

РЕШЕНИЕ МОДЕЛИ ЭЙГЕНА С МОДИФИКАЦИЕЙ СМЕРЛАКА

*В. Суворов*¹

Auriga Inc., Вобурн, Массачусетс, США

Модель Эйгена описывает эволюцию вируса с помощью системы нелинейных дифференциальных уравнений (нелинейное основное уравнение). Эта модель описывает эволюцию вируса, происхождение жизни и некоторые другие явления междисциплинарных исследований. М. Смерлак предложил метод того, как эти нелинейные модели можно связать с марковскими (линейными) моделями. Здесь мы формулируем версию модели Эйгена по Смерлаку и точно решаем динамику в пределе большого генома.

The Eigen (quasispecies) model describes virus evolution via a system of nonlinear differential equations (nonlinear master equation). The model describes virus evolution, the origin of life, and some other phenomena of interdisciplinary research. Recently, M. Smerlak suggested a method by which these nonlinear models can be connected to Markov (linear) models. Here, we formulate Smerlak's version of the Eigen model and solve the dynamics exactly at the large genome limit.

PACS: 87.10.+e

ВВЕДЕНИЕ

В последние десятилетия была обнаружена связь между моделями эволюции и статистической физикой, и такие исследования стали одним из важных направлений современной статистической физики. При рассмотрении моделей, сформулированных эволюционными биологами, были применены почти все методы теоретической физики: квантовая статистическая физика, теория спинового стекла, квантовая теория поля, метод уравнений Гамильтона–Якоби (см. обзор [1]).

Основой этих связей является существование в эволюционных явлениях эквивалента энергии: приспособленности, большого числа степеней свободы и почти линейного характера дифференциальных уравнений. Центральную роль в применении статистической физики к эволюции играет модель Эйгена. Модель Эйгена (квазивидовая модель) была предложена для описания происхождения жизни [3, 4], а позже применена к эволюции вирусов [5] как одна из ключевых моделей в этой области. Модель нашла многочисленные применения, от онкобиологии до искусственной жизни. Фактически, это ключевая модель современной междисциплинарной науки о жизни.

¹E-mail: hellofileandr@gmail.com

Модель хорошо связана со статистической физикой, что впервые было осознано в [6, 7] для случая модели с дискретным временем, где различные генотипы отображаются в конфигурации спинов. В модели Эйгена отбор и мутация связаны. Связанную модель, где мутация и отбор разделены, предложили Кроу и Кимура [8]. В [9] эта модель была сопоставлена с квантовой механикой в мнимом времени, позже средняя фитнес-функция модели Кроу–Кимуры была выведена с использованием алгебраических методов в [10]. В работах [11, 12] средняя фитнес-функция выведена с использованием методов квантовой статистической механики. Если фитнес-функция является функцией среднего поля для случая, когда геном представляет собой цепь из N генов с λ аллелями, функция зависит от доли различных аллелей.

В работах [11, 12] отображение на квантовую статистическую физику позволяет решить динамику для случая одиночной вершины приспособленности, что позже было применено для получения динамики со случайным ландшафтом приспособленности [22], а также динамических ландшафтов приспособленности с одним пиком [14–16].

Устойчивое состояние приспособленности в модели Кроу–Кимуры можно вывести, точно сопоставив проблему эволюции с энергией основного состояния набора одномерных моделей Изинга. Таким образом, в некотором смысле эволюция с отбором и мутацией является квантовомеханической задачей, как впервые было установлено в [9], точнее, задачей квантовой статистической физики, которая может быть отображена на модель квантовой теории поля [17, 18]. С использованием такого отображения была выведена средняя приспособленность модели Эйгена [19].

После установления связи со статистической физикой становится интересным исследовать различные фазы и точки перехода. Уже в первой статье о модели квази-видов Эйген рассматривал фазовый переход в случае увеличения скорости мутаций, анализируя специальный ландшафт приспособленности: один геном имеет высокую приспособленность (скорость репликации), в то время как все остальные — альтернативную, более низкую приспособленность.

При низкой вероятности мутаций подавляющее большинство популяции имеет генотип, близкий к тому, который обладает высокой приспособленностью. При более высоких вероятностях мутаций популяция рассеивается в геномном пространстве, и генотип с высокой приспособленностью составляет лишь незначительную долю.

Позднее фазовая структура была исследована в многомерных моделях, где существует несколько эталонных генотипов, а приспособленность данного генотипа зависит от расстояния Хэмминга до всех эталонных генотипов (расстояние Хэмминга между двумя генотипами — это количество мутаций, необходимых для получения второго фенотипа из первого).

В работах [20, 21] метод Гамильтона–Якоби был предложен для решения как модели Кроу–Кимуры, так и модели Эйгена, и было вычислено распределение в устойчивом состоянии.

Метод Гамильтона–Якоби, возможно, является самым мощным методом для решения задач эволюционной динамики. Он был применен для решения эволюционных моделей с мутациями, рекомбинацией и миграцией. Этот метод был использован для обнаружения нового фазового перехода в эволюционной модели, когда средняя приспособленность остается неизменной, а распределение в устойчивом состоянии резко изменяется [22]. Позднее была выведена динамика обеих эволюционных моде-

лей [23], и в динамике были обнаружены две различные фазы. Средняя приспособленность и энтропия не являются монотонными функциями времени [23].

В работе [24] модель Кроу–Кимуры была введена в многомерное пространство приспособленности (приспособленность является функцией d расстояний от данной последовательности генома до d эталонных последовательностей), и было введено уравнение Гамильтона–Якоби.

В работах [25, 26] была предложена модификация эволюционной модели, которая позволяет рассматривать только линейное уравнение мастера, утверждая, что энтропия является монотонной функцией времени. В работе [27] была исследована модель Кроу–Кимуры с модификацией Смерлака. Модель Эйгена имеет более тесную связь с биологическими данными, чем модель Кроу–Кимуры. Здесь мы исследуем модель Эйгена с модификацией Смерлака и точно решим вопрос динамики.

В разд. 1 мы рассматриваем известные результаты о модели Эйгена, выводя уравнение Гамильтона–Якоби и левый собственный вектор матрицы, используемой в определении модели с симметричным ландшафтом приспособленности. Затем выводим правый собственный вектор этой матрицы. Последний необходим для формулировки версии модели Эйгена по Смерлаку. Выводим уравнение Гамильтона–Якоби для этой модели, точно решаем динамику и проверяем, что энтропия монотонно возрастает со временем.

1. СТАНДАРТНАЯ МОДЕЛЬ ЭЙГЕНА

1.1. Определение модели на общем ландшафте приспособленности. Геном представляет собой цепочку из N генов, каждый из которых имеет два аллеля. Обозначим типы аллелей как ± 1 . Существует $M = 2^N$ различных генотипов. Обозначим через p_i долю i -го генотипа в популяции.

Рассмотрим обычную модель Эйгена. Запишем:

$$\frac{dp_i}{dt} = \sum_{j=1}^{2^N} Q_{ij} r_j p_j - p_i R, \quad R = \sum_{j=1}^{2^N} r_j p_j, \quad (1)$$

где Q_{ij} — это вероятности мутаций из состояния i в состояние j .

$$Q_{ij} = q^{N-d_{ij}} (1-q)^{d_{ij}}, \quad (2)$$

где q — вероятность безошибочного копирования аллеля, а d_{ij} — расстояние Хэмминга между i -м и j -м генотипами, т. е. количество различных аллелей между двумя генотипами. Последний член описывает равномерное разбавление. Благодаря последнему члену выполняется условие баланса вероятностей $\left(\sum_{i=0}^{M-1} p_i = 1 \right)$. R — это средняя приспособленность популяции. Если убрать последний член, то размер популяции будет расти экспоненциально $\sim \exp(Rt)$. r_i — это приспособленность генотипа. Различные выборы приспособленностей r_i описывают разные «ландшафты» приспособленности.

Самый простой выбор ландшафта — это ландшафт с единственной вершиной приспособленности, где 0-й генотип имеет высокую приспособленность A , в то время как

остальные последовательности имеют низкую приспособленность 1. Эйген обнаружил фазовый переход в модели.

При низких скоростях мутаций, когда

$$q^N A > 1, \quad (3)$$

подавляющее большинство популяции сосредоточено вокруг нулевого генотипа. Когда условие (2) не выполняется, популяция рассеивается в геномном пространстве, и $p_0 \sim 1/M$.

Мы можем рассмотреть линейную версию уравнения (1),

$$\frac{d\hat{p}_i}{dt} = \sum_{j=1}^{2^N} Q_{ij} r_j \hat{p}_j, \quad (4)$$

и пересчитать масштаб

$$p_i = \frac{\hat{p}_i}{\sum_j \hat{p}_j}. \quad (5)$$

1.2. Модель на симметричном ландшафте приспособленности. Соберем геномы с одинаковым числом мутаций l в хэмминг-классы, $0 \leq l \leq N$, и обозначим как P_l всю популяцию. Существует $N_l = N!/(l!(N-l)!)$ различных геномов в одном хэмминговом классе, так что $P_l = N_l p_i$, где p_i — из l -го хэммингового класса. Тогда для P_l была выведена следующая система уравнений [4]:

$$\frac{dP_l}{dt} = \sum_{k=0}^N M_{kl} P_k - R P_l, \quad M_{kl} = \hat{Q}_{kl} r_k, \quad R = \sum_{j=0}^N r_j P_j, \quad (6)$$

где r_k — это приспособленность, и существует следующее выражение для новой матрицы мутаций для вероятностей класса \hat{Q}_{ik} [4]:

$$\hat{Q}_{kl} = q^N \sum_{l=0}^m \left(\frac{1-q}{q} \right)^{2j-|k-l|} \binom{N-k}{l + \frac{1}{2}(|k-l| - k + l)} \times \binom{k}{j + \frac{1}{2}(|k-l| + k - l)} \quad (7)$$

и

$$m = \frac{1}{2} [\min(k+l, 2N-k-l) - |k-l|].$$

Матрица M_{kl} в уравнении (6) не является симметричной. Нас интересует вывод как левых, так и правых собственных векторов, соответствующих максимальному собственному значению. У нас есть следующие уравнения для левых и правых собственных векторов матрицы R :

$$\sum_{k=0}^N M_{kl} P_k - R_0 P_l = 0 \quad (8)$$

и

$$\sum_{k=0}^N M_{kl}q_k - R_0q_l = 0, \tag{9}$$

где R_0 — это приспособленность в устойчивом состоянии. Уравнение (8) — версия уравнения (6) для устойчивого состояния, уравнение (9) необходимо для формулировки версии модели Эйгена по Смерлаку.

1.3. Вывод уравнения Гамильтона–Якоби. Выведем уравнение Гамильтона–Якоби, затем, используя его, выведем левый собственный вектор. В следующем разделе выведем выражение для правого собственного вектора.

Мы рассматриваем масштабирование $q = 1 - \gamma/N$ и получаем для $d_{ij} \ll N$

$$Q_{ij} = e^{-\gamma} \left(\frac{\gamma}{N}\right)^{d_{ij}}. \tag{10}$$

Рассмотрим мутации из n -го хэммингового класса, предполагая n_1 мутаций среди n чисел – спинов и n_2 мутаций среди $N - n$ чисел + спинов. Их количество равно

$$n_{\text{mut}} = \frac{n!}{n_1!(n - n_1)!} \frac{(N - n)!}{n_2!(N - n - n_2)!}, \tag{11}$$

так что $d_{ij} = n_1 + n_2$.

Когда $n_1 \ll n, n_2 \ll (N - n)$, имеем

$$n_{\text{mut}} = \frac{n^{n_1}}{n_1!(n - n_1)!} \frac{(N - n)^{n_2}}{n_2!(N - n - n_2)!} = \left(N \frac{1 - x}{2}\right)^{n_1} \left(N \frac{1 + x}{2}\right)^{n_2} \frac{1}{n_1!n_2!}, \tag{12}$$

где $x = 1 - 2n/L$, и мы рассмотрели $n_1 \ll n, n_2 \ll n$.

Сделаем предположение для вероятности одной последовательности

$$p_n = e^{Nu(x,t)}, \tag{13}$$

где $x = 1 - 2n/N$. Тогда для вероятности всего класса имеем, используя формулу

Стирлинга для $\frac{N!}{N \frac{1+x}{2}! N \frac{1-x}{2}!}$,

$$P_n = e^{Nu(x,t) + NG \equiv NU}, \tag{14}$$

$$G = -\frac{1-x}{2} \ln \frac{1-x}{2} - \frac{1+x}{2} \ln \frac{1+x}{2}.$$

Рассмотрим, как изменяется одна последовательность из n -го класса после мутаций. После изменения хэммингового класса n в результате мутаций в $n + l$ класс, где $l = n_1 - n_2$, мы имеем

$$P_{n+l} = \exp \left[Nu \left(x + \frac{2l}{N}, t \right) \right] \approx P_n e^{-2lu'}. \tag{15}$$

Суммируя по мутациям и изменяя масштаб

$$t \rightarrow tN,$$

получаем из уравнения (4) следующее уравнение Гамильтона–Якоби [20]:

$$\begin{aligned} \frac{du}{dt} &= \sum_{n_1, n_2} \frac{r_{n+l}}{n_1! n_2!} e^{-\gamma(\gamma)^{n_1+n_2}} e^{2(n_1-n_2)u'} \left(\frac{1-x}{2}\right)^{n_1} \left(\frac{1+x}{2}\right)^{n_2} = \\ &= f(x) \exp \left[-\gamma + \gamma \frac{1-x}{2} e^{2u'} + \gamma \frac{1+x}{2} e^{-2u'} \right]. \end{aligned} \quad (16)$$

Мы заменили $r_{n+l} \equiv r_{n+n_1-n_2}$ на $r_n \equiv f(x)$ с точностью до $O(1/N)$.

Тогда по выражению (14) уравнение (16) преобразуется в

$$\frac{dU}{dt} = f(x) \exp \left[-\gamma + \gamma \frac{1-x}{2} e^{-2U'} + \gamma \frac{1+x}{2} e^{2U'} \right]. \quad (17)$$

Возьмем следующее представление для асимптотического решения

$$U(x, t) = RT + U_0(x) \quad (18)$$

и получим

$$R = f(x) \exp \left[-\gamma + \gamma \frac{1+x}{2} e^{2U'_0} + \gamma \frac{1-x}{2} e^{-2U'_0} \right]. \quad (19)$$

Ищем решения для действительных значений R , имеем условие

$$R \geq f(x) \exp [-\gamma + \gamma \sqrt{1-x^2}]. \quad (20)$$

Последнее условие должно быть выполнено для всех x , так что

$$R \geq \max [f(x) \exp (-\gamma + \gamma \sqrt{1-x^2})] = f(x_0) e^{-\gamma + \gamma \sqrt{1-(x_0)^2}}. \quad (21)$$

Приняв знак равенства в последнем уравнении, мы получаем среднюю приспособленность устойчивого распределения. Более того, уравнение (19) дает решение для левого собственного вектора матрицы M_{kl} из уравнения (6).

2. РЕШЕНИЕ ВЕРСИИ МОДЕЛИ ЭЙГЕНА ПО СМЕРЛАКУ

2.1. Правый собственный вектор модели Эйгена. Рассмотрим анзац для правого собственного вектора, уравнение (9),

$$P_l = w_l = \exp (Lv_0(x)).$$

Вместо того чтобы рассматривать непосредственно уравнение (9), проще начать с уравнения (4) с

$$\sum_{j=1}^{2^N} Q_{ij} r_j \hat{p}_i. \quad (22)$$

Тогда все аналогично уравнению (16), только $x \rightarrow -x$. Так что вместо уравнения (19) мы получаем

$$\log\left(\frac{f(x)}{R}\right) = \gamma \left[\frac{1-x}{2} e^{2v'_0} + \frac{1+x}{2} e^{-2v'_0} - 1 \right]. \tag{23}$$

Для v' имеем

$$2v'(x) = \ln \frac{(F(x) + 1 \pm \sqrt{(F(x) + 1)^2 - 1 + x^2})}{1 - x}, \tag{24}$$

$$F(x) = \frac{1}{\gamma} \log\left(\frac{R}{f(x)}\right).$$

Мы можем вычислить правый собственный вектор, интегрируя уравнение (23).

Для v'' получаем следующее уравнение:

$$v''(x) = \frac{(e^{-2v'_0} - e^{2v'_0})/2 - F'(x)}{(1+x)e^{2v'_0(x)} + (1-x)e^{-2v'_0(x)}}. \tag{25}$$

2.2. Вывод уравнения Гамильтона–Якоби для модели Смерлака. Рассмотрим новую динамику [25, 26] (рис. 1):

$$\frac{dP_l}{dt} = \sum_n P_n M_{nl} q_n^{-1} q_l - R_0 P_l, \tag{26}$$

где теперь R_0 — это приспособленность в устойчивом состоянии. Суммируя по l , мы получаем условие баланса вероятностей:

$$\sum_l \frac{dP_l}{dt} = P_l R_0 - R_0 P_l = 0. \tag{27}$$

Обозначим $P_l = \exp[LU(x, t)]$. Суммируя различные мутационные термины для динамики, получаем выражение по уравнению (17), только с заменой

$$u' \rightarrow u' - v'.$$

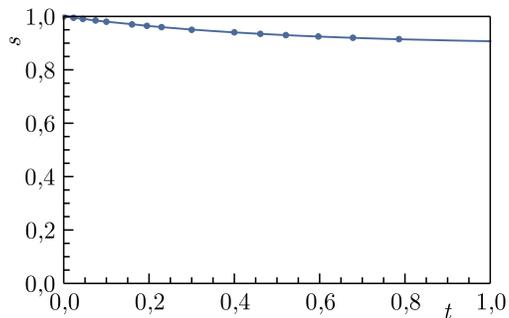


Рис. 1. Динамика модифицированной модели Эйгена при $\gamma = 1$, $f(x) = \exp[2(x - 1)]$, $N = 200$. Сплошная линия — численные результаты, темные точки — наши теоретические результаты по уравнению (31)

Тогда имеем

$$\frac{du}{dt} = -H(x, u'), \quad (28)$$

$$H(x, p) = -f(x) \exp[-\gamma + \gamma(1+x)e^{2p-2v'}/2 + \gamma(1-x)/2e^{-2p+2v'}]$$

или, используя обозначение $q = \partial u/\partial t$,

$$\log\left(\frac{q+R}{f(x)}\right) \equiv F(q, x) = +\frac{1+x}{2}e^{2p+2v'} + \frac{1-x}{2}e^{-2p-2v'} - 1. \quad (29)$$

Уравнение (29) является уравнением Гамильтона–Якоби и может быть решено методами характеристик [28, 29].

Уравнение Гамильтона для x дает

$$\frac{dx}{dt} = 2(q+R)\sqrt{(F+1)^2 - 1 + x^2}, \quad (30)$$

где постоянная q является параметром характеристики. Подставляя $q = 0$ для максимальной точки, интегрируя уравнение (26), получаем следующее уравнение для $x(t)$:

$$t = \int_{x_0}^x dy \frac{1}{2(q+R)\sqrt{(F+1)^2 - 1 + x^2}}. \quad (31)$$

Рассматривая предел $t \rightarrow \infty$ в уравнении (31), получаем для предельной точки $x(t) \rightarrow x_c$.

2.3. Версия модели Эйгена по Смерлаку для дискретного времени. Вместо уравнения (26) теперь рассмотрим распределения в разные моменты времени, $P_i(t)$. Тогда имеем

$$P_l(t+1) = \frac{\sum P_n(t)M_{nl}q_n^{-1}q_l}{R}. \quad (32)$$

Легко проверить условия баланса вероятностей. Обе модели по уравнениям (26), (32) являются марковскими моделями, поэтому в них наблюдается монотонное поведение энтропии. Напротив, обычная энтропия Шеннона модели Эйгена аналогична логарифму дисперсии распределения $\sim \ln(V) + \text{const}$, и рис. 2 иллюстрирует, что последняя не является монотонной функцией времени.

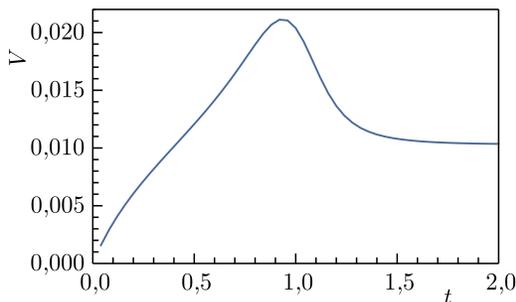


Рис. 2. Динамика дисперсии для модели Эйгена при $\gamma = 1$, $f(x) = \exp[x^2]$, $N = 200$. Мы проверяем, что дисперсия не является монотонной функцией времени

ЗАКЛЮЧЕНИЕ

Мы исследовали некоторые аспекты модели Эйгена, ключевой модели современных междисциплинарных исследований. Хотя к модели были применены различные методы статистической физики, для динамики наиболее успешным является метод уравнения Гамильтона–Якоби, который мы применили в данной статье. Модель с длиной генома N представляет собой цепочку нелинейных дифференциальных уравнений, которые посредством нелинейного алгебраического преобразования могут быть отображены в систему линейных дифференциальных уравнений с несимметричной матрицей. Для модели мы наблюдаем немонотонное поведение энтропии.

Мы вывели как правые, так и левые собственные векторы матрицы. Далее рассмотрели модификацию модели по Смерлаку. В то время как модель Эйгена изначально является нелинейным уравнением мастера, после преобразования по Смерлаку она становится линейной. Здесь мы точно решили предельный случай большого генома для модели Эйгена после модификации по Смерлаку, используя метод характеристик в подходе Гамильтона–Якоби. Мы рассчитали динамику для максимума распределения. Численные результаты хорошо поддерживают наши аналитические результаты. Преимущество модели заключается в том, что она является марковской моделью, поэтому мы можем определить энтропию, которая увеличивается со временем.

Я благодарю Д. Б. Саакяна за обсуждения.

СПИСОК ЛИТЕРАТУРЫ

1. *Saakian D. B., Hu C.-K.* // *J. Theor. Biol.* 2009. V. 259. P. 317–324.
2. *Barton N. H., Coe J. B.* *The Mathematical Models of Quasispecies Theory and Exact Results for the Dynamics* // *Quasispecies: From the Theory to Experimental Systems.* Springer, 2015.
3. *Eigen M.* // *Naturwissenschaften.* 1971. V. 58. P. 465.
4. *Eigen M., McCaskill J. S., Schuster P.* *The Molecular Quasispecies* // *Adv. Chem. Phys.* 1989. V. 75. P. 149.
5. *Eigen M.* // *Proc. Nat. Acad. Sci. USA.* 2002. V. 99. P. 133.
6. *Leuthausser I.* // *J. Stat. Phys.* 1987. V. 48. P. 343.
7. *Tarazona P.* // *Phys. Rev. A.* 1992. V. 45. P. 6038.
8. *Crow J. F., Kimura M.* *An Introduction to Population Genetics Theory.* New York: Harper & Row, 1970.
9. *Baake E., Baake M., Wagner H.* // *Phys. Rev. Lett.* 1997. V. 78. P. 559.
10. *Hermisson J., Redner O., Wagner H., Baake E.* // *Theor. Popul. Biol.* 2002. V. 62. P. 9.
11. *Saakian D. B., Hu C.-K.* *Eigen Model as a Quantum Spin Chain: Exact Dynamics* // *Phys. Rev. E.* 2004. V. 69. P. 021913.
12. *Saakian D. B., Hu C.-K.* // *Ibid.* P. 046121.
13. *Saakian D. B., Fontanari J. F.* *Evolutionary Dynamics on Rugged Fitness Landscapes: Exact Dynamics and Information Theoretical Aspects* // *Phys. Rev. E.* 2009. V. 80. P. 041903.
14. *Ancliff M., Park J.-M.* // *Phys. Rev. E.* 2009. V. 80. P. 061910.
15. *Ancliff M., Park J.-M.* // *Phys. Rev. E.* 2010. V. 82. P. 021904.
16. *Neves A. G. M.* // *Ibid.* P. 031915.
17. *Saakian D. B., Hu C.-K., Khachatryan H.* // *Phys. Rev. E.* 2004. V. 70. P. 041908.
18. *Saakian D. B., Munoz E., Hu C.-K., Deem M. W.* *Quasispecies Theory for Multiple-Peak Fitness Landscapes* // *Phys. Rev. E.* 2006. V. 73. P. 041913.

19. *Saakian D. B., Hu C.-K.* // Proc. Nat. Acad. Sci. USA. 2006. V. 103. P. 4935.
20. *Saakian D. B.* A New Method for the Solution of Models of Biological Evolution: Derivation of Exact Steady-State Distributions // J. Stat. Phys. 2007. V. 128. P. 781.
21. *Sato K., Kaneko K.* Evolution Equation of Phenotype Distribution: General Formulation and Application to Error Catastrophe // Phys. Rev. E. 2007. V. 75. P. 061909.
22. *Saakian D. B., Biebricher Ch. K., Hu C.-K.* Phase Diagram for the Eigen Quasispecies Theory with the Truncated Fitness Landscape // Phys. Rev. E. 2009. V. 79. P. 041905.
23. *Saakian D. B., Rozanova O., Akmetzhanov A.* // Phys. Rev. E. 2008. V. 78. P. 041908.
24. *Saakian D. B., Kirakosyan Z., Hu C.-K.* // Phys. Rev. E. 2012. V. 86. P. 031920.
25. *Smerlak M.* arXiv:1912.05890.
26. *Smerlak M.* // J. Theor. Biol. 2022. V. 522. P. 110699.
27. *Poghosyan R., Saakian D. B.* // Physica A. 2023. V. 617. P. 128652.
28. *Melikyan A.* Generalized Characteristics of First Order PDEs. Boston: Birkhauser, 1998.
29. *Evans L. C.* Partial Differential Equations. Providence: AMS, 2002.

Получено 19 июня 2024 г.