Two-time-scale population evolution on a singular landscape

Song Xu,[1,*] Shuyun Jiao,[2,3] Pengyao Jiang,[4] and Ping Ao[2,5,†]

[1]*Department of Biomathematics, University of California at Los Angeles, Los Angeles, California 90095-1766, USA*

[2]*Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine of Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China*

[3]*Department of Mathematics, Xinyang Normal University, Xinyang 464000, Henan, China*

[4]*Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, Illinois 60637, USA*

[5]*Department of Physics, Shanghai Jiao Tong University, Shanghai 200240, China*

(Received 21 August 2013; revised manuscript received 23 November 2013; published 31 January 2014)

Under the effect of strong genetic drift, it is highly probable to observe gene fixation or gene loss in a population, shown by singular peaks on a potential landscape. The genetic drift-induced noise gives rise to two-time-scale diffusion dynamics on the bipeaked landscape. We find that the logarithmically divergent (singular) peaks do not necessarily imply infinite escape times or biological fixations by iterating the Wright-Fisher model and approximating the average escape time. Our analytical results under weak mutation and weak selection extend Kramers's escape time formula to models with B (Beta) function-like equilibrium distributions and overcome constraints in previous methods. The constructed landscape provides a coherent description for the bistable system, supports the quantitative analysis of bipeaked dynamics, and generates mathematical insights for understanding the boundary behaviors of the diffusion model.

DOI: 10.1103/PhysRevE.89.012724 PACS number(s): 87.23.Kg, 87.23.Cc, 05.10.Gg, 82.20.Uv

I. INTRODUCTION

Interactions among multiple biological effects (genetic drift, mutation, selection, etc.) can cause different evolutionary behaviors of a population. These effects may operate on different time scales [1]. One of the most important issues in the study of evolutionary models is to describe and separate the multi-time-scale dynamics by calculating the corresponding operating times [2]. Of special interests is the calculation of waiting time for rare events to occur [3,4]. Similar problems have been referred to in physics, chemistry, engineering, and biology in close connection to the concept of adaptive landscape or potential energy function [5–7]. In population biology, evolution was found not to be limited by the rate toward local adaptive peaks but by the peak-to-peak transition rate on the landscape [8], as stated by Wright's shifting-balance theory [5]. The peak-to-peak transition was found closely related to the mechanisms of adaptation, divergence, and speciation [9]. Results on the transition rates and the multiple time scales were found to be important for studying biological robustness [10]. Similar ideas and methods have been widely discussed outside biology [11,12].

Typically, a multipeak landscape implies the separation of time scales of uphill (driven by deterministic factors) and downhill [driven by stochastic factors (or noise)] population evolution. In chemistry, where the energy function acts as an adaptive landscape, the known Arrhenius formula estimates the separation factor to be an exponential term of the valley depth [13]. The separation of time scales has also been systematically studied in physical systems where a potential function exists [14,15]. In evolutionary biology, however, lack of theoretical generality of the adaptive landscape itself has made this classical concept and its quantitative utility controversial [16–19]. There is no associated method for the general calculations of the peak-shifting rates on a general landscape. In more recent literature, on the other hand, there is re-emerging interest in studying the evolutionary potential function from a physical point of view [7,11,20,21]. Many have tried to link statistical mechanics to biological evolution by noticing that the process to approach the equilibrium distribution maximizes a free fitness function [22–24].

In this article, we are motivated to apply a dynamical framework proposed by one of the authors of this article [20] to study the biological evolution from a dynamical point of view. It breaks certain constraints of the previous statistical mechanics-based methods as it contains the unbiased steady-state information of the stochastic model while not requiring the existence of equilibrium distribution. The framework constructs the adaptive landscape as a potential function in general stochastic processes, integrating all the biological factors in the formulation. In comparison with Wright's original adaptive landscape that only manifests selection, we call it potential landscape to avoid ambiguity. The potential landscape extends Wright's adaptive landscape to more general contexts and has already been applied to many biological models [7,25–27].

Here we study the classical Wright-Fisher process [28,29]. With the potential landscape, we are able to investigate the genetic drift-induced bistability of the model. In comparison, most previous analyses were based on the selection-induced bistability, which is confined to strong selection cases [8,30]. The genetic drift-induced noise has a nonuniform intensity distribution which vanishes at the boundaries. It causes singular potential peaks on the landscape and makes the calculation of escape time conceptually difficult if confining to the classical conclusions [13]. By computer iterations of the discrete model, we observe a two-time-scale evolution on the singular landscape. We then find an analytical way to revise the classical results to fit in the present types of models and obtain analytical approximations for the escape time from the singular potential peak. We also discuss the connection

*sxu11@ucla.edu

†Corresponding author: aoping@sjtu.edu.cn

 ©2014 American Physical Society

between the mean first passage time (MFPT) and the escape time in this biological model, a key issue [15] often ignored by population geneticists [31,32].

In comparison with previous rate formulas obtained from the diffusion model [14,15], the present results are not based on the assumption of a Gaussian-like probability distribution near a potential peak but come from the direct analysis of the B distribution near the boundary states. Another typical approach is to calculate the eigenvalue of diffusion equations [8] as the transition rate. However, this method does not apply to weak selection cases or the selection-free evolution [4]. A third method from population genetics (appearance rate of the mutation that is destined to fix [1,31]) works for weak selection but fails under more complex situations (e.g., where fixation is not possible or the selection is frequency dependent). To compare, our dynamical approach has no certain biological constraints and is generally applicable in these cases. The constructed landscape extends the fitness landscape, in a way allowing general description of the model under various biological factors. All factors are treated in a consistent manner.

The present article is organized as follows: In Sec. II, we introduce the one-dimensional (1D) diffusion process and construct a potential landscape. In Sec. III, we analyze the uphill and downhill landscape dynamics under strong genetic drift and other factors. In Sec. IV, we iterate the discrete model, get the numerical escape time, and then obtain the analytical approximations under weak mutation and zero or weak selections. In Sec. V, we compare our methods and results to previous work. We also discuss the relation between the MFPT and the escape time. We conclude with the mathematical condition for fixation in singular potential and other boundary behaviors of the diffusion model based on the present results.

II. WRIGHT-FISHER MODEL AND POTENTIAL LANDSCAPE

A. Diffusion process

The 1D Wright-Fisher model considers the evolution of a diploid population at one locus. The number of individuals in a population is a fixed constant N and the generations are nonoverlapping. Denote the alleles of interest as A_1 and A_2; the total number of the gene copies of A_1 and A_2 in the population gene pool is $2N$. In the present article, we mainly study the continuous diffusion approximation of the Wright-Fisher model (assume N is big enough for the continuous approximation). Let the frequency of A_1 gene be x, so the frequency of A_2 is $1-x$. Let $\rho(x,t)$ be the probability distribution of A_1 at time t. The diffusion equation or Fokker-Planck equation (FPE) for the continuous Wright-Fisher model is given by [28,29,33] the following:

$$\partial_t \rho(x,t) = \tfrac{1}{2}\partial_x^2[V(x)\rho(x,t)] - \partial_x[M(x)\rho(x,t)]. \quad (1)$$

$M(x)$ is the average change of the A_1 frequency per generation, corresponding to the deterministic factors of the system. $V(x)$ is the variance of the stochastic factors (noise). For example, under mutation and selection,

$$M(x) = -\mu x + \nu(1-x) + \frac{x(1-x)}{2\overline{\omega}}\frac{d\overline{\omega}}{dx}, \quad (2)$$

where μ is the mutation rate from A_1 to A_2, ν is that from A_2 to A_1, and $\overline{\omega}$ gives the average fitness of the population in that generation, which depends on x. Under random genetic drift,

$$V(x) = \frac{x(1-x)}{2N}. \quad (3)$$

The population size N measures the intensity of genetic drift.

In the 1D model, by assuming zero probability current at system equilibrium ($t = +\infty$), the equilibrium probability distribution of Eq. (1) can be easily obtained as follows [13]:

$$\rho(x,+\infty) = \frac{1}{V(x)}\exp\left[\int^x \frac{2M(y)}{V(y)}dy\right]\Big/ Z = \exp\left[\int^x \frac{2M(y)-V'(y)}{V(y)}dy\right]\Big/ Z, \quad (4)$$

where the normalization constant Z is given by

$$Z = \int_0^1 \exp\left[\int^x \frac{2M(y)-V'(y)}{V(y)}dy\right]dx. \quad (5)$$

B. Potential landscape

Taking the analogy between statistical mechanics and population genetics, we note the equivalent roles played by the potential energy function and the adaptive landscape, both of which give the stability information and the long-term evolution direction of the system. For the FPE in Eq. (1), one of the present authors proposed a decomposition scheme that gives a potential landscape construction [20]. It is derived by the study of the general stochastic different equation (SDE) that is closely connected to Eq. (1) [34]. The potential landscape takes the following form [20]:

$$\Phi(x) = \int^x \frac{2M(y)-V'(y)}{V(y)}dy \doteq \int^x \frac{f(y)}{D(y)}dy. \quad (6)$$

Here we have defined a directed force $f(x)$ and an undirected diffusion term $D(x)$, which are closely related to the system's long-term dynamics,

$$f(x) = M(x) - \tfrac{1}{2}V'(x), \quad (7)$$

$$D(x) = \tfrac{1}{2}V(x). \quad (8)$$

A combination of of Eqs. (4) and (6) immediately suggests the physical validity of the potential function $\Phi(x)$ in the Boltzmann-Gibbs distribution (if the equilibrium distribution exists, $Z < +\infty$) in statistical mechanics [7],

$$\rho(x,+\infty) \propto \exp\left[\frac{\Phi(x)}{\epsilon}\right]. \quad (9)$$

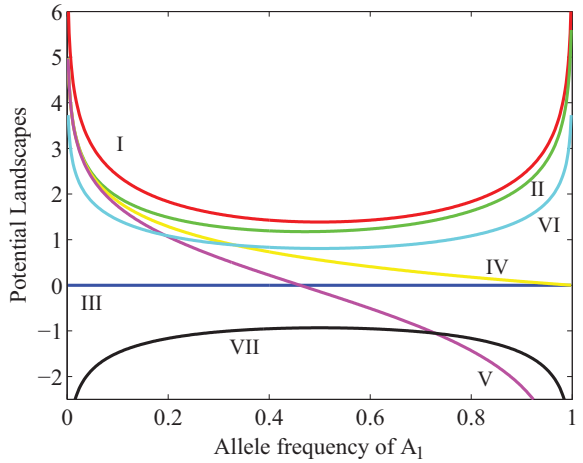
Here ϵ is a non-negative constant, playing the role of temperature. We can set $\epsilon = 1$ because we only care about the relative values of the potential function as a landscape. The potential landscape integrates all biological factors of the model, corresponding exactly to the long-term dynamics: Populations tend to move from the valleys to the peaks on the landscape [see Sec. IIIC]. This is in comparison to

the classical fitness landscape, which cannot give a direct dynamical description for evolution in many cases [19].

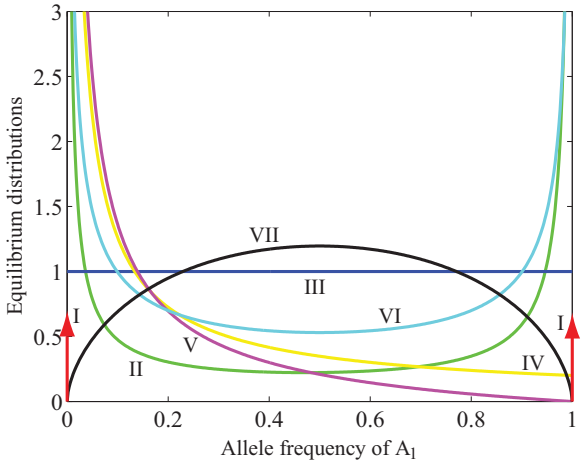
We can specify Eq. (6) in the Wright-Fisher model under the effects of genetic drift, mutation, and selection [33] by considering different forms of Eqs. (2) and (3) as follows:

$$\Phi(x) = -\ln x(1-x) + 4N[\nu \ln x + \mu \ln(1-x)] + 2N \ln \overline{\omega}. \tag{10}$$

Note that $\ln \overline{\omega}$ is the classical fitness landscape, here a part of the potential landscape. In the 1D model (where the detailed-balance condition always holds), Eq. (10) can be derived from the the Boltzmann-Gibbs distribution if in Eq. (5)



(a)Potential landscapes



(b)Equilibrium distributions

FIG. 1. (Color online) Potential landscapes and corresponding equilibrium distributions under different parameter settings in the Wright-Fisher model, differentiated by both the colors and Roman indexes. In all cases (except VII) there is $N = 50$. The following five landscape contours are generated from Eq. (11) under mutation and genetic drift: Red (I): $\mu = \nu = 0$. Green (II): $\mu = 0.0005, \nu = 0.001$. Blue (III): $\mu = \nu = 0.005$. Yellow (IV): $\mu = 0.005, \nu = 0.001$. Magenta (V): $\mu = 0.01, \nu = 0.001$. The last two are generated from Eq. (40) with selection: Cyan (VI): $\mu = \nu = 0.002, s = 0.1$. Black (VII): $\mu = \nu = 0.002, s = 0.1, N = 200$. The two red (I) arrows in (b) denote the Dirac δ functions.

$Z < +\infty$ [24,35]. The present dynamical landscape approach shows salient differences to these methods when $Z = +\infty$ or in higher-dimension models with nongradient flows or even limit cycles [27,36,37].

From this expression, the potential function does not generally change proportionally with the inverse of the population size $1/N$. In this sense, the inverse population size does not directly correspond to the temperature in thermodynamics, which differs from previous work in applying statistical mechanics to population genetics [24]. To show this, we give an example in Fig. 1, where case VI (cyan) and case VII (black) differ only in population sizes ($N = 50$ and $N = 200$). The change of $1/N$ induces the disproportional changes of the potential values and the stabilities of $x = 0$ and $x = 1/2$ states: $x = 0$ is stable in case VI but unstable in case VII; $x = 1/2$ is unstable in case VI but stable in case VII. This property of N is not a desired property for a temperature equivalent.

With the analytical form of $\Phi(x)$, we may classify the Wright-Fisher diffusion models under different parameters according to their long-term behaviors. Several examples of typical Wright-Fisher systems are given in Fig. 1.

III. BISTABLE DYNAMICS UNDER STRONG GENETIC DRIFT

A. Mutation and genetic drift

To start with, we consider the simplest mutation-drift case [we neglect selection by taking $\overline{\omega} \equiv 1$ for all population states $x \in [0,1]$ in Eq. (10)],

$$\Phi(x) = (4N\nu - 1)\ln x + (4N\mu - 1)\ln(1-x). \tag{11}$$

It contains all the linear forms of $M(x)$ and thus covers other typical cases such as the one-island migration [38]. A classification of system dynamics can be easily made by using this expression. We summarize in Table I different dynamical behaviors presented by different landscape shapes for the whole parameter ranges that are biologically meaningful ($\mu, \nu > 0$). Note that all these mutation-drift models correspond to the usually termed "neutral evolution" in population biology, because there is no biological selection in the models. In the sense of probability distribution, the term "neutral evolution" may be reconsidered, as in most of these cases the potential landscapes (as are the equilibrium distributions) are

TABLE I. Classification of the mutation-drift model with changing mutation rates μ and ν (N fixed). $x = a$ is the unique fixed point in the system defined in Eq. (12). Its stability is determined by the sign of $\Phi''(a)$. The boundary values of $\Phi(x)$ is directly calculated from Eq. (11). The landscape configurations are determined by above factors.

μ	ν	Fixed point a	$\Phi(0), \Phi(1)$	Shape of $\Phi(x)$
$< \frac{1}{4N}$	$< \frac{1}{4N}$	(0,1)/Unstable	$+\infty, +\infty$	$\bigcup$ shaped
$> \frac{1}{4N}$	$> \frac{1}{4N}$	(0,1)/Stable	$-\infty, -\infty$	$\bigcap$ shaped
$= \frac{1}{4N}$	$= \frac{1}{4N}$	N/A	0,0	Flat
$< \frac{1}{4N}$	$> \frac{1}{4N}$	Outside (0,1)	$+\infty, -\infty$	Left skewed
$> \frac{1}{4N}$	$< \frac{1}{4N}$	Outside (0,1)	$-\infty, +\infty$	Right skewed

not flat. The evolution is thus directed or "unneutral" in a way. In the bistable cases ($4N\nu, 4N\mu < 1$), for example, the system has a "preference" on the two fixation states $x = 0$ and $x = 1$, shown by the two potential peaks. The long-term evolution is not really "neutral." From Table I, only the special $\mu = \nu = 1/(4N)$ case has a flat landscape (Fig. 1), implying equal probabilities for different states to be occupied by a population.

To maintain a bistable system, we set $4N\nu, 4N\mu < 1$. There is a unique valley state (saddle point) of the landscape in $(0,1)$, here we denote as $x = a$,

$$a = (1 - 4N\nu)/(2 - 4N\mu - 4N\nu), \tag{12}$$

satisfying $\Phi'(a) = 0$ and $\Phi''(a) > 0$ (unstable). Such a valley point defines two attractive basins $(0,a)$ and $(a,1)$. The directed force that drives a population toward a potential peak is

$$f(x) = -\mu x + \nu(1 - x) - \frac{1 - 2x}{4N}. \tag{13}$$

Obviously $x = a$ is the unique zero point of $f(x)$. The diffusion term is

$$D(x) = \frac{x(1 - x)}{4N}. \tag{14}$$

All possible configurations of potential landscapes in the mutation-drift systems are shown in Fig. 1.

B. Two time scales: Single population's view

The movements of a population on a landscape can be typically classified into two fundamentally different types: uphill and downhill processes, often demonstrated to operate on two different time scales [39]. The uphill movements toward potential peak are dominated by the directed force and often considered processing in a shorter time scale (denote as $\mathcal{T}_1$). The stochastically driven downhill process can result in moving away from the peak, crossing over the saddle and escaping to other attractive basins, which is considered to be in a much longer time scale $\mathcal{T}_2$. Figure 2 gives an illustration.

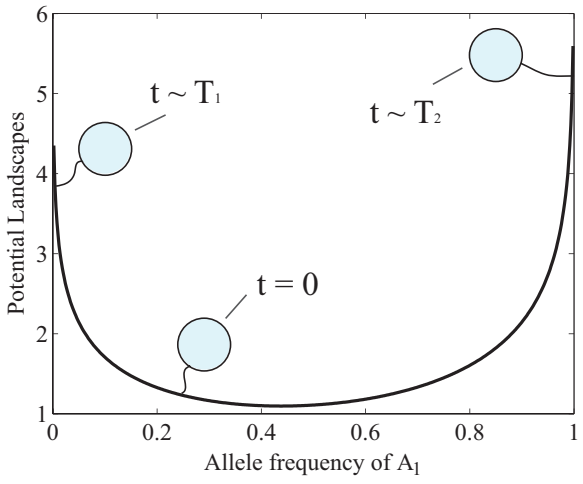


FIG. 2. (Color online) Visualization of the two-time-scale dynamics on a typical bistable landscape in the present model. It gives the most probable state of a population (denoted as a balloon, which always searches for a higher "altitude" to stay) in different time scales ($T_2 \gg T_1$) visualized on the potential landscape. The parameter setting satisfies $\mu < \nu \ll 1/(4N)$.

C. Uphill dynamics

The uphill valley-to-peak evolution is mainly driven by the directed force $f(x)$. To see this, we refer to the Langevin equation or SDE that separates the directed and the undirected (noise) terms. It describes essentially the same evolutionary process with the FPE Eq. (1) but from the point of view of a single population (instead the probability density or the ensemble of infinite many populations)'s evolution [34],

$$\dot{x} = f(x) + \zeta(x,t). \tag{15}$$

Here $\dot{x} = dx/dt$ denotes the change rate of x. The Gaussian white noise $\zeta(x,t)$ has variance $D(x)$. The directed force $f(x)$ is just our Eq. (7). We note that there are infinite possible ways to choose the directed force, which is decided by how we connect an FPE to an SDE [34]. Here our choice of $f(x)$ comes from taking the zero-mass limit (a physical realization) of a Brownian particle system described by Eq. (1) [34]. By averaging out the effect of noise, we obtain the directed rate as follows:

$$\dot{x} = f(x). \tag{16}$$

Note that this differs from taking the infinite-population limit of the model ($N \to +\infty$, a typical way of obtaining the deterministic limit in population genetics), which gives a deterministic force $M(x)$. Actually, $f(x)$ combines the effects of $M(x)$ and $D(x)$ as shown in Eq. (7). Intuitively, the nonuniformly distributed noise has effect on the direction of long-term evolution. The deterministic rate alone may not be able to describe the direction of system evolution.

It is easy to verify that Φ is nondecreasing along the noise-free evolutionary trajectory of a population,

$$\dot{\Phi} = \Phi'(x)\dot{x} = f^2(x)/D(x) \geqslant 0. \tag{17}$$

This manifests Wright's essential idea of a proper landscape to visualize the uphill evolution [5].

Under strong genetic drift ($4N\nu, 4N\mu \ll 1$), the directed force $f(x)$ is near linear. We can always take the approximation form $f \approx -|F|\bar{x}$ ($|F|$ measures the magnitude of $f(x)$; $x - a$ is replaced by $\bar{x}$ to give the distance between x and a). The solution of Eq. (16) takes the approximate form,

$$\bar{x}(t) \approx \bar{x}(0)\exp(-|F|t) \doteq \bar{x}(0)\exp(-t/\mathcal{T}_1), \tag{18}$$

where $\bar{x}(0)$ gives the initial state of the population and $\mathcal{T}_1$ is usually called the relaxation time [40]. Under strong genetic drift, the uphill rate is

$$\dot{x} = f(x) \approx -(1 - 2x)/(4N). \tag{19}$$

For $x < 1/2$, a population is expected to move toward $x = 0$; for $x > 1/2$, it is expected to move toward $x = 1$. This is consistent to the biological expectations, that a population will be fixed at either all A_1A_1 state or all A_2A_2 state, conditioned on the initial population state. The first evolutionary time scale for reaching the local potential peak is then

$$\mathcal{T}_1 \sim |F|^{-1} = 2N\mathcal{O}(1), \tag{20}$$

the typical operating time scale of genetic drift [1].

D. Downhill dynamics

The downhill dynamics is considered as the accumulation of rare downhill movements driven by noise. The process can be characterized by the waiting time τ for such rare effects to grow big enough to cross the valley state. A population thus escapes from the original attractive basin, goes through the potential valley and stays stable in another attractive basin. In a diffusion model with a finite potential barrier (here valley) $\Delta\Phi$, Kramers' classical formula estimates the escape time as [14]

$$\tau \sim \mathcal{T}_1 \exp(\Delta\Phi). \tag{21}$$

The $\exp(\Delta\Phi)$ is also called the Arrhenius term [13].

For the downhill dynamics in the present case, the classical formula would give an estimation of infinite escape time from singular potential peak: From Eq. (11), there is $\Phi(0) = +\infty$, which leads to $\Delta\Phi = \Phi(0) - \Phi(a) = +\infty$. Then, by Eq. (21), we get

$$\tau = +\infty. \tag{22}$$

Under pure genetic drift ($\nu = \mu = 0$), this infinite escape time is expected (see Sec. V D). If, however, there is additional biological factors such as mutation in the system, the infinite escape time may not be a good estimation. Biologically, even though genetic drift (noise) vanishes at the boundary, mutation will constantly push a population away from the fixation state of A_2 by introducing new A_1 mutants, which makes the gene substitutions possible [4]. Mathematically, Eq. (21) causes an unexpected discontinuity of the escape time with changing mutation rate ν, whereby $\tau/\mathcal{T}_1$ changes from $+\infty$ to 1 as $4N\nu \to 1$ (the potential peak at $x = 0$ vanishes then; see Table I). In the next section, we obtain better estimations for τ by simulating the Wright-Fisher model and calculating the MFPT.

IV. ESCAPE TIME IN INFINITE POTENTIAL

A. Two time scales: Distribution's view

If $\tau \neq +\infty$, we expect that $\rho(x,t)$ undergoes two distinct stages of evolution on the bipeaked potential, similarly to the dynamics near a finite potential peak. This is because the divergence of the peak is at a relatively slow rate [log-scale, see Eq. (11)] in a finite space $[0,1]$. The singularities only appear at the two boundary states. The leaking flow of probability density (which is approximately constant in the second stage of evolution) from an attractive basin makes the escape possible [15]. We assume that the flow rate obeys an exponential law as follows:

$$Z_0(t) - Z_0(+\infty) = \Gamma_0 \exp(-\lambda t). \tag{23}$$

Here Γ_0 is the space integral of a time-independent function characterizing the shape of the probability distribution, which approximates a constant value in the second stage evolution. λ is the average leaking rate of probability or the escape rate [15]. We define the cumulative probability density $Z_0(t)$ in the attractive basin $(0,a)$ as

$$Z_0(t) = \int_0^a \rho(x,t)dx. \tag{24}$$

The leaking rate of $Z_0(t)$ changes with time, during which the global equilibrium between different local equilibria is being established [8]. We may characterize this process by the inverse of the average leaking rate $\mathcal{T}_2 = 1/\lambda$, which also gives the time scale to establish the global equilibrium.

Note that the leaking rate is the sum of contributions from both attractive basins ($\lambda = \lambda_0 + \lambda_1 = \tau_0^{-1} + \tau_1^{-1}$). Transitions in the two directions will eventually balance each other as $t \to +\infty$ at the global equilibrium,

$$\tau_0^{-1} Z_0(+\infty) = \tau_1^{-1} Z_1(+\infty). \tag{25}$$

Here $Z_1(t) = 1 - Z_0(t)$ is the cumulative probability density in $(a,1)$. In the long-term evolution, the escape time from the $(0,a)$ basin is given by [15]

$$\tau_0 = \frac{1}{\lambda[1 - Z_0(+\infty)]}. \tag{26}$$

To verify the above theory in the present biological process, we simulate the change rate of the probability vector $P(t)$ in the discrete Wright-Fisher model. We also plotted the solution of $\rho(x,t)$ in the continuous model [28]. The results in Fig. 3 show a clear two-time-scale structure of landscape dynamics. With

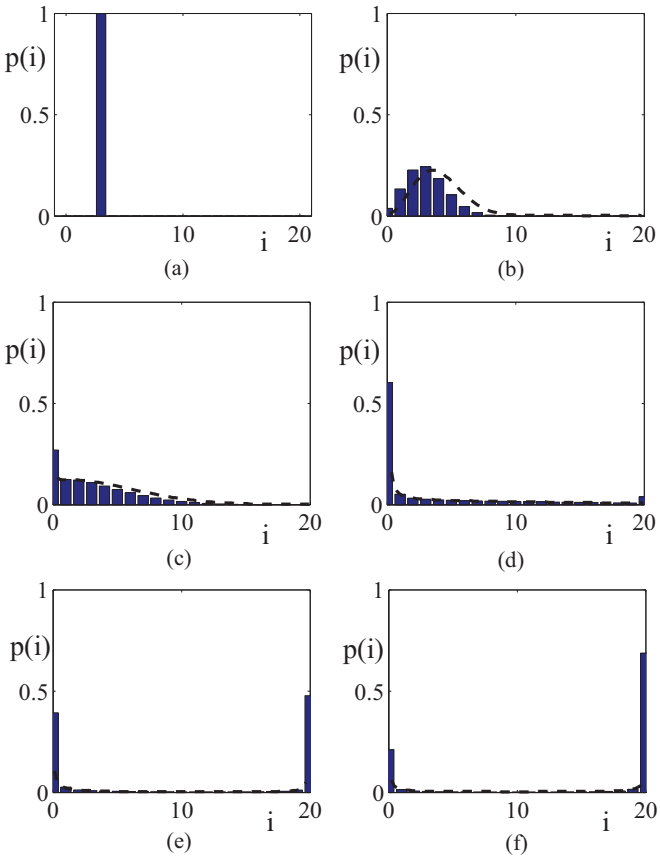


FIG. 3. (Color online) Real-time iterations and continuous solutions of the Wright-Fisher process under mutation and random drift. The bars presents the time iterations of the discrete model. The dotted lines are the analytical solution of $\rho(x,t)$ in the 1D continuous model. The x axis gives the number of A_1 alleles i and the y axis is the probability distribution $p(i)$ (notation in accordance with Appendix C). Parameter settings: $2N = 20$, $\mu = 0.0005$, $\nu = 0.0015$. Panel (a) shows that the initial state is set to $x = 0.2$. Panel (f) shows the establishment of equilibrium distribution after long-enough time.

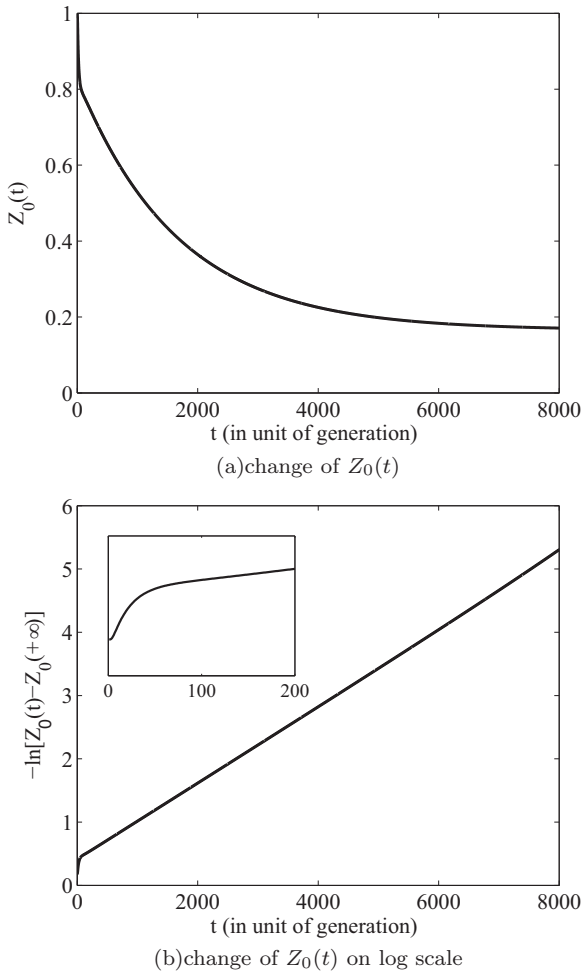


FIG. 4. Simulation of the escape rate from the attractive basin $(0,a)$ under mutation and drift. Parameter settings are as follows: $N = 30$, $\mu = 0.0001$, $\nu = 0.0005$. Panel (a) describes how the cumulative probability density in $(0,a)$ attractive basin $Z_0(t)$ [defined in Eq. (24)] changes with time. Panel (b) gives the change of value $-\ln[Z_0(t) - Z_0(+\infty)]$, whose slope gives the flux rate between the two attractive basins. The inset figure zooms into the first time scale. The simulated values are $\hat{T}_1 \approx 62.77$ [regressing the steady exponential interval (0,N)], $\hat{T}_2 \approx 1464$, $\hat{\tau}_0 \approx 1755$. Under the same setting, the theoretical expectations are as follows: $T_1 = 60$, $T_2 = 1666$, $\tau_0 = 2000$.

the probability densities initialized in $(0,a)$, the decreasing rate of $Z_0(t)$ is strictly exponential [Fig. 4(a)]. Taking the log scale [Fig. 4(b)], the regressed slope gives the escape rate toward the global equilibrium or the sum of the leaking rates in the two directions ($\lambda = \lambda_0 + \lambda_1$). The decay rate shows a rigorous exponential distribution, except for the sudden drop at the beginning period of time ($\sim T_1 \approx 20$ in the example), when the local equilibria have not yet been established. In Fig. 4(b), the inset figure shows that the first period dynamics is also approximately exponential, consistent with our analysis in Sec. IIIC. We obtain the regressed decay rates of the two distinct stages, which are approximately the inverse of the two-time-scale estimations (see the figure caption for details). The simulation validates the assumption of the two-scale dynamics, even though the potential peak at $x = 0$ is singular and the established local equilibrium is not Gaussian. In the next section, to get an analytical estimation of τ, we study the mean first passage time in the infinite potential peak.

B. MFPT in infinite potential

Back to single population's view, one may calculate the mean first passage time (MFPT, or T_{MFP}) between two states x_0 and x_1 by referring to the backward equation of Eq. (1) as follows:

$$[f(x) + \epsilon D'(x)]\partial_x T_{\mathrm{MFP}}(x_0) + \epsilon D(x)\partial_x^2 T_{\mathrm{MFP}}(x_0) = -1. \quad (27)$$

Without loss of generality, we study the stochastic jump out of the attractive basin $(0,a)$. We study a population's MFPT through the valley point $x = a$ to some state $x_1 > a$, starting from $x_0 \approx 0$ in $(0,a)$. The interval of interest is set as $[0,x_1]$, with $x = 0$ the reflecting boundary and $x = x_1$ the absorbing boundary [13],

$$\partial_x T_{\mathrm{MFP}}(x = 0) = 0, \quad (28)$$

$$T_{\mathrm{MFP}}(x = x_1) = 0. \quad (29)$$

Note that the setting of absorbing boundary is reasonable in the sense that we temporarily only care about the first time a population reaches a state x_1. The rejecting boundary is biologically manifested by the nonzero forward mutation ($\nu \neq 0$) but fails when $\nu = 0$. We will show that the escape time under the limit $\nu \to 0$ converges to the same result. The solution of Eq. (27) is

$$T_{\mathrm{MFP}}(x_0 \to x_1) = \int_{x_0}^{x_1} \frac{\exp[-\Phi(y)]}{\epsilon D(y)} dy \int_0^y \exp[\Phi(z)] dz. \quad (30)$$

Here Φ is the potential landscape in Eq. (6). There is no assumption on the configuration of Φ (finite peak, etc.) when applying Eq. (30). The relation between MFPT and escape time depends on the landscape shape (see Sec. IV C) [15]. Here under mutation and random drift,

$$T_{\mathrm{MFP}}(x_0 \to x_1) = 4N \int_{x_0}^{x_1} y^{-4N\nu}(1-y)^{-4N\mu} dy \times \int_0^y z^{4N\nu-1}(1-z)^{4N\mu-1} dz. \quad (31)$$

The integral term $z^{4N\nu-1}$ near $z = 0$ accounts for the infinity of potential peak (and possibly infinite escape time). As $4N\nu, 4N\mu \to 0$, the main contribution of the above integral comes from the inner integral in a small interval $[0,y]$ $(y < a)$, that is, the incomplete B function $B(y; 4N\nu, 4N\mu)$. Under the same limit, we numerically found that it is approximated by $y/(4N\nu)$. Meanwhile, the term $y^{-4N\nu}(1-y)^{-4N\mu}$ in the first integral is approximately constant when $4N\nu, 4N\mu \ll 1$. The whole integral is presumably approximately of a scale $1/\nu$. To make this estimation accurate, we thus expand the incomplete

B function in Eq. (31) near $z = 0$ under $0 < 1 - x_1 < 1 - y < 1 - z < 1$ as follows:

$$\begin{aligned}\mathrm{B}(y;4N\nu,4N\mu) &= \int_0^y z^{4N\nu-1}(1-z)^{4N\mu-1}dz,\\ &= \int_0^y z^{4N\nu-1}\left(\sum_{n=0}^{\infty} z^n \prod_{k=1}^{n}\frac{k-4N\mu}{k}\right)dz,\\ &= \frac{y^{4N\nu}}{4N\nu} + \sum_{n=1}^{\infty}\frac{y^{n+4N\nu}}{n+4N\nu}\prod_{k=1}^{n}\frac{k-4N\mu}{k}.\end{aligned} \tag{32}$$

The convergence of the expansion is obvious given $0 < y < x_1 < 1$. Substituting $B(y;4N\nu,4N\mu)$ and expanding $(1-y)^{-4N\mu}$ in the outer integral of Eq. (31), we obtain

$$\begin{aligned}T_{\mathrm{MFP}}(x_0 \to x_1) = {}& \frac{x_1 - x_0}{\nu} + \frac{4N\mu}{\nu}\sum_{n=1}^{\infty}\frac{x_1^{n+1} - x_0^{n+1}}{n+1}\\ &\times \prod_{k=2}^{n}\left(\frac{k-1+4N\mu}{k}\right) + 4N(1-4N\mu)\\ &\times \sum_{n=1}^{\infty}\frac{x_1^{n+1} - x_0^{n+1}}{(n+1)(n+4N\nu)}\prod_{k=2}^{n}\left(\frac{k-4N\mu}{k}\right).\end{aligned} \tag{33}$$

The expansion converges under $\nu > 0$, $\mu < 1/(4N)$. For the two limiting cases we have the following:

(1) $\nu \to 0$: The expansion of Eq. (32) becomes invalid. The leading term of the expansion changes from $y^{4N\nu}/(4N\nu)$ to $\ln y$, which then becomes sensitive to x_0 near 0. To ensure the convergence of $T_{\mathrm{MFP}}(x_0 \to x_1)$ as $x_0 \to 0$, we need $\nu \neq 0$; this is the condition for the escape problem (from $x = 0$) to be finite (as we have discussed in Sec. III D). On the other hand, we always have $T_{\mathrm{MFP}}(0 \to x_1) \to \infty$ as $\nu \to 0$.

For $\nu = 0$, however, the expansion in Eq. (32) will not be valid. We have instead

$$B(y;4N\nu,4N\mu) = \ln y + \sum_{n=1}^{\infty}\prod_{k=1}^{n}\left(\frac{k-4N\mu}{k}\right)\frac{y^n}{n}, \tag{34}$$

The leading term changes from the polynomial order $\sim y^{-1}$ to the logarithmic scale, which becomes sensitive to the value of x_0 near 0 and approaches infinity at $x_0 = 0$.

(2) $\mu \to 1/(4N)$: The expansion of $(1-y)^{-4N\mu}$ would not converge for $x_1 \to 1$, as the resulted series would then become a divergent harmonic series. This is also illustrated by the vanishing bistability of the system when $4N\mu = 1$ (Fig. 1 case IV, yellow). The escape dynamics then becomes irrelevant. To ensure the convergence of Eq. (33) as $x_1 \to 1$, we need $\mu < 1/(4N)$.

C. Escape time

From the results above, we are able to calculate the MFPT between any two states x_0 and x_1 ($0 \leqslant x_0 \leqslant x_1 \leqslant 1$). The relation between the MFPT and the escape time τ is not always direct. The escape time is defined as the inverse of the average probability rate through the potential valley $x = a$ [14]. In a finite-barrier diffusion process, the escape time was shown to be approximated by $T_{\mathrm{MFP}}(0 \to x_1)$ for $x_1 \approx a, a < x_1 < 1$ and takes the Arrhenius exponential factor [13]. Previous approximation methods are mainly established on the following two assumptions: (1) A "sharp" valley around $x = a$ on the landscape and (2) Gaussian-like local probability distribution around $x = 0$. However, these two assumptions fail in the present type of model, as the landscape usually has "fat" valleys and singular peaks under strong genetic drift (Fig. 1). The local equilibrium established near a potential is not Gaussian [8].

On the present landscape surface, a population may have considerable probability to return back to x_0 immediately (in time $\mathcal{T}_1$) after its first arrival at $x = a$. We may not approximate τ directly by $T_{\mathrm{MFP}}(0 \to a)$ here. In general diffusion cases with an axisymmetric landscape (with axis $x = a$), it is demonstrated that $T_{\mathrm{MFP}}(0 \to a)$ should be compensated by a factor of 2 when approximating the escape time [15]. Under $4N\nu, 4N\mu \ll 1$, we have by Eq. (12) that $a \approx 1/2$ and the approximately axisymmetric landscape. The escape time τ_0 is then approximated by the following (taking $x_0 \to 0$):

$$\begin{aligned}\tau_0 &\approx 2 \times T_{\mathrm{MFP}}(0 \to a)\\ &\approx \frac{1}{\nu} + \frac{4N\mu}{\nu}\sum_{n=1}^{\infty}\frac{2^{-n}}{n+1}\prod_{k=2}^{n}\left(\frac{k-1+4N\mu}{k}\right)\\ &\quad + 4N(1-4N\mu)\sum_{n=1}^{\infty}\frac{2^{-n}}{(n+1)(n+4N\nu)}\\ &\quad \times \prod_{k=2}^{n}\left(\frac{k-4N\mu}{k}\right).\end{aligned} \tag{35}$$

Under $4N\nu, 4N\mu \ll 1$, τ_0 is approximately independent of the initial state x_0 in Eq. (35). The escape time is dominated by the leading term of the series $1/\nu$, added by a remaining term of order $4N\mu/\nu$,

$$\tau_0 \approx \frac{(1 + 1.23N\mu)}{\nu}, \tag{36}$$

The coefficient 1.23 is a numerical approximation of the remaining series in Eq. (35). The error is of order $o(N\mu/\nu)$. Under $4N\nu, 4N\mu \ll 1$, τ_0 is much bigger than the relaxation time ($\sim 2N$) given in Eq. (20). This shows the separation of the two time scales as expected.

Another way to look at the MFPT in Eq. (33) is to set $x_1 = 1$ and obtain the substitution time of A_1 genes. It differs from the escape time above by taking into account the dynamical details in the other attractive basin $(a,1)$ as follows:

$$\begin{aligned}&T_{\mathrm{MFP}}(0 \to 1)\\ &= \frac{1}{\nu} + \frac{4N\mu}{\nu}\sum_{n=1}^{\infty}\frac{1}{n+1}\prod_{k=2}^{n}\left(\frac{k-1+4N\mu}{k}\right)\\ &\quad + 4N(1-4N\mu)\sum_{n=1}^{\infty}\frac{1}{(n+1)(n+4N\nu)}\prod_{k=2}^{n}\left(\frac{k-4N\mu}{k}\right).\end{aligned} \tag{37}$$

The necessary condition for its convergence ($\nu > 0, \mu < 1/(4N)$) has been discussed in Sec. IV B. In Appendix A we show that the condition is also sufficient. Biologically, we

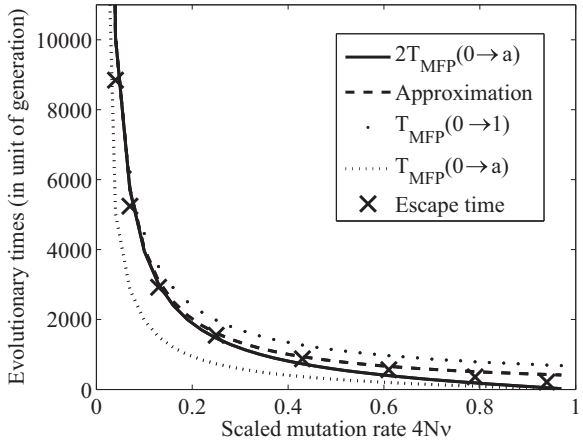


FIG. 5. Comparison between two times the MFPT from 0 to a (solid), our analytical approximation (dashed), the MFPT from 0 to 1 (point), the MFPT from 0 to a (dotted), and the simulated escape time (crosses) of the discrete Wright-Fisher model under mutation and genetic drift. Parameter setting: $N = 100$, $\mu = 0.000\,05$, $0 < 4N\nu < 1$.

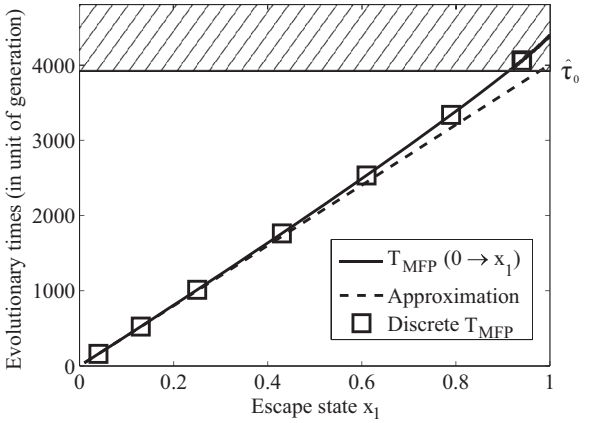


FIG. 6. Comparisons between the MFPT from 0 to x_1 (solid), our analytical approximation of the MFPT (dashed), and the discrete MFPT calculated from the master equation (square). The parameter setting is $N = 100$, $\nu = 0.000\,25$, $\mu = 0.000\,01$. The saddle point $a = 0.4747$. The shadow area above the line $t = \hat{\tau}_0$ (the simulated escape time of the discrete Wright-Fisher model) denotes that the population (expectedly) has already escaped if the time falls into this area.

expect that $T_{\mathrm{MFP}}(0 \to 1) > \tau_0$, as a population would have escaped from $(0,a)$ before it reaches $x = 1$. This relation is also mathematically demonstrated by Eqs. (35) and (37). Under the limit $4N\nu, 4N\mu \ll 1$, the two equations arrive at the same estimation:

$$T_{\mathrm{MFP}}(0 \to 1) \approx \tau_0 \approx 1/\nu. \tag{38}$$

Numerical comparisons of $2T_{\mathrm{MFP}}(0 \to a)$, $T_{\mathrm{MFP}}(0 \to 1)$, our approximation Eq. (36), and the escape time simulated from the discrete model are given in Fig. 5.

In Fig. 5, the escape time is best approximated by $2T_{\mathrm{MFP}}(0 \to a)$. $T_{\mathrm{MFP}}(0 \to 1)$ also approximates the results well but is always bigger. The estimation $T_{\mathrm{MFP}}(0 \to a)$ is obviously not a good estimation: This differs from the escape analysis on a finite landscape, where τ is approximated by $T(0 \to x_1)$ with $x_1 \approx a$. The simulated escape time is always between $2T_{\mathrm{MFP}}(0 \to a)$ and $T_{\mathrm{MFP}}(0 \to 1)$. Figure 6 shows the relations between the simulated escape time and different MFPTs. For the given set of parameters, $T(0 \to 0.9) \approx \hat{\tau}_0$. Another observation is that $2T_{\mathrm{MFP}}(0 \to a)$ is bigger (smaller) than the simulated escape time when $4N\nu$ is small (big). The main reason is that the landscape is no longer symmetric when $\nu \neq \mu$ (see Sec. VB for more discussions).

D. Models with weak selection

With the effects of selection, $M(x)$ is usually not linear dependent on the gene frequency of A_1. The adaptive nature of selection will drive a population monotonously to a fitness peak on the classical fitness landscape. Under mutation and genetic drift, a population is not always expected to evolve towards a fitness peak but towards a potential peak [our Eq. (17)]. We study how the nonlinear selection interacts with mutation and drift on the potential landscape. Particularly, we study the biologically pertinent cases of weak selections where we can also get analytical escape times [41,42], which was not achieved in the previous method [8,30].

Though $f(x)$ no longer takes the linear form, the first time scale can still be estimated by $\mathcal{T}_1 \approx 2N\mathcal{O}(1)$ under weak selection ($4Ns \ll 1$). This constraint makes sense for models where the evolution is almost selection-free or the effective population size is small. The general equation for the MFPT under mutation, drift, and selection is obtained by substituting Eq. (10) into Eq. (30) as follows:

$$\begin{aligned} &T_{\mathrm{MFP}}(x_0 \to x_1) \\ &\quad = 4N \int_{x_0}^{x_1} (1-y)^{-4N\mu} y^{-4N\nu} [\overline{\omega}(y)]^{-2N} dy \\ &\qquad \times \int_0^{y} (1-z)^{4N\mu-1} z^{4N\nu-1} [\overline{\omega}(z)]^{2N} dz. \end{aligned} \tag{39}$$

The inner integral is no longer the standard incomplete B function. If we can expand the fitness term $[\overline{\omega}(y)]^{2N}$ near the singular state $x = 0$, an analytical approximation for the MFPT can be obtained by combining the results with Eq. (37).

An example is the symmetric selection [8]. The fitness setting is $A_1A_1 : A_1A_2 : A_2A_2 = 1 : 1-s : 1$, where s presents the relative heterozygote disadvantage factor. This is a typical bimodal fitness scheme (that both A_1A_1 and A_2A_2 individuals have higher fitness values than A_1A_2 individuals). We assume the Hardy-Weinberg equilibrium, so the effect of selective pressure can be described by the allele frequency [1]. The average rate of change in x per generation by selection alone ($s \ll 1$) is $M_s(x) = -sx(1-x)(1-2x)$. We have then $\overline{w} = 1 - 2sx + 2sx^2$. With two-way mutations and genetic drift, the landscape is as follows:

$$\begin{aligned} \Phi(x) = {} & (4N\nu - 1)\ln x + (4N\mu - 1)\ln(1-x) \\ & - 4Nsx + 4Nsx^2. \end{aligned} \tag{40}$$

The landscape with symmetric mutations is plotted in Fig. 1 case VI (cyan). We study the effects of weak selection on the escape time on the basis of previous discussions. By expanding $e^{-4Nsx+4Nsx^2}$ near $x = 0$, the escape rate $\lambda_0 = \tau_0^{-1}$ is obtained

as

$$\lambda_0 = 1/(2 \times T_{\mathrm{MFP}}(0 \to a)) = 1 \Big/ \left[8N \int_0^a e^{4Nsy-4Nsy^2}(1-y)^{-4N\mu} y^{-4Nv} dy \times \int_0^y e^{-4Nsz+4Nsz^2}(1-z)^{4N\mu-1} z^{4Nv-1} dz \right] \approx v/(1+1.23N\mu+0.67Ns). \quad (41)$$

Here a is the landscape valley of Eq. (40). We give numerical comparisons among Eqs. (39) and (41), and discrete results in Fig. 7. It can be noticed that our approximation also works for $4Ns \approx 1$.

We consider another model with frequency-dependent selection and two-way mutations. We take the fitness scheme $A_1 : A_2 = 1 + s - tx : 1$, where t and s are two arbitrary constants [38]. Here the selective advantage of A_1 over A_2 is dependent on the A_1 frequency x in a simple linear fashion. This setting originally describes the haploid population but the results can be readily extended to general diploid models. The average fitness is $\overline{w} = 1 + sx - tx^2$. The potential landscape under selection, mutation, and genetic drift is

$$\Phi = \phi_s + (4Nv - 1)\ln x + (4N\mu - 1)\ln(1-x), \quad (42)$$

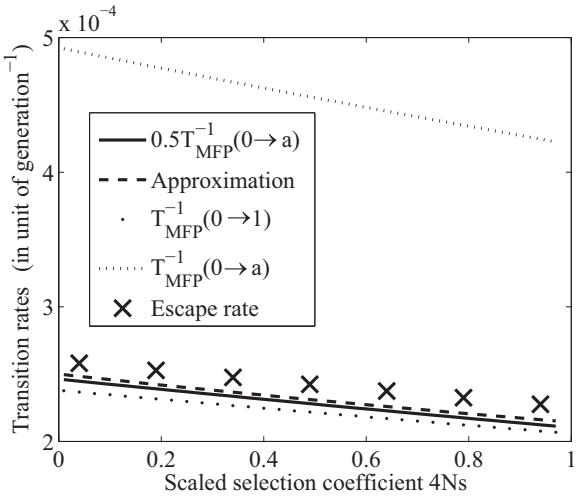


FIG. 7. Comparisons among half the escape rate (corresponding to two times the MFPT) from 0 to a (solid), our analytical approximations (dashed), the transition rate from 0 to 1 (point) and from 0 to a (dotted), and the simulated escape rate (crosses) of the discrete Wright-Fisher model under selection, mutation, and genetic drift. Parameter settings: $N = 50$, $v = 0.000\,25$, $\mu = 0.000\,01$,$0 < 4Ns < 1$.

where ϕ_s denotes the contribution of selection as follows:

$$\phi_s = \begin{cases} 2N\ln|-tx^2+sx+1| + \frac{4Ns}{\sqrt{-4t-s^2}} \arctan\frac{-2tx+s}{\sqrt{-4t-s^2}} & (\text{if } s^2 < -4t) \\ 2N\ln|-tx^2+sx+1| + \frac{2Ns}{\sqrt{4t+s^2}} \ln\left|\frac{-2tx+s-\sqrt{4t+s^2}}{-2tx+s+\sqrt{4t+s^2}}\right| & (\text{if } s^2 > -4t) \\ 4N\left(\ln\left|\frac{s}{2}x+1\right| - \frac{2}{sx+2}\right) & (\text{if } s^2 = -4t) \end{cases} \quad (43)$$

Note that the potential landscape can no longer be obtained from Eq. (10) here under frequency-dependent selection [because now $\phi_s(x) \neq \ln\overline{w}$]. We get $\phi_s(x)$ by calculating the effects of selection on $M(x)$ as $M_s(x) = (s-tx)x(1-x)/(1+sx-tx^2)$. If we further assume $s,t \ll 1$, there is $M_s \approx (s-tx)x(1-x)$, and the landscape is approximately obtained as

$$\Phi = 4Nsx - 2Ntx^2 + (4Nv-1)\ln x + (4N\mu-1)\ln(1-x). \quad (44)$$

We can study the average substitution time of gene A_1, which can be approximated by $T_{\mathrm{MFP}}(0 \to 1)$. To maintain a bistable system, we set $1/(4N) > \mu,v$. To take the expansion we further assume $4Ns,4Nt \ll 1$. The mean first passage time is as follows:

$$T_{\mathrm{MFP}}(0 \to 1) \approx \frac{1}{v}\int_0^1 [1 + (4N\mu - 4Ns)y + 2Nty^2]dy, = v^{-1}\left(1 + 2N\mu - 2Ns + \frac{2}{3}Nt\right). \quad (45)$$

Here if we take $t = 0$ and $\mu = 0$, the system would return to the simplest case with no frequency-dependent selection and reverse mutation [28]. Under this condition, the fixation time of a beneficial or deleterious mutation is often studied. We check the consistency of our results to the previous conclusions. The substitution time in Eq. (45) becomes the following:

$$T_{\mathrm{MFP}}(0 \to 1) \approx \frac{1-2Ns}{v}, \quad (46)$$

From this result, the selective advantage s decreases the substitution time approximately on a linear scale if $4Ns \ll 1$, consistent to the rate of substitution calculated under the same settings without backward mutations ($\mu = 0$) [1],

$$k = \frac{1-e^{-2s}}{1-e^{-4Ns}} \times 2Nv \approx \frac{v}{1-2Ns}, \quad (47)$$

just the inverse of Eq. (46). If instead $\mu \neq 0$ or if there is the frequency-dependent selection, the fixation probability is incalculable and the classical method in Eq. (47) does not work. Our Eq. (46) can still be applied in such cases, carrying the meaning of general transition rate in the direction $A_1 \to A_2$.

V. DISCUSSION

A. Comparisons with previous work

In the present work, we studied the bipeaked landscape dynamics that has emerged from a typical stochastic population model. We identified a two-time-scale evolution structure from the complex probability change [28]. Particularly, we

estimated the escape time from a log-divergent potential peak and applied it to the Wright-Fisher model with weak mutations and weak selections. This singular potential is essentially induced by the strong genetic drift in the present model. The effect of genetic drift is strong at middle states ($x \approx 1/2$) and weak at boundary states ($x \approx 0,1$, even vanishing at $x = 0,1$), a feature that differs substantially from the typical assumption of constant weak noise that would result in Kramers's classical conclusions [14]. The resulted singular potential peak makes it impossible to apply the classical results directly.

The present result shows that the escape time does not necessarily have an exponential dependence on the potential gap between peak and valley but is always dependent on the partial normalization constant Z_i in that attractive basin (see Sec. VD). We note that there is no real conflict among the results. In the finite-peak cases, the effect of Z_i is dominated by the potential gap [15]. In the singular-peak case, the effect of Z_i is distributed in the whole attractive basin and may accumulate near the peak state. Near $x = 0$, the potential peak diverges in a finite phase space $(0,a)$, and the probability area under the probability distribution curve turned out to be finite and sensitive to the small system parameters (here ν or μ). The original Kramers's formula of escape time should be taken special care when dealing with such cases.

Note that the escape rate [the inverse of Eq. (38)] is approximately the mutation rate and is nearly independent of the population size under $4N\nu \ll 1$. The two effects of N (sharpen the potential peak and strengthen the downhill stochastic force) approximately cancels out under this condition. Equation (38) coincides Kimura's rate of gene substitution $2N\nu \times 1/(2N) = \nu$ in the selection-free models [1,31]. Its inverse $1/\nu$ gives the expected time of the appearance of a mutation that is destined to be fixed [38]. In many cases the appearance time is used as the substitution time [1]. This happens under the limit $4N\nu \ll 1$ where the actual time for the fixation process [$\sim\mathcal{O}(2N)$] is negligible. For comparable ν and $1/(2N)$, the population size N will have significant effects on the transition rate, and the substitution time $1/\nu$ is no longer accurate. Our result Eq. (35) generally shows the effect of population size on the escape time. Moreover, Eq. (35) is obtained from the diffusion equation, the functionality of which is not confined by particular biological concerns. It can even be applied under two-way mutations and frequency-dependent selection (Sec. IVD), which makes the fixation probability of a new mutant (and thus Kimura's rate of substitution) incalculable.

Extending the typical rate formulas derived from the diffusion equations in biology [30], the present results are applicable to weak mutation and weak selection models, because we do not make assumptions on the selection type. We only require there to be two peaks on the potential landscape (instead of two peaks on the fitness landscape; see Sec. VB for details) or just the long-term bistability of the model. Another typical method used in diffusion methods is the calculation of the eigenvalue of the diffusion equation [8]. It has a constraint that the deterministic potential $\int^x M(x')dx'$ must be bimodal (so is their "deterministic equilibria"; an example is the first case in our Sec. IVD, where the selection cannot be very weak). The method failed to approximate the transition rate under very weak or zero selection ($s < 4\mu$, or simply $s = 0,4N\nu < 1$) or if the selection is not bimodal. We have shown that the system dynamics (or, to compare, "stochastic equilibria") may still be bimodal if the deterministic potential is not, due to the strong genetic drift-induced noise [Eq. (3)]. Consistently with the eigenvalue estimations in the discrete model [29], our analytical result also gives an estimation for the long-term rate toward the global equilibrium of the system.

B. MFPT and escape time

In general, these are two different concepts. The escape time τ_0 is closely related to the probability flow rate through the saddle. An escaped population from $(0,a)$ is expected to be caught stable in another attractive basin $(a,1)$. We do not count in the cases in which a population comes back to $(0,a)$ immediately after it escapes. However, as the absorbing boundary setting in Eq. (29) is not vigorous, the MFPT is generally not exactly the escape time. There might be considerable returning probability to the initial state after first reaching a specific state $x_1 \in (a,1)$. Based on the present landscape approach, we are able to discuss the MFPT-escape time issue for the strong genetic drift cases.

In Eq. (29), the boundary x_1 is not an ideal absorbing state, and τ_0 and $T_{\mathrm{MFP}}(0 \to a)$ may significantly differ, such as by a factor of 2 (Sec. IVC). An intuitive interpretation is that, after first reaching $x = a$, there is equal probability to go into $(0,a)$ or $(a,1)$ (see Appendix B for more detailed discussions). From our result in Figs. 5 and 7, the factor 2 is also an approximation as the landscape is not axisymmetric to $x = a$.

On a landscape with narrow valley, the escape time is shown to be approximated by some $T_{\mathrm{MFP}}(0 \to x_1)$ with $x_1 \approx a, a < x_1 < 1$ [13]. It is not the case in the present model, as the potential landscape here has a special shape with a wide valley and sharp peaks (Fig. 1). The main difficulty of escape here lies in overcoming the sharp peak (instead of crossing the valley) where most probability densities are concentrated. Addition information that can be read from Fig. 6 is that $T_{\mathrm{MFP}}(0 \to 1)$ is generally bigger than the escape time. A population does not have to reach $x = 1$ at the time of escape, and the escape time from $(0,a)$ does not take into account the actual dynamics in the basin $(a,1)$ [15]. Equation (35) is usually a better approximation. The equivalent escape state turns out to lie somewhere between the saddle and the end state (here near $x = 0.9$, see the intersection of $\hat{\tau}_0$ and the MFPTs in Fig. 6).

C. More on our potential landscape

The presently constructed landscape is consistent to the physical energy function, though opposite in sign. This is in adherence to the convention of biology. The directed force $f(x)$ on the landscape is shown to depend on the nonuniformly distributed noise $D(x)$ [see Eqs. (7) and (8)]. This indicates the effect of random drift on population's long-term evolution, which also results in the N dependency of the potential landscape. The landscape defined in this manner corresponds to the potential function in the Boltzmann-type distribution at steady state of Eq. (1). The temperature coefficient does not correspond to the inverse population size in this sense, which is in comparison with Ref. [24]. The observed two-time-scale dynamics (Figs. 3 and 4) verifies that the population dynamics are faithfully described by the present landscape, even though the biological fitness landscape $\ln \overline{w}$ is not bimodal. Its coherency

is also demonstrated in the limiting case of pure genetic drift as shown in Fig. 1 case I (red). As shown in Sec. IIB, $\Phi(x)$ relates to $\rho(x,t=+\infty)$ through the Boltzmann-Gibbs distribution if $Z<+\infty$ but is essentially a dynamical description of the system. Its validity does not require a normalizable equilibrium distribution (allowing $Z=+\infty$).

The potential landscape Eq. (6) can be compared to the classical fitness landscape, which presents only the effects of selection. Other biological factors may generate various evolutionary mechanisms on the fitness landscape, resulting in nonadaptive movements [43]. It thus constrains the study of bistability to strong selection models [30], excluding the investigation of system bistability induced by other factors (e.g., the strong genetic drift in the present case). The fitness landscape might be corrected by other scalar functions such as an entropy term [24]. The restriction of the fitness landscape is also shown by lack of a unified description of different biological factors, along with other controversies [19]. The confusions include the inconsistency between biology and dynamics. An example is the neutral (selection-free) model, where population probabilities are not evenly distributed in [0,1]. Genetic drift alone can result in directed (non-neutral) evolution. These facts constrain the utility of fitness landscape as a universal law in biology. The present potential landscape may serve as a substitute for Wright's original landscape that more directly visualizes the evolutionary process in a globally coherent way. In even higher-dimensional models, the present landscape is also shown to be applicable [44].

An extension to the fitness landscape is the "deterministic landscape" [8], which integrates all deterministic factors of evolution but does not consider genetic drift. Thus, its shape deviates substantially from the equilibrium distribution when genetic drift is strong. It fails to describe the bistability of the model under strong genetic drift, and the associated approaches fail for such cases (see Secs. VA and IVD). Our landscape approach instead describes the correct long-term dynamics and supports the calculation of multiple time scales. The key point here is that the nonuniform $D(x)$ also contributes to the directed evolution.

Another extension is the free fitness function, which is expected to be maximized in the process of evolution [24]. However, the associated maximum entropy approximation fails under weak mutations as the distribution function diverges near the two boundaries [45]. Moreover, it uses the normalization constant Z as a generating function of system's macroscopic information and requires normal probability distribution. It is not applicable when $Z=\infty$. When the system is bimodal, the average quantities (mean-field method) may not faithfully describe the system dynamics [45]. Our present approach does not have certain constraints, and the validity of our landscape construction and the associated approaches is tested in such singular cases. An application in population genetics is the study of Muller's ratchet under zero backward mutations [46].

D. Normalization constants and fixation

By taking $\nu=0$ in Eq. (36), we have $\tau_0=+\infty$. No escape is expected to happen once a population "traps" into the neighborhood of $x=0$. Biologically, this corresponds to the fixation of A_1 gene as a result of the absence of A_2 mutations. This is also shown by the stationary distribution (see Fig. 1 case I, red) as a combination of the Dirac δ functions under pure genetic drift [47],

$$\rho(x,+\infty)=(1-\langle x_0\rangle)\delta(x)+\langle x_0\rangle\delta(1-x). \tag{48}$$

Here $\langle x_0\rangle=\int_0^1 x\rho(x,t=0)dx$ gives the initial population state. An observation is that the fixation state of the system can be naturally derived from the present results of escape times: Fixation happens when escape is not possible. In Eq. (30), the impossibility of escape comes essentially from the infinity of the incomplete B function $\mathrm{B}(y;4N\nu,4N\mu)$ (which becomes the Dirac δ function) in Eq. (32). If we define a partial normalization constant for each attractive basin as

$$Z_0=\int_0^a \rho(x,+\infty)dx, \tag{49}$$

$$Z_1=\int_a^1 \rho(x,+\infty)dx, \tag{50}$$

Note that Z_0 is a shorthand of $Z_0(t=+\infty)$ defined in Eq. (24); $Z_1=1-Z_0$ is similar. The mathematical condition for the biological fixation at $x=0$ (or $x=1$) should be $Z_0=+\infty, Z_1<+\infty$ (or $Z_1=+\infty, Z_0<+\infty$). If $Z_0=Z_1=+\infty$, both $x=0$ and $x=1$ are fixation states, corresponding to the model with pure genetic drift. When combined with previous discussions, we conclude that $\Phi(0)=+\infty$ (or $\Phi(1)=+\infty$) does not necessarily imply fixation of A_1 (or A_2) genes.

Another observation from the present results is the emerging of absorbing boundaries at the fixation state; the boundary conditions "artificially" set by Ref. [47] are more naturally and generally derived here. Note that there is previous discussions of the boundary conditions based on MFPT alone [32]. Compared to previous work, our discussions are based on the MFPT and the landscape, which gives the physically meaningful escape time from the boundary states. The key issue here is that the choice of the exit state is closely related to the saddle location and landscape shape. Also, the confusing last-step-to-fixation dynamics in the continuous Wright-Fisher model [33] is coherently described by both the discrete and continuous escape-time results here, in a way validating the diffusion approximation even at the boundary states. Our last comment is that unnormalizable distributions in the diffusion model do not generate real problems for understanding the original discrete model. It instead provides important dynamical and equilibrium information of the system. We summarize the above conclusions in Table II.

TABLE II. Summary of the observations in Sec. VD. Z_0 and Z_1 are the partial normalization constants defined in Eqs. (49) and (50). τ_0 and τ_1 are the respective escape times. The "Absorb.Bound." column gives where the absorbing boundary emerges.

Z_0	Z_1	τ_0	τ_1	Fixation	Absorb.Bound.
$<+\infty$	$<+\infty$	$<+\infty$	$<+\infty$	N/A	Neither
$=+\infty$	$<+\infty$	$=+\infty$	$<+\infty$	$x=0$	$x=0$
$<+\infty$	$=+\infty$	$<+\infty$	$=+\infty$	$x=1$	$x=1$
$=+\infty$	$=+\infty$	$=+\infty$	$=+\infty$	$x=0$ or 1	$x=0,1$

E. Comments on the "stochastic tunneling"

In the study of a three-phase transition problem, a "stochastic tunneling" effect was termed that allows transition from one state to another, without passing through the middle state [48]. In light of the present framework of potential landscape and discussions of escape events, we commented that "stochastic tunneling" might not be a proper term. There the essential quantum-mechanical feature of the tunneling effect does not exist. Quantum mechanics is tied to the laws of wave mechanics going through under the potential barrier (here the potential valley) [49,50]. A potential barrier (valley) is not to be overcome but should be tunneled through, which is classically impossible. Furthermore, the tunneling effect is approximately temperature independent, while the "stochastic tunneling" disappears when temperature decreases to zero (stochastic effect vanishes). Actually, the first process of fixation of the deleterious mutation would never happen without noise. In the physical point of view, this process is a classical saddle-passage escape event on a potential landscape. The proper term might be reconsidered as "thermal activation."

ACKNOWLEDGMENTS

The critical comments of D. Waxman and T. Krüger on this work are appreciated. We also thank R. S. Yuan, J. H. Shi, Y. B. Wang, and other members in the laboratory for their constructive comments. We thank X. A. Wang for the technical support. This work was supported in part by the National 973 Project No. 2010CB529200 and by the Natural Science Foundation of China Grant No. NFSC91029738.

APPENDIX A: CONVERGENCE OF EQ. (37)

Under $\nu > 0$, the convergence of Eq. (37) relies on the convergence of the sum

$$S = \sum_{n=2}^{\infty} \prod_{k=2}^{n} \left(\frac{k-1+4N\mu}{k} \right) \frac{1}{n+1}. \tag{A1}$$

We use Raabe's test for series convergence from standard textbooks of real analysis. For $0 \leqslant 4N\mu < 1$, we denote

$$c_n = \prod_{k=2}^{n} \left(\frac{k-1+4N\mu}{k} \right) \frac{1}{n+1}. \tag{A2}$$

Obviously c_n is positive for all $n > 0$. First, we have

$$\lim_{n\to\infty} \frac{c_{n+1}}{c_n} = 1. \tag{A3}$$

We then calculate the Raabe terms

$$R_n = n\left(\frac{c_{n+1}}{c_n} - 1 \right) = (4N\mu - 2)\frac{n}{n+2}. \tag{A4}$$

Here $4N\mu - 2$ is a constant less than -1. By taking the limit $n \to \infty$,

$$\lim_{n\to\infty} R_n = 4N\mu - 2 < -1. \tag{A5}$$

The two conclusions in Eqs. (A3) and (A5) verify the convergence of the partial sum S_n under $0 \leqslant 4N\mu < 1$.

APPENDIX B: INTERPRETATION OF THE FACTOR 2

The choice of factor 2 is because a population will have $1/2$ probability to be caught stable in the $(a,1)$ basin after reaching the valley point a. This is not always the truth, though, as valley $x_1 = a$ is chosen as an ideal absorbing boundary (sink) rather than a smooth distribution of sinks in $(a,1)$ [15]. For the factor of 2 to be exact, we need the landscape to be axisymmetric near the valley. Asymmetry of the landscape far from the valley state will bring higher-order errors to the factor of 2, which is neglected for the present concern and needs further investigations. On the other hand, the error between $2T_{\mathrm{MFP}}$ and the simulated escape time may also come from the diffusion approximation of the discrete model and the approximation of the MFPT.

One detailed interpretation of this factor 2 in a limiting case is given below. Assume that the rates of uphill ($\sim \mathcal{T}_1$) and downhill [$T_{\mathrm{MFP}}(0 \to a)$, denoted here as T_a] movements are well separated (here $4N\nu, 4N\mu \ll 1$ so $\mathcal{T}_1 \ll T_a$). The model can be considered as a three-state-transition process among $0,a,1$. Once a population reaches $x = a$ in T_a, it has probability $1/2$ to fall into either 0 or 1 immediately in $\mathcal{T}_1$. Once falling back to $x = 0$, it will wait another T_a to reach $x = a$ again; it then again has a $1/2$ chance to reach $x = 1$ or return to $x = 0$ immediately. Assume this process continues. The expected time to leave $x = 0$ can then be obtained as (neglect $\mathcal{T}_1$)

$$\tau_0 \approx T_a \times \frac{1}{2} + 2T_a \times \frac{1}{2^2} + \cdots + nT_a \times \frac{1}{2^n} + \cdots = 2T_a, \tag{B1}$$

the same result as Eq. (35).

APPENDIX C: DISCRETE WRIGHT-FISHER PROCESS

The original Wright-Fisher model is discrete both in time (number of generations) and space (number of copies of A_1). It considers the evolution of the probability distribution function P_t (a vector of $2N+1$ elements) with time t. The tth generation sampled $2N$ times to give the $t+1$th generations. The probability that these $2N$ trials of sampling a population with i copies of A_1 gene will give j copies of A_1 gene is given by the (i,j)th element of the transition probability matrix G, defined by the binomial distribution as follows:

$$G_{ij} = C_j^{2N} p(i)^j (1-p(i))^{2N-j}. \tag{C1}$$

C_j^{2N} is the number of combinations to choose j genes from a gene pool of size $2N$. $p(i)$ is the probability of choosing an A_1 gene from the pool. To give the explicit form of $p(i)$, we denote $y = i/(2N)$. $p(i)$ is determined by the biological factors considered in the model. Here under genetic drift, mutation, and selection,

$$\begin{aligned} p(i) = {} & [(W_{11}y^2 + W_{12}y(1-y))(1-\mu) \\ & + (W_{12}y(1-y) + W_{22}(1-y)^2)\nu]/[W_{11}y^2 \\ & + 2W_{12}y(1-y) + W_{22}(1-y)^2] \end{aligned} \tag{C2}$$

is the probability of sampling an A_1 gene in the population gene pool. W_{11}, W_{12}, W_{22} denotes the fitness values of A_1A_1, A_1A_2, A_2A_2 individuals, respectively. The system evolves as

$$P_{t+1} = P_t G. \tag{C3}$$

In the simulation in Sec. IVC, the model is under mutation and genetic drift, and the fitness values are chosen as $W_{11} : W_{12} : W_{22} = 1 : 1 : 1$. In the first example of Sec. IVD, the fitness values are $W_{11} : W_{12} : W_{22} = 1 : 1 - s : 1$.

[1] J. H. Gillespie, *Population Genetics: A Concise Guide* (The Johns Hopkins University Press, Baltimore, 1998).
[2] J. H. Gillespie, Evolution **38**, 1116 (1984).
[3] M. Kimura, Ann. Math. Stat. **28**, 882 (1957).
[4] M. Kimura, Genetics **47**, 713 (1962).
[5] S. Wright, in *Proceedings of the Sixth International Congress on Genetics*, edited by D. F. Jones, Vol. 1 (Genetics Society of America, Austin, 1932), p. 356.
[6] S. J. Arnold, M. E. Pfrender, and A. G. Jones, Genetica **112–113**, 9 (2001).
[7] P. Ao, Phys. Life Rev. **2**, 117 (2005).
[8] N. H. Barton and S. Rouhani, J. Theor. Biol. **125**, 397 (1987).
[9] S. Gavrilets, Evolution **57**, 2197 (2003).
[10] P. Ao, J. Genet. Genom. **36**, 63 (2009).
[11] H. Qian, J. Phys. Condens. Matter **17**, S3783 (2005).
[12] T. Krüger, *Random Graphs and Random Walks An Introduction to the Stochastic Analysis of Complex Networks and Databases* (Springer, Berlin, 2010).
[13] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985).
[14] H. A. Kramers, Physica **7**, 284 (1940).
[15] P. Hanggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[16] S. Wright, Am. Nat. **131**, 115 (1988).
[17] J. A. Coyne, N. H. Barton, and M. Turelli, Evolution **51**, 643 (1997).
[18] S. Rice, *Evolutionary Theory: Mathematical and Conceptual Foundations* (Sinauer Associates, Sunderland, MA, 2004).
[19] J. Kaplan, Biol. Philos. **23**, 625 (2008).
[20] P. Ao, J. Phys. A: Math. Gen. **37**, L25 (2004).
[21] J. Wang, L. Xu, and E. Wang, Proc. Natl. Acad. Sci. USA **105**, 12271 (2008).
[22] Y. Iwasa, J. Theor. Biol. **135**, 265 (1988).
[23] G. Sella and A. E. Hirsh, Proc. Natl. Acad. Sci. USA **102**, 9541 (2005).
[24] N. H. Barton and J. B. Coe, J. Theor. Biol. **259**, 317 (2009).
[25] X. M. Zhu, L. Yin, L. Hood, and P. Ao, Funct. Integr. Genom. **4**, 188 (2004).
[26] P. Ao, D. Galas, L. Hood, and X. M. Zhu, Med. Hypotheses **70**, 678 (2008).
[27] Y. Tang, R. S. Yuan, and Y. A. Ma, Phys. Rev. E **87**, 012708 (2013).
[28] M. Kimura, J. Appl. Probab. **1**, 177 (1964).
[29] R. A. Blythe and A. J. McKane, J. Stat. Mech. Theor. Exp. (2007) P07018.
[30] R. Lande, Proc. Natl. Acad. Sci. USA **82**, 7641 (1985).
[31] M. Kimura and T. Ohta, Genetics **61**, 763 (1969).
[32] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Vol. 2 (Academic Press, San Diego, 1981).
[33] W. J. Ewens, *Mathematical Population Genetics* (Springer-Verlag, New York, 2004).
[34] P. Ao, C. Kwon, and H. Qian, Complexity **12**, 19 (2007).
[35] R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation* (Wiley, Chincester, UK, 2000).
[36] X. M. Zhu, L. Yin, and P. Ao, Int. J. Mod. Phys. B **20**, 817 (2006).
[37] R. S. Yuan, X. A. Wang, Y. A. Ma, B. Yuan, and P. Ao, Phys. Rev. E **87**, 062109 (2013).
[38] J. Felsenstein, *Theoretical Evolutionary Genetics* (2011), online book. Available at: http://evolution.genetics.washington.edu/pgbook/pgbook.html
[39] D. Zhou and H. Qian, Phys. Rev. E **84**, 031907 (2011).
[40] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007).
[41] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1984).
[42] J. G. Kingsolver, H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli, Am. Nat. **157**, 245 (2001).
[43] P. A. P. Moran, Ann. Hum. Genet. **27**, 383 (1964).
[44] C. Kwon, P. Ao, and D. J. Thouless, Proc. Natl. Acad. Sci. USA **102**, 13029 (2005).
[45] N. H. Barton and H. P. de Vladar, Genetics **181**, 997 (2009).
[46] S. Jiao and P. Ao, BMC Syst. Biol. **6**, S10 (2012).
[47] A. J. McKane and D. Waxman, J. Theor. Biol. **247**, 849 (2007).
[48] Y. Iwasa, F. Michor, and M. A. Nowak, Genetics **166**, 1571 (2004).
[49] E. V. Anslyn and D. A. Dougherty, *Modern Physical Organic Chemistry* (University Science Books, Mill Valley, CA, 2006).
[50] P. Ao and D. J. Thouless, Phys. Rev. Lett. **72**, 132 (1994).