

Ա.Ի.Ալիխանյանի անվան ԱԶԳԱՅԻՆ ԳԻՏԱԿԱՆ ԿԱՌՈՐԱՏՈՐԻԱ
Երևանի ֆիզիկայի ինստիտուտ

Կիրակոսյան Զատա Ալեքսանդրի

Էվոլյուցիայի մոդելների լուծումը վիճակագրական ֆիզիկայի
մեթոդներով

Ա.04.02 - "Տեսական ֆիզիկա" մասնագիտությամբ
ֆիզիկամաթեմատիկական գիտությունների թեկնածուի գիտական
ասպիրանտի հայցման արեւնախոսության

ՄԵՂՄԱԳԻՐ

Երևան 2013

НАЦИОНАЛЬНАЯ НАУЧНАЯ ЛАБОРАТОРИЯ им. А.И.Алиханяна
(Ереванский физический институт)

Киракосян Зара Александровна

Решение моделей эволюции методами статистической
физики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени кандидата
физико-математических наук по специальности
01.04.02 - "Теоретическая физика"

Ереван 2013

Արեւմախոսության թեման հասարակված է Ա.Ի.Ալիխանյանի անվան Ազգային Գիտական Լաբորատորիայի գիտական խորհրդում:

- Գիտական ղեկավար - Ֆիզմաթ. գիտ. դոկտոր
Ղ. Բ. Սահակյան (ԱԱԳԼ)
- Պաշտոնական ընդդիմախոսներ - Ֆիզմաթ. գիտ. դոկտոր, պրոֆեսոր
Ն. Ս. Անանիկյան (ԱԱԳԼ)
- Ֆիզմաթ. գիտ. դոկտոր
Ա.Ս.Գևորգյան(ԳԱԱ ԻԱՊԻ)
- Առաջարար կազմակերպություն - Երևանի պետական համալսարան

Պաշտպանությունը կկայանա 2013թ. մայիսի 21-ին ժամը 14.00-ին, Ա.Ի.Ալիխանյանի անվան ազգային գիտական լաբորատորիայում գործող ԲՆՀ-ի "Միջուկի և փարրական մասնիկների ֆիզիկա" 024 մասնագիտական խորհրդի նիստում (0036, Երևան, Ալիխանյան Եղբայրների փող. 2):

Արեւմախոսությանը կարելի է ծանոթանալ Ա.Ի.Ալիխանյանի անվան Ազգային Գիտական Լաբորատորիայի գրադարանում:

Սեղմագիրն առաքված է ապրիլի 19-ին 2013թ.

Մասնագիտական խորհրդի գիտ. քարտուղար,
Ֆիզմաթ. գիտ. դոկտոր,

 Դ. Ն. Շիրախանյան

Тема диссертации утверждена ученым советом Национальной Научной Лаборатории им. А.И. Алиханяна.

- Научный руководитель - доктор физ.-мат. наук.
Д.Б.Саакян(ННЛ им.А.И.Алиханяна)
- Официальные оппоненты - доктор физ.-мат. наук, профессор
Н.С.Ананикян(ННЛ им.А.И.Алиханяна)
- доктор физ.-мат. наук
А.С.Геворкян(НАН ИИПА)
- Ведущая организация - Ереванский государственный университет

Защита диссертации состоится 21 мая 2013 года в 14.00 часов на заседании специализированного совета ВАК "Физика ядра и элементарных частиц" 024, действующего при Национальной научной лаборатории им. А.И. Алиханяна (0036, г. Ереван, ул. Братьев Алиханян 2).

С диссертацией можно ознакомиться в библиотеке ННЛ им. А.И. Алиханяна.

Автореферат разослан 19 апреля 2013г.

Ученый секретарь спец. совета, д.ф.м.н.  Д.Б. Саакян

Общая характеристика работы

Актуальность темы. Помимо применения к известным задачам физики с большим количеством степеней свобод, статфизика в последние годы широко используется при решении современных проблем, возникающих на стыке наук. Одним из таких плодотворных ее применений является решение моделей эволюции методами статфизики. Для решения эволюционных моделей вирусов и молекул приходится использовать всю мощь современной теоретической физики: статфизика спиновых стекол, квантовая статистическая физика, уравнения Гамильтона-Якоби, квантовая теория поля. Применение статфизики к решению моделей эволюции не дань моде, а жизненная необходимость-нет альтернативных методов.

Чем обусловлена связь между моделями эволюции и статфизикой?

В эволюционных моделях есть множество степеней свободы, как в статфизике, что в какой-то степени эквивалентно одномерной модели Изинга. Второй важный аспект: в эволюционных моделях существует нечто похожее на энергию, это фитнес - количество отпрысков.

В области мутационно-селекционных моделей эквивалентность с классической двумерной моделью Изинга известна уже около двадцати лет. В моделях эволюции вирусов и молекул существуют две фазы - селективная и неселективная. В селективной фазе большинство вирусов имеет генотип близкий к генотипу с максимальным фитнесом, а в неселективной фазе все генотипы заселены примерно одинаково. Порог ошибок (error threshold)- это точка фазового перехода. Существование этого явления, сильно напоминающего фазовый переход ферромагнетик-парамагнетик в статфизике, в основном и привлекло внимание экспертов статфизики к изучению моделей эволюции.

В данной диссертационной работе впервые была дана строгая математическая теория проблемы летальных мутаций (мутации приводят к смерти особи)[1].

Также рассматривается изначально длинный геном, как коллекция конечного числа блоков, определяя среднюю "магнетизацию" в каждом таком блоке и фитнес как функцию от магнетизации блоков. Затем используется метод уравнения Гамильтона-Якоби для решения модели[2].

Математически гораздо сложнее изучение моделей, когда длина генома меняется. Если модель с базовыми заменами и селекцией похожа на квантовую статистическую механику, то модель вставок и делеций похожа на квантовую теорию поля. В диссертационной работе эта

задача решена для случая модели Эйгена с однопиковым фитнес ландшафтом [3].

В диссертации также рассматривается модель Эйгена при непуассоновском распределении общего числа мутаций, исследуется феномен порога ошибок в случае высокого уровня мутаций [4].

Модели гаплоидной эволюции описываются системой нелинейных уравнений, которые путем простого преобразования переходят в линейные уравнения. В случае диплоидной эволюции соответствующие уравнения сильно нелинейны. В данной диссертационной работе [5] точно решена модель диплоидной эволюции в случае общего гладкого фитнес ландшафта, вычислен средний фитнес и распределение в устойчивом состоянии.

В некоторых эволюционных задачах важна не только задача для бесконечной длины генома, но и эффекты конечной длины генома (эффекты конечного объема. Эта исключительно сложная в смысле техники проблема тоже была решена в данной работе [6].

Основными предметами изучения в данной работе являются модели Эйгена и Кроу-Кимуры. Рассматривается случай так называемой бесконечной популяции: система описывается с помощью детерминистических дифференциальных уравнений. Генотип отождествляется с одномерной цепочкой длины N спинов s_i . Спины могут принимать значения ± 1 , $1 \leq i \leq N$. Различие между двумя конфигурациями S_i и $S_j \equiv (s_j^1, \dots, s_j^N)$ описывается расстоянием Хамминга $d_{ij} = (N - \sum_k s_i^k s_j^k)/2$. Состояние системы определяется 2^N относительными вероятностями P_i , $1 \leq i \leq 2^N$. В простой модели Кроу-Кимуры вероятности P_i удовлетворяют следующим уравнениям:

$$\frac{dP_i}{dt} = P_i \left(r_i - \sum_{j=1}^{2^N} r_j P_j \right) + \sum_{j=1}^{2^N} m_{ij} P_j. \quad (1)$$

где r_i - фитнес, m_{ij} - коэффициент мутации и $\sum_i m_{ij} = 0$ для баланса вероятностей. m_{ij} определяются следующим образом: $m_{ii} = -\gamma_0 N$; для $d_{ij} = 1$ $m_{ij} = \gamma_0$ и $m_{ij} = 0$ в остальных случаях. В модели Эйгена уравнения имеют следующий вид:

$$\frac{dP_i}{dt} = \sum_{j=1}^{2^N} [Q_{ij} r_j - \delta_{ij} D_j] P_j - P_i \left[\sum_{j=1}^{2^N} (r_j - D_j) P_j \right]. \quad (2)$$

где D_i - коэффициент разрушения, а $Q_{ij} = q^N \frac{d(i,1)}{d(i,1)} (1-q)^{d(i,1)}$, где $\exp[-\gamma] \equiv q^N$ вероятность точного копирования и $\gamma = N(1-q)$.

Целью диссертационной работы является исследование и решение моделей эволюции, а именно в случае каждой модели - нахождение среднего фитнеса и точки фазового перехода из селективной фазы (ферромагнетик) в неселективную фазу (парамагнетик). Целью также является вычисление поправок конечной длины генома при общем фитнес ландшафте. В рамках данной работы исследованы и решены следующие модели:

1. модель эволюции в случае летальных мутаций
2. модель биологической эволюции при многомерном фитнес ландшафте
3. модель Эйгена с мутациями, вставками и делециями
4. модель Эйгена с несколькими коррелированными мутациями
5. диплоидная модель эволюции с общим гладким фитнес ландшафтом

Научная новизна.

Определена и впервые решена эволюционная модель с летальными мутациями в общем случае симметричного фитнес ландшафта. Показано, что существование летальных мутантов может существенно изменить порог ошибок (около 50 %).

Впервые сформулирована и решена (вычислен средний фитнес) многомерная эволюционная модель.

Впервые решена Эйгеновская модель с мутациями, вставками и делециями.

Предложен простой вариант решения парадокса катастрофы ошибок Эйгена путем предположения существования корреляций между мутациями в геноме.

Впервые точно решена диплоидная модель эволюции при общем фитнес ландшафте.

Впервые посчитаны поправки конечной длины генома для эволюционной модели с рекомбинацией и диплоидной эволюционной модели. Полученные результаты позволяют аналитически посчитать свойства моделей с короткой длиной генома.

Практическая и теоретическая ценность.

Общее решение моделей эволюции для летальных и многомерных фитнес ландшафтов может быть применено в конкретных ситуациях

при исследовании эволюционных моделей с точки зрения статфизики, а также в математической теории эволюции, вирусологии и биомедицине.

Метод, предложенный в данной диссертации для вычисления поправок конечной длины генома может быть применен также к другим нелинейным задачам теории вероятности.

Метод, предложенный в данной диссертации для вычисления среднего фитнеса диплоидной модели может быть использован для вычисления асимптотики уравнения Гамильтона-Якоби.

Полученные результаты могут быть использованы в статфизических группах институтов Stanford University (California, U.S.A.), Rice University (Houston, U.S.A.), University of Naples Federico II (Naples, Italy), University of Cologne (Cologne, Germany).

Основные положения, выносимые на защиту.

На защиту выносятся следующие результаты по исследованию моделей эволюции методами статистической физики:

1. Для эволюционной модели с летальными мутациями при симметричном или случайном фитнес ландшафте используя метод отображения в уравнение Шредингера получены выражения для среднего фитнеса и найдена точка фазового перехода из ферромагнитической фазы в парамагнитическую (порог ошибок)[1].
2. Методом уравнения Гамильтона-Якоби решена модель биологической эволюции при многомерном фитнес ландшафте. Получены выражения для среднего фитнеса и найден порог ошибок [2].
3. Для эволюционной модели Эйгена с мутациями, вставками и делециями получены выражения для среднего фитнеса и найден порог ошибок. Построена кривая для фазового перехода[3].
4. Предложено решение парадокса катастрофы ошибок Эйгена, предполагая, что между мутациями есть какая-то корреляция. Методом Гамильтона-Якоби дано решение модели Эйгена с несколькими коррелированными мутациями. Продемонстрировано существование ферромагнитной фазы при длине генома 8000 в случае модели с коррелированными мутациями, тогда как в случае оригинальной модели Эйгена при этой длине вирусы находятся в парамагнитической фазе. Построен график отношения фитнесов в ферромагнитической и парамагнитической фазах от величины мутаций[4].

5. Решена диплоидная модель эволюции с общим гладким фитнес ландшафтом методом уравнения Гамильтона-Якоби. Получены выражения для среднего фитнеса и найден порог ошибок. Построена зависимость среднего фитнеса от параметра и показано совпадение численных и аналитических результатов. Построен график для распределения вероятностей в случае квадратичной фитнес функции и показано совпадение численных результатов с аналитическими[5].

6. Посчитаны поправки конечной длины генома для среднего фитнеса и ген-вероятностей в эволюционных моделях. Построены графики для сравнения численных и аналитических значений вероятностей с поправками и без, показано хорошее совпадение в случае с поправками[6].

Апробация работы. Основные результаты диссертации докладывались диссертантом на семинаре в ННЛ им А.И. Алиханяна.

Публикации. По теме диссертационной работы опубликовано 6 научных работ, список которых приводится в конце автореферата.

Структура и объем диссертации. Диссертационная работа изложена на 110 страницах, состоит из введения, 6 глав, заключения, приложения и списка цитированной литературы, включающего 112 наименований.

Содержание работы

Во введении (Глава 1) обоснована актуальность темы и сделан краткий обзор по проблемам, затронутым в диссертации. Изложены практическая ценность и краткое описание работы.

Во второй главе рассмотрены эволюционные модели с летальными мутациями при симметричном или случайном фитнес ландшафте. В случае дискретной модели Эйгена $P_{i,n}$ рассматривается как вероятность i -го типа в n -ый момент времени. Получается:

$$P_{i,n+1} = \frac{\sum_j Q_{ij} \hat{r}_j P_{j,n}}{\sum_j \hat{r}_j P_{j,n}} \quad (3)$$

Летальные мутанты определяются, как последовательности, для которых $\hat{r}_j = 0$. От дискретной модели Эйгена делается переход к непрерывной модели Киммуры и для летальных мутантов получается $r_i = \ln \hat{r}_i / \tau = -\infty$.

Для нахождения среднего фитнеса и порога ошибок используется квантовомеханический подход. Изначальное уравнение отображается в уравнение Шредингера в мнимом времени используя квантовый Гамильтониан

$$H = -\gamma(1 - \sum_i \sigma_i^x / N) + f_0(\sigma_1^z, \dots, \sigma_N^z). \quad (4)$$

где σ обозначает оператор Паули, а средний фитнес R может быть выражен как

$$R = \lim_{\beta \rightarrow \infty} \frac{\ln Z(\beta)}{N\beta}, \quad Z(\beta) = \text{Tr} e^{-\beta H}. \quad (5)$$

здесь β является обратной температуры. Для получения энергии основного состояния берется нулевой предел температуры. Для преобразования проблемы квантовой статистической механики в проблему классической механики, используется тождество $\text{Tr} e^{C+H} = \text{Tr} [e^{C/L} e^{H/L}]^L$ при $L \rightarrow \infty$ (C и H некоммутирующие операторы) и затем получается выражение, которое действует для любого фитнеса.

$$Z = \sum_{s_i^l = \pm 1} \prod_{l=1}^L e^{-\beta N + \frac{N\beta}{L} \sum_i f_0(s_1^l \dots s_N^l)} e^{[\beta \sum_n (s_n^l s_{n+1}^l - 1)]}. \quad (6)$$

Предлагается следующий вид фитнес-функции: для нелетальной последовательности фитнес является функцией расстояния Хамминга от базисной последовательности $f_0(s_1, \dots, s_N) = f(\frac{1}{N(1-\lambda)} \sum_i s_i)$, а для летальных - $f_0 \rightarrow -\infty$. Получается следующее выражение для среднего фитнеса R :

$$R \equiv \frac{\ln Z(\beta)}{N\beta} = \max_m [f(m) - \mu_0 + (1-\lambda)\mu_0 \sqrt{1-m^2}]. \quad (7)$$

Значение коэффициента мутации, при котором $R = 0$, и есть порог ошибок. Результаты главы опубликованы в работе [1].

В третьей главе изучается биологическая эволюция при многомерном фитнес ландшафте. Порогом ошибки является "намагниченность" m_n , определяемая как

$$m_n = 1 - \frac{l_n}{2L_n}, \quad (8)$$

в n -ой части генома для всех n с $1 \leq n \leq H$. Геном состоит из H частей (блоков), с длинами $L_n, 1 \leq n \leq H$, такими, что $\sum_{n=1}^H L_n = L$, а L

- длина генома. Фитнес γ определяется, как $r_{l_1, \dots, l_H} \equiv Lf(m_1, \dots, m_H)$. Затем используется метод уравнения Гамильтона-Якоби для решения уравнения. В итоге с точностью $1/L$ получается следующее уравнение Гамильтона-Якоби:

$$\begin{aligned} \frac{\partial u(m_1, \dots, m_H)}{\partial t} &= -H(m_1, \dots, m_H; \frac{\partial u}{\partial m_1}, \dots, \frac{\partial u}{\partial m_H}), \\ -H(m_1, \dots, m_H; \hat{P}_1, \dots, \hat{P}_H) &= f(m_1, \dots, m_H) \\ -1 + \sum_{1 \leq n \leq H} \frac{L_n}{L} &(\frac{1+m_n}{2} e^{2\hat{P}_n} + \frac{1-m_n}{2} e^{-2\hat{P}_n}), \end{aligned} \quad (9)$$

где опускаются все члены порядка $O(1/L)$ и импульсы представляются, как $\hat{P}_n = \partial u / \partial m_n$. Затем рассматривая асимптотическое решение $u(m_1, \dots, m_H, t) = Rt + u_0(m_1, \dots, m_H)$, получается уравнение для среднего фитнеса:

$$\begin{aligned} R &\geq \max [U(m_1, \dots, m_H)]_{m_1, \dots, m_H}, \\ U(m_1, \dots, m_H) &= f(m_1, \dots, m_H) - 1 \\ &+ \sum_{1 \leq n \leq H} \frac{L_n}{L} \sqrt{1-m_n^2}. \end{aligned} \quad (10)$$

Рис.1 дает сравнение аналитических результатов для Ур.(10) с численными результатами в случае трехмерной модели.

Результаты главы опубликованы в работе [2].

В четвертой главе вычисляется средний фитнес и порог ошибки модели Эйгена с мутациями, инсерциями и делециями для однопикового фитнес ландшафта.

Существует пиковая конфигурация со всеми плюсовыми спинами $N = N_0$, фитнес которой равен $r(N_0, N_0) = J$, а для остальных конфигураций фитнес равен 1.

В модели Эйгена важно то, что мутации происходят только во время воспроизведения особей, а также то, что базовые замены, вставки и делеции возможны в одно и то же время: вероятность базовой замены для одного спина - a/N_0 , вероятность делеции для одного спина - b/N_0 , вероятность вставки для одного спина - c/N_0 . Таким образом, получается уравнение

$$\frac{dP_i}{dt} = \sum_j \hat{Q}_{ij} r_j P_j - P_i (\sum_j r_j P_j), \quad (11)$$

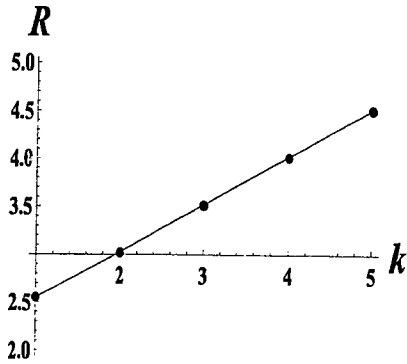


Рис. 1: Сравнение аналитических результатов (непрерывная линия) с численными (точки) для трехмерной модели с $L_1 = L_2 = L_3 = 20$. Мутационная скорость для всего генома равна 1. Фитнес первой части $f_1(m_1) = km_1^2/2$, во второй части все мутации летальны. В третьей части $J = 3$ для пиковой подпоследовательности и ноль для остальных. Фитнес-функция определяется как $f(m_1, m_2, m_3) = km_1^2/2 - [1 - \delta(m_2 - 1)] * \infty + J\delta(m_3 - 1)$, где дискретная $\delta(x) = 1$, если $x = 0$ и 0 в остальных случаях. Средний фитнес задается как $k(1 - 1/(3k))^2/2 + 3 - 2/3$.

где \hat{Q}_{ij} описывает комбинацию всех трех процессов. В стационарном состоянии для $P(N_1 + l, N_1 + l) \equiv p_l$ получается следующее уравнение:

$$Rp_l = e^{-a-b-c} \sum_n \sum_{h-m=n} p_{l-n} [1 + (J-1)\delta_{l,l_0}] \frac{b^m c^h}{2^h h! m!}, \quad (12)$$

где $l_0 = N_0 - N_1$. После сложных комбинаторных вычислений для среднего фитнеса получается уравнение

$$-\frac{1}{J-1} = \frac{1}{L} \sum_{l=1}^L \left(1 + \frac{R}{Q \exp[bq(l) + \frac{c}{2q(l)}] - R} \right). \quad (13)$$

которое в пределе $L \rightarrow \infty$ дает

$$\frac{J}{J-1} = -\frac{1}{2\pi i} \int \frac{dz}{z} \left[\frac{\frac{R}{Q}}{\exp[(bz + c/(2z))] - \frac{R}{Q}} \right]. \quad (14)$$

В неселективной фазе популяция размещается у пиковой последовательности, поэтому средний фитнес равен 1. Исходя из этого получается выражение для порога ошибок:

$$\frac{J}{J-1} = -I(bc, \frac{1}{Q}) \quad (15)$$

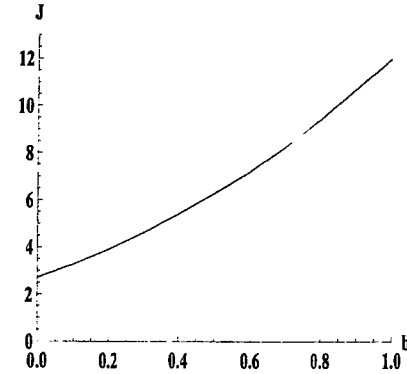


Рис. 2: Значение порога ошибки для параметра J в случае модели однопикового фитнес ландшафта с $a = 1, b = c = b$, которые определяются формулой Ур.(15).

$$I(x, y) = -1 - \sum_{l=1}^{\infty} \frac{1}{y^l} \sum_{n=0}^{\infty} \frac{(\sqrt{\frac{x}{2}} l)^{2n}}{n!} = -1 - \sum_{l=1}^{\infty} \frac{I_0(\sqrt{2xl})}{y^l} \quad (16)$$

Результаты главы опубликованы в работе [3].

В пятой главе рассматривается модель, которая задается Ур.(2), но с коррелированными мутациями. В этом случае мутационная матрица изменяется следующим образом

$$Q_{ij} \equiv Q_n = \hat{Q} q^{-n} (1-q)^n g_n, \quad (17)$$

где $n \equiv d(i, j)$, g_n коэффициенты, которые увеличиваются с увеличением n , и \hat{Q} определяет нормализацию вероятностей Q_{ij} . Рассматривается симметричное распределение $p_i = P_l$ и для P_l выводится следующее выражение:

$$\frac{dP_l}{dt} = \sum_{n_1=0}^l \sum_{n_2=0}^{L-l} \frac{l!}{n_1!(l-n_1)!} \frac{(L-l)!}{n_2!(L-l-n_2)!} Q_{n_1+n_2} P_{l-n_1-n_2}, \quad (18)$$

где $n_1 - n_2 = n$.

Затем используются комбинаторные формулы, в случае $N - l \gg 1$,

$l \gg 1$ получается уравнение Гамильтона-Якоби

$$\frac{\partial u(m, t)}{\partial t} + H(m, u') = 0, \\ -H = \sum_{n=0}^L \left[\left(\frac{1+m}{2} e^{-2u'} + \frac{1-m}{2} e^{2u'} \right) \gamma \right]^n \frac{g_n}{n!} f(m) \hat{Q}. \quad (19)$$

Средний фитнес выводится следуя идеи в [5]. Рассматривается простой выбор g_n :

$$g_1 = 1, \quad n \leq 2; \\ g_n = \frac{1}{(n+1)(n+2)(n+3)}, \quad n > 2. \quad (20)$$

В этом случае для среднего фитнеса выводится следующее выражение

$$R = \max \left[\frac{f(m) \gamma^3}{e^\gamma - 1 - \gamma - \frac{1}{2} \gamma^2 + \frac{5}{6} \gamma^3 + \frac{23}{24} \gamma^4 + \frac{59}{120} \gamma^5} \right. \\ \left. \times \left(\frac{e^w - 1 - w - \frac{w^2}{2}}{w^3} + \frac{5}{6} + \frac{23}{24} w + \frac{59}{120} w^2 \right) \right]_{|m}, \\ w = \gamma \sqrt{1 - m^2}, \quad (21)$$

В селективной фазе для среднего фитнеса, при однопиковом фитнес ландшафте, получается:

$$R = A \frac{\gamma^3}{e^\gamma - 1 - \gamma - \frac{1}{2} \gamma^2 + \frac{5}{6} \gamma^3 + \frac{23}{24} \gamma^4 + \frac{59}{120} \gamma^5}. \quad (22)$$

а для того, чтобы получить условие для порога ошибок, подставляется $R = 1$ в Ур.(22).

$$1 = A \frac{\gamma^3}{e^\gamma - 1 - \gamma - \frac{1}{2} \gamma^2 + \frac{5}{6} \gamma^3 + \frac{23}{24} \gamma^4 + \frac{59}{120} \gamma^5}. \quad (23)$$

На Рис.4 представлено сравнение критического значения A для разных значений γ в случае модели Эйгена по уравнению $Ae^{-\gamma} = 1$, и в случае модели по Ур.(23). Видно, что при заданных значений A и $(1-q)$, Ур.(23) дает большую критическую длину для цепочки, чем определяется уравнением $Ae^{-\gamma} = 1$.

Результаты главы опубликованы в работе [4].

В шестой главе решаются модели диплоидной эволюции. В этом случае уравнение имеет следующий вид:

$$\frac{dP_l}{dt} = P_l F_l - P_l \left(1 + \sum_k F_k P_k \right) \\ + \frac{1}{N} [(N-l+1)P_{l-1} + (l+1)P_{l+1}]. \quad (24)$$

где $F_l = \sum_{n=0}^N f(1-2l/N, 1-2n/N) P_n$. С точностью $1/N$, F_l может быть заменена на $F_l = f(m, m_0)$. Подразумевая, что $N \rightarrow \infty$, получается

$$\frac{N du(m, t)}{dt} = f(m, m_0) + \frac{1+m}{2} \exp[2u'(m, t)] \\ + \frac{1-m}{2} \exp[-2u'(m, t)] - 1 - f(m_0, m_0), \quad (25)$$

Максимального значения $\exp[Nu(m, t)]$ достигает в точке m_0 . Большая постоянная N исчезает в правой стороне Ур.(25). Средний фитнес R определяется как $R = \sum_i F_i P_i$. В итоге R задается следующим выражением:

$$R = \max_m [U(m, M_0)], \\ U(m, M_0) = \sqrt{1 - m^2} - 1 + f(m, M_0). \quad (26)$$

Рис.5 видно, что аналитические результаты, полученные для среднего фитнеса R с квадратичной фитнес функцией совпадают с численными результатами.

Результаты главы опубликованы в работе [5].

В седьмой главе вычисляются поправки $1/L$. Уравнение в случае модели Кроу-Кимуры имеет следующий вид

$$\frac{dP_l}{dt} = P_l (L f(m_l) - L) + (L-l+1)P_{l-1} \\ + (l+1)P_{l+1}. \quad (27)$$

С помощью анзаца $P_l = \exp[LU(m, t)]$ и формулы $P_{l \pm 1} \approx P_l e^{-(\pm 2U'_m)}$, где $U' \equiv \frac{\partial U(m, t)}{\partial m}$, Ур.(27) преобразовывается в уравнение Гамильтона-Якоби

$$\frac{\partial U}{\partial t} = H(m, U'), \\ H(m, y) \equiv f(m) - 1 + \frac{1+m}{2} e^{2y} + \frac{1-m}{2} e^{-2y}, \quad (28)$$

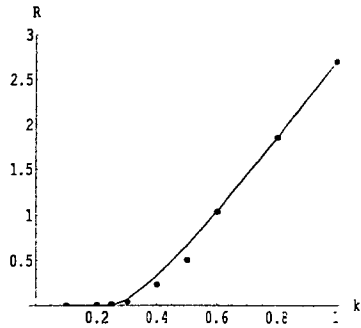


Рис. 3: Средний фитнес $R = f(M_0, M_0)$ в зависимости от k для модели Кроу-Кимуры при квадратичном фитнесе $f(m_1, m_2) = k[a(m_1^2 + m_2^2)/2 + bm_1m_2]$ с $a = 4$, $b = 0.5$ и $N = 100$; сплошная линия: аналитические результаты, точки: численные результаты. Теоретический порог ошибок находится в точке $k = 0.246$.

где $y \equiv U'$. Предполагая асимптотику $U(m, t) = kt + u(m)$, получаем простое дифференциальное уравнение

$$k = f(m) - 1 + \frac{1+m}{2}e^{2u'} + \frac{1-m}{2}e^{-2u'}$$

$$2u' = \ln \frac{k - 1 + f(m) \pm \sqrt{(k - 1 + f(m))^2 - 1 - m^2}}{1 + m}. \quad (29)$$

Чтобы вычислить поправки высокого порядка, $\ln P_i$ представляется как

$$\ln P_i = L(k + \frac{k_1}{L})t + Lu(m) + u_1(m) + O(1). \quad (30)$$

Уравнения (27) и (30) дают:

$$k + \frac{k_1}{L} = f(m) - 1 + \frac{1+m}{2}e^{2u'} + \frac{1-m}{2}e^{-2u'}$$

$$+ \frac{1}{L}(e^{2u'} + e^{-2u'}) + \frac{2u''}{L}(\frac{1+m}{2}e^{2u'} + \frac{1-m}{2}e^{-2u'})$$

$$+ \frac{2u_1'}{L}[\frac{1+m}{2}e^{2u'} - \frac{1-m}{2}e^{-2u'}]. \quad (31)$$

В Ур.(31), k является основным выражением для среднего фитнеса и k_1/L - это поправка первого порядка к нему. Имея значение k_1 , можно вычислить u_1' . k_1 можно определить из Ур.(31) в точке $m = m_0$, где

коэффициент u_1' становится равным нулю. Затем, имея k_1 , вычисляем u_1 .

Имея u_1 , вычисляется с точностью $1/L$ выражение для вероятностей гаплоидных классов P_i (они определены как $p_{i1} \equiv (Lu(m) + u_1(m))$), также вычисляются вероятности $p_{0i} \equiv e^{Lu(m)}$, Ур.(30).

Результаты главы опубликованы в работе [6].

В заключении представлены основные результаты работы:

1. Решена модель эволюции в случае летальных мутаций: фитнес получен, как симметричная функция Хамминг расстояния от пиковой последовательности, и есть конечная вероятность, что фитнес станет нулевым. Такая модель описывает практически важную ситуацию эволюции вирусов под воздействием противовирусных препаратов. Точно определяется порог ошибок в этом случае [1].
2. Сформулирована и решена модель эволюции в многомерном пространстве, как в случае модели Кроу Кимуры, так и для моделей с рекомбинацией. Это важно для применений, поскольку фитнес сложная функция и не может быть описана как простая функция Хамминг расстояния от одной последовательности [2].
3. Сформулирована и решена модель Эйгена в случае вставок и делеций. Найден порог ошибок [3].
4. Решена модель Эйгена в случае коррелированных мутаций. определен порог катастрофы ошибок. Это простой случай явления, когда мутации слегка управляются. Предложена простая схема для решения парадокса катастрофы ошибок Эйгена при возникновении жизни [4].
5. Решена модель диплоидной эволюции с гладким фитнес ландшафтом, когда есть еще и рекомбинация. определен и средний фитнес популяции, и распределение вероятностей [5].
6. Посчитаны поправки конечного объема в эволюционных моделях. Помимо модели Кроу-Кимуры были рассмотрены модели диплоидной эволюции и модель с рекомбинацией. Полученные результаты позволяют аналитически посчитать свойства моделей с короткой длиной генома [6].

Работы автора по теме диссертации

- [1] Z.A. Kirakosyan, D. B. Saakian, and C.-K. Hu, "*Evolution models with lethal mutations on symmetric or random fitness landscapes*", Phys. Rev. E **82**, 011904 (2010).
- [2] D.B. Saakian, Z.A.Kirakosyan and C.-K. Hu, "*Biological evolution in a multidimensional fitness landscape*", Phys.Rev.E **86** 031920 (2012).
- [3] Z.A. Kirakosyan, "*Eigen evolution model with mutations, insertions and deletions*" , Chinese journal of physics, **50**, 101 (2012).
- [4] Z.A.Kirakosyan, D.B. Saakian, C.-K. Hu, "*Eigen model with correlated multiple mutations and solution of error catastrophe paradox in the origin of life*", Journal of the physical society of Japan, **81** (2012).
- [5] D. B. Saakian, Z.A. Kirakosan and C.-K. Hu, "*Diploid biological evolution model with general smooth fitness landscape*", Phys. Rev. E **77** 061907 (2008).
- [6] Z.A. Kirakosyan, D. B. Saakian and C.-K. Hu, "*Finite genome length corrections for the mean fitness and gene probabilities in evolution models*", J. Stat. Phys **144** 198 (2011).

ԱՄՓՈՓՎԱԳԻՐ

Արենախոսությունում ուսումնասիրված են էվոլյուցիոն մաթեմատիկական մոդելներ, որոնք լուծվում են վիճակագրական ֆիզիկայի մեթոդներով: Այս կերպ արացվել են որոշ մոդելների առաջին լուծումները, քանի որ վիճակագրական ֆիզիկայի մեթոդներին չկա այլընտրանք: Ոչ գծային դիֆերենցիալ հավասարումների համակարգերն արտապարկելով են վիճակագրական ֆիզիկայի մոդելներին, ու դինամիկայի փարբեր անալիտիկ լուծումներ համապարասխանում են վիճակագրական ֆիզիկայի մոդելների փարբեր փուլերի: Ներագրվել են մոդելների փուլային անցման կետերը: Աշխարանքում սրացված հիմնական արդյունքներն են`

1. Լուծվել են էվոլյուցիայի մոդելները լեպալ մուտացիաների դեպքում: Վիրուսի ֆիթնեսը կամ գրո է ինչ-որ հավանականությամբ, կամ որոշվում է որպես սիմետրիկ ֆունկցիա, կախված սպինային հաջորդականության միջին մագնիսականությունից: Լուծվել է պարահական էներգիաների մոդելի դեպքը: Մենք նախ ձևակերպել ենք մոդելը դիսկրետ ժամանակի դեպքում, հետո անցել Կրոու-Կիմուրայի մոդելին: Վերջինիս համար լեպալ դեպքը համապարասխանում է մինուս անսահմանություն արժեքի ֆիթնեսին: Նաշվել են համակարգի միջին ֆիթնեսը և փուլային անցման կետը: Ընդհանուր դեպքում միջին ֆիթնեսն արտահայտվել է այն դեպքի միջին ֆիթնեսի արտահայտությամբ, երբ չկան լեպալ մուտացիաներ: Սրացված արտահայտությունները ցույց են փայխ, որ լեպալ մուտացիաները լրջորեն կարող են փոխել սխալների շեմի արժեքը:

2. Ուսումնասիրվել են էվոլյուցիայի մոդելները բազմաչափ փարածությունում: Այս դեպքում գենոմը բաղկացած է մի քանի կոորդներից, ամեն կոորն ունի սպինային միջին մագնիսականություն, ու ֆիթնեսը ֆունկցիա է այդ մագնիսականությունից: Այս մոդելը շար ավելի իրարեսական է, քան լայն կիրառում գրած մոդելը, որտեղ ֆիթնեսը կախված է միայն մի հաջորդականությունից ունեցած Նամինգի հեռավորությունից: Ձևակերպվել է մոդելը, այնուհետև լուծվել Նամիլոն-Յակոբիի հավասարման միջոցով, գրվել է մոդելի միջին ֆիթնեսը թե՛ թվային, թե՛ անալիտիկ մեթոդներով: Այս մոդելի մասնավոր դեպքն է ֆիթնես լանդշաֆտը, երբ ֆիթնեսը ֆունկցիա է մի քանի հաջորդականություններից ունեցած Նամինգի հեռավորությունից: Դիփարկվել են նաև այն դեպքը, երբ կան լեպալ մուտացիաներ: Ներագրվել են մի շարք երկչափ և եռաչափ դեպքեր: Ձևակերպվել և լուծվել է մոդելը նաև այն դեպքի համար, երբ կա ռեկոմբինացիա, և դուրս են բերվել Նամիլոն-Յակոբիի հավասարումը և կարևոր թեորեմա դինամիկայի վերաբերյալ: Եթե սկզբնական բաշխումը սիմետրիկ է, և ֆիթնես ֆունկցիան ունի միջին դաշտի տեսք, ապա ռեկոմբինացիան չի կարող փոխել միջին ֆիթնեսը մինչ շար երկար գենոմի երկարության կարգի ժամանակի ընթացքում: Այս թեորեմը ճիշտ չէ ասիմետրիկ ռեկոմբինացիայի

դեպքում: Մոդելի դինամիկան կարելի է մոտավոր հաշվել, օգտագործելով Նամիլյուն-Յակոբիի հավասարումը:

3. Ուսումնասիրվել է էվոլյուցիայի մոդելն այն դեպքում, երբ գենոմում բացի մուտացիաներից ինչ-որ հավանականությամբ կամ ջնջվում են սպիններ. կամ նոր սպիններ են ավելանում: Մեկ զագաթանի ֆիթնեսի էվոլյուցիոն մոդելը լուծվել է էյզենի մոդելի փարբերակի դեպքում: Խնդիրը լուծվել է գենոմի անվերջ երկարության դեպքում, օգտագործելով բարդ կոմբինատորիկա: Հաշվվել են միջին ֆիթնեսը, ինչպես նաև փուլային անցման կետը:

4. Ուսումնասիրվել է էյզենի մոդելն այն դեպքի համար, երբ կա կորելյացիա գենոմի փարբեր մասերում փեղի ունեցող մուտացիաների միջև, այնպես որ մուտացիաների ընդհանուր թիվն ունի ոչպուասոնյան բաշխում: Գրնվել է սխալների կապասարոֆայի շեմը: Մենք օգտագործել ենք Նամիլյուն-Յակոբիի հավասարման մեթոդը մոդելը լուծելու համար, հաշվել ենք միջին ֆիթնեսը և սխալների շեմը: Այս կերպ կարելի է լուծել էյզենի սխալների կապասարոֆայի պարադոքսը, եթե համարենք, որ մի քանի մուտացիաների հավանականությունը մեծ է, իսկ շատ մուտացիաները քիչ հավանական են:

5. Օգտագործելով Նամիլյուն-Յակոբիի հավասարումը, մենք դուրս ենք բերել անալիտիկ հավասարումներ միջին ֆիթնեսի և կայուն վիճակի բաշխման համար Կրոու-Կիմուրայի և էյզենի էվոլյուցիայի դիպլոիդ մոդելների դեպքում, երբ ֆիթնես ֆունկցիան սիմետրիկ է: Լուծվել է դիպլոիդ էվոլյուցիայի մոդելը նաև ռեկոմբինացիայի դեպքի համար: Այս մոդելի համիլյունիանը կախված է ոչ միայն կոորդինատից և գործողության ածանցյալից, այլև այն M_0 կետից, որտեղ գործողությունն ունի մաքսիմում: Եթե ասեքսուալ էվոլյուցիայի դեպքում ունենք մեկ լուծում R միջին ֆիթնեսի համար, ապա դիպլոիդ էվոլյուցիայի դեպքում հնարավոր են փարբեր լուծումներ, քանի որ $U(m, M_0)$ պոպուլացիալ կախված է M_0 պարամետրի արժեքից: Քանի որ $U(m, M_0)$ պոպուլացիալ էքսպրեսիոնի պրոբլեմն ունի մի քանի լուծում, ապա կարող են լինել մի քանի կայուն լուծումներ:

6. Օգտագործել ենք Նամիլյուն-Յակոբիի հավասարման մեթոդը հեթազոպելու համար մոդելները գենոմի L երկարության դեպքում: Մենք հաշվել ենք $1/L$ կարգի ճշգրտումները միջին ֆիթնեսի և կայուն վիճակի բաշխման համար խիստ ոչզծային էվոլյուցիոն մոդելների դեպքում գեների հորիզոնական փոխանցման և դիպլոիդ էվոլյուցիայի դեպքերն ընդհանուր սիմետրիկ ֆիթնեսի դեպքում: Միջին ֆիթնեսը սեփական մաքսիմալ արժեքի (զծային հանրահաշվի դեպք) էկվիվալենտն է ոչզծային մոդելների դեպքում:

SUMMARY

In this thesis we studied mathematical evolution models with the methods of statistical physics. This way we get the first solutions of some models, as there are no alternative methods. We mapped the system of nonlinear differential equations to the models of statistical physics and found a connection between different analytical solutions of dynamics of that equations and different phases of statistical physics models. We also studied the phase transition points. The main scientific results presented in this thesis are:

1. It has been solved the evolution model with lethal mutations. In that case, the fitness is equal to 0 with some probability or it is defined as a symmetric function from mean magnetization. We also solved the Random Energy model. First we defined the lethal mutations using the discrete time Eigen model, then we used the mapping into the continuous time Crow-Kimura model. We get the fitness equals minus infinity for the lethal mutants. We calculated the mean fitness and phase transition points for the considered models. In the general case, the mean fitness of the model with lethal mutations is expressed by the mean fitness of the model without mutations. According to our results, the existence of lethal mutants can substantially change the error threshold (about 50 %).
2. It has been studied and solved the evolution model in the multidimensional fitness space, where the genome is considered as a collection of several pieces (each piece has its mean magnetization), and the total fitness is the function of these magnetizations. Such model is more general and more realistic than the widely applied model, where the fitness is a function of the Hamming distance from only one sequence. We formulated the model, then solved it using Hamilton-Jakobi equation method, and calculated the mean fitness both analytically and numerically. The particular version of this model corresponds to the fitness as a function of the Hamming distances from several sequences. We considered the case with lethal mutation. Several two- and three dimensional cases have been studied. We formulated the recombination model in the multidimensional fitness space, derived the Hamilton-Jacobi equation for the dynamics of population and deduced an important theorem about the dynamics. For the initially narrow initial distribution and mean-field fitness landscape, the recombination does not affect the dynamics of the population for a rather long period of times. This theorem is not valid in the case of asymmetric recombination.

3. It has been studied the evolution model in case when, besides the base substitution, with some probability there are insertions and deletions in the genome. We solved the Eigen model for the single-peak fitness with infinite genome length, using the involved combinatorics. The mean fitness of the model and the phase transition point have been calculated.
4. We investigated the Eigen model when there is a correlation between the mutations at different parts of the genome, and the total number of mutations has non-Poisson distribution. We applied Hamilton-Jacobi equation method to solve the model and calculated the mean fitness and error threshold. The suggested mechanism gives a simple solution of error catastrophe paradox in the origin of life.
5. Using Hamilton-Jacobi equation approach, we obtain analytic equations for steady state population distributions and mean fitness for Crow-Kimura and Eigen type diploid biological evolution models with general smooth symmetric fitness landscapes. We also study the parallel diploid model for the simple case of recombination and calculate the variance of distribution. The Hamiltonians of these models depend on the value M_0 of the point of maximum of distribution. In case of asexual evolution, we have a single solution for the mean fitness R , while in diploid case several solutions are possible, as potential $U(m, M_0)$ depends on the value of parameter M_0 . The existence of several solution for the extremum problems of a potential $U(m, M_0)$ indicates the possibility of multi-stability.
6. Using Hamilton-Jacobi equation approach to study the models with the genomes of length L , we calculated $1/L$ corrections to the mean fitness and steady state distribution for a strongly nonlinear evolution models: horizontal gene transfer model and diploid parallel evolution model for general fitness function. The mean fitness is the equivalent of the maximum eigenvalue (for linear algebra) in case of nonlinear models.

