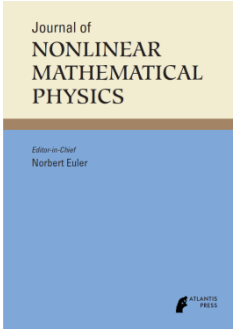

Journal of Nonlinear Mathematical Physics

ISSN (Online): 1776-0852 **ISSN (Print): 1402-9251**
Journal Home Page: https://www.atlantis-press.com/journals/jnmp

Nonlinear Deterministic Equations in Biological Evolution

Kavita Jain, Sarada Seetharaman

To cite this article: Kavita Jain, Sarada Seetharaman (2011) Nonlinear Deterministic Equations in Biological Evolution, Journal of Nonlinear Mathematical Physics 18:Supplement 2, 321–338, DOI: https://doi.org/10.1142/S1402925111001556

To link to this article: https://doi.org/10.1142/S1402925111001556

Published online: 04 January 2021

ARTICLE

World Scientific
www.worldscientific.com

ATLANTIS PRESS

Journal of Nonlinear Mathematical Physics, Vol. 18, Suppl. 2 (2011) 321–338

DOI: 10.1142/S1402925111001556

NONLINEAR DETERMINISTIC EQUATIONS IN BIOLOGICAL EVOLUTION

KAVITA JAIN*,‡ and SARADA SEETHARAMAN†

Theoretical Sciences Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
Jakkur P.O., Bangalore, India
*jain@jncasr.ac.in
†saradas@jncasr.ac.in

Received 24 November 2010
Revised 17 March 2011
Accepted 18 March 2011

We review models of biological evolution in which the population frequency changes deterministically with time. If the population is self-replicating, although the equations for simple prototypes can be linearised, nonlinear equations arise in many complex situations. For sexual populations, even in the simplest setting, the equations are necessarily nonlinear due to the mixing of the parental genetic material. The solutions of such nonlinear equations display interesting features such as multiple equilibria and phase transitions. We mainly discuss those models for which an analytical understanding of such nonlinear equations is available.

Keywords: Deterministic equations; biological evolution; multiple equilibria; phase transitions.

1. Introduction

A population evolves when the changes that happen during a generation are passed on to the subsequent generations. These changes may happen in the somatic immune cells in order to adapt to a microbe attack or in the germline cells. Though in both the cases the genome is altered, in the former, it also manifests as changes in the composition of the protein coded by that part of the genome. Therefore one defines the models describing biological evolution in genotype or protein space [44].

The quantity of interest is the population frequency of a genotype which changes under the action of two elementary processes namely selection and mutation. In the simplest setting, the time-dependent equations for the population fraction are nonlinear but they can be linearised and the steady state solution obtained at long times can be shown to be unique. In more complex situations such as when subpopulations are coupled to each other or when the growth rate of a genotype depends on its current frequency, nonlinear evolution equations give rise to multiple equilibria. In the cases where the solution is unique,

‡Also at Evolutionary and Organismal Biology Unit.

phase transition may occur in the steady state. If the process of sexual reproduction is also included, the resulting equations are bilinear in population and such inherently nonlinear equations exhibit multiple solutions in the steady state and dynamic phase transitions.

In this review, we will focus on infinite populations which obey deterministic equations of evolution. Although the real populations are finite and evolve stochastically, phenomena observed in deterministic setting may survive in the presence of stochasticity as well [40], and deterministic solutions can also be utilised to get insight in the corresponding stochastic problem [24] and to develop stochastic theories [20]. For a discussion of topics not covered in this article, we refer the reader to several excellent textbooks [15, 11, 35] and other review articles on the subject [23].

The article is organised as follows. In the next section, we introduce some basic concepts and definitions. This is followed by a discussion of models for asexually reproducing populations in Sec. 3 and sexually reproducing ones in Sec. 4. Finally a summary and outlook is presented in Sec. 5.

2. Basic Definitions

In this section, we explain some basic concepts and definitions which are relevant to the discussion in the following sections.

Sequence and sequence space: A sequence $\sigma = \{\sigma_1, \ldots, \sigma_L\}$ is a string of L letters which are chosen from an alphabet of size a. It represents a protein if σ_i denotes one of the $a = 20$ amino acids and a genotype when the letters are one of the four nucleotides. The total sequence space consists of all possible strings of length L and thus has a size $n = a^L$ which increases exponentially with L. For computational ease, it is useful to lump some of the information in a single letter. For example, instead of working with all the four nucleotides in a genotype, one can classify them as purines (adenine and guanine) and pyrimidines (thymine and cytosine) thus reducing a to two. Similarly instead of considering all possible mutations at a locus, one may differentiate between genotypes by the absence or presence of a mutation which again corresponds to $a = 2$ [50]. In this article, we will work with binary sequences unless specified otherwise. Such $n = 2^L$ sequences can be arranged on a Hamming space, an example of which is shown in Fig. 1 for binary sequence of length $L = 3$. Two sequences σ and σ' are said to be at Hamming distance $d(\sigma, \sigma')$ if they differ

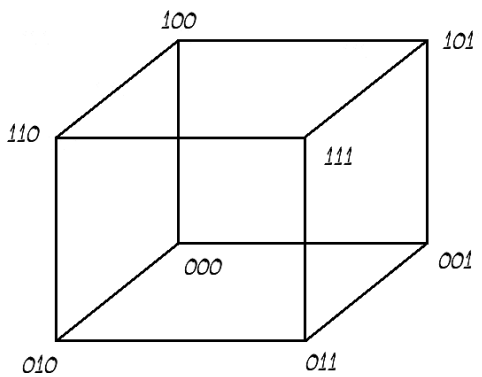

Fig. 1. The sequence space for $L = 3$ represented on a Hamming cube.

at d loci. For a binary sequence in which $\sigma_i = 0$ or 1, one may write

$$d(\sigma, \sigma') = \sum_{i=1}^{L} (\sigma_i - \sigma'_i)^2. \tag{1}$$

Fitness: The fitness $W(\sigma)$ of a sequence σ is a measure of its reproductive success in a given environment. It represents the replication rate of a genotype or the functionality of a protein. The sequence space along with the fitness of each sequence comprises the *fitness landscape*. The choice of fitness landscape plays an important role in determining the course of evolution and can be made according to the biological situation that one wishes to model and the available experimental data or the analytical tractability of the problem. A fitness landscape can be simple in that the fitness of a sequence depends only on its distance from a given sequence. More often however the fitness landscapes are complex and one has to specify all the a^L fitnesses. These fitnesses can be assumed to be independent random variables [22] or they may have correlations [39, 43]. An important feature of generic fitness landscapes is the presence of epistasis which is a measure of the nonlinear contribution of locus fitness to the sequence [29]. If each locus contributes independently to the sequence fitness, a fitness landscape is said to be nonepistatic. Fitness can also depend on time as in the case of changing environment [32, 33, 53] or it can be a function of the concentration of the genotype frequency. In this review, we will employ various types of fitness landscapes.

Mutation: Stochastic changes known as mutations may happen in the genome of an individual. These may insert, delete or change the nucleotides in the genome and thus create a new sequence with a different fitness. If the fitness of the mutant is higher, the change may propagate in the population and the population evolves towards a higher fitness value, otherwise it is eliminated. In this review, we will consider only *point mutations* that change a locus σ_i to one of the other $a - 1$ possibilities with a certain probability and thus preserve the length of the sequence.

Recombination: A sequence genetically different from the parents can be produced by the recombination process in which two parent sequences mix to produce a new offspring sequence thus producing genetic variation within a population. Recombination occurs not only during gamete formation in sexually reproducing multicellular organisms but in unicellular organisms such as bacteria and fungi as well [16]. We will consider a recombination scheme (one-point crossover) in which the parent sequences σ and σ' break at a point i and exchange the genetic material with a certain probability resulting in offspring sequences $\{\sigma_1, \ldots, \sigma_i, \sigma'_{i+1}, \ldots, \sigma'_L\}$ and $\{\sigma'_1, \ldots, \sigma'_i, \sigma_{i+1}, \ldots, \sigma_L\}$.

3. Asexually Reproducing Populations

We first describe the equations governing the evolution of self replicating populations. Although the time-dependent equations for the population frequency of such asexual populations are nonlinear in general, they can be linearised by a transformation of variables in some simple cases [49, 25]. We will mainly discuss the steady state properties of these models in the following subsections.

3.1. *Haploid population*

In a haploid population, each individual carries a single copy of its genome sequence σ. In the presence of selection and mutation, the population frequency $X(\sigma, t+1)$ of a sequence σ

at generation $t+1$ can be obtained from each sequence σ' that makes $W(\sigma')$ copies of itself in one generation and mutates to sequence σ with a probability $M(\sigma \leftarrow \sigma')$. This gives the discrete time evolution equation as

$$X(\sigma, t+1) = \frac{\sum_{\sigma'} M(\sigma \leftarrow \sigma') W(\sigma') X(\sigma', t)}{\sum_{\sigma'} W(\sigma') X(\sigma', t)} \tag{2}$$

where the denominator on the right-hand side (RHS) is the average fitness $\mathcal{W}(t)$ and ensures that the normalisation $\sum_\sigma X(\sigma, t) = 1$ is satisfied at all times. If the mutation probability per locus per generation is μ and the point mutations occur independently at each locus, the probability that a sequence σ' mutates to sequence σ at Hamming distance $d(\sigma, \sigma')$ is given by

$$M(\sigma \leftarrow \sigma') = \mu^{d(\sigma,\sigma')}(1-\mu)^{L-d(\sigma,\sigma')} \tag{3}$$

It is evident that Eq. (2) is nonlinear due to the presence of denominator. However in terms of an unnormalised population variable defined as

$$Z(\sigma, t) = X(\sigma, t) \prod_{\tau=0}^{t-1} \sum_{\sigma'} W(\sigma') X(\sigma', \tau) \tag{4}$$

we find that the unnormalised variables $Z(\sigma, t)$ obey a linear equation given by

$$Z(\sigma, t+1) = \sum_{\sigma'} M(\sigma \leftarrow \sigma') W(\sigma') Z(\sigma', t). \tag{5}$$

On writing

$$X(\sigma, t) = \frac{Z(\sigma, t)}{\sum_{\sigma'} Z(\sigma', t)} \tag{6}$$

in (5), Eq. (2) is obtained. In matrix notation, (5) can be written as $\mathbf{Z}(t+1) = A\mathbf{Z}(t)$ where the σ, σ' element of matrix A is given by $M(\sigma \leftarrow \sigma')W(\sigma')$ and $\mathbf{Z}(t)$ is the population vector at time t. Since the fitness $W(\sigma) \geq 0$, the matrix A is non-negative and it follows from the Perron–Frobenius theorem that the largest eigenvalue of matrix A is real, positive and nondegenerate with the corresponding eigenvector real and positive [4]. Using this eigenvector in (6) and taking the infinite time limit, the normalised frequencies in the steady state can be obtained. However in some cases, it is possible to work directly with the nonlinear equation (2) in the steady state (see the discussion below).

In continuous time, one can write down the equation for the rate of change $\dot{X}(\sigma, t) = \partial X(\sigma, t)/\partial t$ of the fraction $X(\sigma, t)$ of the population with sequence σ as

$$\dot{X}(\sigma, t) = \sum_{\sigma'} M(\sigma \leftarrow \sigma') W(\sigma') X(\sigma', t) - \left(\sum_{\sigma'} W(\sigma') X(\sigma', t) \right) X(\sigma, t) \tag{7}$$

where the last term on the RHS is the death term which accounts for the normalisation $\sum_\sigma X(\sigma, t) = 1$. Note that (7) is *not* the continuous time limit of (2) although both equations have the same steady state.

The Eqs. (2) and (7) define respectively the discrete and continuous time versions of *Eigen's quasispecies model* [12, 13]. The main result of the quasispecies theory is that in

the steady state, for several choices of fitness landscapes, there exists a critical mutation rate below which the population forms a quasispecies consisting of the fittest sequence and its closely related mutants. Above this error threshold, the population is homogeneously distributed over the entire sequence space. To illustrate this, we consider the *sharp peak fitness landscape* defined by

$$W(\sigma) = W_0\delta_{\sigma,\mathbf{0}} + (1 - \delta_{\sigma,\mathbf{0}}), \quad W_0 > 1 \tag{8}$$

where $\mathbf{0} = \{0, 0, \ldots, 0\}$ is the sequence with all zeros. Using this choice for $W(\sigma)$ in (2) for the sequence $\mathbf{0}$ in the steady state, we get

$$X(\mathbf{0}) = \frac{W_0(1-\mu)^L X(\mathbf{0}) + \sum_{\sigma' \neq \mathbf{0}} M(\mathbf{0} \leftarrow \sigma')\mathbf{X}(\sigma')}{W_0 X(\mathbf{0}) + 1 - X(\mathbf{0})} \tag{9}$$

In the scaling limit $\mu \to 0, L \to \infty$ with $U = \mu L$ finite, the terms in the numerator on RHS arising due to mutations to sequence $\mathbf{0}$ vanish and we obtain [34]

$$X(\mathbf{0}) = 1 - \frac{U}{U_c}, \quad U < U_c = \ln W_0 \tag{10}$$

Thus the master sequence $\mathbf{0}$ supports a finite fraction of population below U_c. Above the critical probability U_c, the population is homogeneously distributed over the sequence space.

Not all fitness landscapes exhibit error threshold transition [51]. One such example is the non-epistatic *multiplicative fitness landscape* defined by

$$W(\sigma) = \prod_{i=1}^{L} (1-s)^{\sigma_i} \tag{11}$$

where the $0 < s < 1$ is a selection parameter. It can be checked that the exact steady state frequency is given by [54],

$$X(\sigma) = \prod_{i=1}^{L} x_0^{1-\sigma_i} x_1^{\sigma_i} \tag{12}$$

where x_0, x_1 are the solutions of (2) for the corresponding one locus model. For a discussion of error threshold transition on other fitness landscapes, we refer the reader to [23].

If the replication and mutation are treated as independent processes unlike in (2) and (7), we obtain the *Crow–Kimura model* [9, 8] in which it is assumed that the replication process is error-free and mutations occur due to external factors such as radiation. Then the equation for the rate of change $\dot{X}(\sigma, t)$ can be written as [9, 1]

$$\dot{X}(\sigma, t) = \left[W(\sigma) - \sum_{\sigma'} W(\sigma') X(\sigma', t) \right] X(\sigma, t) + \sum_{\sigma'} M(\sigma \leftarrow \sigma') X(\sigma', t). \tag{13}$$

where the mutation matrix is given by

$$M(\sigma \leftarrow \sigma') = \begin{cases} 0, & d(\sigma, \sigma') > 1 \\ \mu, & d(\sigma, \sigma') = 1 \\ -L\mu, & d(\sigma, \sigma') = 0 \end{cases} \tag{14}$$

since $\sum_\sigma M(\sigma \leftarrow \sigma')$ should be zero. As in the Eigen's model, the nonlinearity in (13) can be eliminated by passing to unnormalised population variables $Z(\sigma, t)$ defined by

$$Z(\sigma, t) = X(\sigma, t) \exp\left[\sum_{\sigma'} W(\sigma') \int_0^t d\tau X(\sigma', \tau)\right] \tag{15}$$

The error threshold transition for various fitness landscapes has been demonstrated using the Crow–Kimura equation (13) also [2, 41].

3.2. *Diploid population*

Higher organisms such as humans are diploid as they carry two copies of their genome and we represent an individual of a diploid population by (σ, σ'). A sequence is said to be homozygous if σ and σ' are identical and heterozygous otherwise. Selection-mutation equations analogous to the haploid case can be written for the population frequency $X(\sigma, t)$ of the sequence σ. For the Crow–Kimura model, the evolution equation reads as [52, 3]

$$\dot{X}(\sigma, t) = \left[\tilde{W}(\sigma, t) - \sum_{\sigma''} \tilde{W}(\sigma'', t) X(\sigma'', t)\right] X(\sigma, t) + \sum_{\sigma'} M(\sigma \leftarrow \sigma') X(\sigma', t). \tag{16}$$

where $\tilde{W}(\sigma, t) = \sum_{\sigma'} W(\sigma, \sigma') X(\sigma', t)$ is the marginal fitness of sequence σ and $W(\sigma, \sigma')$ is the fitness of genotype (σ, σ'). A transformation similar to (15) which can render the above system of nonlinear equations linear is not known and the steady solution may not be unique.

The existence of multiple steady state solutions can be illustrated by a diploid analogue of the sharp fitness landscape defined as [52]

$$W(\mathbf{0}, \mathbf{0}) = f_0 = 1 + 2s \tag{17}$$

$$W(\mathbf{0}, \sigma) = W(\sigma, \mathbf{0}) = f_1 = 1 + 2hs, \quad \sigma \neq \mathbf{0} \tag{18}$$

$$W(\sigma, \sigma') = 1, \quad \sigma, \sigma' \neq \mathbf{0} \tag{19}$$

where $s, h > 0$. In the above equations, s is a selection coefficient and h is a dominance parameter which controls the contribution of the master sequence to the fitness of the heterozygote. When $h = 1$, since the fitness $W(\mathbf{0}, \sigma) = W(\mathbf{0}, \mathbf{0})$, the master sequence $\mathbf{0}$ is dominant. On the other hand, when $h = 0$, the fitness $W(\mathbf{0}, \sigma) = W(\sigma, \sigma') = 1$ and therefore the master sequence acts recessively. The dominance is absent when $h = 1/2$ as the heterozygote fitness $W(\mathbf{0}, \sigma) = 1 + s$ is the average of the master fitness and the mutant fitness.

Using the above equation, the marginal fitness can be written as

$$\tilde{W}(\sigma) = \begin{cases} f_0 X(\mathbf{0}) + f_1(1 - X(\mathbf{0})), & \sigma = \mathbf{0} \\ f_1 X(\mathbf{0}) + (1 - X(\mathbf{0})), & \sigma \neq \mathbf{0} \end{cases} \tag{20}$$

and the average fitness as

$$\sum_{\sigma''} \tilde{W}(\sigma'') X(\sigma'') = X(\mathbf{0})[f_0 X(\mathbf{0}) + f_1(1 - X(\mathbf{0}))] + (1 - X(\mathbf{0}))[f_1 X(\mathbf{0}) + 1 - X(\mathbf{0})]. \tag{21}$$

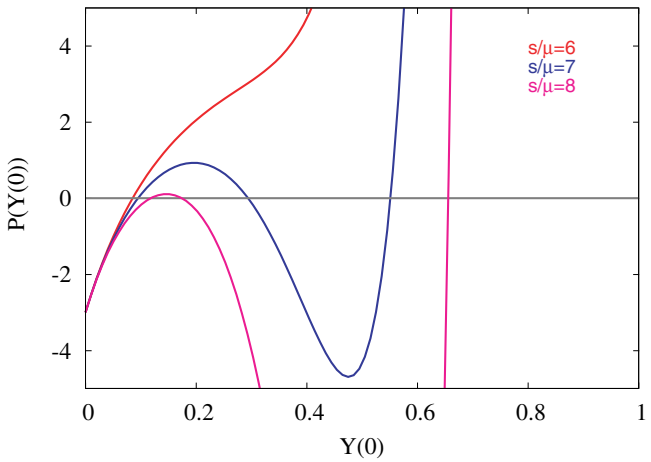

Fig. 2. Plot of the polynomial $P(Y(0))$ as a function of $Y(0)$ for various s/μ (see Subsec. 3.2). The equilibrium frequency $Y(0)$ is obtained when $P(Y(0)) = 0$.

Since the fitness landscape (17)–(19) depends only on the Hamming distance from the master sequence $\mathbf{0}$, one can work with the error class frequencies $Y(d)$ which are obtained by summing over the population fractions at Hamming distance d from the master sequence. Specialising to $h = 0$, the steady state equation in terms of Y's reads as [52]

$$\sum_{d'=0}^{L} \tilde{M}(d \leftarrow d')Y(d') = 2sY^2(0) \begin{cases} Y(0) - 1, & d = 0 \\ Y(d), & d \neq 0 \end{cases} \tag{22}$$

where the mutation matrix $\tilde{M}$ can be found using (14). The frequency $Y(0)$ obeys a polynomial equation of degree at most $2(L+1)$. For small L, the above set of nonlinear equations can be straightforwardly solved. For $L = 4$, the fraction $Y(0)$ obeys a polynomial equation $P(Y(0)) = 0$ of degree 9 [52]. The polynomial $P(Y(0))$ is plotted against $Y(0)$ in Fig. 2 for various s/μ to show the occurrence of multiple steady state solutions. Which of these multiple solutions occur depends on the initial conditions. For example, an initial distribution with $Y(0) = 1$ gives different steady state fitness from the initial condition $Y(L) = 1$ [52].

3.3. *Concentration-dependent fitness*

The fitness of a sequence is not always a constant and may depend on the concentration of other sequences. In such cases, one ends up with nonlinear dynamical equations which cannot be linearised. An example of this scenario is the evolution of grammar in a population [28]. It has been proposed [46, 36] that a set of grammars $G_1, G_2, \ldots, G_n$ are innately available to a learner and the language is learnt by just listening to the sentences and choosing the correct grammar.

A grammar that is easily understandable has a greater probability of being propagated than the others and hence the fitness indicates its prevalence in the population. This is equal to the fraction of sentences and their corresponding meanings that is common between that grammar and all others multiplied by the population fraction using each grammar. If $w(i, j)$ is the probability that a speaker of grammar G_j can understand a sentence by a user of

grammar G_i, the fitness $W(\{X(i)\})$ of grammar G_i can be given as [28]

$$W(\{X(i)\}) = \frac{1}{2}\sum_{j=1}^{n}[w(i,j) + w(j,i)]X(j). \tag{23}$$

If the probability that a person learning from a teacher speaking grammar G_i ends up with grammar G_j is $M(j \leftarrow i)$, the rate of change of the population speaking G_j can be written as

$$\dot{X}(j,t) = \sum_{i=1}^{n} M(j \leftarrow i)W(\{X(i)\})X(i,t) - \left(\sum_{i=1}^{n} W(\{X(i)\})X(i,t)\right)X(j,t). \tag{24}$$

The interpretation of the terms in the above equation is similar to (2) or (7). However an important difference is that the fitness $W(\{X(i)\})$ of the grammar G_i now depends on the frequency of the other grammars as well. Such a selection-mutation equation with concentration-dependent fitness is known as replicator-mutator equation [35].

Assuming that the error to any grammar is equally likely, it follows that $M(j \leftarrow i) = q\delta_{i,j} + [(1-q)/(n-1)](1-\delta_{i,j})$ where $q = 1-\mu$ is the learning accuracy. A detailed analysis of the above equation is possible for the fitness choice [28]:

$$w(i,j) = w(j,i) = w \quad \text{for } i \neq j \tag{25}$$

$$w(i,i) = 1. \tag{26}$$

The stable fixed points for the system of equations given by (24) can be found by setting the left-hand side to be zero and choosing all grammars except one, say $X(1) = X$, to be equally used so that, $X(i) = (1-X)/(n-1),\ i \neq 1$. This reduces the equation for $X(1)$ to

$$X^3 - X^2 q + \frac{(1-X)^2}{n-1}\left(X - \frac{1-q}{n-1}\right) + \frac{(1-q)w(nX-1)}{(1-w)(n-1)} = 0. \tag{27}$$

The above cubic equation for X has three solutions namely X_0, X_+ and X_- as shown in Fig. 3. The solution X_0 corresponds to the case in which all the grammars are equally used and exists for all $0 \leq q \leq 1$. The other two solutions $X_\pm$ appear beyond a critical learning

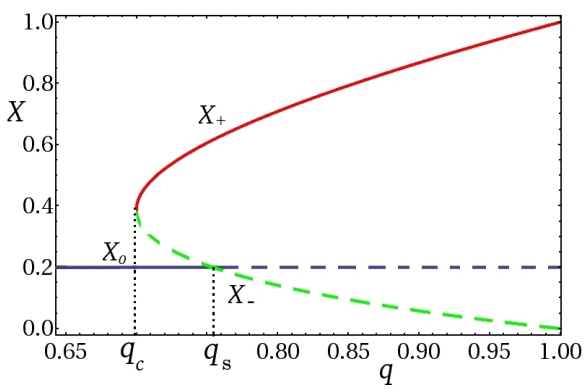

Fig. 3. Multiple solutions of population frequency X for $w = 0.2$ and $n = 5$ (see Subsec. 3.3). Unstable solutions are indicated by broken lines and stable ones by solid lines.

accuracy q_c and correspond to the most used (X_+) and the least used (X_-) grammars. Using a linear stability analysis it can be shown that the stability of these solutions falls in three regimes depending on the learning accuracy q: when $q < q_c$, the fraction X_0 is the only solution and is stable, whereas in the range $q_c \leq q < q_s$ all the three solutions exist but X_- is unstable and finally when $q \geq q_s$, the fraction X_0 also loses stability and X_+ is the only stable solution.

Concentration based fitness is confined not just to languages but is also seen in other systems such as host-parasite [6, 42] and immune system-pathogen interactions [18, 26]. In these systems, the evolution is not based on the concentration of the same species populations but on the concentration of other species. Thus their evolution equations are coupled and this is dealt with in the next section.

3.4. *Coupled quasispecies models*

A class of models in which the growth of a population depends on another population constitute an example of a set of nonlinear evolution equations. Below we discuss two such models in some detail.

Coevolution of quasispecies: When an organism is infected by a virus, the immune receptors of the host cell counterattack the virus. There is a one-to-one mapping between the virus and the immune receptors so that a viral sequence σ is attacked only by its corresponding receptor sequence $\tilde{\sigma}$, σ' only by $\tilde{\sigma}'$ and so on. In order to escape the immune system, the virus adapts and in response, the immune system adapts to counter the new viral strain (see Fig. 4) and this cycle repeats over a time period τ. Thus the viral species and the immune receptors are involved in a dynamic evolutionary race but may coexist under certain conditions as explained below.

Assuming that both the receptor and viral sequences have the same length L, the evolution equations for the frequency $X(\tilde{\sigma}, t)$ of immune receptor sequence $\tilde{\sigma}$ and $x(\sigma, t)$ of the corresponding viral sequence σ can be written as [26]:

$$\dot{X}(\tilde{\sigma}, t) = \sum_{\tilde{\sigma}'} M_\nu(\tilde{\sigma} \leftarrow \tilde{\sigma}') W(x(\sigma', t)) X(\tilde{\sigma}', t) - \tilde{D} X(\tilde{\sigma}, t) \tag{28}$$

$$\dot{x}(\sigma, t) = \sum_{\sigma'} M_\mu(\sigma \leftarrow \sigma') w(\sigma', t) x(\sigma', t) - D(X(\tilde{\sigma}, t)) x(\sigma, t) \tag{29}$$

where the subscripts in the sequence mutation probability M (see (3)) denote the mutation probability per locus and the death term of the immune receptor $\tilde{D} = \sum_{\tilde{\sigma}'} W(x(\sigma', t)) X(\tilde{\sigma}', t)$. As the immune receptor population moves in response to the viral population, the fitness $W(x(\sigma, t))$ of the receptor $\tilde{\sigma}$ depends on the concentration of the corresponding viral sequence σ. In the above equations, the death terms are different for the two populations as the number of immune receptors is conserved while the virus number is not. For simplicity, one can choose the death rate of the virus as

$$D(X(\tilde{\sigma}, t)) = \begin{cases} \delta, & \text{if } \tilde{\sigma} = \text{immune receptor master sequence} \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

A time-dependent sharp peak fitness landscape is assumed for both immune receptor and virus as their master sequences move through the sequence space. Since the viral fitness w

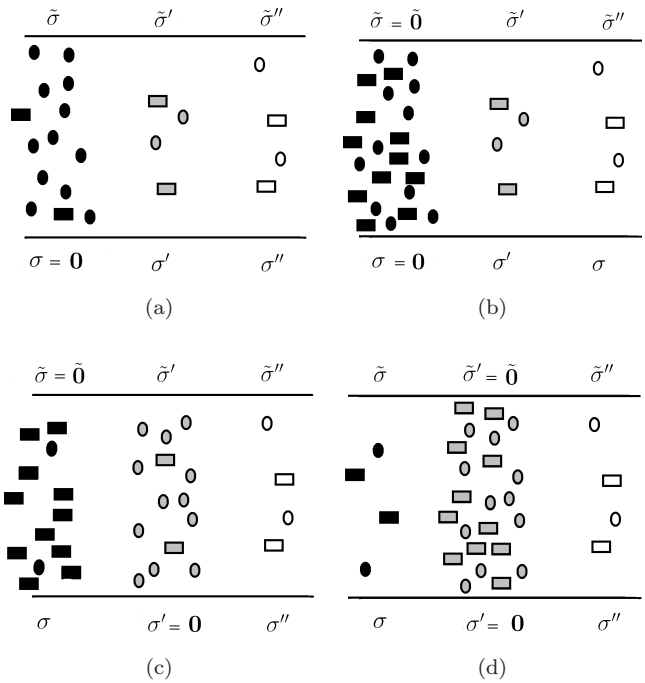

Fig. 4. Dynamics of the coevolution of the viral (ovals) and the immune receptor sequences (rectangles): (a) The viral quasispecies is initially formed around a master sequence $\mathbf{0}$ surrounded by its mutant sequences. (b) The receptor sequence $\tilde{\mathbf{0}}$ corresponding to $\mathbf{0}$ proliferates and forms receptor quasispecies around $\tilde{\mathbf{0}}$. (c) To escape the attack of immune system, the viral master sequence randomly shifts to one of its one mutant neighbours. (d) In response, the master sequence of the immune receptor also migrates to the location of the new viral master sequence.

is independent of $X(\sigma)$, we can write

$$w(\sigma, t) = \begin{cases} w_0, & \text{if } \sigma = \text{viral master sequence at time } t \\ 1, & \text{otherwise} \end{cases} \tag{31}$$

where $w_0 > 1$. Similarly $W(x(\sigma, t)) = W_0 > 1$ if σ is the viral master sequence and unity otherwise.

In periodically changing fitness landscapes such as being considered here, there is no steady state as the population keeps migrating with the fitness landscape. However one can still define an error threshold in the large time limit analogous to that on static fitness landscapes as the maximum mutation rate above which the population gets uniformly distributed over the sequence space. A possible way to determine the critical mutation rate is to consider the behavior of relative frequency κ of the new master sequence to the frequency of a sequence far away from the current master sequence at the time period τ of the fitness landscape [32]. At large times, it is a good approximation to assume that the far-off sequences have reached a quasi-equilibrium and therefore their unnormalised frequency grows exponentially fast with the growth rate given by the respective fitness. However such an equilibrium is not reached for the populations in the vicinity of the (migrating) master sequence and the growth at such sequences depends on the mutational contribution from the current master sequence. If the mutation probability or the time period is too small, the

population cannot build up at the new master sequence and the relative frequency $\kappa < 1$. On the other hand, the new master sequence grows for $\kappa > 1$. Thus $\kappa = 1$ marks the transition point between the extinction and survival phases of the quasispecies on periodically changing fitness landscapes.

Following the arguments sketched above, the fraction κ_μ for the virus can be found and is given by [26]

$$\kappa_\mu = \frac{(e^{((1-\mu)^L w_0 - 1)\tau} - e^{((1-\mu)^L - 1)\tau})\mu w_0}{(w_0 - 1)(1-\mu)}. \tag{32}$$

The relative frequency of the immune receptors κ_ν is obtained on replacing μ by ν and w_0 by W_0 in the above expression. Setting κ_μ and κ_ν equal to one gives a phase diagram in $\mu - \nu$ plane which shows that while both the populations exhibit the classical error catastrophe at high mutation rates (as discussed in Subsec. 3.1), the viral population has an additional transition point when its mutation rate is too low to escape the immune response and in between these values the two populations coexist [32, 26]. The predicted mutation rates of the B-cells that produce the immune receptors and the receptor lengths that maximise both regimes of viral error catastrophe for optimal immune response are seen to match the experimental observations [26].

Evolution of a mixed population: As discussed in Subsec. 3.1, there exists an error threshold above which the mutational load is too high to be compensated by selection. For this reason, and because most mutations are known to have deleterious effect [47, 10], the spontaneous mutation rate is expected to be small [27]. However small subpopulations of strains with high mutation rates have been observed in natural isolates [30] and in experiments [45, 5].

Consider such a mixed population with nonmutator and mutator strains with mutation probability μ and $\nu = \lambda\mu, \lambda > 1$ respectively. Due to the damage in error repair systems, the mutation rate of normal strains can rise and hence a nonmutator can convert to a mutator with probability f. Then the average fraction $x(\sigma, t)$ and $X(\sigma, t)$ of the nonmutator and the mutator respectively at generation t evolves according to the following coupled nonlinear difference equations [31]:

$$x(\sigma, t+1) = \frac{(1-f)\sum_{\sigma'} M_\mu(\sigma \leftarrow \sigma')W(\sigma')x(\sigma', t)}{\mathcal{W}(t)} \tag{33}$$

$$X(\sigma, t+1) = \frac{\sum_{\sigma'} M_\nu(\sigma \leftarrow \sigma')W(\sigma')X(\sigma', t)}{\mathcal{W}(t)} + \frac{f\sum_{\sigma'} M_\mu(\sigma \leftarrow \sigma')W(\sigma')x(\sigma', t)}{\mathcal{W}(t)} \tag{34}$$

where the average fitness $\mathcal{W}(t) = \sum_\sigma W(\sigma)[x(\sigma, t) + X(\sigma, t)]$ and the subscripts in the mutation matrix refer to the mutation probability per locus per generation. For the reasons mentioned above, the mutators are selected against and their number is expected to be low. But with increasing f, mutators are continually generated thus increasing their frequency and at sufficiently high f, the mutator population can reach unity. Thus a phase transition can occur at a critical probability f_c between the mixed phase with both nonmutator and mutator population and a pure mutator phase (see Fig. 5). In the steady state, such a phase transition has been shown to occur on single peak fitness landscapes [48] and multiplicative fitness landscapes [31].

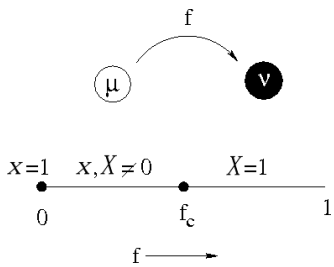

Fig. 5. Schematic phase diagram of the quasispecies model with nonmutator and mutator populations. The pure nonmutator phase occurs when $f = 0$, pure mutator phase for $f \geq f_c$ and the system is in the mixed phase for $0 < f < f_c$ (see Subsec. 3.4).

To see this transition for fitness choice (11), we first observe that $x(\sigma) = 0$ is a solution of Eq. (33) and thus corresponds to a phase in which the entire population consists of mutators and the total mutator fraction $X = \sum_\sigma X(\sigma) = 1$. As (34) reduces to (2) in this phase, using the exact solution (12), the average fitness $\mathcal{W}_>$ in the $f > f_c$ phase can be found. If, on the other hand, the total nonmutator fraction $x = \sum_\sigma x(\sigma)$ is nonzero, on summing over all the sequences on both sides of Eq. (33), we find that the average fitness $\mathcal{W}_<$ in the mixed phase corresponding to $f < f_c$ does not depend on the mutator fraction and can be written as

$$\mathcal{W}_< = \frac{(1-f)\sum_\sigma W(\sigma)x(\sigma)}{\sum_\sigma x(\sigma)}, \quad x \neq 0 \tag{35}$$

thus leading to an uncoupled nonlinear equation for $x(\sigma)$. On eliminating $\mathcal{W}$ from Eq. (33) using the above equation, we see that $x(\sigma)/\sum_{\sigma'} x(\sigma')$ obeys the quasispecies equation (2) and one can find the average fitness $\mathcal{W}_<$ as well. Equating the fitnesses $\mathcal{W}_<$ and $\mathcal{W}_>$ at the critical point, the phase boundary in the $f - \lambda$ plane is obtained,

$$(1-f_c)^{1/L} = \frac{(2-s)(1-\nu_c) + \sqrt{4\nu_c^2(1-s) + s^2(1-\nu_c)^2}}{(2-s)(1-\mu) + \sqrt{4\mu^2(1-s) + s^2(1-\mu)^2}}. \tag{36}$$

Using the above analysis, it is also possible to calculate the average mutator fraction as a function of f and λ. The results are seen to be in good agreement with the experiments on *E. coli* [31].

4. Sexually Reproducing Populations

In this section, we mainly consider a recombining haploid population with sequence length two. As discussed in Sec. 2, due to recombination, the sequences $\{0,0\}$ and $\{1,1\}$ can give rise to offspring sequences $\{0,1\}$ or $\{1,0\}$. Similarly the recombination between $\{0,1\}$ and $\{1,0\}$ can result in $\{0,0\}$ and $\{1,1\}$. In the following, for brevity we denote the population at the sequences $\{0,0\}, \{0,1\}, \{1,0\}$ and $\{1,1\}$ by X_0, X_1, X_2 and X_3 and their respective fitness by W_0, W_1, W_2 and W_3. If such a population undergoes recombination alone, the

frequency $X_i(t)$ evolves according to the following equation:

$$X_i(t+1) = \sum_{j,k=0}^{3} R(i \leftarrow j,k) X_j(t) X_k(t) \tag{37}$$

where $R(i \leftarrow j,k)$ is the probability that a sequence i is obtained by recombining sequences j and k. The recombination process between suitable sequences is assumed to occur with probability r and does not occur with $1-r$. For example, for the offspring sequence 0, we have

$$R(0 \leftarrow 0,0) = 1, \quad R(0 \leftarrow 0,1) = R(0 \leftarrow 0,2) = \frac{r}{2} + \frac{1-r}{2} = \frac{1}{2} \tag{38}$$

$$R(0 \leftarrow 0,3) = \frac{1-r}{2}, \quad R(0 \leftarrow 1,2) = \frac{r}{2} \tag{39}$$

and the rest of the probabilities are zero. On writing the recombination probabilities in a similar manner for other sequences and using the normalisation $\sum_{i=0}^{3} X_i = 1$, we find that the population fractions evolve according to [15]

$$X_0(t+1) = X_0(t) + r(X_1(t)X_2(t) - X_0(t)X_3(t)) \tag{40}$$

$$X_1(t+1) = X_1(t) + r(X_0(t)X_3(t) - X_1(t)X_2(t)) \tag{41}$$

$$X_2(t+1) = X_2(t) + r(X_0(t)X_3(t) - X_1(t)X_2(t)) \tag{42}$$

$$X_3(t+1) = X_3(t) + r(X_1(t)X_2(t) - X_0(t)X_3(t)). \tag{43}$$

Thus the population fractions obey a set of nonlinear equations when recombination is present and it is not known if these equations can be linearised.

The bilinear frequency combination $X_1(t)X_2(t) - X_0(t)X_3(t)$ is called *linkage disequilibrium* $D(t)$ at time t and is a measure of the correlation between the frequency at the two loci. Using (40)–(43) we have $D(t+1) = (1-r)D(t)$ so that the linkage disequilibrium vanishes in the steady state i.e. $X_1X_2 = X_0X_3$ and as a consequence, the frequency of the sequence $\{\sigma_1, \sigma_2\}$ equals the product of frequency of sequences $\{\sigma_1\}$ and $\{\sigma_2\}$. For example, the frequency of zero sequence at first locus equals $X_0 + X_1$ and that at the second locus is $X_0 + X_2$. Using $D = 0$, it follows that the product $(X_0 + X_1)(X_0 + X_2) = X_0$, the frequency of the sequence $\{0,0\}$. Although the linkage disequilibrium is zero when only recombination is present, it is usually nonzero when selection and/or mutation are also included.

We now discuss the situation when selection, mutation and recombination are present. We will consider the fitness scheme in which two fitness peaks are separated by a fitness valley and assume that $W_3 > W_0 = 1 > W_1 = W_2$. In a population initially localised at $\{0,0\}$, a mutation in $\{0,0\}$ to $\{0,1\}$ or $\{1,0\}$ is deleterious but the fitness loss can be compensated by acquiring another mutation resulting in the sequence $\{1,1\}$. In the absence of recombination and for small mutation rates, the population will eventually localise around the fittest $\{1,1\}$ sequence (see Subsec. 3.1). However due to nonlinear evolution equations, multiple steady states may result [8, 17]. As discussed below, there exists a critical recombination rate r_c below which the population can cross the intervening valley and reach the fittest peak at $\{1,1\}$. But above r_c, the population can remain trapped at the initial sequence with low fitness and thus the sexual reproduction can affect the adaptation process adversely. We now describe the population behavior for two schemes of mutation rates.

Multiple equilibria in steady state: If the mutation matrix is symmetric and given by (3), the evolution equations can be written as [38]

$$X_0(t+1) = X_0'(t) - r(1-2\mu)^2 \frac{D(t)}{\mathcal{W}^2(t)} \tag{44}$$

$$X_1(t+1) = X_1'(t) + r(1-2\mu)^2 \frac{D(t)}{\mathcal{W}^2(t)} \tag{45}$$

$$X_2(t+1) = X_2'(t) + r(1-2\mu)^2 \frac{D(t)}{\mathcal{W}^2(t)} \tag{46}$$

$$X_3(t+1) = X_3'(t) - r(1-2\mu)^2 \frac{D(t)}{\mathcal{W}^2(t)} \tag{47}$$

where $\mathcal{W}(t) = \sum_{k=0}^{3} W_k X_k(t)$ is the average fitness of the population, the linkage disequilibrium $D(t) = W_0 W_3 X_0(t) X_3(t) - W_1 W_2 X_1(t) X_2(t)$ and the primed fractions are given by the left hand side of (2):

$$X_i'(t) = \frac{\sum_{j=0}^{3} M(i \leftarrow j) W_j X_j(t)}{\mathcal{W}(t)}. \tag{48}$$

To arrive at the set of Eqs. (44)–(47), it has been assumed that recombination occurs after selection and mutation. Thus in the set of Eqs. (40)–(43), the frequency on the right hand side refers to $X_i'(t)$ upon using which (44)–(47) are obtained.

In the steady state, for the fitness landscape described above, the fractions X_i's can be expressed in terms of the fitness W_i's and the average fitness $\mathcal{W}$. On using the resulting expressions for X_i's in the equation for $\mathcal{W}$, a quartic equation for $\mathcal{W}$ is obtained. An analysis [38] of this equation shows that for $r < r_c$, the fittest sequence is always populated while for $r > r_c$, there are two stable solutions: either the population stays at the initial sequence $\{0, 0\}$ or moves to the fittest sequence $\{1, 1\}$.

Time to fixation: If the mutations are unidirectional with the probability to mutate from 0 to 1 being μ and zero for the back mutation, the whole population occupies the fittest sequence and the sequence $\{1, 1\}$ is said to be fixed. In such a case, it is interesting to study the dynamics of the population and more specifically, one can find the time T to fixation.

For the one-way mutation scheme in which first selection takes place followed by recombination and finally mutation, the time evolution occurs according to the following nonlinear equations [21]:

$$X_0(t+1) = \frac{(1-\mu)^2 W_0 X_0(t) - r(1-\mu)^2 D(t)}{\mathcal{W}(t)} \tag{49}$$

$$X_1(t+1) = \frac{\mu(1-\mu) W_0 X_0(t) + (1-\mu) W_1 X_1(t) + r(1-\mu)^2 D(t)}{\mathcal{W}(t)} \tag{50}$$

$$X_2(t+1) = \frac{\mu(1-\mu) W_0 X_0(t) + (1-\mu) W_2 X_2(t) + r(1-\mu)^2 D(t)}{\mathcal{W}(t)} \tag{51}$$

$$X_3(t+1) = \frac{\mu^2 W_0 X_0(t) + \mu(W_1 X_1(t) + W_2 X_2(t)) + W_3 X_3(t) - r(1-\mu)^2 D(t)}{\mathcal{W}(t)}. \tag{52}$$

Here $D(t) = (W_0W_3X_0(t)X_3(t) - W_1W_2X_1(t)X_2(t))/\mathcal{W}(t)$ is the linkage disequilibrium at time t and $\mathcal{W}(t) = \sum_{k=0}^{3} W_kX_k(t)$ is the average fitness of the population. The above equations can be written down in a manner analogous to the above cases. Since selection occurs before recombination, on replacing $X_i(t)$ by $W_iX_i(t)/\mathcal{W}(t)$ on the RHS of (40)–(43), the evolution equations with selection and recombination are obtained. Finally the unidirectional mutation scheme is implemented.

The equations for the corresponding unnormalised populations Z_k's defined by (4) can also be written. But due to the recombination term, the equations for Z_k's also remain nonlinear. An approximate method to handle these dynamical nonlinear equations can be developed by noting that at any instant, for small mutation rates, only one of the four populations dominate. Then the dynamics of population Z_k's can be divided in following three dynamical phases [21] : (i) $Z_0 \gg Z_1, Z_3$ (phase I) (ii) $Z_1 \gg Z_0, Z_3$ (phase II) and (iii) $Z_3 \gg Z_0, Z_1$ (phase III). Thus one can expand the equations for unnormalised populations in powers of $Z_1/Z_0, Z_3/Z_0$ in phase I, $Z_0/Z_1, Z_3/Z_1$ in phase II and similarly, $Z_0/Z_3, Z_1/Z_3$ in phase III. The time at which a phase ends is obtained by matching the solutions of the relevant populations in the two phases. The fixation time is then obtained by summing over these phase times.

As mentioned above, there exists a critical recombination fraction r_c beyond which a population initially located at $\{0, 0\}$ cannot cross the intermediate fitness valley and reach the double mutant fitness peak [8, 14]. The inset of Fig. 6 shows that the fixation time diverges as r approaches critical recombination probability $r_c = (w_4 - 1)/w_4$. A simple calculation using the method described above but ignoring the nonlinearities shows that the fixation time diverges as $1/(r_c - r)$. However a more careful analysis that takes the nonlinear terms into account shows that the fixation time is well approximated by [21]

$$T \approx \frac{1}{(r_c - r)W_3}\left[\ln\ln\left(\frac{W_3(1-W_1)(r_c-r)^2}{(1-W_1+r_cW_3)W_1^2K^2}\right) - \ln\left(\frac{2r_c^2W_1^2K}{(1-W_1)(r_c-r)^2}\right)\right] \tag{53}$$

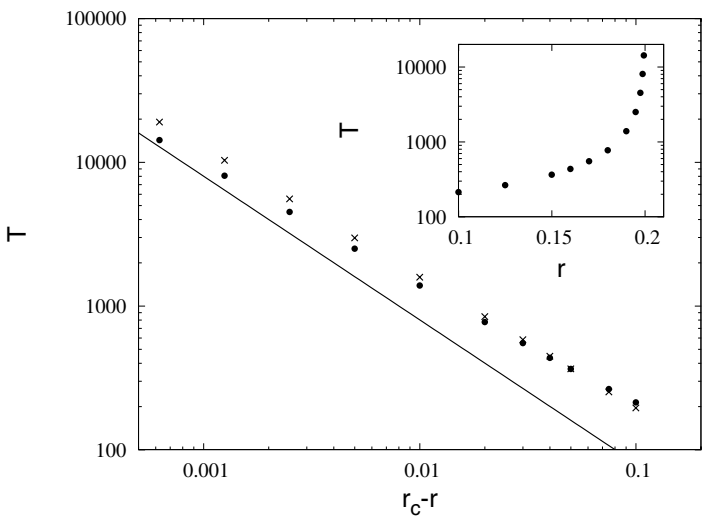

Fig. 6. Fixation time as a function of r obtained using exact iteration (•) and analytical result (×) given by (53). The solid line has a slope equal to −1.

where the constant $K \sim (r_c - r)^{-1}$. Thus the fixation time decays slower than $1/(r_c - r)$ due to the logarithmic corrections (see Fig. 6).

The population frequencies and fixation time can be analysed for other fitness schemes as well and a discussion can be found in [14, 21]. Although we have discussed the haploid case, the diploid problem has also been studied [7]. For studies on models that consider more than two loci, the reader may refer to [37, 19].

5. Summary

In this review, we have presented a brief (and incomplete) overview of evolutionary processes and models in deterministically evolving populations. As we have discussed, these systems are inherently nonlinear and difficult to analyse analytically. The nonlinearity of these systems that makes them so difficult to handle, is also responsible for the complex behaviour of their solutions. The existence of multiple steady states and dynamic phase transitions are some of the interesting features displayed by these models.

While these theoretical models of evolutionary biology have garnered interest amongst physicists and mathematicians, they have also been successful in predicting biological properties and explaining the experimental results quantitatively. It is hoped that the work integrated from various disciplines will take us closer to an understanding of the complex and continuous process of the evolution of life.

References

[1] E. Akin, *The Geometry of Population Genetics* (Springer, New York, 1979).
[2] E. Baake and H. Wagner, Mutation-selection models solved exactly with methods of statistical mechanics, *Genet. Res. Camb.* **78** (2001) 93–117.
[3] E. Baake and T. Wiehe, Bifurcations in haploid and diploid sequence space models, *J. Math. Biol.* **35** (1997) 321–343.
[4] R. Bellman, *Introduction to Matrix Analysis* (Society for Industrial and Applied Mathematics, Philadelphia, 1997).
[5] L. Boe, M. Danielsen, S. Knudsen, J. B. Petersen, J. Maymann and P. R. Jensen, The frequency of mutators in populations of Escherichia coli, *Mut. Res.* **448** (2000) 47–55.
[6] Y. Brumer and E. I. Shakhnovich, Host-parasite coevolution and optimal mutation rates for semiconservative quasispecies, *Phys. Rev. E* **69** (2004) 061909.
[7] R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation* (Wiley, Chichester, 2000).
[8] J. F. Crow and M. Kimura, Evolution in sexual and asexual populations, *Am. Nat.* **99** (1965) 439–450.
[9] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper and Row, New York, 1970).
[10] J. W. Drake, B. Charlesworth, D. Charlesworth and J. F. Crow, Rates of spontaneous mutation, *Genetics* **148** (1998) 1667–1686.
[11] R. Durrett, *Probability Models for DNA Sequence Evolution* (Springer, New York, 2002).
[12] M. Eigen, Selforganization of matter and evolution of biological macromolecules, *Naturwissenchaften* **58** (1971) 465–523.
[13] M. Eigen and P. Schuster, The hypercycle, *Naturwissenchaften* **64** (1977) 541–565.
[14] I. Eshel and M. W. Feldman, On the evolutionary effect of recombination, *Theo. Pop. Biol.* **1** (1970) 88–100.
[15] W. J. Ewens, *Mathematical Population Genetics* (Springer, Berlin, 1979).

[16] D. L. Hartl and E. W. Jones, *Genetics: Analysis of Genes and Genomes* (Jones and Barlett Publishers, 2009).
[17] P. G. Higgs, Compensatory neutral mutations and the evolution of RNA, *Genetica*, **102–103** (1998) 91–101.
[18] N. Sh. Izmailian, Vl. V. Papoyan, V. B. Priezzhev and C. K. Hu, Self-organizing behavior in a lattice model for co-evolution of virus and immune systems, *Phys. Rev. E* **75** (2007) 041104.
[19] M. N. Jacobi and M. Nordahl, Quasispecies and recombination, *Theo. Pop. Biol.* **70** (2006) 479–485.
[20] K. Jain, Loss of least-loaded class in asexual populations due to drift and epistasis, *Genetics* **179** (2008) 2125.
[21] K. Jain, Time to fixation in the presence of recombination, *Theo. Pop. Biol.* **77** (2010) 23.
[22] K. Jain and J. Krug, Evolutionary trajectories in rugged fitness landscapes, *J. Stat. Mech. Theory Exp.* (2005) P04008.
[23] K. Jain and J. Krug, Adaptation in simple and complex fitness landscapes, in *Structural Approaches to Sequence Evolution: Molecules, Networks and Populations*, eds. U. Bastolla, M. Porto, H. E. Roman and M. Vendruscolo (Springer, Berlin, 2007), pp. 299–340.
[24] K. Jain and J. Krug, Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes, *Genetics* **175** (2007) 1275.
[25] B. Jones, R. Enns and S. Rangnekar, On the theory of selection in coupled macromolecular systems, *Bull. Math. Biol.* **38** (1976) 15.
[26] C. Kamp and S. Bornholdt, Coevolution of quasispecies: B-cell mutation rates maximize viral error catastrophes, *Phys. Rev. Lett.* **88**(6) (2002) 068104.
[27] M. Kimura, On the evolutionary adjustment of spontaneous mutation rates, *Genet. Res.* **9** (1967) 23–34.
[28] N. L. Komarova, P. Niyogi and M. A. Nowak, The evolutionary dynamics of grammar acquisition, *J. Theor. Biol.* **209** (2001) 43–59.
[29] R. D. Kouyos, O. K. Silander and S. Bonhoeffer, Epistasis between deleterious mutations and the evolution of recombination, *Trends Ecol. Evol.* **22** (2007) 308–315.
[30] J. E. LeClerc, B. Li, W. L. Payne and T. A. Cebula, High mutation frequencies among Escherichia coli and Salmonella pathogens, *Science* **274** (1996) 1208–1211.
[31] A. Nagar and K. Jain, Exact phase diagram of quasispecies model with mutation rate modifier, *Phys. Rev. Lett.* **102** (2009) 038101.
[32] M. Nilsson and N. Snoad, Error thresholds for quasispecies on dynamic fitness landscapes, *Phys. Rev. Lett.* **84** (2000) 191–194.
[33] M. Nilsson and N. Snoad, Quasispecies evolution on a fitness landscape with a fluctuating peak, *Phys. Rev. E* **65** (2002) 031901.
[34] M. A. Nowak and P. Schuster, Error thresholds of replication in finite populations: mutation frequencies and the onset of Muller's ratchet, *J. Theor. Biol.* **137** (1989) 375–395.
[35] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, 2006).
[36] M. A. Nowak, N. L. Komarova and P. Niyogi, Evolution of universal grammar, *Science* **291** (2001) 114–118.
[37] S. P. Otto and M. W. Feldman, Deleterious mutations, variable epistatic interactions, and the evolution of recombination, *Theo. Pop. Biol.* **51** (1997) 134–147.
[38] S.-C. Park and J. Krug, Bistability in two-locus models with selection, mutation, and recombination, *J. Math. Biol.* **62** (2011) 763–788.
[39] A. S. Perelson and C. A. Macken, Protein evolution on partially correlated landscapes, *Proc. Natl. Acad. Sci. USA* **92** (1995) 9657–9661.
[40] J. Quer, R. Huerta, I. S. Novella, L. Tsimring, E. Domingo and J. J. Holland, Reproducible nonlinear population dynamics and critical points during replicative competitions of RNA virus quasispecies, *J. Mol. Biol.* **264** (1996) 465–471.

[41] D. B. Saakian and C. K. Hu, Solvable biological evolution models with a parallel mutation-selection scheme, *Phys. Rev. E* **69** (2004) 046121.
[42] J. Sardanyes and R. V. Sole, Chaotic stability in spatially-resolved host-parasite replicators: The red queen on a lattice, *Internat. J. Bifurcation Chaos* **17** (2007) 589–606.
[43] S. Seetharaman and K. Jain, Evolutionary dynamics on strongly correlated fitness landscapes, *Phys. Rev. E* **82** (2010) 031109.
[44] J. Maynard Smith, Natural selection and concept of a protein space, *Nature* **225** (1970) 031109.
[45] P. D. Sniegowski, P. J. Gerrish and R. E. Lenski, Evolution of high mutation rates in experimental populations of Escherichia coli, *Nature* **387** (1997) 703–705.
[46] A. Sorace, C. Heycock and R. Shillcock, *Language Acquisition: Knowledge Representation and Processing* (North Holland, 1999).
[47] A. H. Sturtevant, Essays on evolution. I. On the effects of selection on the mutation rate, *Q. Rev. Biol.* **12** (1937) 464–476.
[48] E. Tannenbaum, E. Deeds and E. I. Shakhnovich, Equilibrium distribution of mutators in the single fitness peak model, *Phys. Rev. Lett.* **91** (2003) 138105.
[49] C. J. Thompson and J. L. McBride, On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules, *Math. Biosci.* **21** (1974) 127.
[50] J. A. G. M. de Visser, S.-C. Park and J. Krug, Exploring the effect of sex on an empirical fitness landscape, *Am. Nat.* **174** (2009) S15–S30.
[51] T. Wiehe, Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models, *Genet. Res. Camb.* **69** (1997) 127–136.
[52] T. Wiehe, E. Baake and P. Schuster, Error propagation in reproduction of diploid organisms: A case study on single peaked landscapes, *J. Theo. Biol.* **177** (1995) 1–15.
[53] C. O. Wilke, C. Ronnewinkel and T. Martinetz, Dynamic fitness landscapes in molecular evolution, *Phys. Rep.* **349** (2001) 395–446.
[54] G. Woodcock and P. G. Higgs, Population evolution on a multiplicative single-peak fitness landscape, *J. Theor. Biol.* **179** (1996) 61–73.