

Модели предбиологической и ранних стадий биологической эволюции

НИИ системных исследований РАН

Редько Владимир Георгиевич
vgredko@gmail.com

План доклада

- 1. Модель квазивидов: модель эволюции информационных молекул РНК (М. Эйген)**
- 2. Модель гиперциклов: РНК и ферменты кооперируются (М. Эйген, П. Шустер)**
- 3. Сайзеры: модель самовоспроизводящихся систем (SYstem of SELft-Reproduction) – на пути к протоклеткам (В.А. Ратнер, В.В. Шамин)**
- 4. Адаптивный сайзер: модель возникновения системы управления на предбиологическом уровне**
- 5. Модель взаимодействия между обучением и эволюцией**

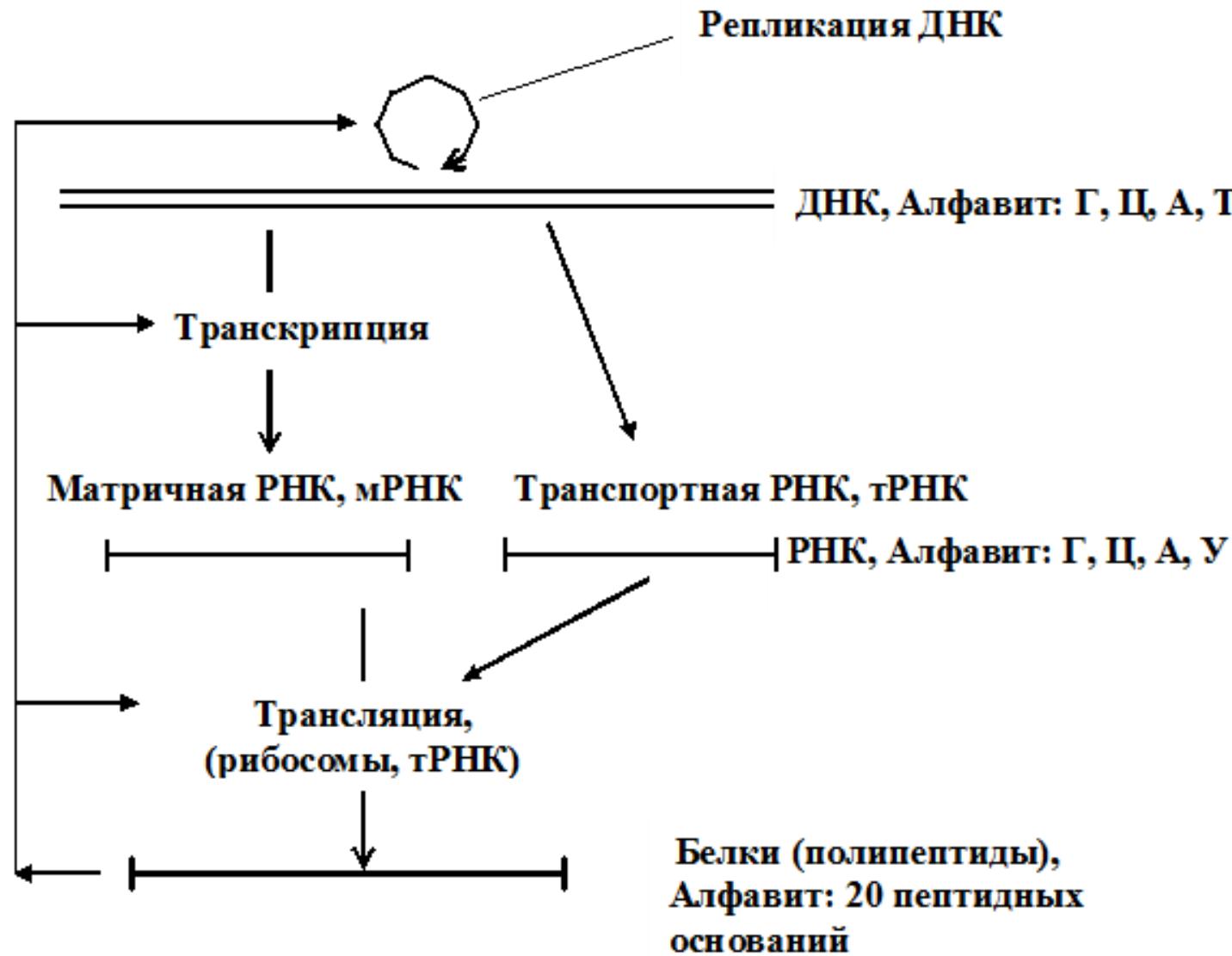
Ключевая литература

- 1. Эйген М.** Самоорганизация материи и эволюция биологических макромолекул. М.: Мир, 1973.
- 2. Эйген М., Шустер П.** Гиперцикл. Принципы самоорганизации макромолекул. М.: Мир, 1982.
- 3. Ратнер В.А., Шамин В.В.** Сайзеры: моделирование фундаментальных особенностей молекулярно-биологической функции. Соответствие общих свойств генетических процессов и конструктивных особенностей коллективов макромолекул // Журн. общ. биологии. 1983. Т. 44, № 1. С. 51–61.
- 4. Ратнер В.А.** Молекулярно-генетическая система управления // Природа. 2001, № 3. С. 16-22.

Ключевая литература

5. Редько В.Г. Эволюция, нейронные сети, интеллект. Модели и концепции эволюционной кибернетики. М.: УРСС, 2005.
6. Редько В.Г. Моделирование когнитивной эволюции: На пути к теории эволюционного происхождения мышления. М: ЛЕНАНД/URSS, 2015.
7. Редько В.Г. Интернет-лекции «Эволюционная кибернетика» (1999). URL: <http://www.keldysh.ru/pages/BioCyber/Lectures.html>
8. Red'ko V.G. Mathematical modeling of evolution. Principia Cybernetica Project, 1999. URL: <http://pespmc1.vub.ac.be/MATHME.html>

К чему идти: схема самовоспроизводящейся молекулярно-генетической системы живой клетки



Гипотетический процесс предбиологической эволюции

Квазивиды → Гиперциклы → Сайзеры →
→ Протоклетки → Простейшие организмы

Модель квазивидов

(М. Эйген, Г. Кун, Дж. Эдельман)

Содержательно модель квазивидов описывает возможный процесс того, как могли бы эволюционно возникнуть самые простейшие макромолекулы (РНК), кодирующие наследственную информацию. При наличии притока энергии в виде богатых энергией молекул типа АТФ цепочки РНК могут реплицироваться.

Эволюция такой популяции приводит к отбору **квазивида** – такого распределения цепочек РНК, в которое входит как «наилучшая РНК» (размножающаяся с максимальной скоростью), так и близкие к ней цепочки, отличающиеся от этой наилучшей мутационными заменами.

Формальная схема модели квазивидов

Рассматривается **Дарвиновская эволюция** популяции особей (цепочек символов, аналогов ДНК) S_1, S_2, \dots, S_n .

$$S_{ki} = 0, 1; \quad k = 1, 2, \dots, n; \quad i = 1, 2, \dots, N.$$

Длина цепочек N и численность популяции n велики:
 $N, n \gg 1$. $N = \text{const}$, $n = \text{const}$.

Имеется **оптимальная цепочка S_M** , имеющая максимальную приспособленность. **Приспособленность** произвольной особи S определяется расстоянием по Хеммингу $\rho(S, S_M)$ между S и S_M (числом несовпадающих символов в соответствующих позициях цепочек):

$$f(S) = \exp[-\beta \rho(S, S_M)], \quad (1)$$

β – интенсивность отбора, $\beta > 0$.

Расстояние по Хеммингу

$$\mathbf{S_M} = (0, 1, 1, 0, 1)$$

$$\mathbf{S_1} = (0, 0, 1, 1, 0), \quad \rho(\mathbf{S_1}, \mathbf{S_M}) = 3.$$

$$\mathbf{S_2} = (1, 1, 1, 1, 1), \quad \rho(\mathbf{S_2}, \mathbf{S_M}) = 2.$$

Алгоритм модели

Шаг 0. *Формирование начальной популяции $\{S_k(0)\}$.* Для каждого $k = 1, \dots, n$, и для каждого $i = 1, \dots, N$, выбираем случайно символ S_{ki} , полагая его равным 0 либо 1.

Шаг 1. Отбор

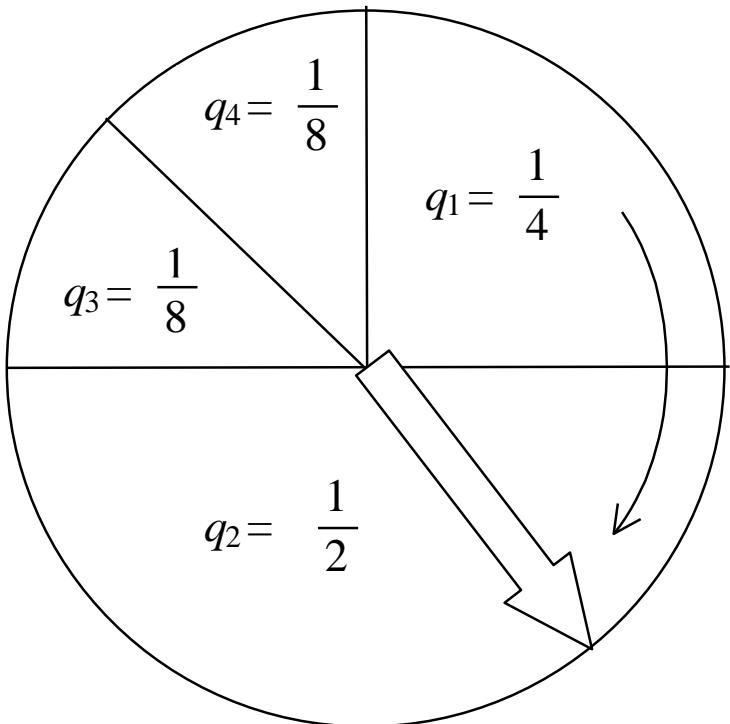
Подшаг 1.1. *Расчет приспособленностей.* Для каждого $k = 1, \dots, n$, вычисляем величину $f(S_k)$.

Подшаг 1.2. *Формирование новой популяции $\{S_k(t+1)\}$.* Отбираем n особей в новую популяцию $\{S_k(t+1)\}$ с вероятностями, пропорциональными $f(S_k)$.

Шаг 2. Мутации особей в новой популяции. Для каждого $k = 1, \dots, n$, для каждого $i = 1, \dots, N$, случайно заменяем символ $S_{ki}(t+1)$ на 0 либо 1 с вероятностью P_m .

P_m – интенсивность мутаций.

Отбор пропорционально-вероятностный (рулеточный метод)



Доля k -го сектора рулетки

$$q_k = f_k [\sum_l f_l]^{-1}, \quad f_k = f(S_k).$$

Ровно n раз крутим рулетку,
номер сектора определяет номер
особи, выбираемой в популяцию
следующего поколения.

Для каждого вращения
вероятность k -й особи попасть в
следующее поколение
пропорциональна ее
приспособленности f_k .

Показан пример, для которого $n = 4$,

$$f_1 = 1/2, f_2 = 1, f_3 = 1/4, f_4 = 1/4.$$

Модель квазивидов. Детерминированный случай: $n \gg 2^N$

При $n \gg 2^N$ динамику популяции характеризуем так:

$$dx_k/dt = W_k x_k + \sum_l f_{kl} x_l - E x_k,$$

x_k – численность особей k -го вида, их W_k – селективная ценность, f_{kl} – мутационные потоки, E – параметр, характеризующий общее разбавление популяции.

Считаем, что

$$\sum_k x_k = n = \text{const}$$

Модель квазивидов.

Детерминированный случай: $n \gg 2^N$

При отсутствии мутаций ($f_{kl} = 0$):

$$dx_k/dt = (W_k - \langle W \rangle) x_k, \quad \langle W \rangle = (1/n) \sum_k W_k x_k$$

Общее решение есть

$$x_k(t) = x_k(0) n \exp(W_k t) [\sum_k x_k(0) \exp(W_k t)]^{-1}$$

Выживает вид, имеющий наибольшую селективную ценность W_k

При наличии мутаций выживает квазивид

Стохастический случай: $2^N \gg n$

Параметры модели квазивидов:

N – длина цепочки, длина модельного генотипа

n – численность популяции

β – интенсивность отбора

P_m – интенсивность мутаций

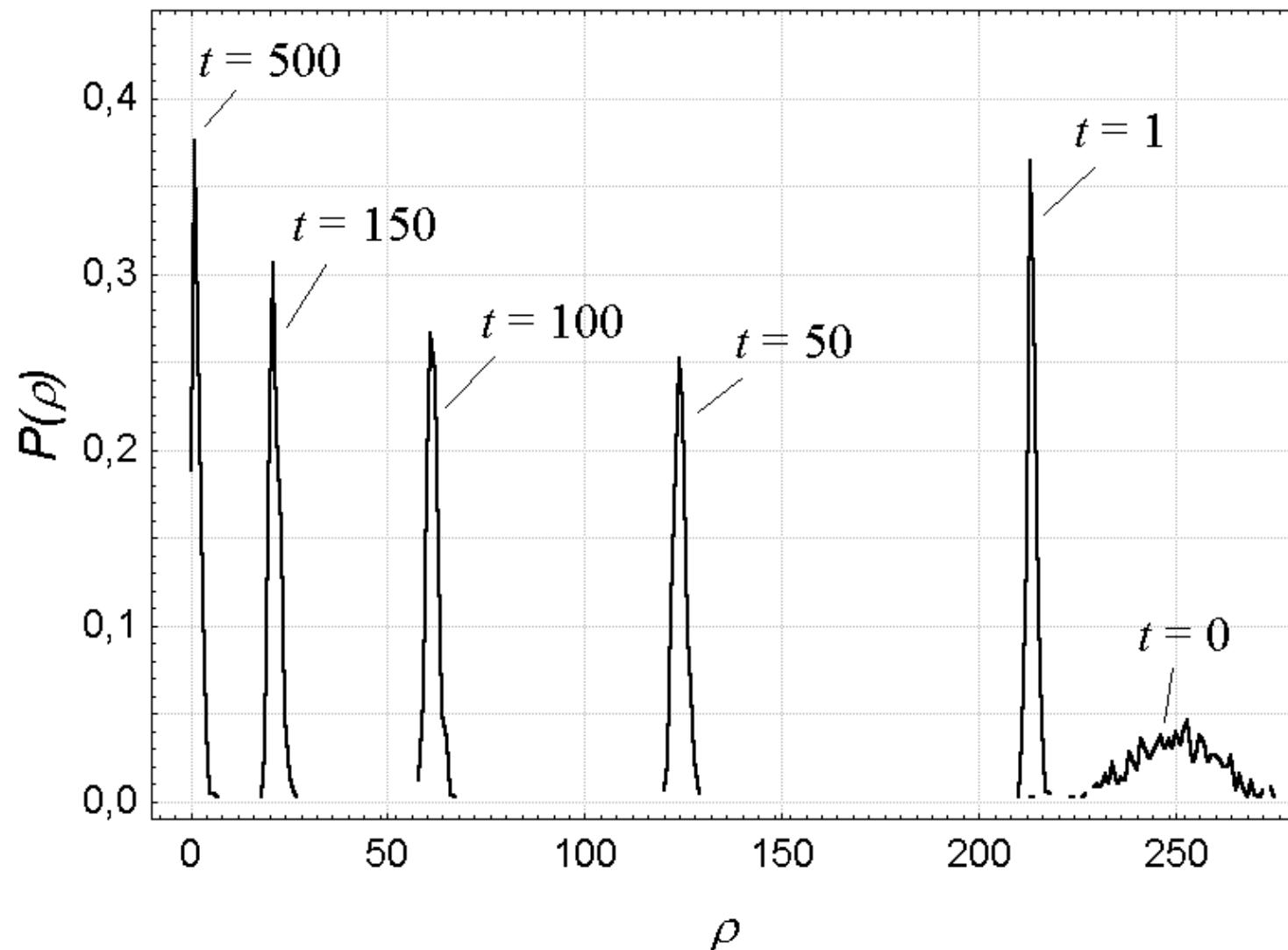
Предположения модели: $N, n \gg 1, 2^N \gg n$

Обычно будем считать: $\beta \sim P_m N, P_m N \sim 1$

Достоинство модели квазивидов

Можно использовать одну существенную переменную –
расстояние до оптимума ρ

Качественная схема эволюции



$$N = 500, n = N, \beta = 1, P_m = 0.002$$

Роль нейтрального отбора

Если численность популяции n сравнительно невелика, то эволюционный процесс существенно стохастический и особи могут фиксироваться в популяции случайно, независимо от их приспособленностей.

Роль нейтрального отбора исследовалась в работах М. Кимуры, было показано, что характерное время T_n (число поколений) нейтрального отбора составляет порядка численности популяции n

Кимура М. Молекулярная эволюция: теория нейтральности. М.: Мир, 1985.

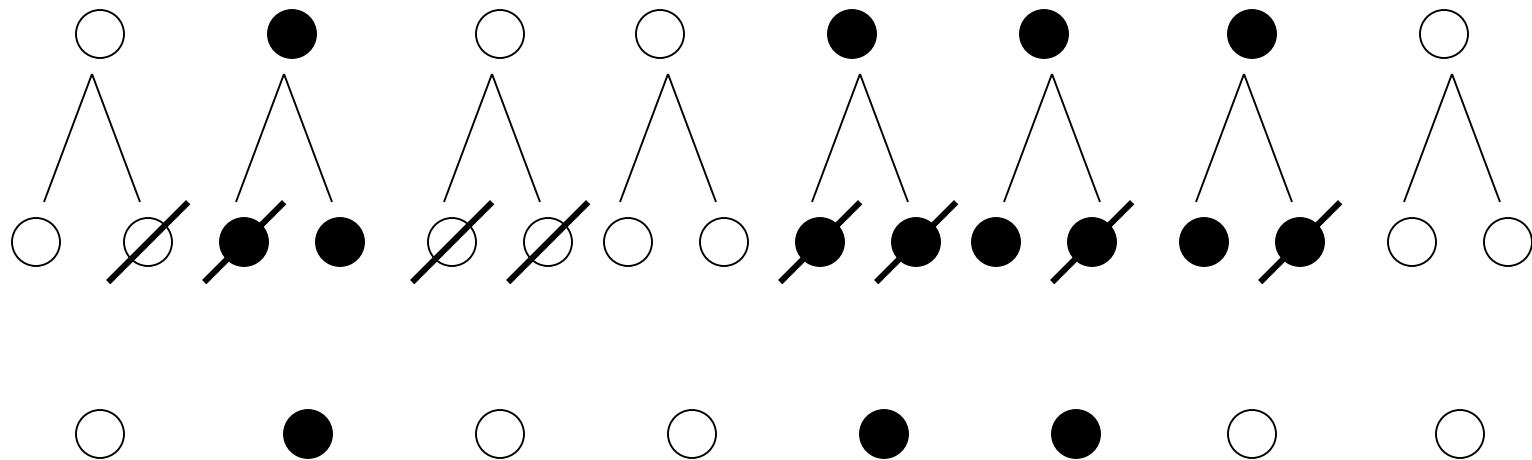
Чисто нейтральная игра

1. Имеется популяция черных и белых шаров, общее количество шаров в популяции равно n .
2. Эволюция состоит из последовательности поколений. Каждое поколение состоит из двух шагов.
 - 2.1. На первом шаге дублируем все шары, сохраняя их цвета: черный шар имеет два черных потомка, белый шар имеет два белых потомка.
 - 2.2. На втором шаге мы случайным образом (независимо от цвета) удаляем из популяции ровно половину шаров.

Игра определяет Марковский процесс, для которого показано, что:

- 1) рассматриваемый процесс всегда сходится к одному из поглощающих состояний: все шары белые либо все шары черные.
- 2) при больших n характерное число поколений T_n , требуемое для сходимости к какому-либо из поглощающих состояний, равно $2n$

Чисто нейтральная игра



Чисто нейтральная игра

Популяция находится в l -состоянии, если число черных и белых шаров равны l и $n - l$.

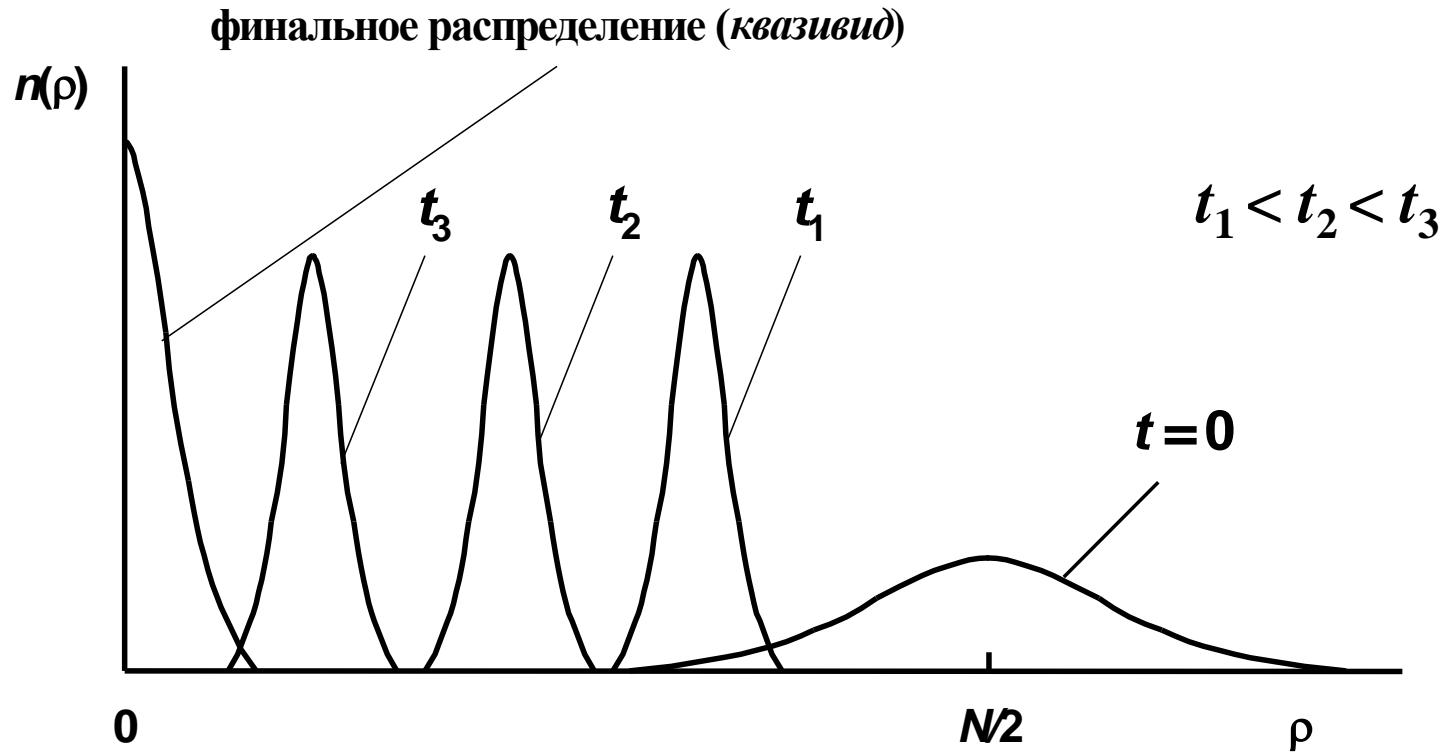
Вероятности переходов P_{lm} есть:

$$P_{lm} = \begin{cases} \binom{2l}{m} \times \binom{2n-2l}{n-m} \Big/ \binom{2n}{n}, & \text{если } 2l-n \leq m \leq 2l \\ 0, & \text{если } m > 2l \text{ или } m < 2l-n \end{cases}$$

Матрица P_{lm} задает Марковский процесс, для которого известно, что при больших n характерное время сходимости T_n к какому-либо из поглощающих состояний равно $2n$:

$$T_n = 2n.$$

Качественная схема эволюции



Оценка скорости эволюции определяется интенсивностью мутаций.

Предположения модели: $N, n \gg 1, \beta \sim P_m N, P_m N \sim 1, 2^N \gg n$

Аналитическая оценка скорости эволюции

Предполагаем, что **нейтральным отбором можно пренебречь**:

$$T_n > \sim T, \quad (2)$$

T – характерное время всей эволюции, $T_n \sim n$.

Характерное время t_{-1} , за которое среднее по популяции расстояние до оптимума $\langle \rho \rangle$ уменьшается на 1, составляет:

$$t_{-1} \sim t_m + t_{\text{от}}, \quad t_m \sim (NP_m)^{-1}, \quad t_{\text{от}} \sim \beta^{-1}$$

t_m – характерное время мутаций, $t_{\text{от}}$ – характерное время отбора.

Отсюда для $T \sim t_{-1} N$ имеем:

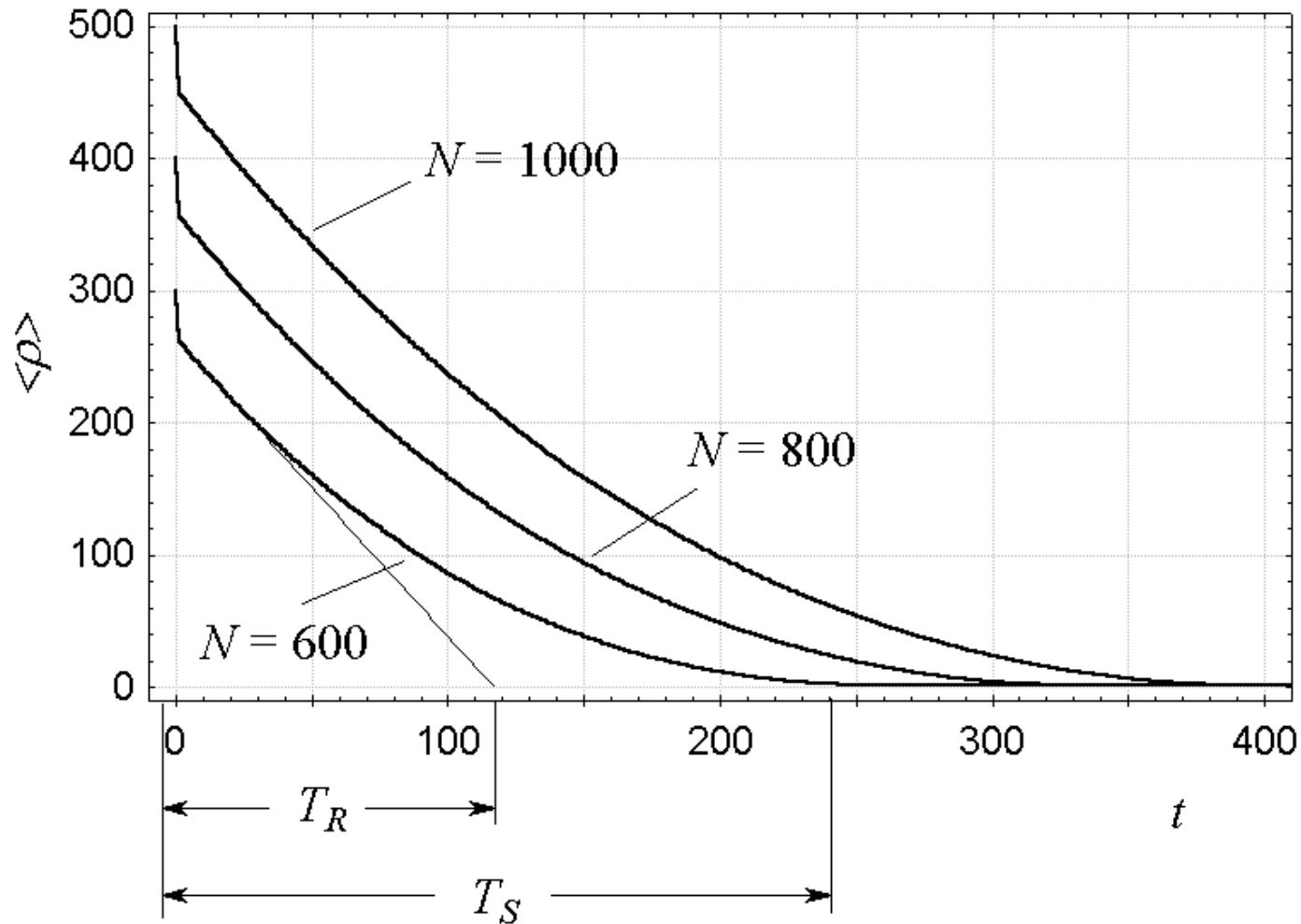
$$T \sim (P_m)^{-1} + N\beta^{-1}. \quad (3)$$

Считаем отбор достаточно интенсивным: $T \sim (P_m)^{-1}$ и мутации «оптимальными» (одна мутация на генотип) $P_m \sim N^{-1}$. Тогда $T \sim N$.

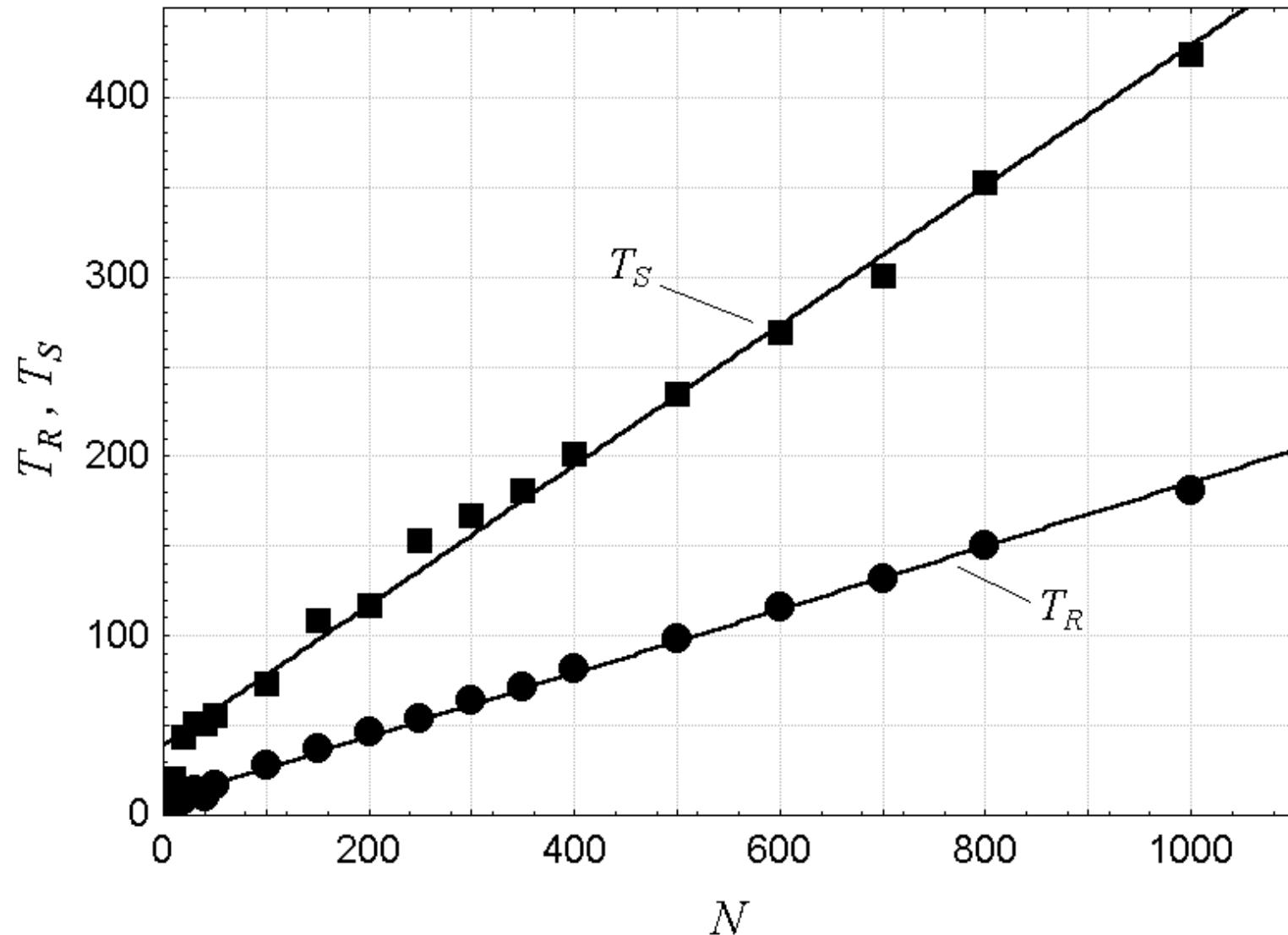
Пусть (2) выполняется на пределе $T_n \sim T$, тогда $n \sim T_n \sim T \sim N$.

Общее число особей, участвующих в эволюции, равно $n T \sim N^2$

Зависимость среднего расстояния до оптимума от номера поколения ($n = N$, $P_m = N^{-1}$, $\beta = 1$), (Ю.Р. Цой, 2005)



Зависимости времени релаксации T_R и времени выхода на стационар T_S от N (Ю.Р. Цой, 2005)



Близкие эволюционные модели

1. Модель узкого канала. Идти к оптимуму можно только последовательно меняя по одному символу. Например:

(0000) → (1000) → (1100) → (1110) → (1111)

2. Спин-стекольная модель эволюции

Приспособленность определяется энергией спинового стекла:

$$f(\mathbf{S}_k) = \exp[-\beta E(\mathbf{S}_k)], \quad E(\mathbf{S}) = - \sum_{i,j=1; i < j}^N J_{ij} S_i S_j, \quad S_i = +1, -1.$$

Оценки скорости и эффективности эволюции для этих моделей близки к оценкам для рассмотренной модели квазивидов с Хемминговой мерой близости.

Итог оценок

Число поколений: $T \sim N$

Численность популяции: $n \sim N$

Общее число особей, участвующих эволюционном поиске:

$$n_{\text{общ}} = n T \sim N^2$$

Оценки были сделаны при разумном выборе параметров:

1) $\beta > \sim NP_m$ – интенсивность отбора достаточно велика

2) $P_m \sim N^{-1}$ – мутации «оптимальны» (одна мутация на генотип за поколение), $\Rightarrow \beta > \sim 1$

3) $n \sim N$ – условие пренебрежения нейтральным отбором выполняется «на пределе»

Связь с биологической эволюцией

С биологической точки зрения модель квазивидов может рассматриваться только как **очень грубое приближение**. Тем не менее, сопоставим цифры.

Для бактерии *E-coli* число нуклеотидов равно $4 \cdot 10^6$, т.е. порядка $N = 10^7$ бит, для человека число нуклеотидов равно $3 \cdot 10^9$, т.е. порядка $N = 10^{10}$ бит.

Жизнь бактерии ~ 1 час.

«Оптимальную» бактерию в популяции 10^7 особей можно найти за 1000 лет.

Жизнь млекопитающего ~ 1 год.

«Оптимального» млекопитающего в популяции 10^{10} особей можно найти за 10^{10} лет.

Полученные оценки дают аргументы в пользу возможности происхождения организмов путем дарвиновской эволюции.

Основные публикации по оценкам скорости эволюции

**Редько В.Г. Оценка скорости эволюции в моделях Эйгена и
Куна // Биофизика. 1986. Т. 31. №3. С. 511-516.**

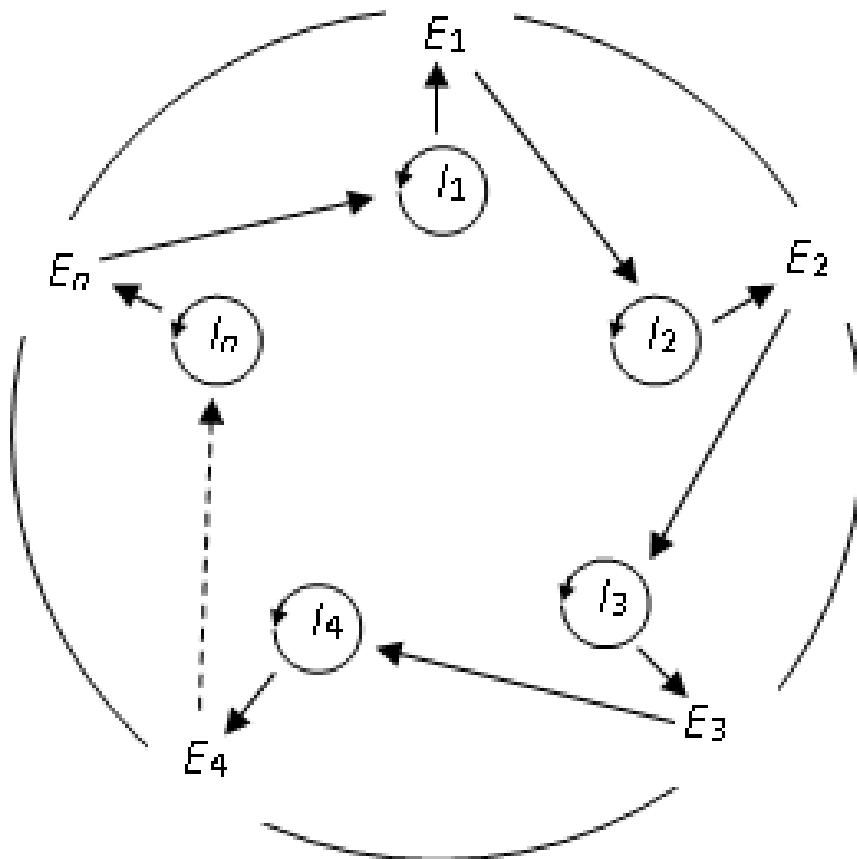
**Редько В.Г., Цой Ю.Р. Оценка эффективности
эволюционных алгоритмов // Доклады АН, 2005. Т. 404. №. 3.
С. 312-315.**

**Редько В.Г., Цой Ю.Р. Оценка скорости и эффективности
эволюционных алгоритмов // Бионические
информационные системы и их практические применения.
(Под ред. Зинченко Л.А., Курейчика В.М., Редько В.Г.). М.:
Физматлит, 2011. С. 109-126.**

Модель гиперциклов: РНК и ферменты кооперируются

Эйген М., Шустер П. Гиперцикл. Принципы самоорганизации макромолекул. М.: Мир, 1982.

Структура гиперцикла



Циклическая организация гиперцикла обеспечивает его структурную стабильность

I_i – РНК матрицы,
 E_i – ферменты
репликации.

**Ферменты
циклически
катализируют
репликацию РНК**

**Макромолекулы
кооперативно
обеспечивают
примитивную
трансляцию**

Модель гиперцикла

Для эффективной конкуренции различные гиперциклы должны быть помещены в отдельные компартменты. Это обеспечивает отбор гиперциклов в соответствии с их селективными ценностями. Иначе селекция определяется начальными условиями, а не катализитическими свойствами макромолекул гиперцикла

Эйген М., Шустер П. Гиперцикл. Принципы самоорганизации макромолекул. М.: Мир, 1982

Файстель Р., Романовский Ю.М., Васильев В.А. Эволюция гиперциклов Эйгена, протекающих в коацерватах // Биофизика. 1980. Т.25. N.5. С. 882-887.

Модель гиперцикла

Полагаем, что 1) гиперциклы помещены в коацерватные капли; 2) каждый коацерват включает только один тип гиперциклов; 3) объем любого коацервата пропорционален числу макромолекул внутри данной коацерватной капли; 4) процесс трансляции значительно быстрее процесса репликации (последнее означает, что достаточно проанализировать только динамику РНК, а динамика ферментов будет подстраиваться под динамику РНК).

Тогда:

$$dN_i/dt = Vf_i, \quad V = c^{-1}\sum_i N_i, \quad x_i = N_i/V, \quad i = 1, 2, \dots, n,$$

N_i и x_i – число молекул i -й РНК и ее концентрация в коацервате; V – объем коацервата; $c = \sum_i x_i$; f_i – скорость синтеза i -й РНК.

Модель гиперцикла

Скорости синтеза РНК:

$$f_i = a x_i x_j, \quad i = 1, 2, \dots, n, \quad j = i - 1 + \delta_{i1},$$

Динамика концентраций РНК:

$$dx_i/dt = a x_i x_j - x_i c^{-1} \sum_l f_l, \quad f_i = a x_i x_j, \quad i = 1, 2, \dots, n, \quad j = i - 1 + n \delta_{i1}$$

$$dV/dt = V c^{-1} \sum_l f_l$$

Предельные концентрации РНК: $x_i^0 = c/n$, или $\langle x_i \rangle = c/n$

Выживают гиперциклы с наибольшими W_k

Если есть несколько гиперциклов и их ферменты имеют различные катализитические способности, то объемы их коацерватов увеличиваются пропорционально селективным ценностям гиперциклов:

$$W_k = \langle c_k^{-1} \Sigma_l f_{kl} \rangle$$

Предполагая, 1) что каждый коацерват расщепляется на две части, когда его объем превышает некоторую критическую величину, и 2) что суммарный объем коацерватов всех типов гиперциклов постоянен V_T , получим:

$$dV_k/dt = W_k V_k - E V_k, \quad E = V_T^{-1} \Sigma_k W_k V_k$$

Гиперциклы или сайзеры?

Завершая книгу «Гиперцикл», М. Эйген и П. Шустер приводят цитату из Платона: **«И еще: "Если кто-нибудь может назвать более прекрасный треугольник, лежащий в основе вещей, мы будем приветствовать его не как соперника, но как друга правды"»**

Перейдем к рассмотрению другого «треугольника» – модели сайзеров, которая была предложена В.А. Ратнером и В.В. Шаминым, как альтернатива модели гиперциклов

Модель сайзеров

Конвергенция идей

В начале 1980-х годов несколько авторов независимо предложили фактически одну и ту же модель

Ратнер В.А., Шамин В.В. Сайзеры: моделирование фундаментальных особенностей молекулярно-биологической организации // *Математические модели эволюционной генетики*. Новосибирск: ИЦИГ, 1980. С. 66 - 126.

White D.H. A theory for the origin of a self-replicating chemical system. I: Natural selection of the autogen from short random oligomers // *Journal of Molecular Evolution*, 1980. V.16. N.2. PP. 121-147.

Feistel R. On the evolution of biological macromolecules. Precellular organization. 4. Holobiotic competition // *Studia biophysica*. 1983. V.93. N.2. PP. 113-128.

Авторы модели сайзеров



Ратнер

Вадим Александрович

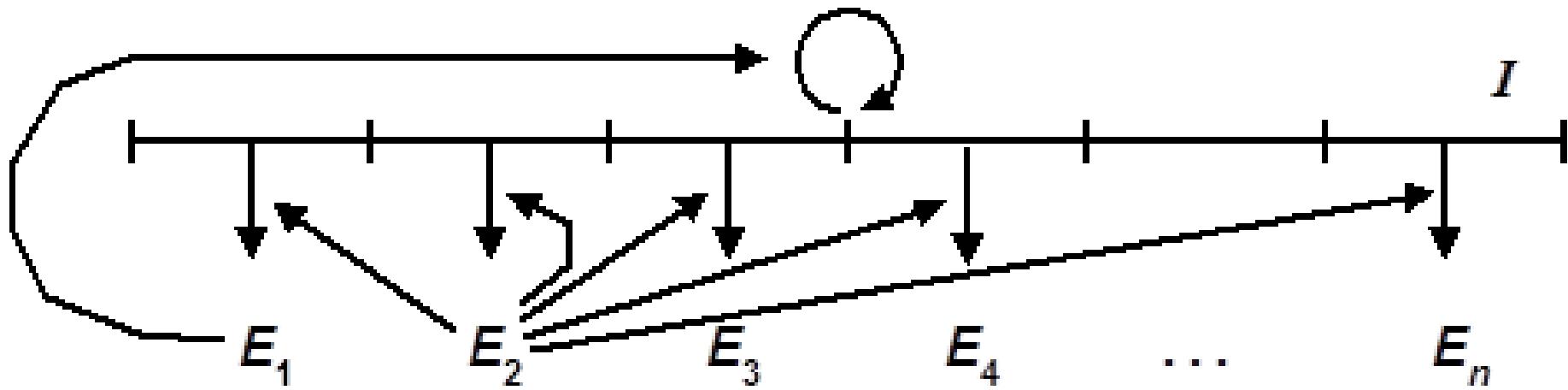
1932 – 2002

**Институт цитологии и
генетики РАН,
Новосибирск**

Rainer Feistel

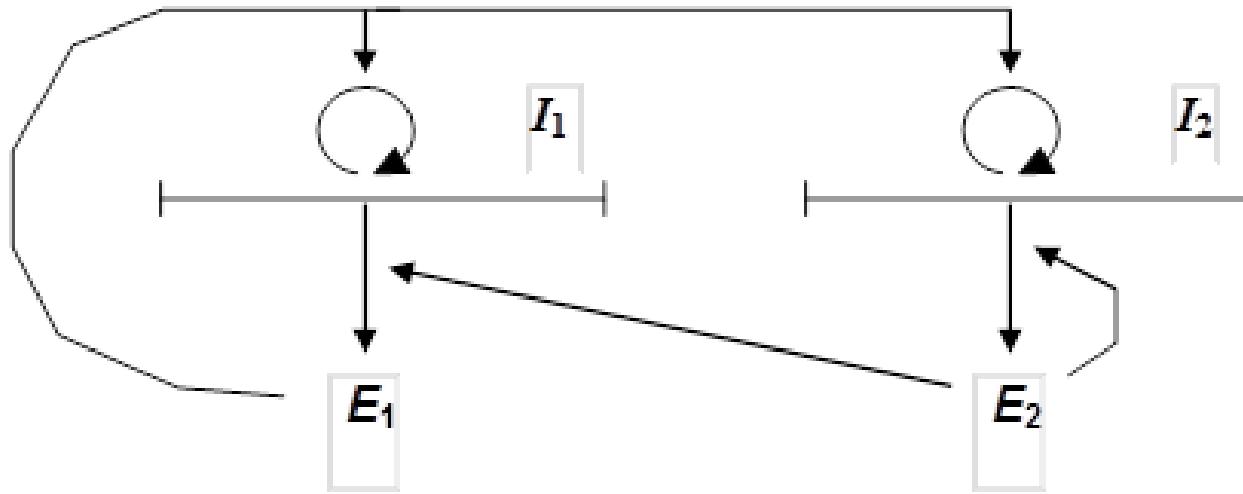
**Leibniz Institute for
Baltic Sea Research,
Warnemünde
(near Rostok)**

Общая схема сайзеров



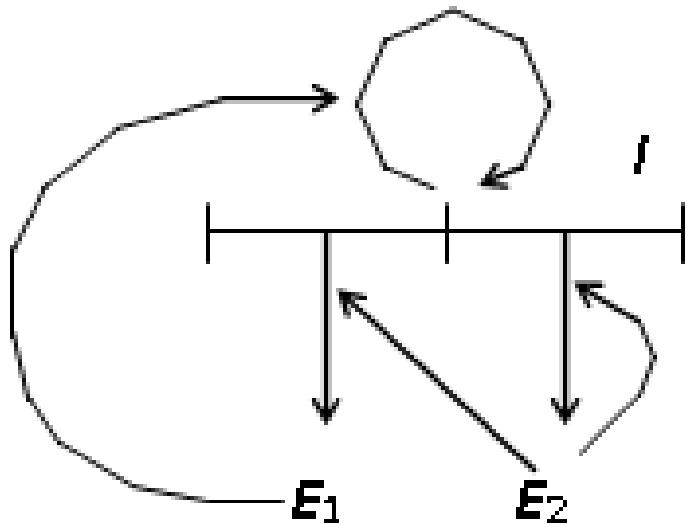
I – полинуклеотидная матрица, E_1 – фермент репликации, E_2 – фермент трансляции

Структурно неустойчивый сайзер



Для структурной устойчивости такого сайзера скорости репликации полинуклеотидных матриц должны быть строго равны друг другу, иначе концентрация одной из матриц будет все время уменьшаться по сравнению с другой

Сайзеры в гомогенной среде – «эволюционный застой»



Для простоты рассматриваем мини-сайзер, динамика макромолекул такова:

$$dx_k/dt = \sum_l a_l y_l x_k - \varphi_I x_k$$

$$dy_k/dt = \sum_l b_l z_l x_k - \varphi_E y_k$$

$$dz_k/dt = \sum_l b_l z_l x_k - \varphi_E z_k$$

$$\sum_k x_k = c_I = \text{const} ,$$

$$\sum_k (y_k + z_k) = c_E = \text{const}$$

x_k – концентрации матриц,

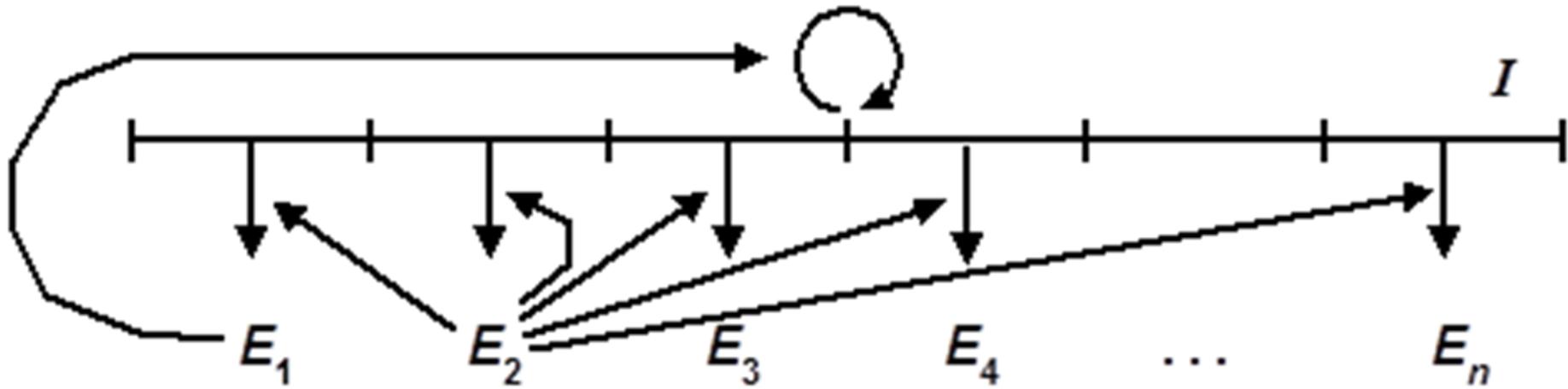
y_k, z_k – концентрации ферментов

Так как ферменты репликации всех видов сайзеров универсальны, то относительные концентрации полинуклеотидных матриц разных видов не меняются, а концентрации ферментов становятся пропорциональными концентрациям матриц

Нет отбора лучших сайзеров

Сайзеры в коацерватах – «эволюционный прогресс»

Как и для гиперциклов, используем метод, предложенный Файстелем, Романовским и Васильевым



- 1) Сайзеры помещены в коацерваты, каждый из которых есть капля, имеющая полупрозрачную оболочку: оболочка прозрачна для малых молекул (АТФ), но непрозрачна для макромолекул. 2) Каждый коацерват включает только один тип сайзеров. 3) Объем коацервата пропорционален числу макромолекул в нем.

Динамика концентраций

$$\frac{dN_i}{dt} = Vf_i, \quad V = \frac{1}{c} \sum_{i=0}^n N_i, \quad x_i = \frac{N_i}{V}, \quad i = 0, 1, 2, \dots, n,$$

$$f_0 = a_0 x_0 x_1, \quad f_i = a_i x_0 x_2, \quad i = 1, \dots, n,$$

N_i и x_i – число и концентрация макромолекул i -го типа; V – объем коацервата; c – суммарная постоянная концентрация макромолекул, индекс $i = 0$ относится к матрице I , остальные i – к ферментам

$$\frac{dx_i}{dt} = f_i - \frac{x_i}{c} \sum_{l=0}^n f_l, \quad i = 0, \dots, n, \quad \frac{dV}{dt} = \frac{V}{c} \sum_{l=0}^n f_l.$$

Анализ этой системы показывает, что концентрации макромолекул сходятся к устойчивой особой точке

Селективные ценности сайзеров

Координаты особой точки:

$$x_0^0 = \frac{ca_0a_1}{a_0a_1 + a_2(a_1 + \dots + a_n)}, \quad x_i^0 = \frac{ca_i a_2}{a_0a_1 + a_2(a_1 + \dots + a_n)}, \quad i = 1, \dots, n.$$

Скорость роста объемов коацерватов сайзера пропорциональна

$$W = \frac{1}{c} \sum_{l=0}^n f_l \quad W^0 = \frac{ca_0a_1a_2}{a_0a_1 + a_2(a_1 + \dots + a_n)}$$

Величина W определяется скоростями синтеза макромолекул

Чем больше W , тем больше преимущества данного сайзера перед остальными

Конкуренция сайзеров

Для анализа конкуренции сайзеров в явном виде считаем, что 1) каждый коацерват делится пополам, когда его объем превышает некоторую критическую величину и 2) полный объем всех коацерватов постоянен. Тогда для k -го сайзера имеем:

$$\frac{dV_k}{dt} = W_k V_k - EV_k \quad E = \frac{1}{V_T} \sum_k W_k V_k$$

Эти уравнения совпадают с уравнениями модели квазивидов в отсутствии мутаций

Выживает сайзер с наибольшей селективной ценностью W_k

Сайзеры и самовоспроизводящиеся автоматы

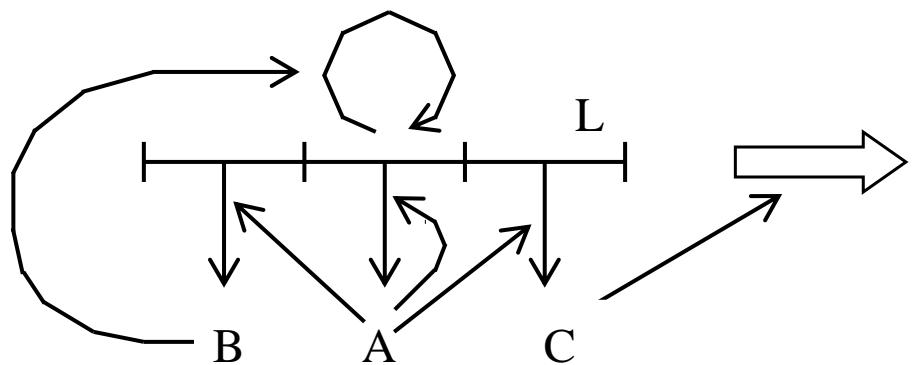
Дж. фон Неймана

Дж. фон Нейман предложил и исследовал модель самовоспроизводящихся автоматов. Эти автоматы состоят из следующих основных компонент:

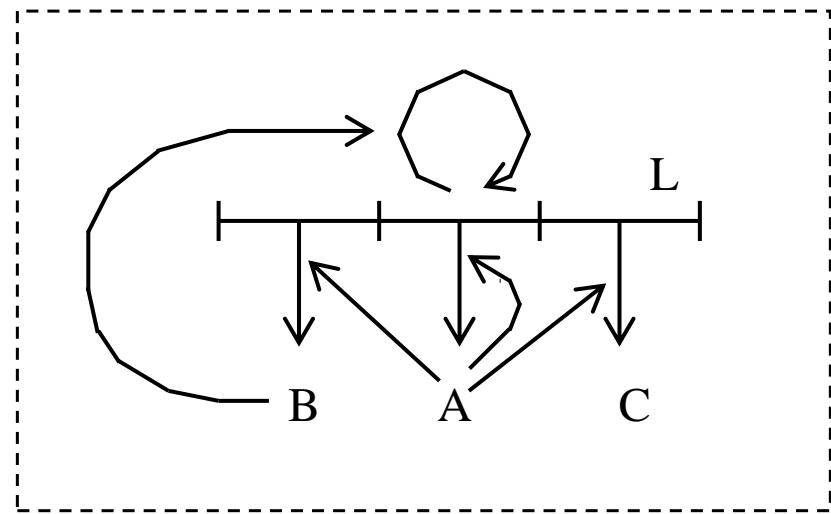
- 1) лента L , хранящая информацию
- 2) автомат A , предназначенный для изготовления произвольного автомата согласно информации, закодированной в ленте L
- 3) автомат B , предназначенный для копирования ленты L
- 4) автомат C , координирующий процесс отделения нового изготовленного автомата-потомка от автомата-родителя

Могут быть и дополнительные автоматы, конструкция которых кодируется лентой L

Сайзеры и самовоспроизводящиеся автоматы Дж. фон Неймана



Автомат-родитель



Автомат-потомок

Сайзеры и самовоспроизводящиеся автоматы

Дж. фон Неймана

Самовоспроизводящиеся автоматы Дж. Фон Неймана	Сайзеры
Запоминающая лента L	Полинуклеотидная матрица <i>I</i>
Автомат A , предназначенный для изготовления произвольного автомата согласно информации, закодированной в ленте L	Фермент трансляции E_2
Автомат B , копирующий ленту L	Фермент репликации E_1
Автомат C , координирующий процесс отделения автомата- потомка от автомата-родителя	Деление коацерватов в процессе роста сайзеров

Адаптивный сайзер

Редько В.Г. Адаптивный сайзер // Биофизика. 1990. Т.35. № 6.
С.1007-1011

Адаптивный сайзер

Рассматриваем схему адаптивного модельного организма, имеющего макромолекулярную систему управления, которая «включает» и «выключает» синтез определенного фермента в ответ на изменения внешней среды.

Схема этой молекулярной адаптации подобна классической схеме регулирования синтеза белков в бактериях E. Coli по Ф. Жакобу и Ж. Моно.

Jacob F., Monod J. Genetic regulatory mechanisms in the synthesis of proteins // Journal of Molecular Biology. 1961. Vol. 3. P. 318-356.

Чем хорош адаптивный сайзер

Это модель простейшей системы управления, которая могла возникнуть на предбиологическом уровне или на очень ранних этапах биологической эволюции.

Рассмотрена схема адаптивного модельного организма, имеющего макромолекулярную систему управления, которая «включает» и «выключает» синтез определенного фермента в ответ на изменения внешней среды; схема этой адаптации подобна классической схеме регулирования синтеза белков в бактериях по Ф. Жакобу и Ж. Моно.

Такая система управления могла быть одной из первых, возникших в процессе эволюции в биосфере Земли.

Биологический прототип

Транскрипция происходит, когда нет глюкозы и есть лактоза

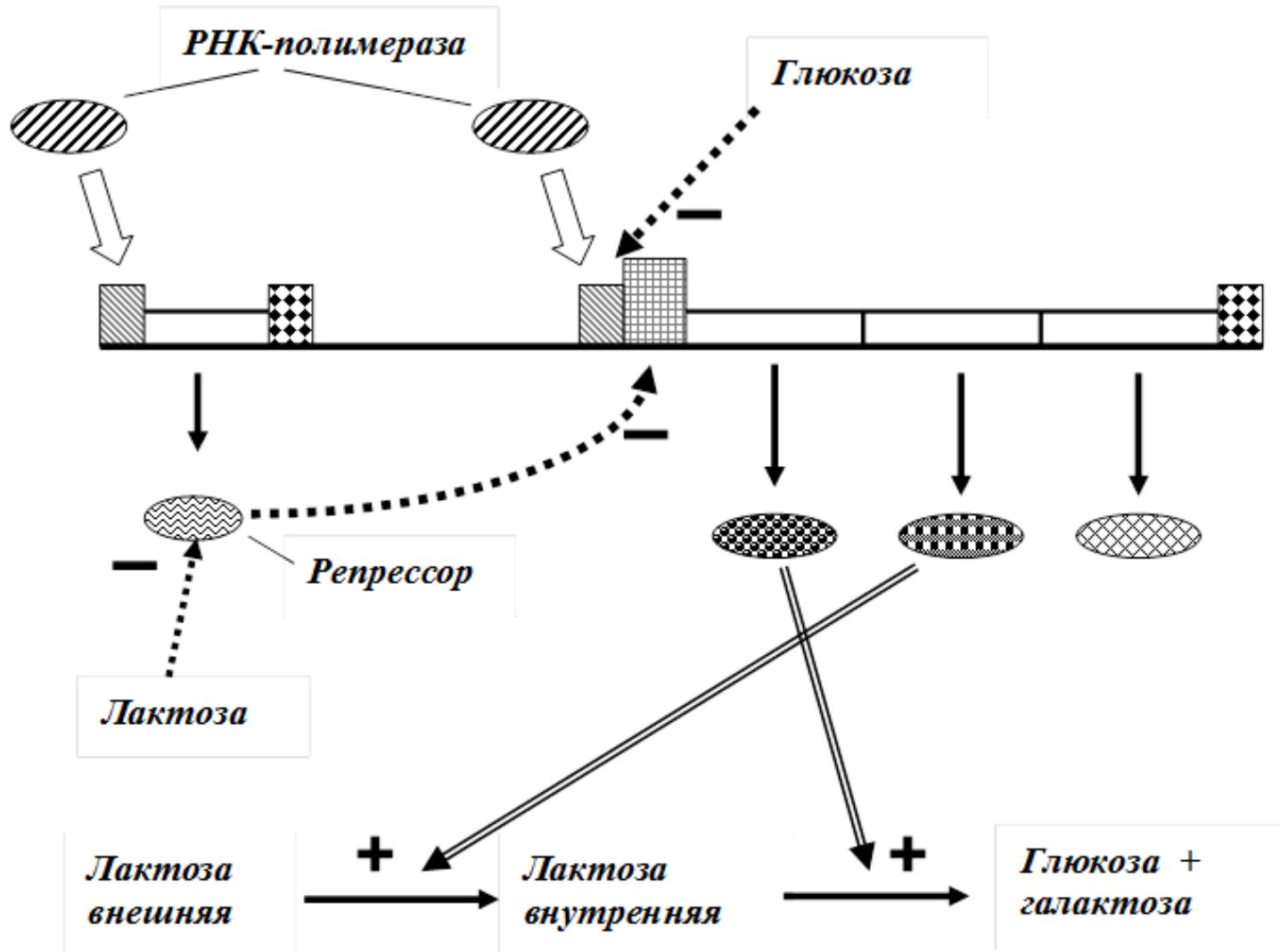
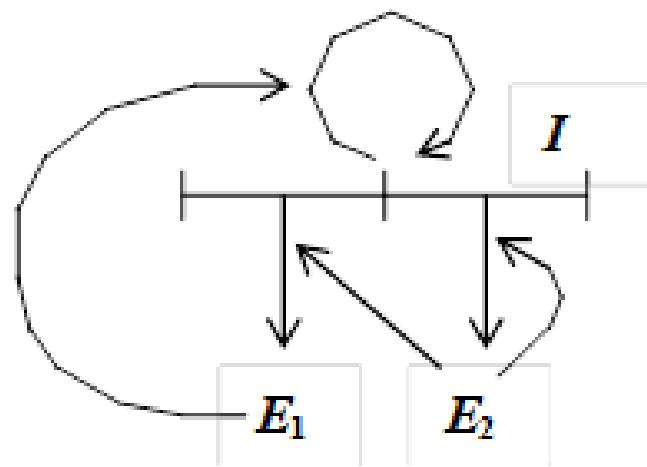
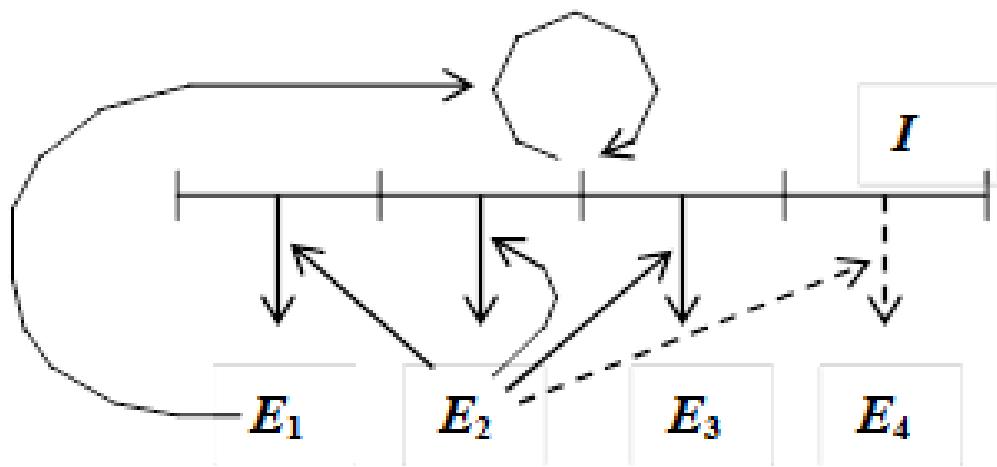


Схема адаптивного сайзера



Мини-сайзер



Адаптивный сайзер

I – полинуклеотидная матрица, E_1 и E_2 – ферменты репликации и трансляции

E_3 – фермент-регулятор, E_4 – фермент-адаптер

Фермент-регулятор E_3 отслеживает состояние внешней среды и включает или выключает синтез фермента-адаптера E_4 в соответствии с изменениями окружающей среды

Схема управления адаптивного сайзера

Имеется два возможных состояния внешней среды: среда *A* и среда *B*

Среда *A* – обычная, в которой оба сайзера обладают способностью самовоспроизведения

Среда *B* – измененное состояние, в котором макромолекулярный синтез имеет место только в адаптивном сайзере и только при наличии фермент-адаптера E_4

Фермент-регулятор E_3 в адаптивном сайзере синтезируется постоянно с малой интенсивностью в обеих средах и включает синтез адаптера E_4 в среде *B* и выключает этот синтез в окружающей среде *A*. Адаптер E_4 обеспечивает макромолекулярный синтез в адаптивном сайзере в среде *B*

Конкуренция мини-сайзера и адаптивного сайзера

Какой сайзер будет выживать в переменной среде?

Адаптивный сайзер имеет селективные преимущества по сравнению с мини-сайзером в окружающей среде B .

Для переменной среды ситуация сложнее, так как адаптивный сайзер должен синтезировать дополнительный фермент E_3 и это – дополнительная нагрузка для адаптивного сайзера по сравнению с мини-сайзером. Более того, адаптивному сайзеру необходимо определенное время для релаксации от одной среды к другой, так что конкуренция сайзеров должна зависеть и от частоты смены сред. Поэтому необходим анализ конкуренции сайзеров.

Конкуренция мини-сайзера и адаптивного сайзера

Анализ конкуренции был произведен путем анализа системы уравнений, определяющих динамику концентраций макромолекул в сайзерах и динамику объемов соответствующих коацерватов. Это было сделано для медленной либо частой смены сред А и В (характерное время смены сред значительно больше либо значительно меньше времени релаксации концентраций к равновесным значениям).

Динамика макромолекул

$$\frac{dx_i}{dt} = f_i - \frac{x_i}{c} \sum_{l=0}^n f_l, \quad i = 0, \dots, n, \quad \frac{dV}{dt} = \frac{V}{c} \sum_{l=0}^n f_l.$$

Мини-сайзер:

$$f_0 = a_0 x_0 x_1, \quad f_i = a_i x_0 x_2, \quad i = 1, 2, \quad \text{в среде } A$$

$$f_0 = f_1 = f_2 = 0, \quad \text{в среде } B$$

Адаптивный сайзер:

$$f_0 = a_0 x_0 x_1, \quad f_i = a_i x_0 x_2, \quad i = 1, 2, 3, \quad \text{в среде } A$$

$$f_0 = b_0 x_0 x_1, \quad f_i = b_i x_0 x_2, \quad i = 1, 2, 3, 4, \quad \text{в среде } B$$

Конкуренция мини-сайзера и адаптивного сайзера

$$\frac{dV_{lm}}{dt} = W_{lm} V_{lm} - \frac{W_{1m} V_{1m} + W_{2m} V_{2m}}{V_T} V_{lm}$$

l ($=1,2$) – тип сайзера; m ($=A,B$) – тип среды. $V_T = V_{1m} + V_{2m}$

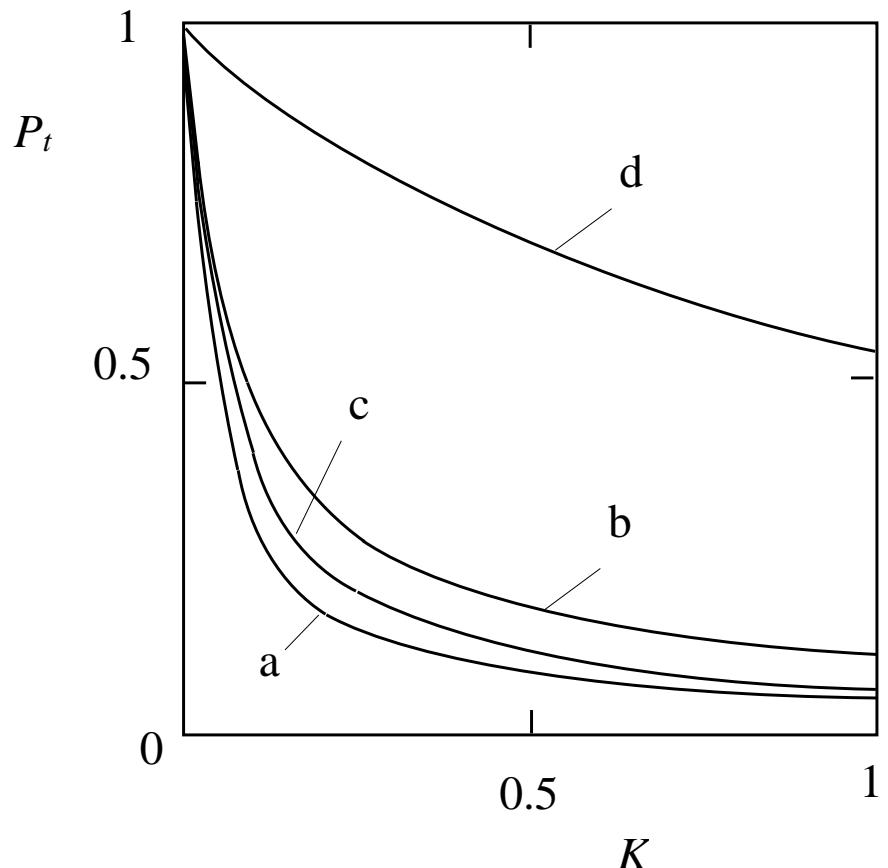
Рассматривались

- 1) редкая смена сред – длительности сред T_A, T_B намного больше характерных времен τ_A, τ_B сходимости к равновесным концентрациям: $T_A \gg \tau_A, T_B \gg \tau_B$
- 2) частая смена сред: длительности сред $T_A \ll \tau_A, T_B \ll \tau_B$

Сравнение мини-сайзера и адаптивного сайзера

В постоянной среде А выживает мини-сайзер, так как в адаптивном постоянно синтезируется фермент-регулятор E_3

В постоянной среде В выживает адаптивный сайзер



В переменной среде
адаптивный сайзер выживает
выше соответствующей кривой

P_t — пороговая частота среды B

$K = b_i/a_i$.

Редкая (a, b) либо частая (c, d) смена
сред.

Малая (a,c) либо большая (b,d)
скорость синтеза фермента-
регулятора E_3

Резюме по модели адаптивного сайзера

Адаптивный сайзер имеет селективные преимущества по сравнению с мини-сайзером, однако, не всегда, а только тогда, когда «расходы» (необходимость постоянного синтеза фермента-регулятора E_3), требуемые для обеспечения функционирования системы управления, достаточно малы

Новое «изобретение» эволюции полезно не всегда, а только тогда, когда расходы на поддержание работы этого «изобретения» меньше, чем польза от этого изобретения

Рассмотренная система управления могла быть первой системой управления, возникшей в процессе эволюции

Гипотетический процесс предбиологической эволюции

Квазивиды → Гиперциклы → Сайзеры →
→ Протоклетки → Простейшие организмы

Модель взаимодействия между обучением и эволюцией

План этой части доклада

- 1. Предпосылки моделирования взаимодействия между обучением и эволюцией**
- 2. Модели предшественников (Hinton, Nowlan, 1987; Mayley, 1997)**
- 3. Исследованные механизмы взаимодействия:**
 - 3.1. Механизм генетической ассимиляции**
 - 3.2. Эффект экранирования**
 - 3.3. Влияние нагрузки на обучение**
- 4. Сопоставление с моделью Hinton & Nowlan**

Предпосылки

После появления дарвиновской теории эволюции возник вопрос: **могут ли за счет случайного мутационного поиска возникать весьма нетривиальные полезные навыки живых организмов?**

Появились концепции: обучение может существенно способствовать эволюционному процессу (эффект Болдуина).

Baldwin J.M. A new factor in evolution. American Naturalist. **1896.** V. 30. P. 441-451.

Morgan C.L. On modification and variation. Science. **1896.** V. 4. P. 733-740.

Osborn H.F. Ontogenetic and phylogenetic variation. Science. **1896.** V. 4. P. 786-789.

Предпосылки

Возможный механизм влияния обучения на эволюцию

Имеются два этапа

- 1) Эволюционирующие организмы (благодаря мутациям) приобретают **свойство обучаться некоторому полезному навыку**. Приспособленность таких организмов увеличивается, они распространяются по популяции. Но обучение имеет свои недостатки, так как оно требует энергии и времени.
- 2) Поэтому возможен второй этап, который называют *генетической ассимиляцией*: приобретаемый полезный навык может быть «повторно изображен» генетической эволюцией, в результате чего он записывается непосредственно в генотип и становится наследуемым. Второй этап длится множество поколений.

Таким образом, полезный навык, который первоначально был приобретаемым, может стать наследуемым, хотя эволюция имеет Дарвиновский характер.

Предшественники

Был ряд моделей, анализирующих взаимодействие между обучением эволюцией.

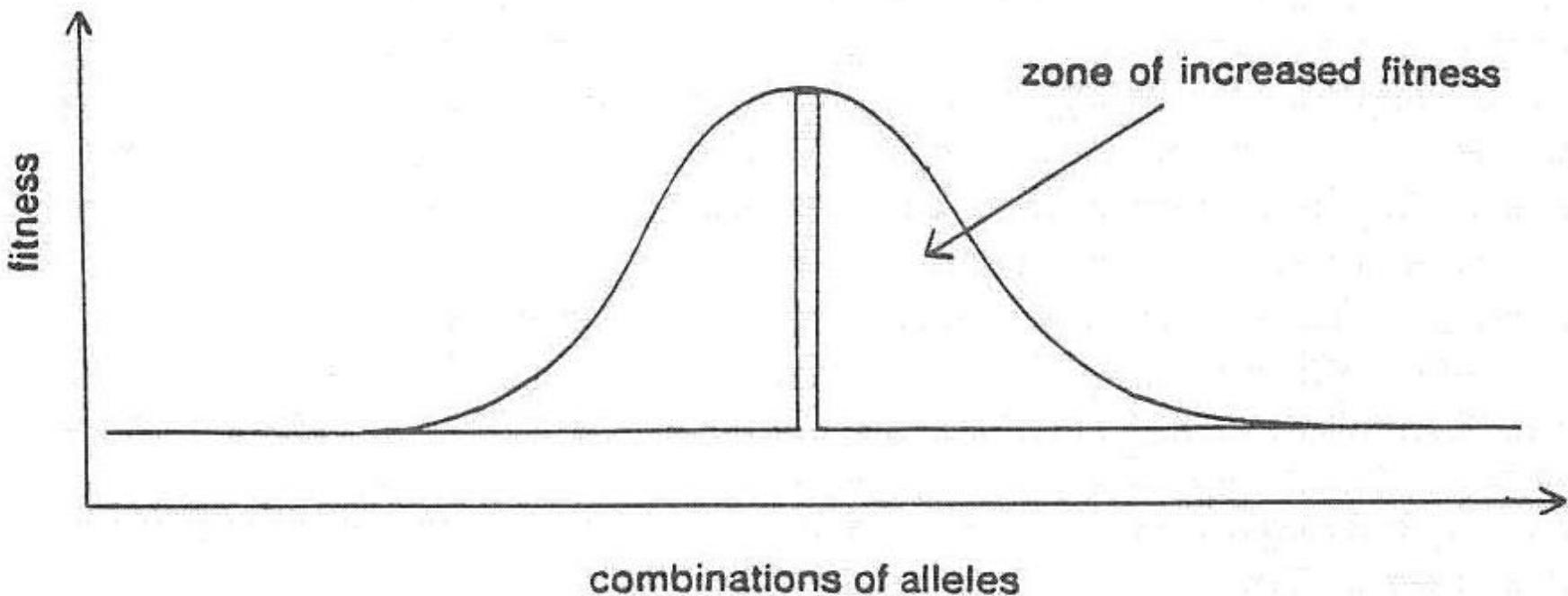
Hinton G.E., Nowlan S.J. How learning can guide evolution // Complex Systems. 1987. V. 1. P. 495-502.

Mayley G. Guiding or hiding: Explorations into the effects of learning on the rate of evolution // Proceedings of the Fourth European Conference on Artificial Life. 1997. P. 135-144.

Однако, эти модели использовали довольно сложный генетический алгоритм со скрещиванием генотипов, поэтому были продемонстрированы эффекты взаимодействия, но плохо раскрыты механизмы эффектов.

В настоящей модели используется четкая схема эволюции, основанная на отборе и мутациях – модель квазивидов (М. Эйген, Редько В.Г., Цой Ю.Р.)

Hinton G.E., Nowlan S.J. How learning can guide evolution



Генотип имеет длину $N = 20$, аллели равны 0 либо 1. Имеется единственный оптимальный генотип G_{max} (вертикальная полоска). Обучение происходит путем проб и ошибок. Если при обучении особи произошло приближение к G_{max} , то приспособленность ее возрастает, особь рождает потомков.

Mayley G. Guiding or hiding

Схема генетической асимиляции

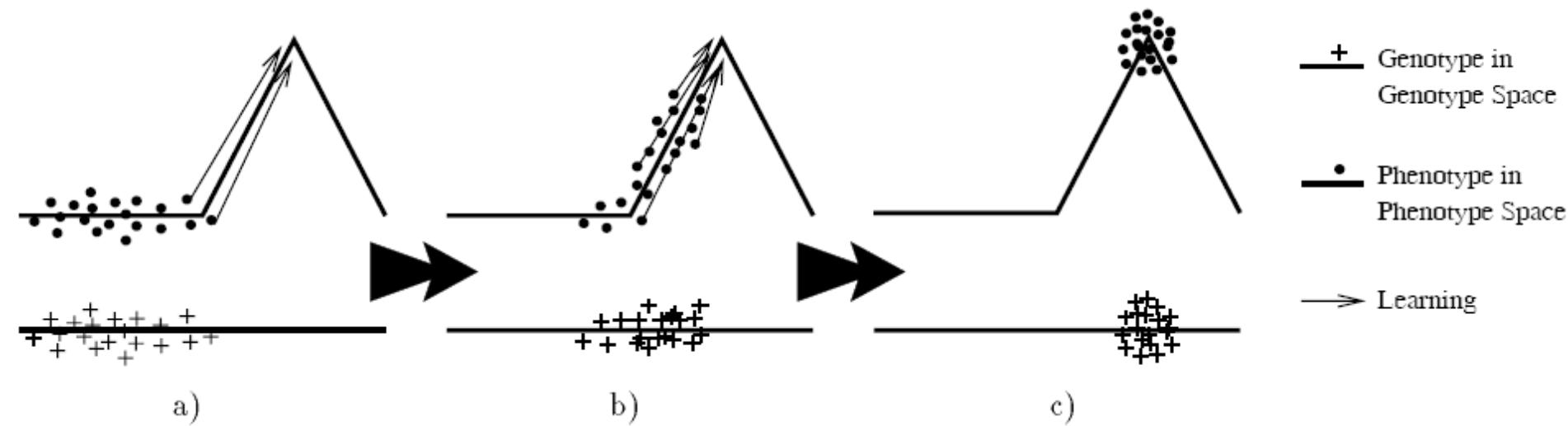
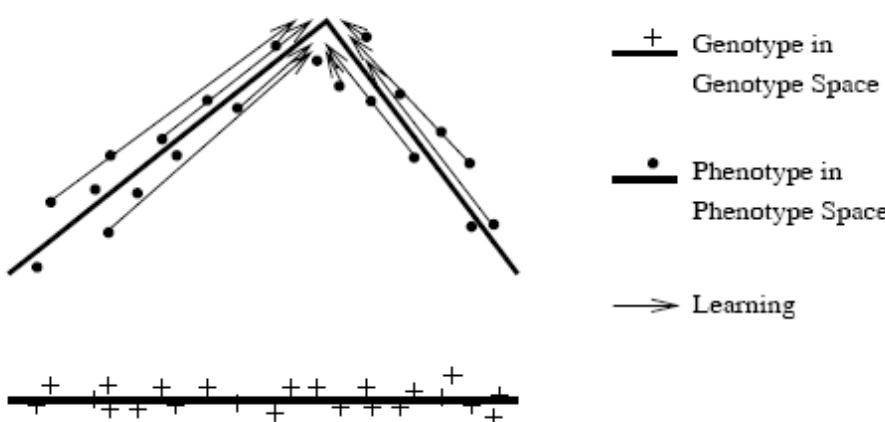


Схема экранирования



При сильном обучении
хороший фенотип
находится независимо
от генотипа,
отбора генотипов
не происходит

Каноническая модель эволюции: модель квазивидов (М. Эйген, Г. Кун, Дж. Эдельман)

Рассматривается **Дарвиновская эволюция** популяции особей (цепочек символов, аналогов ДНК) S_1, S_2, \dots, S_n .

$$S_{ki} = 0, 1; \quad k = 1, 2, \dots, n; \quad i = 1, 2, \dots, N.$$

Длина цепочек N и численность популяции n велики:
 $N, n \gg 1$. $N = \text{const}$, $n = \text{const}$.

Имеется **оптимальная цепочка S_M** , имеющая максимальную приспособленность. Приспособленность произвольной особи S определяется расстоянием по Хеммингу $\rho(S, S_M)$ между S и S_M (числом несовпадающих символов в соответствующих позициях цепочек):

$$f(S) = \exp[-\beta \rho(S, S_M)], \quad (1)$$

β – интенсивность отбора, $\beta > 0$.

Достоинство модели квазивидов

Можно использовать одну существенную переменную –
расстояние до оптимума ρ

Итог оценок скорости эволюции

Число поколений: $T \sim N$

Численность популяции: $n \sim N$

Общее число особей, участвующих эволюционном поиске:

$$n_{\text{общ}} = n T \sim N^2$$

Оценки были сделаны при разумном выборе параметров:

1) $\beta > \sim NP_m$ – интенсивность отбора достаточно велика

2) $P_m \sim N^{-1}$ – мутации «оптимальны» (одна мутация на генотип за поколение), $\Rightarrow \beta > \sim 1$

3) $n \sim N$ – условие пренебрежения нейтральным отбором выполняется «на пределе»

Модель взаимодействия между обучением и эволюцией

Основная модель

Эволюционирующая популяция состоит из n особей.

Особь имеет **генотип** S_{Gk} и **фенотип** S_{Pk} , которые закодированы **одинаковыми** по форме **цепочками символов**, k – номер особи, $k = 1, 2, \dots, n$.

Длина цепочек генотипа и фенотипа N одинакова. **Символы цепочек S_{Gk} и S_{Pk} равны 0 либо 1.**

Длина цепочек и численность популяции велики: $N, n \gg 1$, $2^N \gg n$.

Имеется случайная **оптимальная цепочка S_M** той же формы. **S_M ищется в процессе обучения и эволюции.**

Пример генотипа – модельная ДНК, фенотипа – веса синапсов нейронной сети.

Основная модель

При рождении (в начале поколения) потомки наследуют (с мутациями) генотипы S_{Gk} своих родителей.

В начале поколения коды фенотипов S_{Pk} равны $S_{Pk}(t=0) = S_{Gk}$.

В течение поколения генотипы S_{Gk} не меняются.

В течение поколения фенотипы S_{Pk} модифицируются путем обучения методом проб и ошибок: случайно меняются отдельные символы, и если происходит приближение S_{Pk} к оптимуму S_M , то измененный символ сохраняется, иначе возвращается старый символ.

Отбор особей в новое поколение осуществляется в соответствии с приспособленностями особей, которые определяются фенотипами особей S_{Pk} в конце поколения.

Основная модель

Приспособленность особи определяется расстоянием по Хеммингу $\rho = \rho(S_{FPk}, S_M)$ между оптимумом S_M и фенотипом S_{Pk} в конце поколения:

$$f(S_k) = \exp[-\rho(S_{FPk}, S_M)] + \varepsilon, \quad (*)$$

где $S_{FPk} = S_{Pk}$ ($t = T$), $0 < \varepsilon \ll 1$. Слагаемое ε характеризует влияние случайных факторов окружающей среды на приспособленность особей.

Отбор осуществляется с вероятностями, пропорциональными приспособленностям особей (рулеточным методом).

Итак, потомки наследуют (с мутациями) генотипы S_{Gk} отбираемых особей, отбор осуществляется в соответствии с конечными фенотипами S_{Pk} .

Что исследовалось

- 1. Сопоставление: эволюция с обучением и чистая эволюция.**
- 2. Анализ механизма генетической ассимиляции: как приобретаемые путем индивидуального обучения навыки могут стать наследуемыми в течение ряда поколений дарвиновской эволюции?**
- 3. Анализ механизма эффекта экранирования: при сильном обучении оптимальный фенотип может быть найден независимо от генотипа.**
- 4. Анализ роли эффекта нагрузки на обучение: обучение может иметь и недостатки – оно требует времени и ресурсов.**

Учет эффекта нагрузки на обучение – приспособленность:

$$f(S_k) = \exp\{-\rho[S_{Pk}(t=0), S_{Pk}(t=T)]\} \{ \exp[-\rho(S_{FPk}, S_M)] + \varepsilon \} \quad (**)$$

Параметры расчетов

Длина генотипа и фенотипа $N = 100$

Численность популяции $n = N = 100$

Интенсивность мутаций $P_m = N^{-1} = 0.01$

Параметр стохастичности среды $\epsilon = 10^{-6}$.

Усреднение по 1000 -10000 расчетам. Ошибка $<\sim 1\%$.

Параметры обучения

Вероятность случайной замены символа P_l .

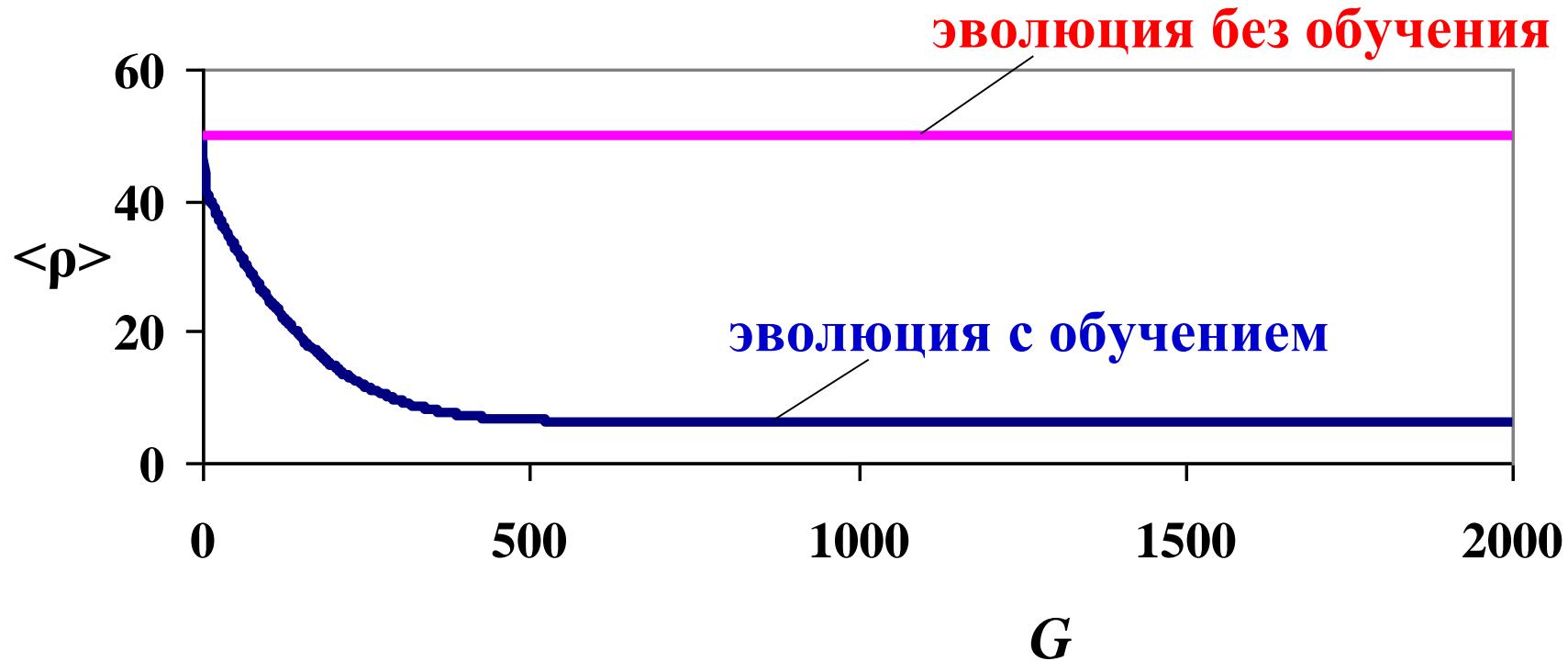
Обычно $P_l = 1$, при слабом обучении $P_l = 0.5$.

Длительность поколения $T = 2$.

В течение поколения фенотипы S_{Pk} модифицируются путем обучения методом проб и ошибок: случайно меняются отдельные символы, и если происходит приближение S_{Pk} к оптимуму S_M , то измененный символ сохраняется, иначе возвращается старый символ.

Эволюция с обучением и эволюция без обучения

Зависимость среднего по популяции расстояния генотипа до оптимума $\langle \rho \rangle = \langle \rho(S_{Gk}, S_M) \rangle$ от номера поколения G

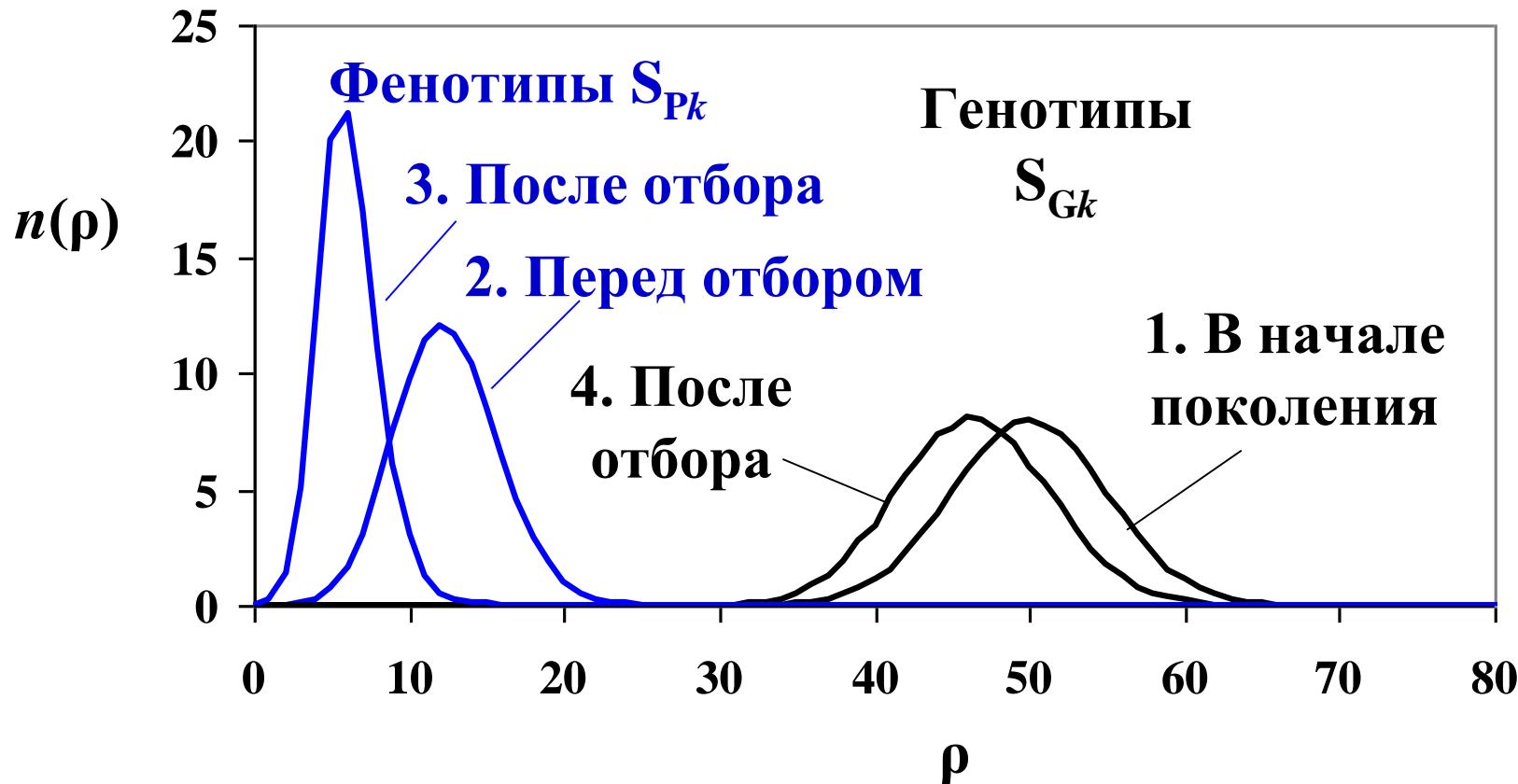


Параметры расчетов: $N = 100$, $n = 100$, $P_m = 0.01$, $\varepsilon = 10^{-6}$.

Почему чистая эволюция «не работает»: $f(S_k) \approx \varepsilon$.

Механизм генетической ассимиляции

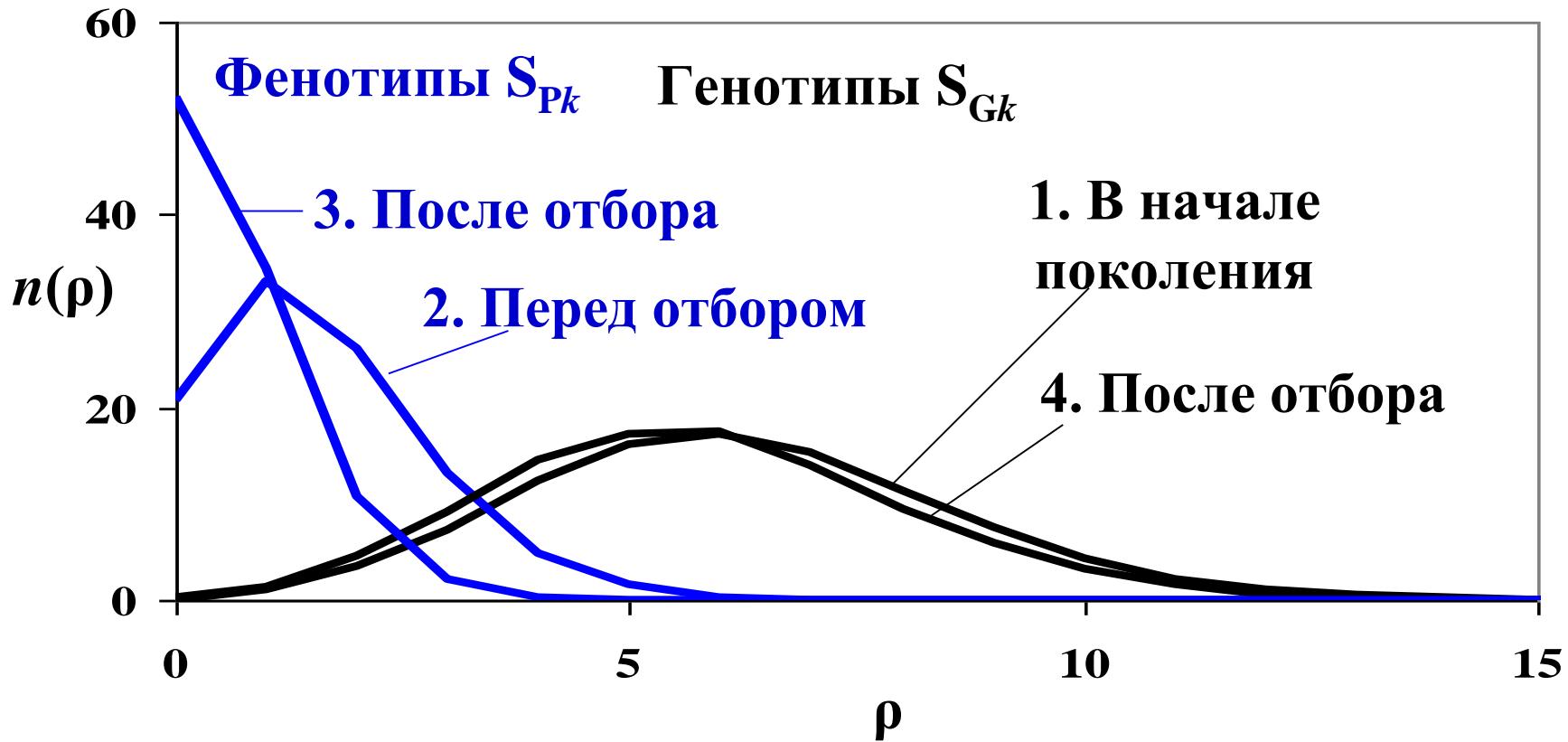
Распределение особей $n(\rho)$ по величинам ρ в первом поколении эволюции



Генотипы отобранных особей (4) достаточно близки к фенотипам этих особей (3)

Механизм эффекта экранирования

Распределение особей $n(\rho)$ по величинам ρ в конце эволюции (при $G = 2000$)

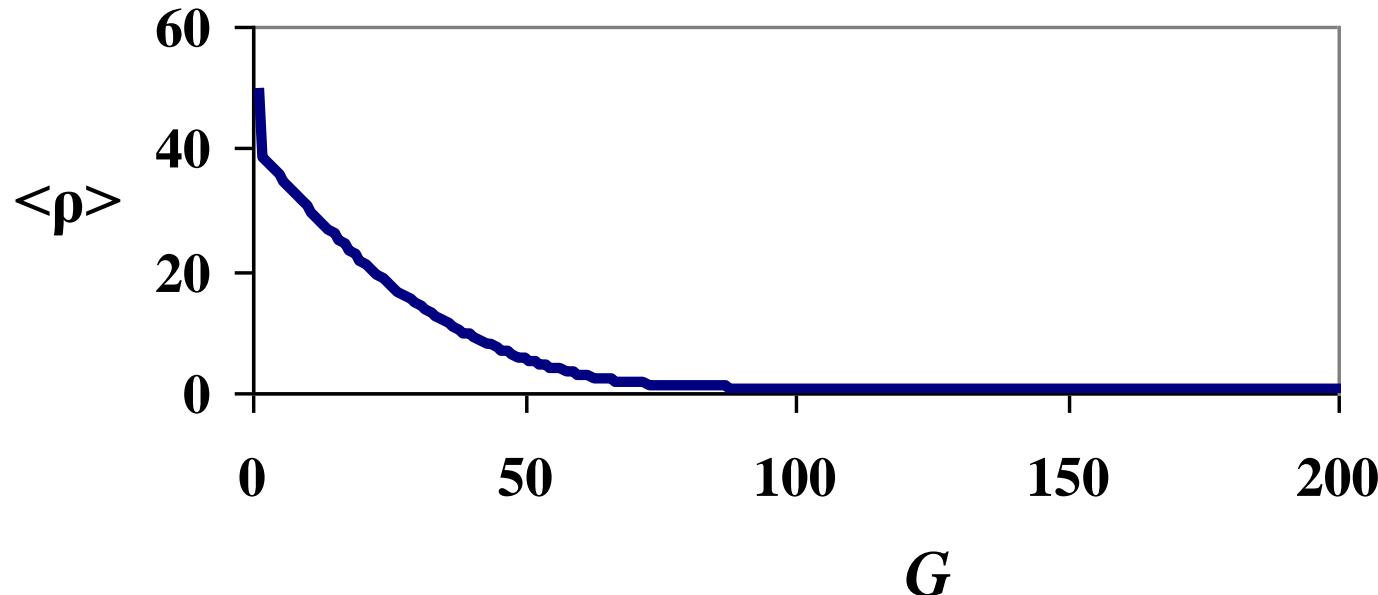


При сильном обучении оптимум фенотипа находится независимо от генотипа

Роль эффекта нагрузки на обучение

Зависимость среднего по популяции расстояния генотипа до оптимума $\langle \rho \rangle = \langle \rho(S_{Gk}, S_M) \rangle$ от номера поколения G .
Приспособленность особи равна:

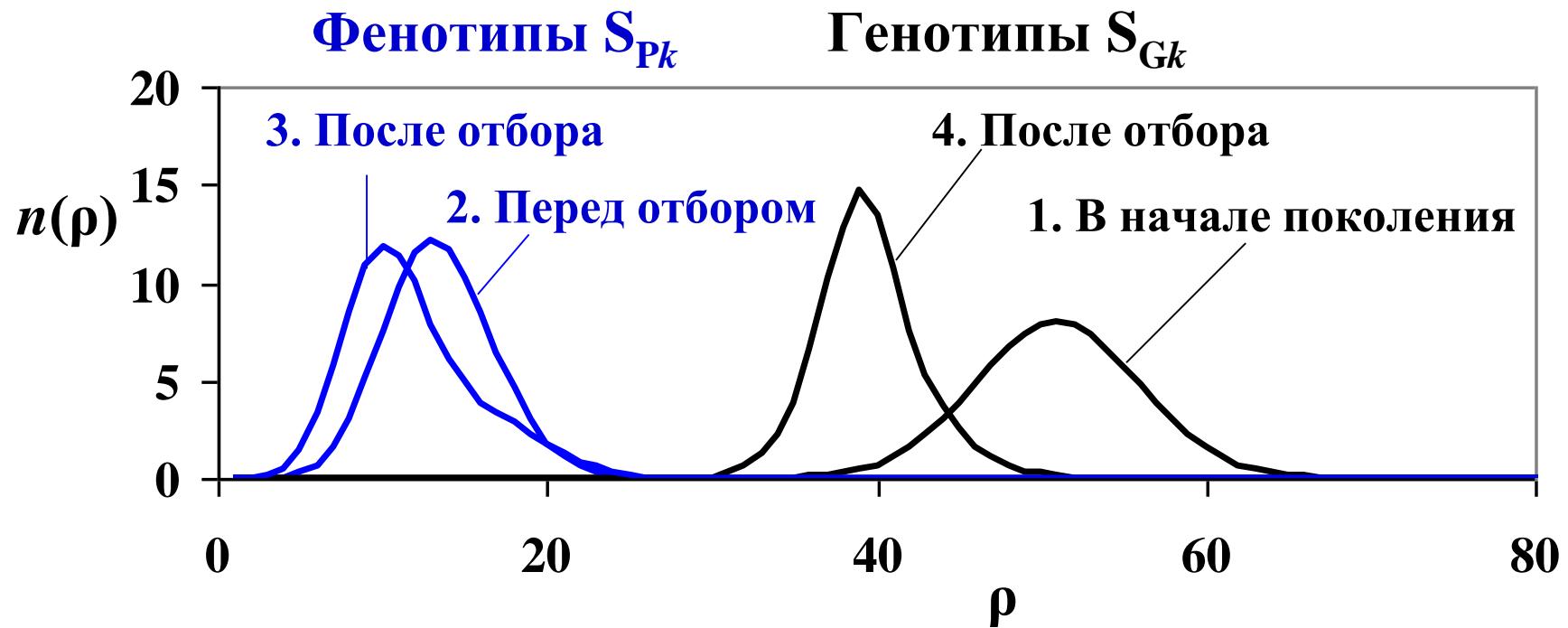
$$f(S_k) = \exp\{-\rho[S_{Pk}(t=0), S_{Pk}(t=T)]\} \{ \exp[-\rho(S_{FPk}, S_M)] + \varepsilon \}$$



Оптимум находится. Поиск происходит примерно в 10 быстрее, чем без учета эффекта нагрузки

Механизм эффекта нагрузки на обучение

Распределение особей $n(\rho)$ по величинам ρ в первом поколении эволюции

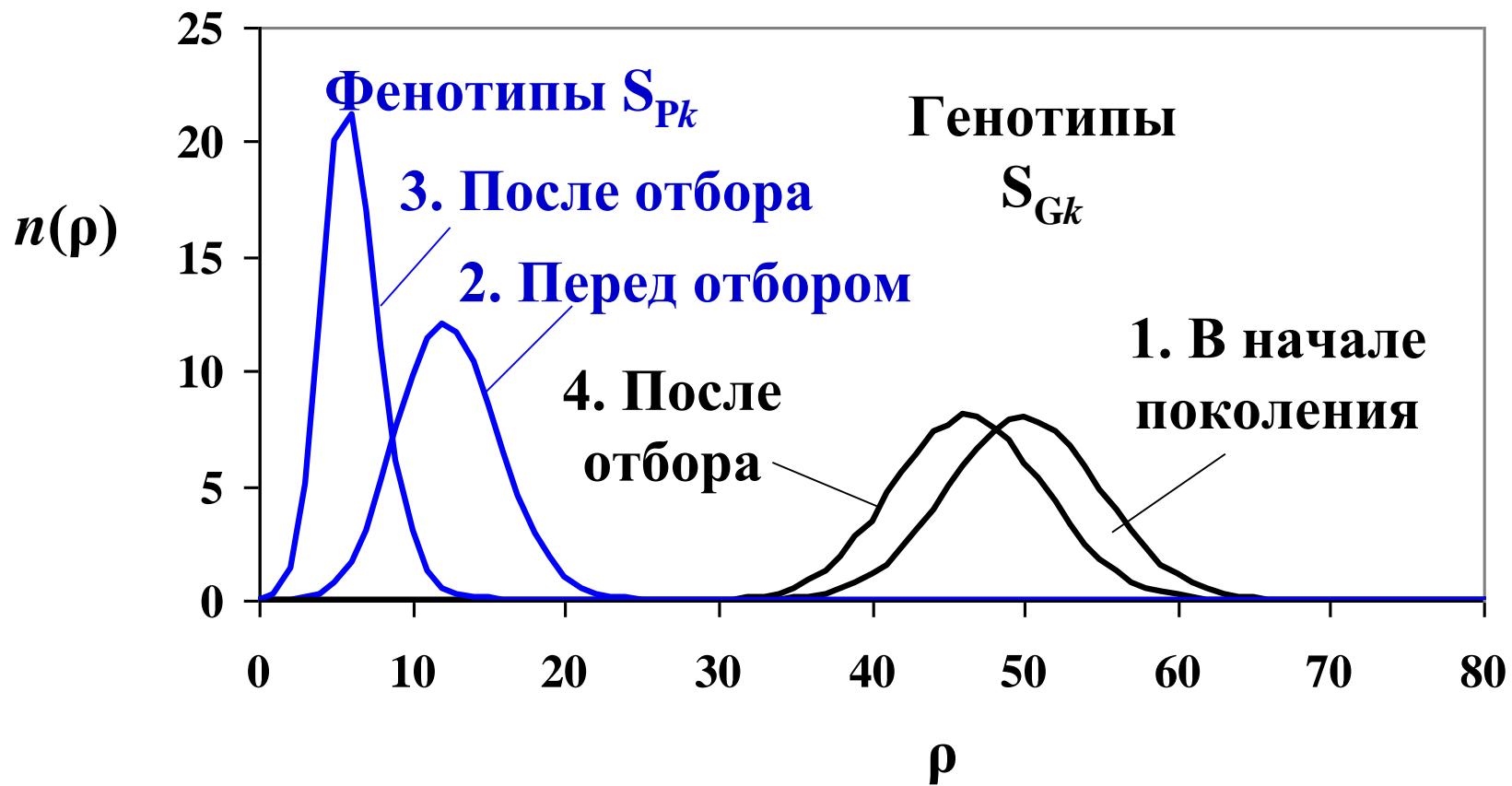


Распределения (3) и (4) резко сближаются. Это приводит к эффективному ускорению поиска

Механизм генетической ассимиляции

(без учета нагрузки, повтор)

Распределение особей $n(\rho)$ по величинам ρ в первом поколении эволюции



Сопоставление с моделью Hinton & Nowlan (1987)

Дополнительная модель

Строится так же, как и основная.

Эволюционирует популяция из n особей, особь имеет генотип S_{Gk} и фенотип S_{Pk} , длина цепочек генотипа и фенотипа равна N . Символы цепочек S_{Gk} и S_{Pk} равны 0 либо 1. Ищется оптимум S_M путем обучения и эволюции.

Меняется только приспособленность особей:

$$f(S_k) = \exp[-\rho(S_{FPk}, S_M)]$$

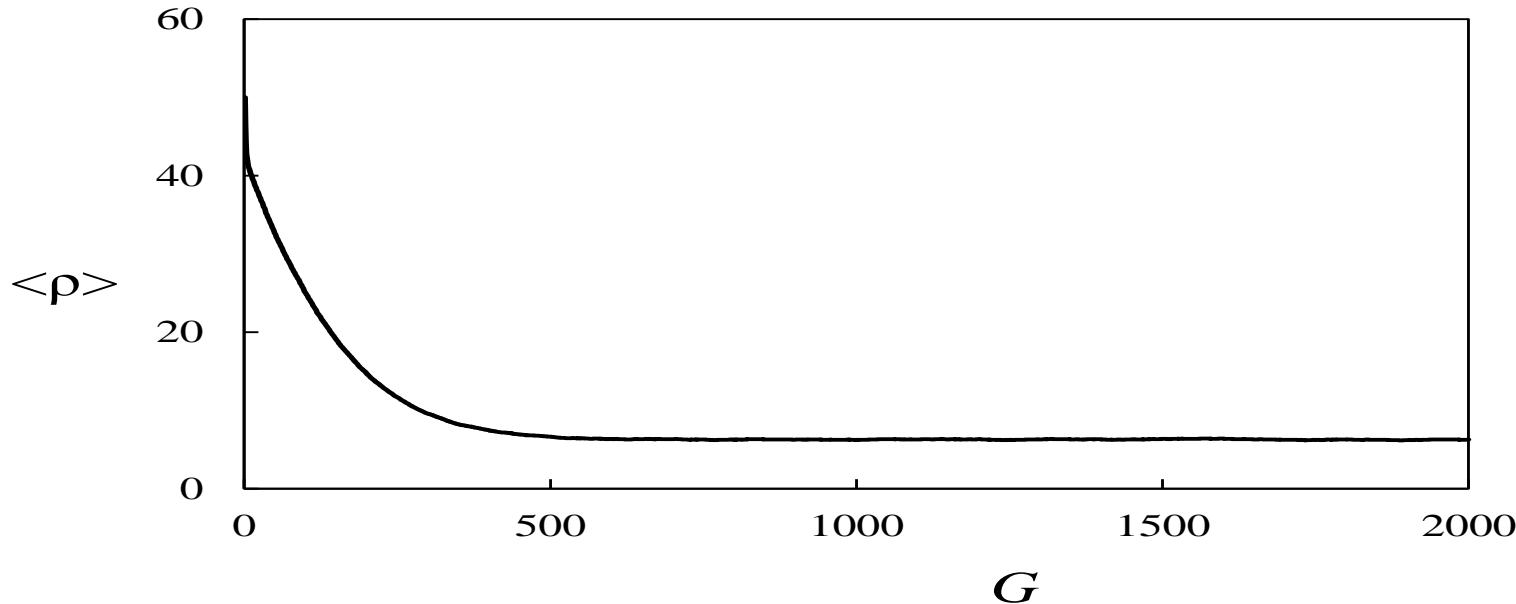
Результаты по дополнительной модели почти такие же, как и для основной

Совпадение в двух моделях почти полное

В дополнительной модели наблюдаются:

- 1) Генетическая ассимиляция.
- 2) Эффект экранирования. Распределения по ρ в конце эволюции такие же, как и для основной модели.
- 3) Нагрузка на обучение. Только в этом случае имеется небольшое различие в распределениях по ρ в первом поколении. Скорость сходимости такая же, как и для основной модели. В остальных случаях совпадение в моделях с точностью $\sim 1\%$.

Зависимость среднего по популяции расстояния генотипа до оптимума $\langle \rho \rangle = \langle \rho(S_{Gk}, S_M) \rangle$ от номера поколения G

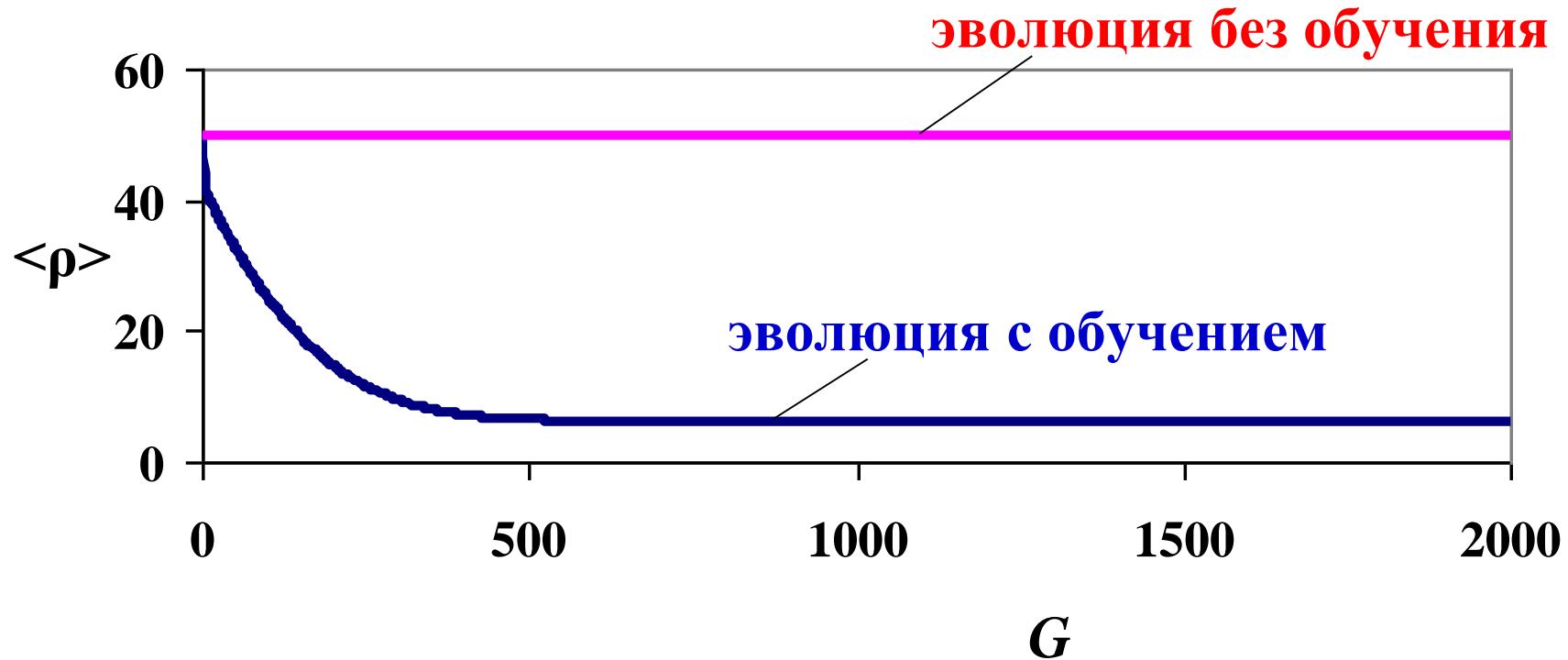


Параметры расчета: $N = 100$, $n = 100$, $P_m = 0.01$.

Наблюдается эффект экранирования: конечное $\langle \rho \rangle \approx 6.2$

Эволюция с обучением и эволюция без обучения (повтор)

Зависимость среднего по популяции расстояния генотипа до оптимума $\langle \rho \rangle = \langle \rho(S_{Gk}, S_M) \rangle$ от номера поколения G



Параметры расчетов: $N = 100$, $n = 100$, $P_m = 0.01$, $\varepsilon = 10^{-6}$.

Почему чистая эволюция «не работает»: $f(S_k) \approx \varepsilon$.

Выводы по модели взаимодействия между обучением и эволюцией

1. Показано, что обучение может радикально ускорять эволюционный поиск
2. Проанализирован механизм генетической ассилияции: показано, что ассилияция может происходить за счет того, что генотипы отобранных особей достаточно близки к фенотипам обученных и отобранных особей
3. Роль эффекта экранирования состоит в том, что при сильном обучении оптимальный фенотип может быть найден независимо от генотипа
4. Показано, что эффект нагрузки на обучение может значительно ускорить генетическую ассилияцию

Основные публикации по модели взаимодействия между обучением и эволюцией

1. Редько В.Г. Модель взаимодействия между обучением и эволюционной оптимизацией // Математическая биология и биоинформатика (электронный журнал), 2012. Т.7. № 2. С. 676-691.

URL: http://www.matbio.org/2012/Redko_7_676.pdf

2. Vladimir G. Red'ko. Model of Interaction between Learning and Evolution // In Cornell University Library Archive (<http://arxiv.org/>):

<http://arxiv.org/ftp/arxiv/papers/1411/1411.5053.pdf>
<http://arxiv.org/abs/1411.5053>

Гипотетический процесс предбиологической эволюции

Квазивиды → Гиперциклы → Сайзеры →
→ Протоклетки → Простейшие организмы

Спасибо за внимание!