0 \leqslant i \leqslant M-1$ and the i-th configuration S_i is assigned a number r_i to represent the reproduction rate or fitness of that configuration and another number p_i to represent the probability in that configuration. Such p_i satisfies the normalization condition: $\sum_{i=0}^{M-1} p_i = 1$. The coupled differential equations satisfied by p_i for the Crow-Kimura model [5] and the Eigen model [3,4] are given in appendix A and appendix B, respectively. However, such coupled differential equations are valid only in the limit of infinite population size N, which is not the case in many real systems, *e.g.* a virus in a given environment. Thus the study of the finite population size problem has attracted much attention in recent decades [24–31]. While the case of two alleles (types of genes) in the Wright-Fisher model [1] and the Moran model [2] can be analytically solved [6], the realistic case of evolution with many sequences (genomes) stays intractable by traditional methods. In [25] the additive fitness landscape has been considered, when the contributions of different alleles to the fitness are random numbers and in [26] a finite population was considered in which the finesses of different sequences are independent random numbers.

The purpose of the present letter is to formulate the Crow-Kimura model and the Eigen model for finite population size and solve them for the single-peak fitness landscape, popular in quasispecies literature. In such a landscape, the fitness of a configuration, say S_0, is larger than the fitness of other configurations, *i.e.* $r_0 > r_i$ for $i > 0$ and all r_i are equal for $i > 0$. We first formulate the iteration procedure for the finite population models, then derive HJE to describe the dynamic of the probability distribution. The steady-state solution of HJE gives the variance of the mean fitness. Our results are useful for understanding the population sizes of a virus in which the coupled differential equations can give reliable

(a) E-mail: saakian@phys.sinica.edu.tw
(b) E-mail: huck@phys.sinica.edu.tw

results for biological evolution problems. Our results are exact derivations in comparison with the numerics or uncontrolled approximations of the vast majority of finite population articles.

Consider the case when the total number of different genotypes N_g is either $N_g \sim 4^L$, where L is the number of nucleotides in the virus genome, or $N_g \sim 2^L$, where L is the number of two types of alleles or (spins), located in different places (loci). The infinite population case is when the population size N is large enough to have a large number of viruses of any type. The convergence of evolutionary dynamics with population size depends on the mutation rate and the fitness landscape. In the infinite population limit the evolution equations are deterministic, and, for molecular evolution models [3–7], there are many exact results [7–18]. It is possible even to find exact solutions for the steady state and dynamics [15–18]. In biology, the populations are often relatively small. In the case of viruses, the population size can be rather large, even 10^{12}. Then the collective characteristics of the evolving population will fluctuate.

Finite population Crow-Kimura model with symmetric fitness landscape. – We consider the symmetric fitness landscape, popular in virology. The genome is a collection of letters (spins) ± 1, denoting the alleles type. Thus a sequence is identified with a spin configuration of a one-dimensional Ising model. By mutation any letter may randomly change to the other value.

An important characteristics of the sequence space is the number of neighbors L and the number of sequences at the Hamming distance d (the sequences have d alleles different from the reference sequence) $N_d = \frac{L!}{d!(L-d)!} \sim \frac{L^d}{d!}$ for the two-loci case at $d \ll L$. We consider a simple fitness landscape, popular in population genetics, when the sequence from the l-th Hamming class has a fitness w_l.

As evolution is a stochastic process, we should work with probabilities. We are interested in the steady-state properties of the evolution model. We consider the finite population version of the Crow-Kimura model, see the appendix. In this case we consider the evolution as a Markov process, where the state (the state of the population is characterized by the number of individuals for the different possible numbers of mutations) is defined by $L+1$ integers, the numbers of different Hamming types. During one evolution step, there are three processes: birth of new individuals proportional to the fitness of the corresponding Hamming class, transitions between Hamming classes proportional to l/L to the lower Hamming class and transitions proportional to $(L-l)/L$ to the higher Hamming class [8,19]. These factors l/L and $(L-l)/L$ have been derived in [8,19] for the infinite population models, and should be applied to the discrete-time scheme of the finite population models as well. The iteration step is completed by the random reduction of the population to the initial size, N. This evolution dynamics described here corresponds to the Moran model with many alleles. Compared to the ordinary multi-allele Moran model [2], in our case there is a non-trivial geometry in the sequence space, defined by the Hamming distance.

We first define our model for the case of a general symmetric fitness landscape with Wrightian fitness $\hat{r}_l = e^{Ur_l}$, in the l-th Hamming class, $0 \leqslant l \leqslant L$, where r_l is the fitness defined earlier. Here the average number of mutations of genome per one replication period is $U \equiv \gamma_0 \tau$, where γ_0 is a mutation rate per genome in the continuous-time parallel model, and τ is the time step. At any moment the state of our model is characterized via $L+1$ integers n_l. We choose $\gamma_0 = 1$, therefore $U = \tau$. We consider the $U \ll 1$ limit. In this case the steady-state results and dynamics are U independent (U gives just the scale).

During the iteration step we consider the following processes:

– A. Random growth with $n_l \to n_l + \delta n_l$. The δn_l is a random binomial process with probability $U\hat{r}_l$ and n_l trials.

– B. Mutations.

There are f_l forward mutations from the class l.

We consider random integer number f_l with binomial distribution with the probability parameter $(1 - \frac{l}{L})U$ and n_l trials. There are b_l back mutations from the class l.

We consider a random integer number b_l with binomial distribution with the probability parameter $\frac{l}{L}U$ and n_l trials. Due to back mutations we have the following change of n_l: $n_l \to n_l - b_l + b_{l+1}$.

– C. We randomly remove $\sum_l \delta n_l$ individuals from the population to keep a fixed population size.

Sharp peak (single-peak) model. – Consider the Wrightian fitness with $\hat{r}_0 = e^{\epsilon J}$ for the peak sequence, $U \ll 1$ is the number of mutations per generation, J is a fitness gap in the corresponding continuous-time parallel selection-mutation model, and $r_i = 1$ for $i \geqslant 1$. Our goal is to investigate how the mean fitness depends on the finite population size.

In the case of infinite population, one can calculate the number of viruses with the peak sequence using a single equation, with $1/L$ accuracy. Assume that there is some probability distribution $\rho(n)$ for the number n of viruses with the peak sequence, which satisfies the normalization condition $\sum_{n=0}^{N} \rho(n) = 1$. Then we can derive both the steady-state distribution, which is a rather simple function, and even the exact dynamics, which is a complicated expression for $\rho(n)$.

We consider a discrete-time scheme of evolution with small U. During each iteration we consider the steps A, B, C. In step A, there are δn new viruses with the peak

sequence. The number δn is derived via a binomial n sampling with a small probability UJ. During the step C of reduction to keep a constant population size anyone of these δn viruses could be removed from the system. The total number of viruses removed from the peak sequence m is calculated via the binomial distribution with δn trials and probability $x \equiv n/N$. Therefore, the result of the A and C steps should be a sampling of n particles with a probability $UJ(1-x)$. Thus after steps A and C the original n changes as $n \to n+h$, where h has a binomial distribution with a probability parameter $p = UJ(1-x)$, and the number of trials is n. During the step B of mutation, $n \to n-m$, where m has a binomial distribution with a probability parameter U and the number of trials is n. Thus after one iteration $n \to n+h-m$.

If we have a distribution $\rho(t,n)$ at the t-th moment of time, then after an iteration with the period of time U we have a distribution

$$\rho(t+U,n) = \langle \rho(t, n-h+m) \rangle \tag{1}$$

when the averaging is over the (binomial) distributions of h and m.

Let us assume the following anzats for the probability distribution at time t:

$$\rho(t,n) = \exp[N\phi(t,x)],\ x = n/N. \tag{2}$$

After an iteration

$$e^{N\phi(t+U,x)} = \int \mathrm{d}t e^{N\phi(x)} \langle e^{-(h-m)\phi'(t,x)} \rangle|_{h,m}, \tag{3}$$

where $\phi' \equiv \frac{\partial \phi(t,x)}{\partial x}$. As we used binomial probability distributions in the iteration step, we should perform an average via the binomial distribution in eq. (3). We use the following formula of the binomial distribution of h with success probability p and M trials:

$$\begin{aligned} \langle e^{hk} \rangle &\equiv \sum_{h=0}^{M} e^{hk} p^h (1-p)^{M-h} \frac{M!}{h!(M-h)!} \\ &= (1 + p(e^k - 1))^M \\ &\approx e^{pM(e^k-1)}. \end{aligned} \tag{4}$$

We consider the case of $p \ll 1$.

Taking $k = -\phi', M = Nx$ and $p = UJ(1-x)$ in eq. (4) (see the definition of iteration steps A, C) we find

$$\begin{aligned} \langle e^{-h\phi'} \rangle &= [(1-x)UJe^{-\phi'} + 1 - UJ(1-x)]^{Nx} \\ &\approx e^{JNUx(1-x)(e^{-\phi'}-1)}. \end{aligned} \tag{5}$$

In the same way we consider the mutation, taking $k = \phi', p = U, M = xN$ we derive

$$\langle e^{\phi' m} \rangle = [Ue^{\phi'} + 1 - U]^{xN} = \exp[NUx(e^{\phi'} - 1)]. \tag{6}$$

Combining eqs. (5), (6) and holding only the linear terms in U, we obtain the following expression:

$$\phi(t+U,x) = \phi(t,x) + UxJ(1-x)(e^{-\phi'} - 1) + Ux(e^{\phi'} - 1) \tag{7}$$

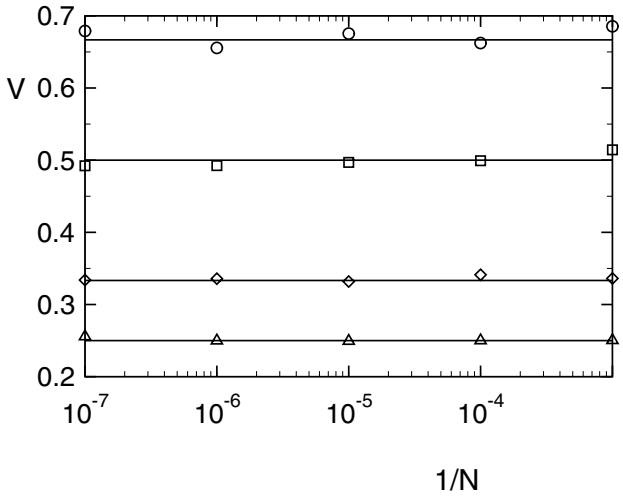

Fig. 1: Verification of eq. (10) for variance: $V = 1/J$. The horizontal lines from top to bottom are the analytic results for $J = 1.5, 2, 3$ and 4, respectively. Circles, squares, diamonds, and triangles represent numerical data for $J = 1.5, 2, 3$ and 4, respectively. Analytic and numerical results are quite consistent with each other.

or

$$\frac{\partial \phi(t,x)}{\partial t} = xJ(1-x)(e^{-\phi'} - 1) + x(e^{\phi'} - 1). \tag{8}$$

In the steady state we just have an ordinary differential equation for ϕ. We derive the following non-trivial solution $\phi_0(x) = \phi(\infty, x)$ and the corresponding distribution

$$\begin{aligned} \phi_0(x) &= \int_{x_0}^{x} \mathrm{d}x \ln J(1-x) = (x - x_0)\ln J \\ &\quad + (1-x)(1 - \ln(1-x)) - (1-x_0)(1 - \ln(1-x_0)), \\ \rho(x) &= \sqrt{\frac{NJ}{2\pi}} \exp[N\phi(x)], \end{aligned} \tag{9}$$

where we added the pre-factor $\sqrt{\frac{NJ}{2\pi}}$ to ensure the condition that the total probability is 1. Here the distribution has a maximum at $x = x_0 \equiv (1 - 1/J)$, see [9], and $\phi(x_0)'' = -J$. Then we derive for the variance:

$$V \equiv (\langle x^2 \rangle - \langle x \rangle^2)N = \frac{1}{J}. \tag{10}$$

Thus we derived the whole steady-state distribution via eqs. (2), (8), and the expression for the variance, eq. (10).

Equation (10) is verified numerically in fig. 1 for $J = 1.5, 2, 3, 4$. One could follow the method used in [18] to solve eq. (8) and get the time evolution of $\phi(t,x)$.

Finite population version of the Eigen model. – Consider now the finite population version of the Eigen model with zero degradation. There are n viruses at the peak sequence.

At any discrete moment of time we consider three processes:

A. The number of viruses in the class l grows with a probability Ur_l. There are mutations. New viruses mutate with a finite mutation probability $1-Q$,

C. There is a dilution of the whole population, keeping strictly the total population size as N.

Consider again the single-peak fitness, $r_0 = A$, and for $l > 0, r_l = 1$. n is the number of viruses with the peak sequence, and $x = n/N$.

Let us give the details of the processes A and B.

A1. Reproduction in the peak sequence S_0: We randomly choose l elements from a pool of n elements and each element is chosen independently with a probability UA. Thus the probability to get l elements is

$$\rho_1(l) = \frac{n!}{l!(n-l)!}(UA)^l(1-UA)^{n-l}. \tag{11}$$

l is the number of new sequences at the peak sequence.

A2. Reproduction in the other sequences, *i.e.* S_i for $i > 0$: We randomly choose k elements from a pool of $(N-n)$ elements and each element is chosen independently with probability U. Thus the probability distribution to get k elements is

$$\rho_2(k) = \frac{(N-n)!}{k!(N-n-k)!}U^k(1-U)^{N-n-k}. \tag{12}$$

After the A1 and A2 steps there are $n+l$ viruses at the peak sequences and $N-n+k$ sequences at other sequences.

B. We randomly choose m elements from a pool of l elements in S_0 and each element is chosen independently with probability $Q = \exp[-\gamma]$ to be in S_0. Thus the probability to get m elements in S_0 is

$$\rho_3(m) = \frac{l!}{m!(l-m)!}Q^m(1-Q)^{l-m}. \tag{13}$$

After the step B, there are $n+m$ viruses in the peak sequence S_0 and $N-n+k+(l-m)$ viruses in other sequences. Thus there are $N+k+l$ sequences in S_i for $0 \leqslant i \leqslant M-1$. In the next step, we will uniformly remove $l+k$ sequences so that the total population is still N.

C. We randomly choose h elements from a pool of $l+k$ elements in S_0 and each element is chosen independently with probability x. Thus the probability to remove h elements from S_0 is

$$\rho_4(h) = \frac{x^h(1-x)^{(l+k-h)}}{h!(l+k-h)!}. \tag{14}$$

Besides, we remove $l+k-h$ elements from S_i for $i > 0$. We have that during one iteration step $n \to n+m-h$, therefore we need to find the average $\langle e^{-\phi'(m-h)}\rangle$ via the distributions $\rho_1(l)\rho_2(k)\rho_3(m)\rho_4(h)$. We consider

$$\langle e^{-\phi'(m-h)}\rangle = \sum_{l,k,m,h} \frac{n!(UA)^l(1-UA)^{n-l}}{l!(n-l)!}\frac{l!Q^m(1-Q)^{l-m}e^{-\phi' m}}{m!(l-m)!} \times \frac{(N-n)!U^k(1-U)^{N-n-k}}{k!(N-n-k)!}\frac{(l+k)!x^h(1-x)^{(l+k-h)}}{h!(l+k-h)!}e^{\phi' h}. \tag{15}$$

First we transform

$$\sum_m \frac{l!Q^m(1-Q)^{l-m}e^{-\phi' m}}{m!(l-m)!} = (Qe^{-\phi'}+1-Q)^l. \tag{16}$$

Using the transformation

$$\sum_h \frac{(l+k)!x^h(1-x)^{(l+k-h)}}{h!(l+k-h)!}e^{\phi' h} = (1-x+xe^{\phi'})^{l+k}, \tag{17}$$

we obtain

$$\sum_{l,k} \frac{n!(UA)^l(1-UA)^{n-l}}{l!(n-l)!}\frac{(N-n)!U^k(1-U)^{N-n-k}}{k!(N-n-k)!} \times (Qe^{-\phi'}+1-x)^l(1-x+xe^{\phi'})^{l+k}. \tag{18}$$

The sum over l, k gives an equation

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = F(\phi'), \tag{19}$$

where

$$F(\phi') = xA[(Qe^{-\phi'}+1-Q)(xe^{\phi'}+1-x)-1] + x(1-x)(e^{\phi'}-1). \tag{20}$$

We need to consider the first two terms in the ϕ' expansion

$$F(\phi') \approx -x[(QA-1)-(A-1)x]\phi' + x[QA(1-2x)+(A-1)x+1]\frac{\phi'^2}{2}. \tag{21}$$

In the steady state we consider $F(\phi') = 0$. We expand eq. (21) in powers of $y \equiv x - \frac{(QA-1)}{(A-1)}$ and find the following steady-state solution:

$$\phi' = -2\frac{(A-1)y}{Q(1-Q)\frac{2A^2}{(A-1)} - (2QA+1-A)y}. \tag{22}$$

Therefore,

$$\phi''(0) = -\frac{(A-1)^2}{Q(1-Q)A^2}, \tag{23}$$

and eventually we obtain for the variance V of distribution

$$V = N\langle y^2\rangle \equiv N(\langle p_0^2\rangle - \langle p_0\rangle^2) = \frac{Q(1-Q)A^2}{(A-1)^2}. \tag{24}$$

In appendix C, we derive the steady-state distribution and the variance for the Eigen model with degradation eq. (C.6).

Discussion. – The investigation of the finite population problem is the hardest mathematical problem in evolution theory. In this article we solved exactly some aspects of the finite population version of the Crow-Kimura and Eigen model with degradation. We calculated the variance of the distribution for the mean fitness at equilibrium. Our equation could be applied to calculate the dynamics of the distribution as well.

The quasispecies model, especially the one with single-peak fitness and its simple generalizations, has a lot of applications in the virus evolution [32], cancer modeling [33] and molecular evolution [34]. Therefore any rigorous results here should be welcomed.

In this article we considered just one aspect of convergence of finite population result to the results in infinite population considering the variance of the mean fitness. According to this criterion, $N \sim L^2$ is large enough to have the same mean fitness as the infinite population with accuracy $1/L$. Actually an important open problem is to investigate the equilibrium here (mutation-selection), like the equilibrium in thermodynamics, and how the equilibrium is affected by the finite size of the population.

* * *

DBS thanks U.S. Defense Advanced Research Projects Agency DARPA grant No. HR00110510057, DARPA Prophecy Program and Academia Sinica for financial support. This work was supported in part by NSC 100-2112-M-001-003-MY2 and NCTS in Taiwan.

Appendix A. Crow-Kimura model. – We here consider the infinite population model. The $M \equiv 2^L$ genome configurations (sequences) are defined as chains of L spins $s_k, 1 \leqslant k \leqslant L$ having values ± 1. There is a reference sequence S_0 with all spins $+1$. We define the Hamming distance between the given sequence and the reference sequence by $\sum_k (1-s_k)/2 \equiv N(1-m)/2$, where m is an overlap.

The state of the system is specified by the M relative frequencies $p_i, 0 \leqslant i \leqslant M-1$:

$$\begin{aligned} \frac{\mathrm{d}p_i}{\mathrm{d}t} &= \sum_j A_{ij} p_j - p_i \sum_j p_j r_j, \\ A_{ij} &= \delta_{ij} r_j + m_{ij}. \end{aligned} \tag{A.1}$$

Here m_{ij} is the rate of mutation from the sequence j to the sequence i, and r_i is the fitness. Two sequences have a Hamming distance $d_{ij} = (L - \sum_k s_i^k s_j^k)/2$. Here $m_{ii} = -\gamma_0$. When $d_{ij} = 1$, $m_{ij} = \gamma_0/L$, $m_{ij} = 0$ for $d_{ij} > 1$ [7]. We are interested in the symmetric fitness landscape with

$$r_i = f(1 - 2d_{i0}/L). \tag{A.2}$$

We choose the first $L+1$ sequences such that the l-th sequence has a Hamming distance l from the reference sequence, $0 \leqslant l \leqslant L$. Then our r_l are connected with the $\hat{r}_l$ in the main text as

$$\hat{r}_l = \exp[r_l U], \tag{A.3}$$

where

$$U = \tau \gamma_0, \tag{A.4}$$

where τ is the iteration duration and l in eq. (A.3) is the Hamming class of the i-th sequence. In the main text we consider the discrete-time evolution with minimal time interval $\tau = U$, choosing $\gamma_0 = 1$.

Appendix B. Eigen model. – In Eigen's quasispecies theory [3,4], the i-th sequence produces offspring of the type j with probability $Q_{ij} = q^{L-d_{ij}}(1-q)^{d_{ij}}$, where $1-q$ is the average number of errors per site and d_{ij} is the Hamming distance.

Eigen proposed that p_i satisfy [3,4]

$$\begin{aligned} \frac{\mathrm{d}p_i}{\mathrm{d}t} &= \left\{ Q_{ii} r_i - D_i \sum_k \hat{r}_k p_k(t) \right\} p_i(t) \\ &\quad + \sum_{k \neq i} Q_{ik} r_k p_k(t). \end{aligned} \tag{B.1}$$

Here D_i describes the degradation. It is convenient to work with the error rate $\gamma \equiv L(1-q)$, leading to $Q = e^{-\gamma}$.

Appendix C. Eigen model with degradation. – We now consider the Eigen model when there is a degradation D in the peak sequence S_0, and zero degradation for other sequences S_i for $i > 0$. In this case we should add the random sampling for the degradation case. The calculation procedure is similar to the case dealt with in the section on the Eigen model. We should just modify the iteration sub-steps from that section, after the point B.

C. There is a dilution of the population with the peak sequence. We randomly choose t elements from a pool of n elements and each element is chosen independently with probability UD.

D. There is a dilution of the whole population, keeping strictly the total population size as N.

We randomly choose h elements from a pool of $l+k-t$ elements and each element is chosen independently with probability $U(1-x)$.

Now after one iteration step $n \to n+m-h-t$. Thus we should calculate $\langle e^{-\phi'(m-t-h)} \rangle$. We get the following expression:

$$\begin{aligned} &\langle e^{-(m-t-h)\phi'} \rangle = \\ &\sum_{l,k,m,t,h} \frac{n!(UA)^l (1-UA)^{n-l}}{l!(n-l)!} \frac{l! Q^m (1-Q)^{l-m} e^{-\phi' m}}{m!(l-m)!} \\ &\times \frac{n! e^{\phi' t} e^{\phi' h)} (UD)^t (1-UD)^{n-t}}{t!(n-t)!} \\ &\times \frac{(N-n)! U^k (1-U)^{N-n-k}}{k!(N-n-k)!} \frac{(l+k-t)! x^k (1-x)^{(l+k-t-h)}}{h!(l+k-t-h)!}. \end{aligned} \tag{C.1}$$

We first perform the sum over h:

$$\begin{aligned} &\sum_h \frac{(l+k-t)! x^h (1-x)^{(l+k-t-h)}}{h!(l+k-t-h)!} e^{\phi' h} = \\ &(1 + x(e^{\phi'} - 1))^{l+k-t}, \end{aligned} \tag{C.2}$$

then perform the sum over t:

$$\sum_t \frac{n! e^{\phi' t}(UD)^t (1-UD)^{n-t}}{t!(n-t)!}(1+x(e^{\phi'}-1))^{-t} = \left(U\frac{de^{\phi'}}{(1+x(e^{\phi'}-1))}+1-U\right)^{Nx} = \exp\left[Ux\left(\frac{e^{\phi'}}{1+x(e^{\phi'}-1)}-1\right)\mathrm{d}N\right]. \quad \text{(C.3)}$$

Comparing our formulas with the expression of $F(\phi')$ from the section on the finite population Eigen model, we find just a new additional term to those of eq. (20). Eventually we have

$$\frac{\mathrm{d}\phi'}{\mathrm{d}t} = xA[(Qe^{-\phi'}+1-Q)(xe^{\phi'}+1-x)-1] + x\left(\frac{e^{\phi'}}{1+x(e^{\phi'}-1)}-1\right)D + (1-x)x(e^{\phi'}-1). \quad \text{(C.4)}$$

We expand in powers of ϕ':

$$F(\phi') \approx -[(QA-1-D)-(A-1-D)x]\phi' + \frac{\phi'^2}{2}[QA(1-2x)+(A-1)x+1+D(1-x)(1-2x)]. \quad \text{(C.5)}$$

Putting the value of $x = \frac{AQ-D-1}{A-D-1}$, we derive

$$F(\phi') \approx -[(QA-1-D)-(A-1-d)x]\phi' + \frac{\phi'^2}{2}\frac{(2a(-1+Q)((D+D^2+(-1+a)aQ-2aDQ)))}{(1+D-a)^2}.$$

and obtain for the variance

$$V = \frac{A(1-Q)((A-1)AQ+2AQd-d-d^2)}{(A-1-d)^3}. \quad \text{(C.6)}$$

For $D=0$, eq. (C.6) reduces to eq. (20) for the Eigen model without degradation.

REFERENCES

[1] Wright S., in *Proceedings of the Sixth International Congress of Genetics*, Vol. **1** (Genetics Society of America) 1932, p. 356.
[2] Moran P. A. P., *The Statistical Processes of Evolutionary Theory* (Clarendon, Oxford) 1962.
[3] Eigen M., *Naturwissenschaften*, **58** (1971) 465.
[4] Eigen M., McCaskill J. and Schuster P., *Adv. Chem. Phys.*, **75** (1989) 149.
[5] Crow J. F. and Kimura M., *An Introduction to Population Genetics Theory* (Harper Row, New York) 1970.
[6] Ewens W. J., *Mathematical Population Genetics* (Springer-Verlag, New York) 2004.
[7] Baake E., Baake M. and Wagner H., *Phys. Rev. Lett.*, **78** (1997) 559.
[8] Baake E. and Wagner H., *Genet. Res.*, **78** (2001) 93.
[9] Saakian D. B. and Hu C.-K., *Phys. Rev. E*, **69** (2004) 021913; (2004) 046121.
[10] Saakian D. B., Khachatryan H. and Hu C.-K., *Phys. Rev. E*, **70** (2004) 041908.
[11] Saakian D. B. and Hu C.-K., *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006) 4935; Saakian D. B., Munoz E., Hu C.-K. and Deem M. W., *Phys. Rev. E*, **73** (2006) 041913.
[12] Park J. M. and Deem M. W., *J. Stat. Phys.*, **125** (2006) 975.
[13] Park J. M. and Deem M. W., *Phys. Rev. Lett.*, **98** (2007) 058101.
[14] Munoz E. T., Park J. M. and Deem M. W., *Phys. Rev. E*, **78** (2008) 061921.
[15] Saakian D. B., *J. Stat. Phys.*, **128** (2007) 781.
[16] Sato K. and Kaneko K., *Phys. Rev. E*, **75** (2007) 061909.
[17] Saakian D. B., Kirakosan Z. and Hu C.-K., *Phys. Rev. E*, **77** (2008) 061907.
[18] Saakian D. B., Rozanova O. and Akmetzhanov A., *Phys. Rev. E*, **78** (2008) 041908.
[19] Woodcock H. and Higgs P. G., *J. Theor. Biol.*, **179** (1996) 61.
[20] Saakian D. B., Biebricher C. K. and Hu C.-K., *PLoS ONE*, **6** (2011) 21904.
[21] Kirakosyan Z., Saakian D. B. and Hu C.-K., *J. Stat. Phys.*, **144** (2011) 198.
[22] Peng C.-K. *et al.*, *Nature*, **356** (1992) 168.
[23] Nowak M. and Schuster P., *J. Theor. Biol.*, **137** (1989) 375.
[24] Alves D. and Fontanari J. F., *Phys. Rev. E*, **57** (1998) 7008.
[25] Wahl L. M. and Krakauer D., *Genetics*, **156** (2000) 1437.
[26] Jain K. and Krug J., *Genetics*, **175** (2007) 1275.
[27] Rouzine I. M., Wakeley J. and Coffin J. M., *Proc. Natl. Acad. Sci.U.S.A.*, **100** (2003) 587.
[28] Keightley P. D. and Otto S. P., *Nature*, **443** (2006) 89.
[29] Desai M. M., Fisher D. S. and Murray A. W., *Curr. Biol.*, **17** (2007) 385.
[30] Rouzine I. M., Brunet E. and Wilke C. O., *Theor. Popul. Biol.*, **73** (2008) 24.
[31] Park A. C., Simon D. and Krug J., *J. Stat. Phys.*, **138** (2010) 381.
[32] Eigen M., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 13374.
[33] Sole R. and Deisboeck T. S., *J. Theor. Biol.*, **228** (2004) 45.
[34] Woo H. J. and Wallqvist A., *Phys. Rev. Lett.*, **106** (2011) 060601.