

ВВЕДЕНИЕ В ИНФОРМАЦИОННУЮ БИОЛОГИЮ

Н.А. Колчанов, С.А. Лашин

Электронно-лекционный курс
разработан в рамках реализации Программы развития НИУ-НГУ
2012 год

Лекция 10
***Математическое и компьютерное
моделирование эволюционно-
популяционных процессов***

Что такое эволюция?

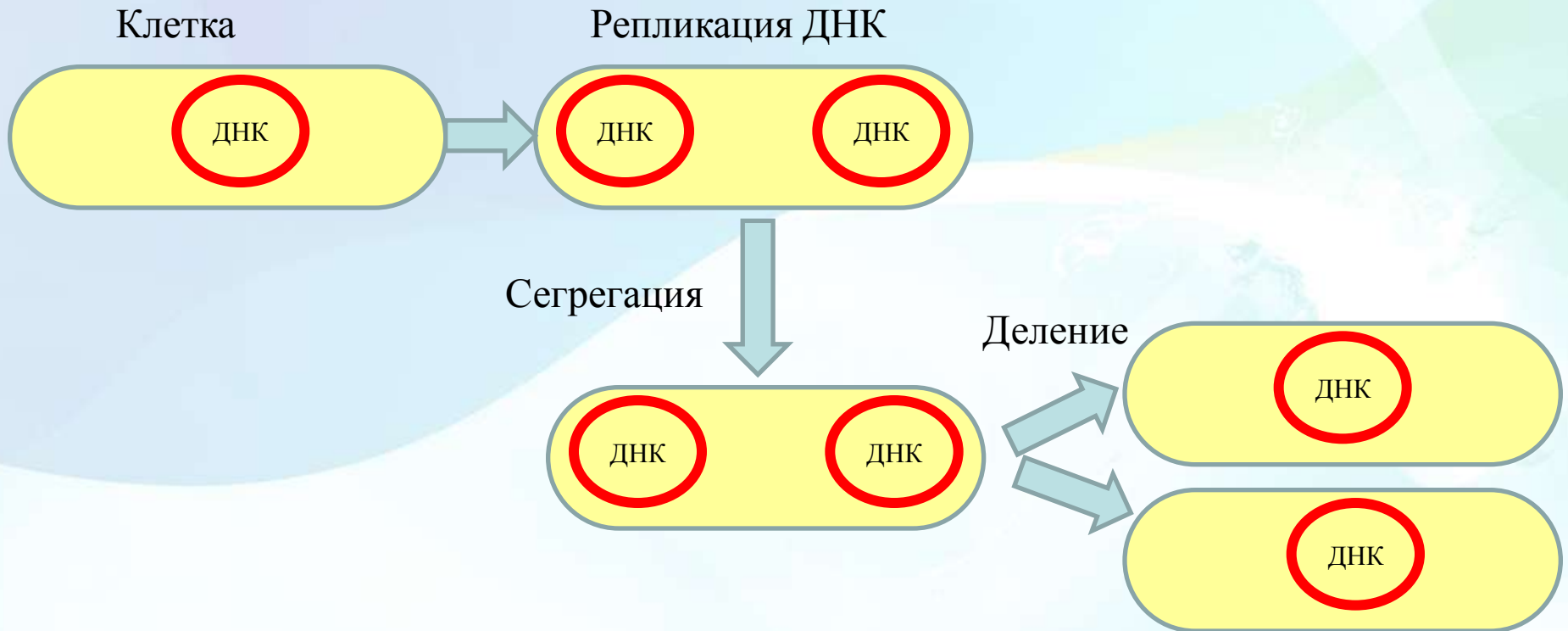
- **Эволюция** (от лат. evolutio-развертывание) – процесс изменения (развития) системы
- Эволюция состоит из постепенных изменений (в противовес революции)
- Эволюция относительно детерминирована исходным состоянием системы (эволюция звёзд, планет, химических элементов, климата, ландшафта и т.д.)
- Внешние события могут менять траекторию такого развития, но такие изменения относительно предсказуемы, если известны исходное состояние системы и характеристики модифицирующих воздействий
- Эволюционный процесс – детерминистически воспроизводим.

Что такое биологическая эволюция?

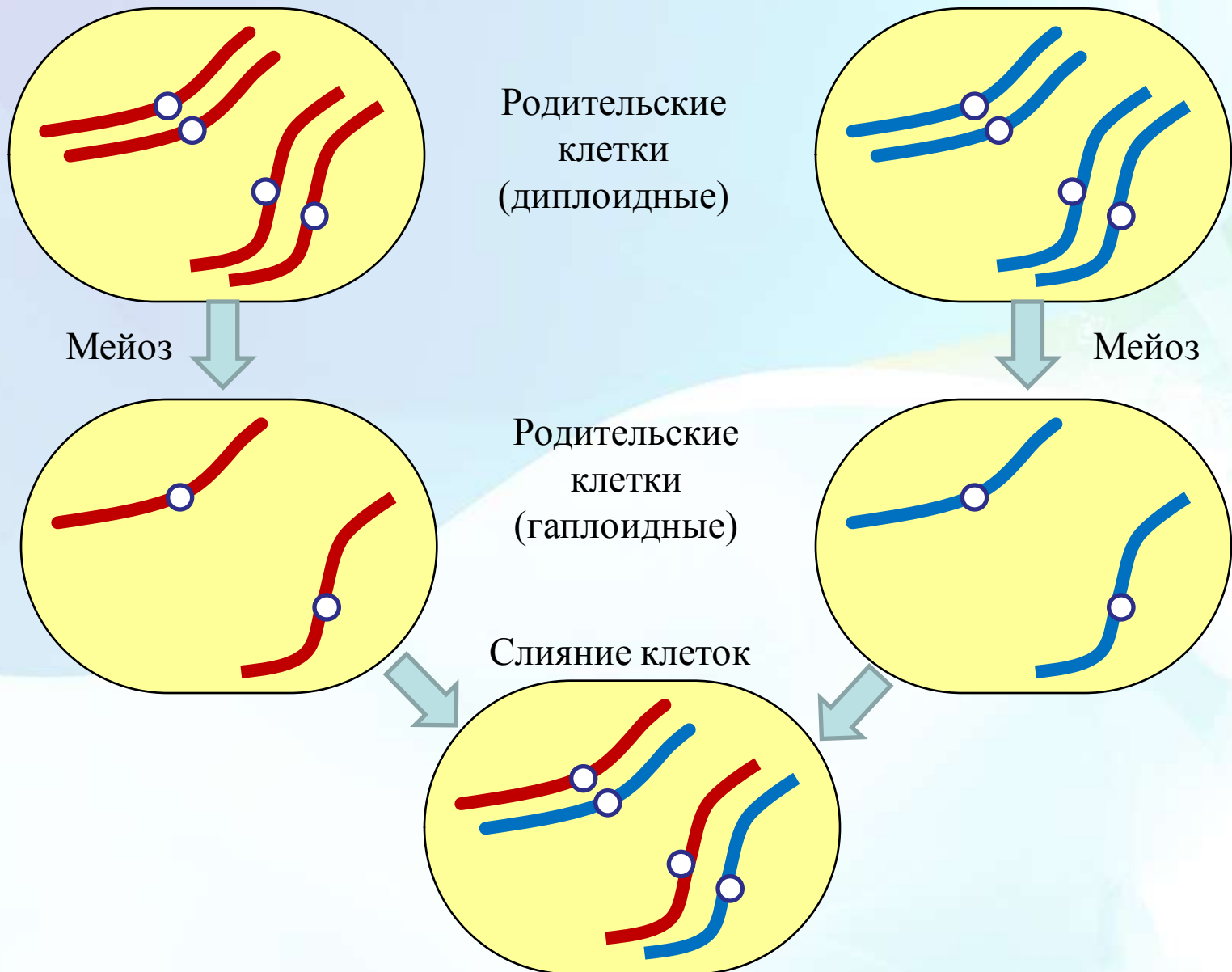
- Эволюция – естественный процесс развития жизни на Земле
- Эволюция включает:
 - Изменения и адаптации
 - Образование и вымирание видов
 - Преобразование экосистем
 - Преобразование биосферы в целом
- В биологии чётко различают два процесса:
 - *Онтогенез* – реализация уже существующей генетической программы в ходе развития организма
 - *Филогенез* – процесс возникновения и становления новой генетической программы
- Под *эволюцией* подразумевают именно *филогенез*.

Как размножаются живые организмы?

Бесполое размножение (деление, вегетативное размножение)



Как размножаются живые организмы? Половое размножение



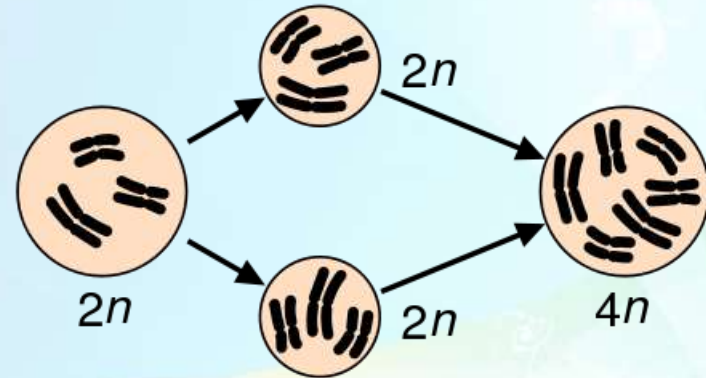
Как появляется материал для эволюции: мутации

Классификации мутаций

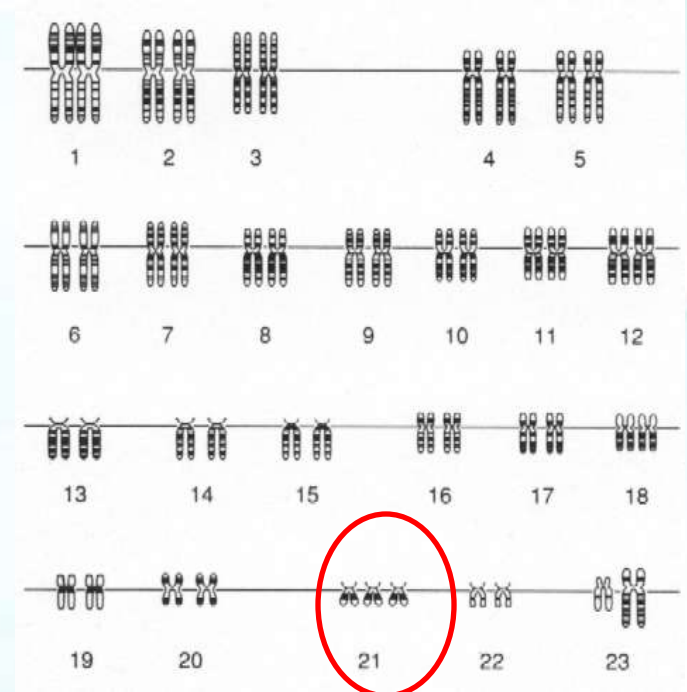
- По структурному происхождению
 - Геномные (полиплоидизация, анеуплоидия)
 - Хромосомные (перестройки: делеции, дупликации, инверсии, транслокации)
 - Генные (замены, делеции, инсерции и т.д.)

Как появляется материал для эволюции: геномные мутации

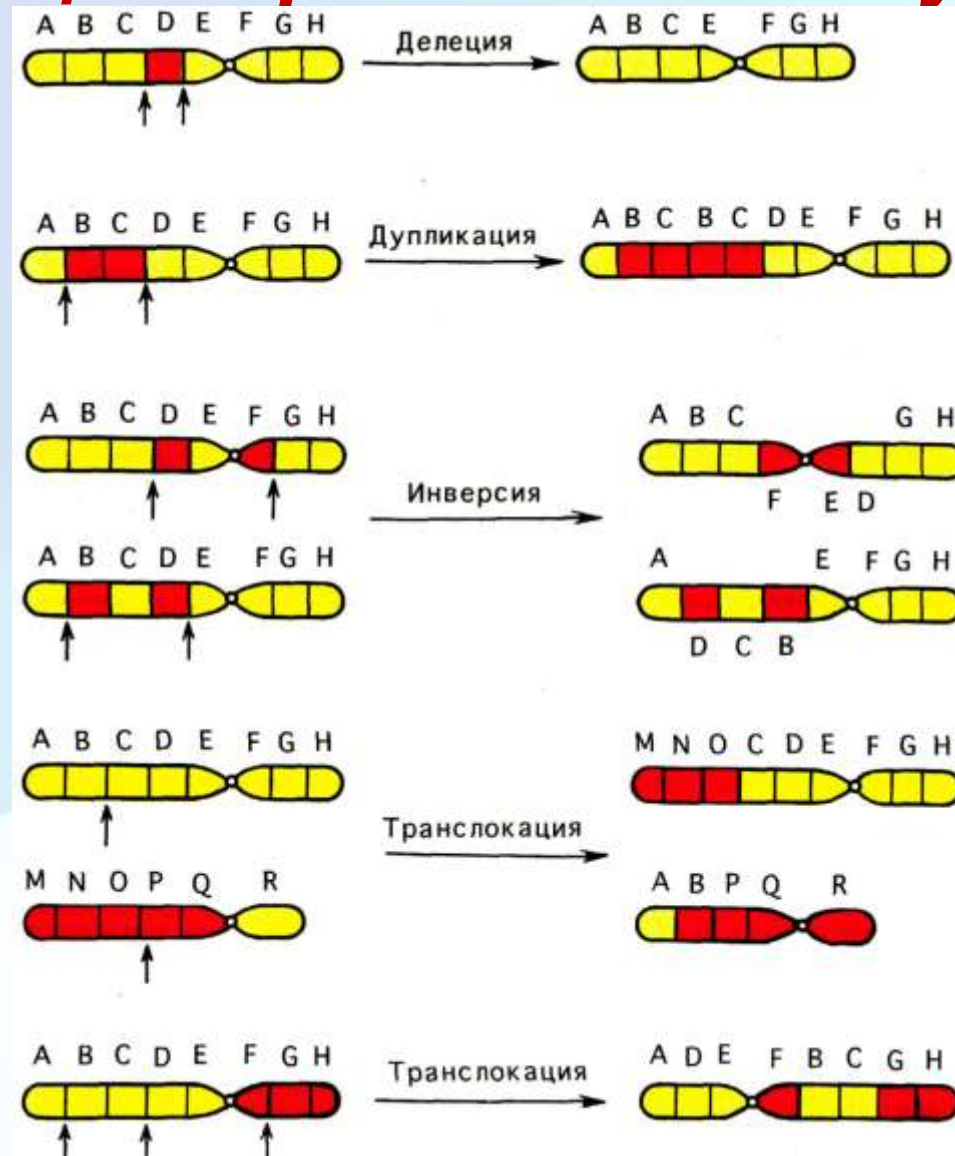
Полиплоидизация



Анеуплоидия
Приводит, в
частности, к
болезни Дауна



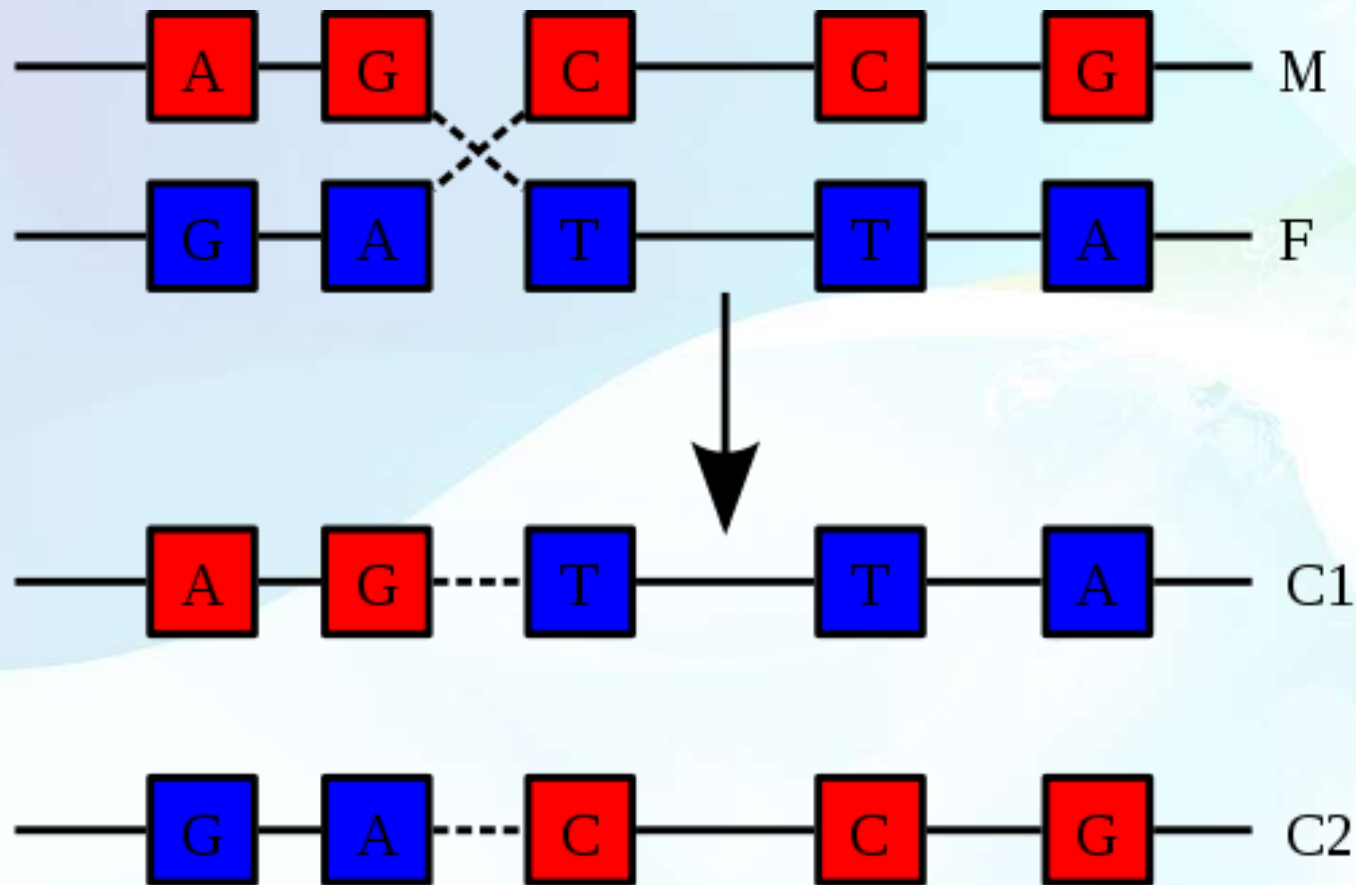
Как появляется материал для эволюции: хромосомные мутации



Как появляется материал для эволюции: генетические мутации

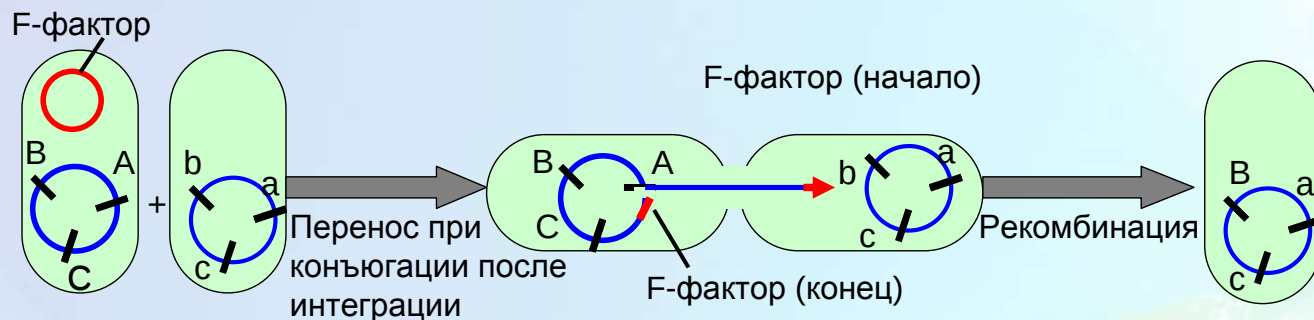
- В кодирующей области
 - Сохраняющие рамку считывания
 - Замены (синонимичные или несинонимичные)
 - Нарушающие рамку считывания
 - Инсерции (фрагменты длиной некратной 3)
 - Делеции (фрагменты длиной некратной 3)
- В некодирующей области
 - Изменяющие регуляцию (в промоторах, интронах и т.д.)
 - Нейтральные

Как появляется материал для эволюции: рекомбинация

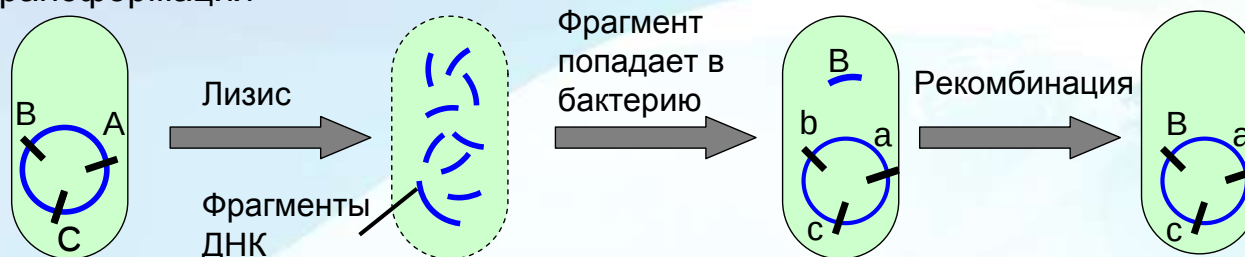


Как появляется материал для эволюции: горизонтальный перенос генов

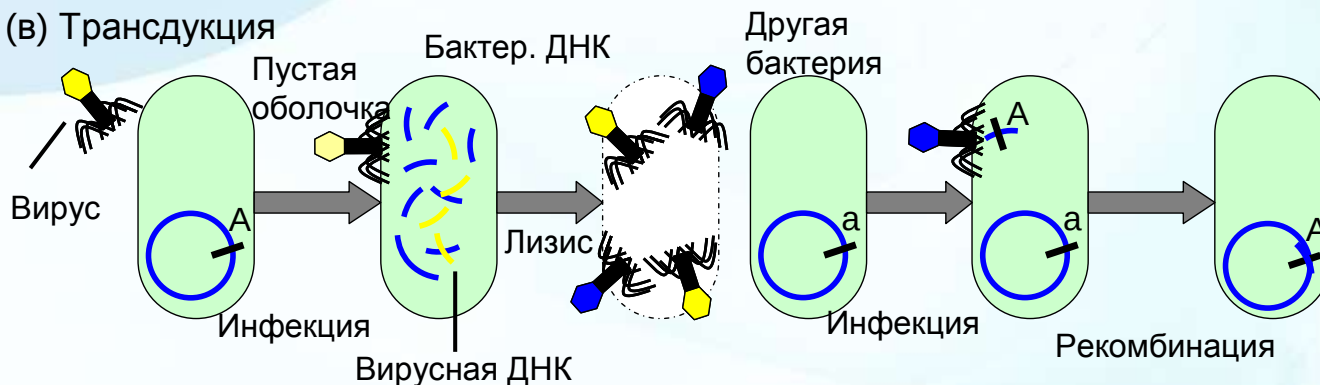
(а) Конъюгация



(б) Трансформация



(в) Трансдукция



Горизонтальный перенос генов и эволюция прокариот

- Исследования полных геномов показали, что порядка **20%** белок-кодирующих генов в геномах современных прокариот появились в результате горизонтального переноса [1].
- Тесная ассоциация бактерий в сообществах (биоплёнках) повышает вероятность горизонтального переноса, позволяя приобретать новые функции [2, 3].

[1]: Beiko R.G., Harlow T.J., Ragan M.A. Highways of gene sharing in prokaryotes // Proc. Natl. Acad. Sci. USA. - 2005. – Vol. 1002, N 40. – P. 14332-14337.

[2]: Ochman H., Lawrence J.G., Groisman E.A. Lateral gene transfer and the nature of bacterial innovation // Nature. 2000. Vol. 405, N 6784. p. 299-304.

[3]: Шестаков С.В. О ранних этапах биологической эволюции с позиции геномики // Палеонтологический журнал. – 2003а. - № 6. - С. 50-57.

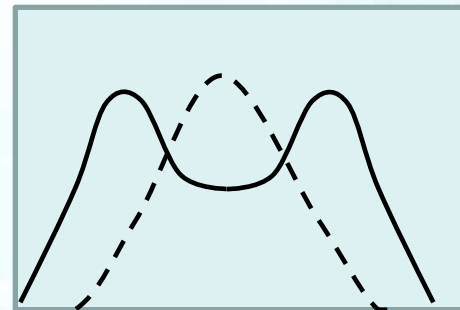
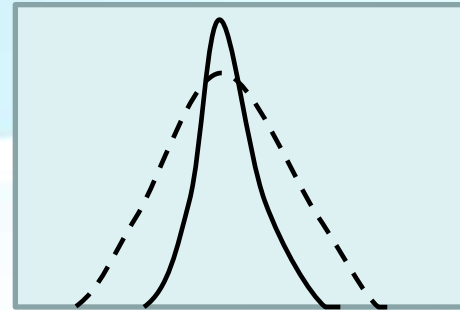
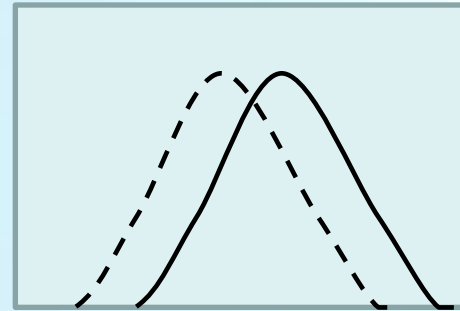
А судьи кто?

Естественно, отбор

- Естественный отбор – выживает (и размножается) сильнейший
- ЕО – по Дарвину, и далее в СТЭ – главный оценщик эволюционного материала
- ЕО – оценивает прежде всего **фенотипы!**
 - **Генотипы** оцениваются в виде комбинаций аллелей

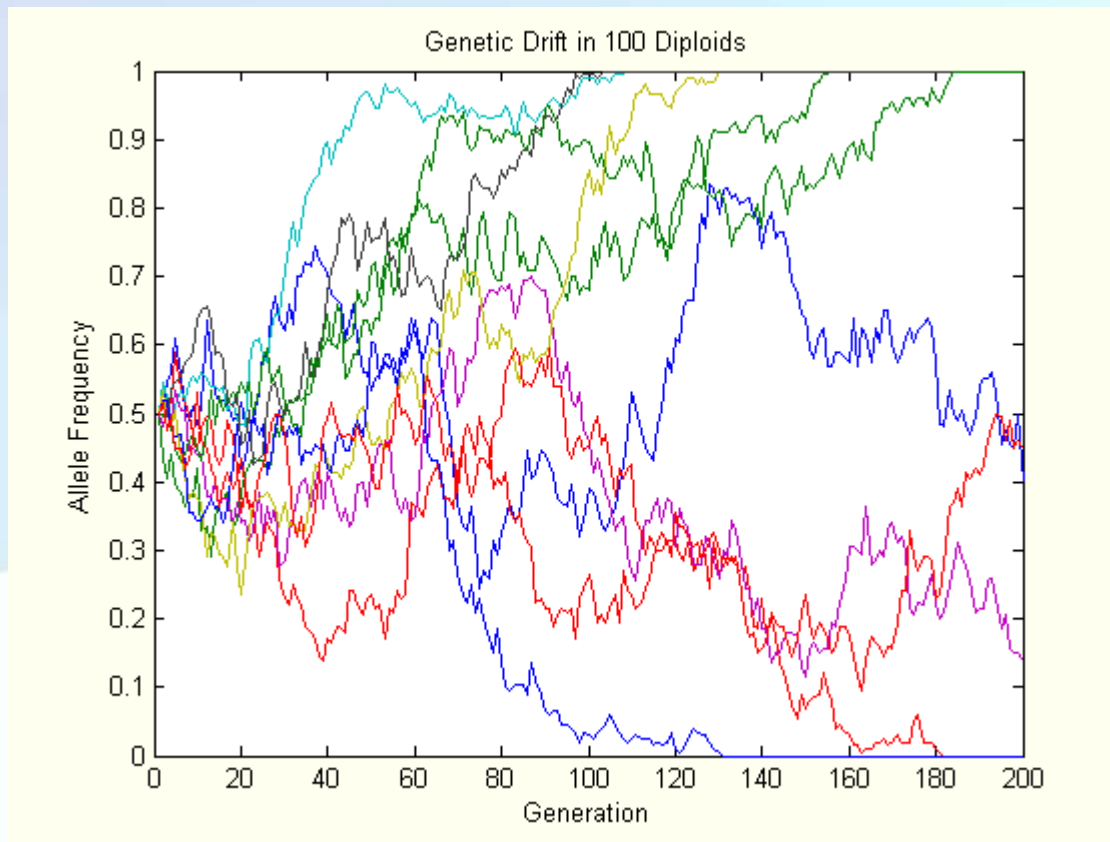
Формы естественного отбора

- Движущий отбор – при **направленном** изменении среды преимущество получают особи с признаками, **сдвинутыми** относительно среднего значения, что ведёт к сдвигу среднего значения
- Стабилизирующий отбор – направлен против особей, имеющих крайние проявления признака (отбор в пользу средних значений)
- Дизруптивный отбор – благоприятствует двум крайним проявлениям признака, но против среднего значения. Лежит в основе дивергенции видов (Дарвин)
- Половой отбор – прежде отбираются признаки, ответственные за успех в **размножении**, а не в **жизнеспособности** особи



Дрейф генов

- Дрейф генов — **ненаправленное** изменение частот аллельных комбинаций в популяции, обусловленное **стохастическими** причинами



Макро- и микроэволюция

- *Микроэволюция* — это протекание эволюционных процессов внутри популяций, рас и других объединений особей, вплоть до вида.
 - Такие объединения особей имеют общий перемешиваемый генофонд, т.е. некоторую степень внутренней связанности и внешней изоляции. В больших популяциях отбор является главным фактором микроэволюции.
- Процесс микроэволюции определяется также *случайными факторами*, например генетическим дрейфом в малых популяциях и молекулярным драйвом, унифицирующим повторяющиеся последовательности в геномах за счёт неравного кроссинговера при мейозе, генной конверсии и транспозиции [Дубинин, 1986].
- В ходе микроэволюции мутантные аллели могут фиксироваться, элиминироваться (гибнуть) либо, когда отбор идет в пользу гетерозигот, возникает *сбалансированный полиморфизм* [Ли, 1978; Левонтин, 1978].

Макро- и микроэволюция

- *Макроэволюция* — напротив, протекание эволюционных процессов в репродуктивно изолированных таксонах (начиная от вида и выше). Макроэволюционные процессы *на много порядков более длительны*, чем микроэволюционные.
- На сегодняшний день *не существует* однозначного мнения по поводу сущности и механизмов макроэволюции [Иорданский, 2004].

Что нас интересует в эволюционном процессе?

- Механизмы эволюционных изменений
- Их функциональное значение для организма
- Изменение популяционного состава
- Влияние эволюционных изменений на популяцию, сообщество, экосистему, биосферу в целом

Математическое описание эволюционного процесса

- Дифференциальные уравнения
- Алгебраические уравнения
- Рекуррентные соотношения
- Стохастическое моделирование
- Сетевое моделирование
- Имитационное моделирование
- ...

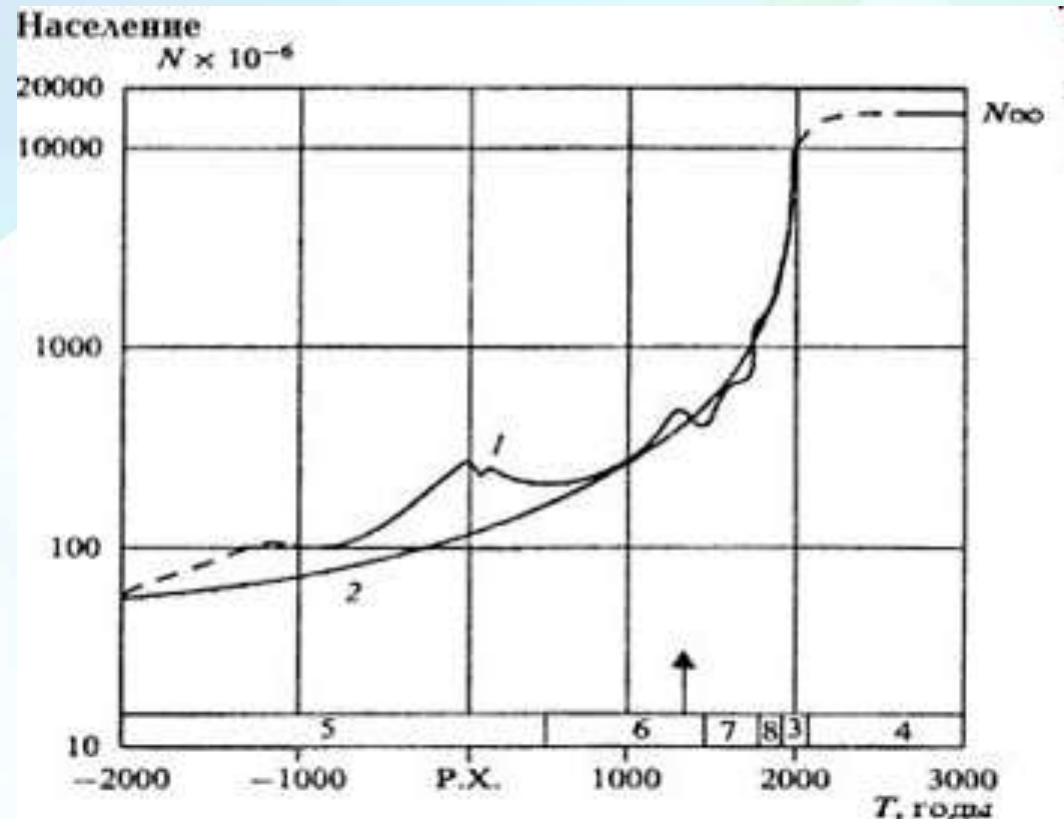
Модели динамики популяций

Модель неограниченного роста (Мальтуса)

$$\frac{dN}{dt} = aN, \quad a > 0$$

N – численность популяции,

a – коэффициент прироста
численности популяции



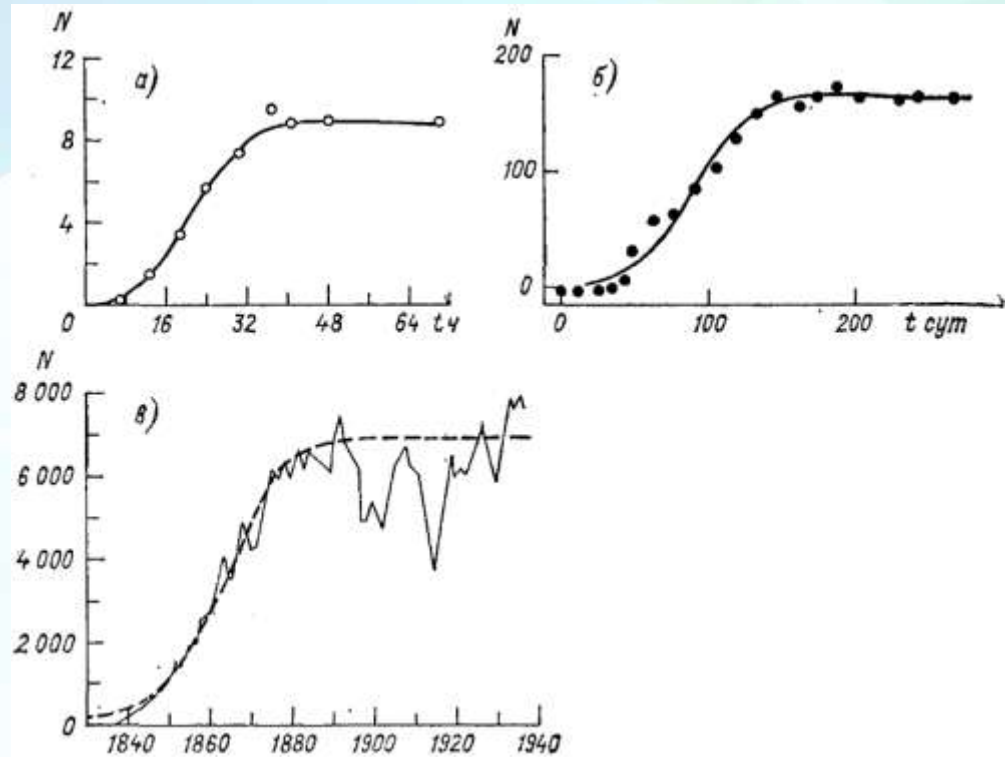
Модели динамики популяций

Модель ограниченного роста (Ферхюльста)

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right), \quad r > 0, K > 0$$

r – коэффициент прироста
численности популяции,

K – максимальная численность
популяции



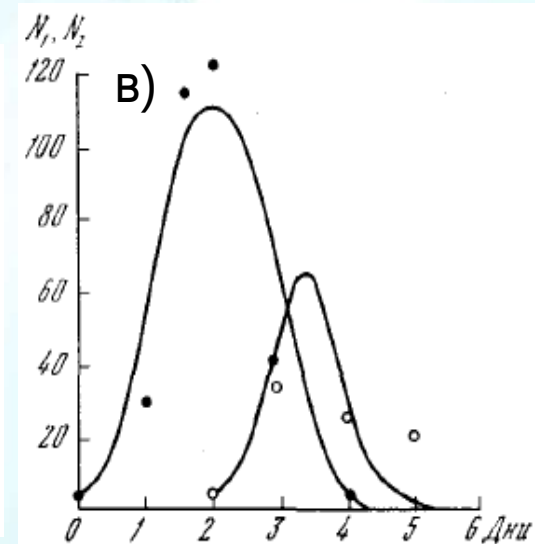
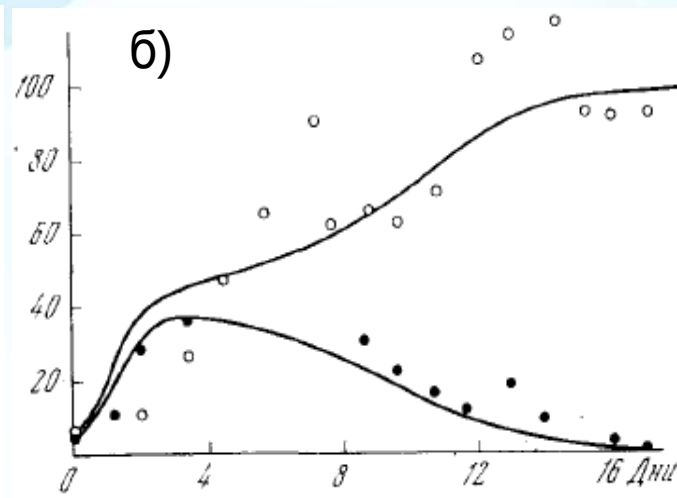
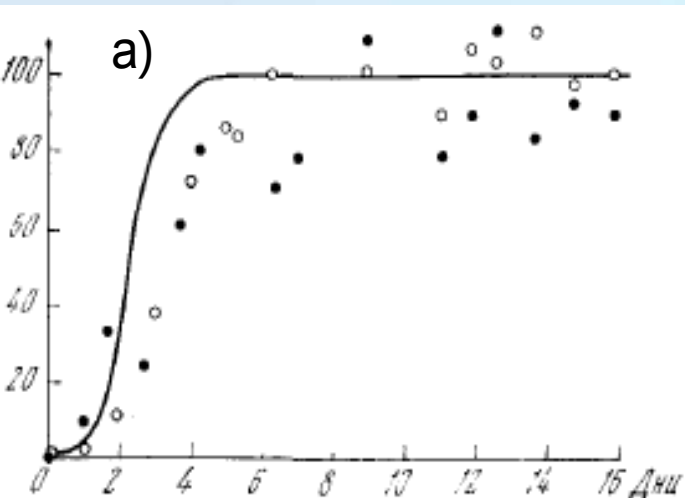
Модели динамики популяций

Модель «хищник-жертва»

$$\frac{dx}{dt} = x(\varepsilon_x - \gamma_{xy}y)$$

$$\frac{dy}{dt} = y(\gamma_{yx}x - \varepsilon_y)$$

x – жертвы, y – хищники,
 ε_x – естественный прирост жертв,
 ε_y – естественная убыль хищников
 γ_{xy} – влияние хищников на жертв,
 γ_{yx} – влияние жертв на хищников



Взято из [Гаузе, 1934]

Модели популяционной генетики, описанные с помощью ДУ

Модель менделевской бесполой диплоидной панмиктической популяции с одним диаллельным локусом

Один диаллельный локус (A_1, A_2)

Приспособленность гамет: w_1, w_2

Их разница:

$$s = w_1 - w_2$$

Частота аллеля A_1 :

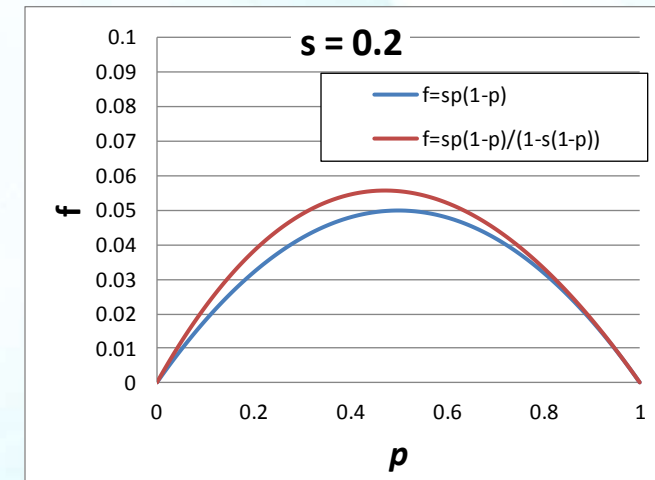
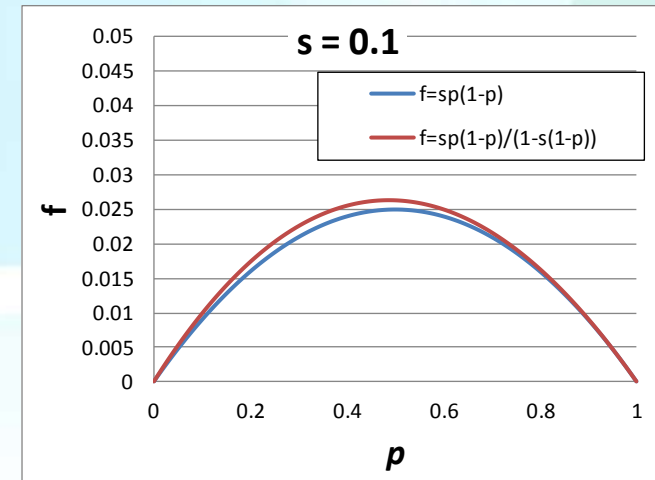
$$p$$

Частота аллеля A_2 :

$$1-p = q$$

Уравнение динамики частоты p в популяции:

$$\frac{dp}{dt} = \frac{sp(1-p)}{1-s(1-p)} \approx sp(1-p)$$



Модели популяционной генетики, описанные с помощью ДУ

Модель менделевской бесполой диплоидной
панмиктической популяции с одним диаллельным локусом

Один диаллельный локус (A_1, A_2)

Приспособленность гамет: w_1, w_2

Их разница:

$$s = w_1 - w_2$$

Частота аллеля A_1 :

$$p$$

Частота аллеля A_2 :

$$1 - p = q$$

Уравнение динамики частоты p в популяции:

$$\frac{dp}{dt} = \frac{sp(1-p)}{1-s(1-p)} \approx sp(1-p)$$

Равновесные точки:

1) $p = 0, q = 1$

2) $p = 1, q = 0$

3) ???

Модели популяционной генетики, описанные с помощью ДУ

Модель менделевской бесполой диплоидной
панмиктической популяции с одним диаллельным локусом

Один диаллельный локус (A_1, A_2)

Приспособленность гамет: w_1, w_2

Их разница:

$$s = w_1 - w_2$$

Частота аллеля A_1 :

$$p$$

Частота аллеля A_2 :

$$1 - p = q$$

Уравнение динамики частоты p в популяции:

$$\frac{dp}{dt} = \frac{sp(1-p)}{1-s(1-p)} \approx sp(1-p)$$

Равновесные точки:

1) $p = 0, q = 1$

2) $p = 1, q = 0$

3) ???

При $s > 0$ устойчива 1)

При $s < 0$ устойчива 2)

Модели популяционной генетики, описанные с помощью ДУ

Модель менделевской бесполой диплоидной
панмиктической популяции с одним диаллельным локусом

Один диаллельный локус (A_1, A_2)

Три генотипа: A_1A_1 , A_1A_2 , A_2A_2

Приспособленность диплоидов: w_{11}, w_{12}, w_{22}

Частота аллеля A_1 : p

Частота аллеля A_2 : $1-p = q$

Уравнение динамики частоты p в популяции:

$$\frac{dp}{dt} = \frac{p(1-p)}{w} (w_{1\bullet} - w)$$

$$w_{1\bullet} = w_{11}p + w_{12}(1-p)$$

$$w_{2\bullet} = w_{21}p + w_{22}(1-p)$$

$$w = w_{1\bullet}p + w_{2\bullet}(1-p) =$$

$$= w_{11}p^2 + 2w_{12}p(1-p) + w_{22}(1-p)^2$$

Модели популяционной генетики, описанные с помощью ДУ

Модель менделевской бесполой диплоидной
панмиктической популяции с одним диаллельным локусом

Один диаллельный локус (A_1, A_2)

Три генотипа: A_1A_1 , A_1A_2 , A_2A_2

Приспособленность диплоидов:

w_{11}, w_{12}, w_{22}

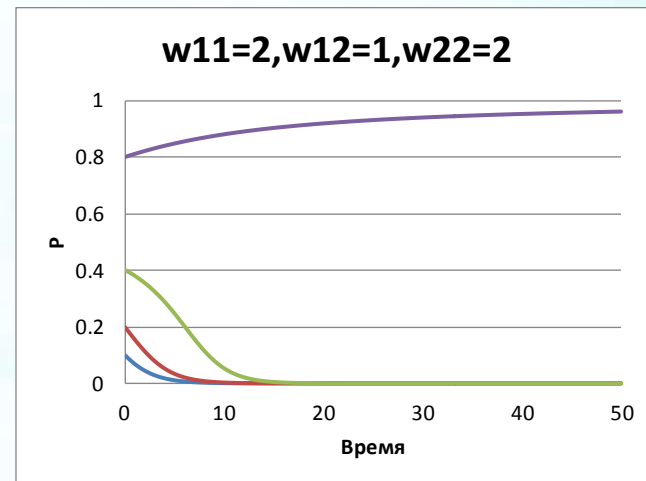
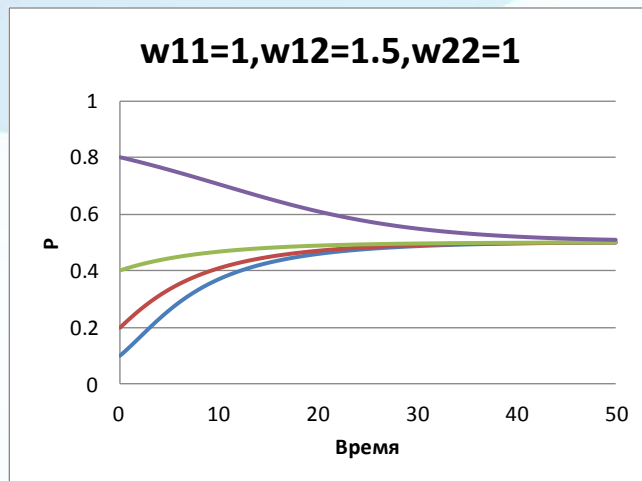
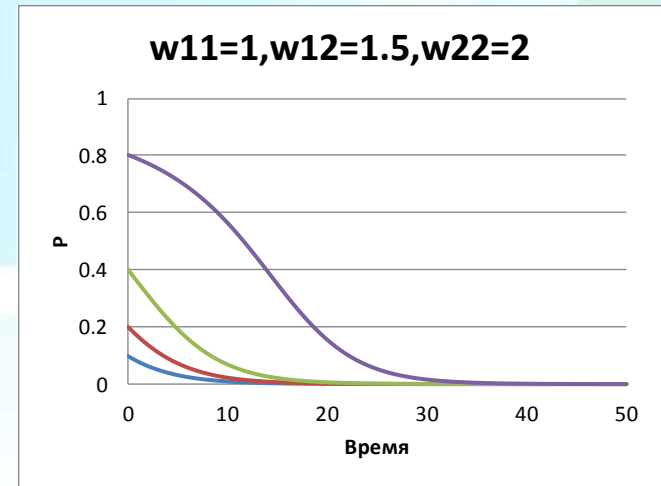
Частота аллеля A_1 :

p

Частота аллеля A_2 :

$1-p = q$

Уравнение динамики частоты p в популяции:



Рекуррентные модели популяционной генетики

Рекуррентные уравнения динамики частот гамет:

Частота гена A у взрослых особей в n -м поколении - p_n

Частоты зигот в $(n+1)$ поколении:

$$p_n^2 AA : 2p_n (1 - p_n) Aa : (1 - p_n)^2 aa$$

Приспособленность особей (вероятность дожить до половозрелого состояния):

- для AA и $Aa = S$
- для $aa = S(1-k)$

Тогда относительные частоты взрослых особей в $(n+1)$ -м поколении:

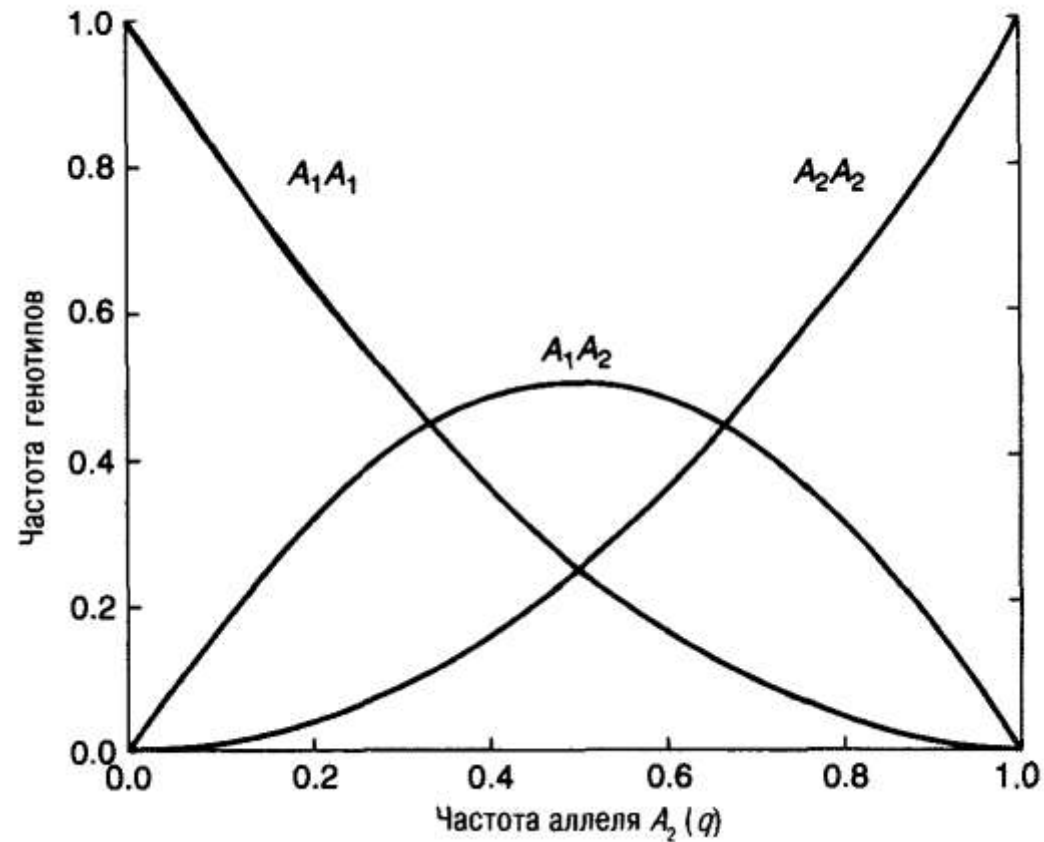
$$Sp_n^2 AA : 2Sp_n (1 - p_n) Aa : S (1 - p_n)^2 (1 - k) aa$$

Модели популяционной генетики

Закон Харди-Вайнберга

Типы скрещиваний и потомство в равновесной панмиктической популяции [644]

Тип скрещивания	Частота скрещивания	Потомство		
		AA	Aa	aa
$AA \times AA$	p^4	p^4	0	0
$AA \times Aa$	$4p^3q$	$2p^3q$	$2p^3q$	0
$Aa \times Aa$	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
$AA \times aa$	$2p^2q^2$	0	$2p^2q^2$	0
$Aa \times aa$	$4pq^3$	0	$2pq^3$	$2pq^3$
$aa \times aa$	q^4	0	0	q^4
Для популяции в целом	1,00	p^2	$2pq$	q^2



Компьютерное моделирование эволюции

- Имитационные (*портретные, структурные, индивидуально-ориентированные*, англ. *simulation, individual-based*) подходы к моделированию - методы исследования, в котором изучаемая система заменяется моделью с достаточной точностью описывающей **объекты** и **процессы** реальной системы. С моделью проводятся эксперименты (*проигрываются сценарии*, жарг. *проводятся имитации/симуляции*) с целью получения информации об этой системе.
- Имитационная модель — логико-математическое описание объекта, которое может быть использовано для экспериментирования на компьютере в целях проектирования, анализа и оценки функционирования объекта.

Простейшая модель: дрейф генов

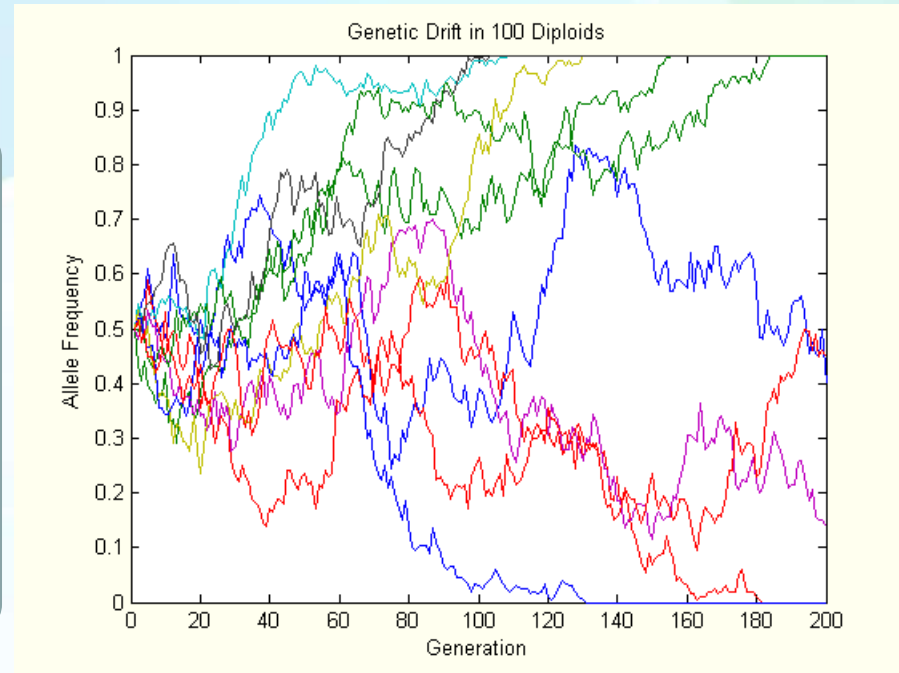
Геном особи = пара (A_1, A_2) , $A_i = \{0, 1\}$

Генерация геномов
особей

Размножение:

1. Случайно выбрать «маму»
2. Случайно выбрать аллель
3. Случайно выбрать «папу»
4. Случайно выбрать аллель
5. Сгенерировать новую особь

повторить



Имитационные модели динамики и эволюции популяций

Модель «AEvol» (Beslon et al.)

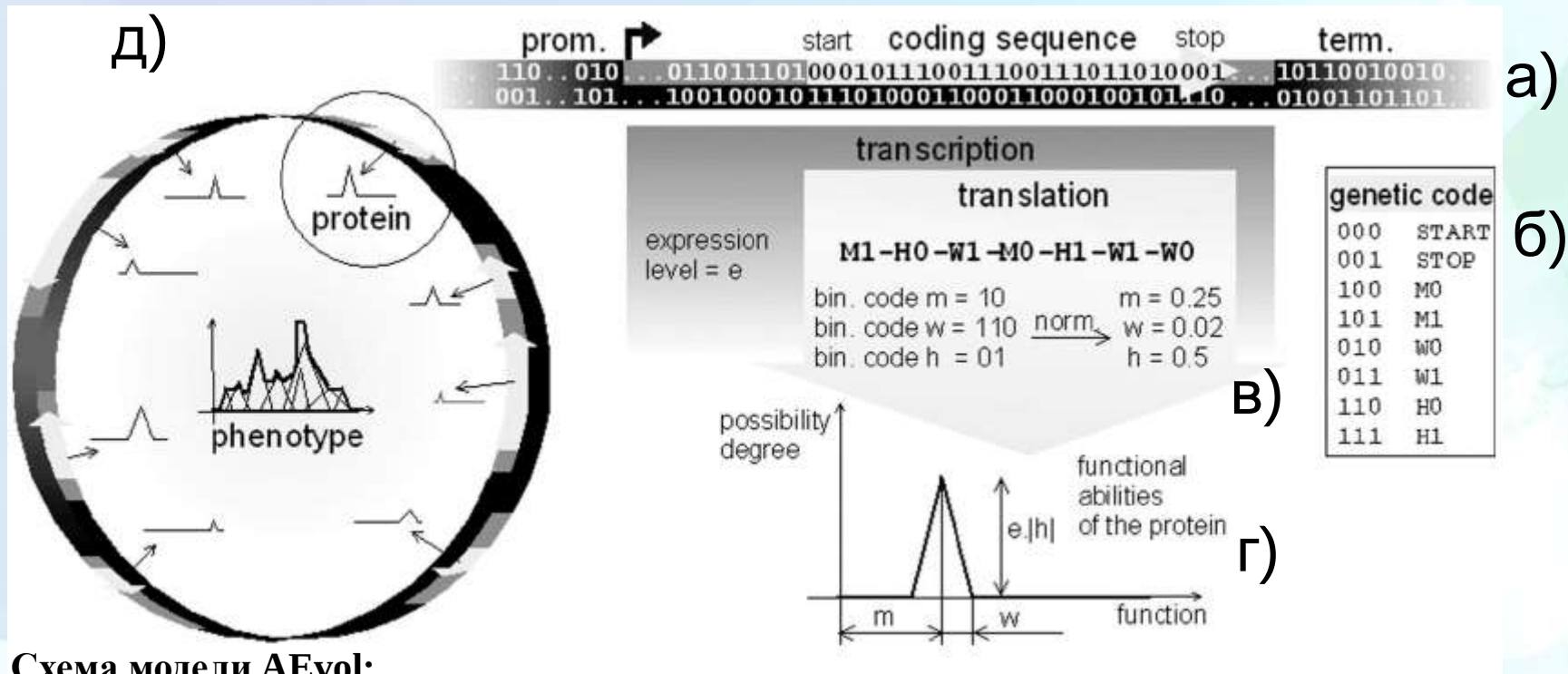


Схема модели AEvol:

- а) кодирующая последовательность – «ДНК»;
- б) «генетический код», согласно которому последовательность «гена» перекодировается в последовательность «белка» (в);
- г) фенотипический эффект «белка»;
- д) формирование фенотипа за счёт интерференции «фенотипов» «протеома». Взято из: [Knibbe et al., 2007].

Эволюционно-популяционные процессы: Переход от «одноуровневых» моделей к «многоуровневым»



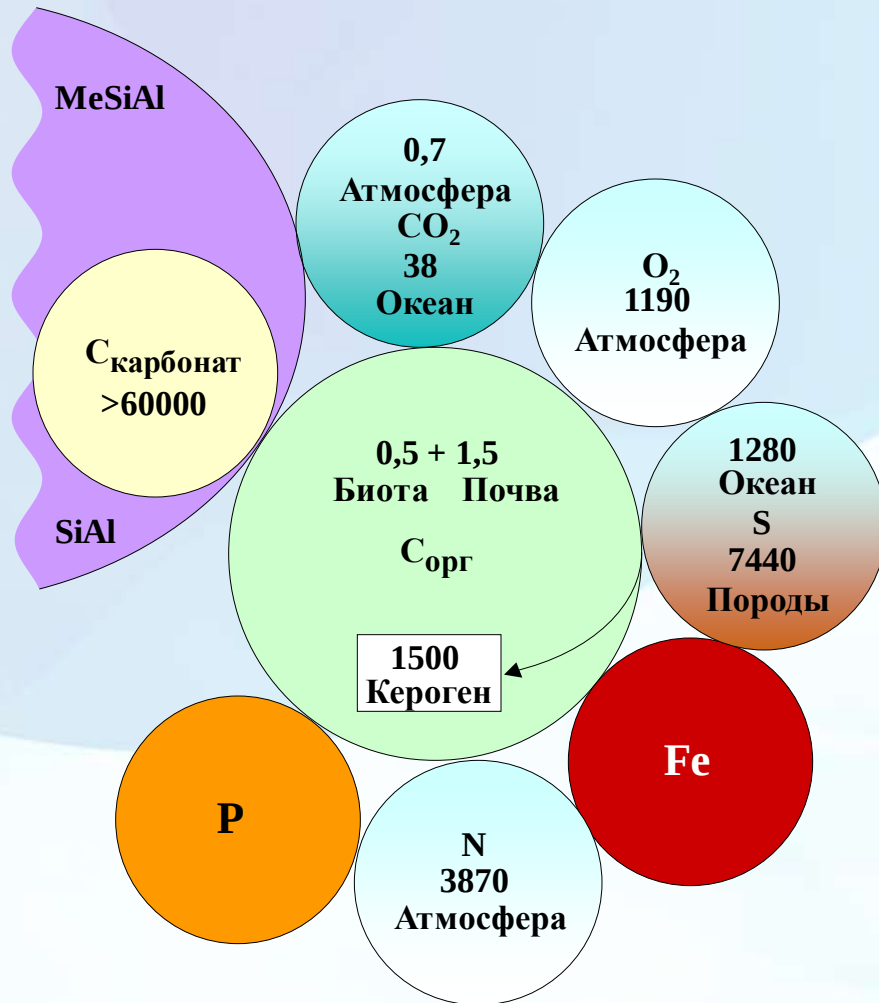
Схема построения комплексной модели эволюционного процесса

Многослойная архитектура модели: каждый слой представляет собой подмодель, ответственную за описание определённого уровня биологической организации:

- **генетический**
- **метаболический**
- **популяционный**
- **экоценотический**

Интерфейсы взаимодействия между моделями разных слоёв (входные и выходные данные) чётко специфицированы, но сами модели сохраняют относительную самостоятельность и могут различаться.

Роль прокариотических сообществ в функционировании биосферы

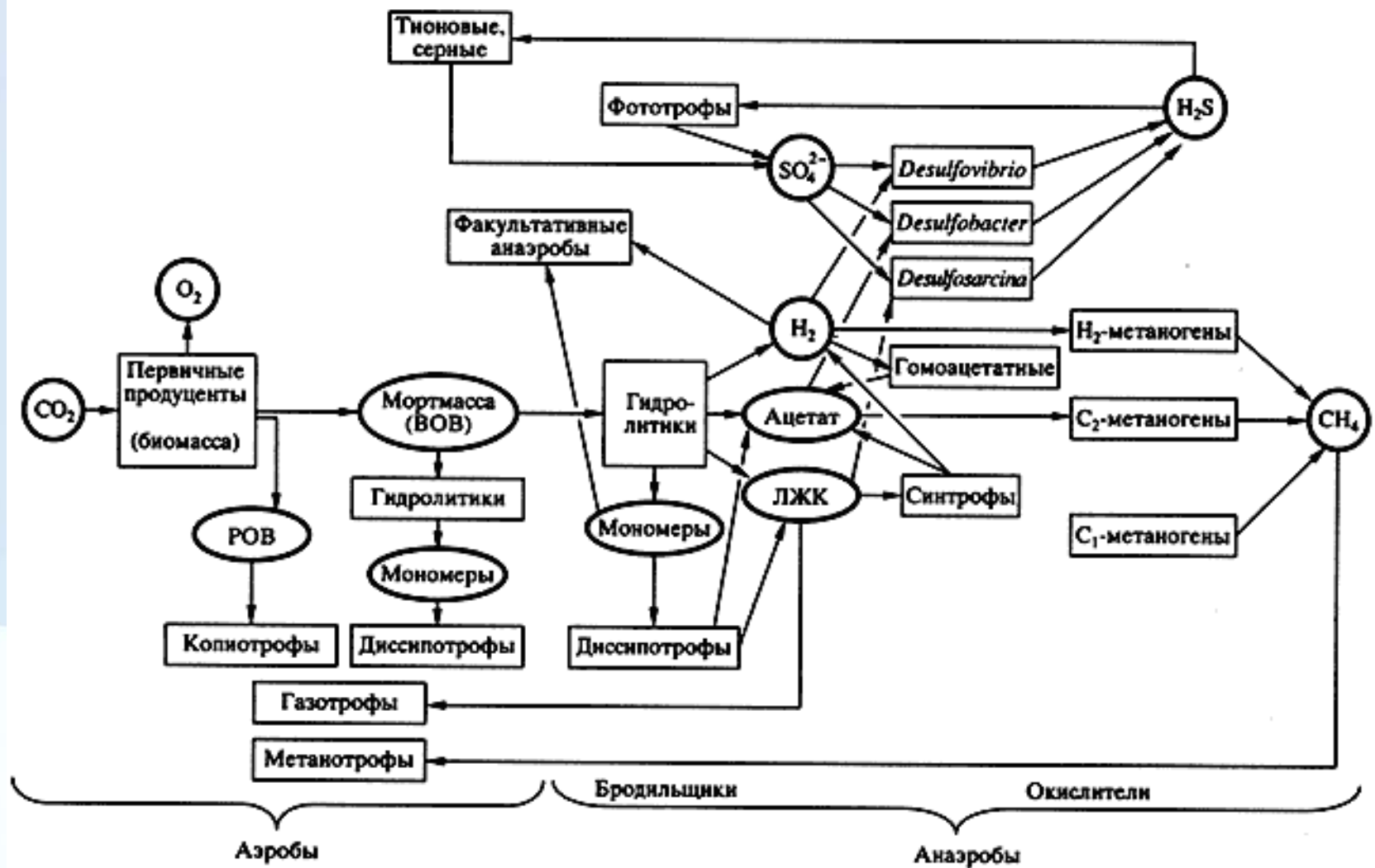


Сопряжение биогеохимических циклов

Все величины даны в 10^{18} г

(Заварзин Г.А., 2001,
Становление биосферы)

Граф трофических отношений в бактериальном сообществе



(Заварзин Г.А., 2001, Становление биосферы)

Недостатки основных подходов моделирования динамики бактериальных популяций

- Классические методы популяционной динамики (дифференциальные, разностные и алгебраические уравнения, стохастические подходы)
 - Полиморфизм невозможен или очень ограничен. Нет генетической структуры популяций.
 - «Предопределённая» структура модели: невозможно появление новых видов, заранее не описанных, невозможно появление новых аллелей.

Недостатки основных подходов к моделированию динамики популяций

- Классические методы теоретической популяционной генетики (разработаны для моделирования достаточно простой популяционной динамики)
 - как правило, фиксированный размер популяции
 - не моделируются трофические связи

Недостатки основных подходов к моделированию динамики популяций

- Портретное моделирование
 - Огромная вычислительная сложность
(миллиарды объектов для бактериальных популяций)

Имитационное моделирование

- Имитационные (*портретные, структурные, индивидуально-ориентированные*, англ. *simulation, individual-based*) подходы к моделированию - методы исследования, в котором изучаемая система заменяется моделью с достаточной точностью описывающей **объекты** и **процессы** реальной системы. С моделью проводятся эксперименты (*проигрываются сценарии*, жарг. *проводятся имитации/симуляции*) с целью получения информации об этой системе.
- Имитационная модель — логико-математическое описание объекта, которое может быть использовано для экспериментирования на компьютере в целях проектирования, анализа и оценки функционирования объекта.

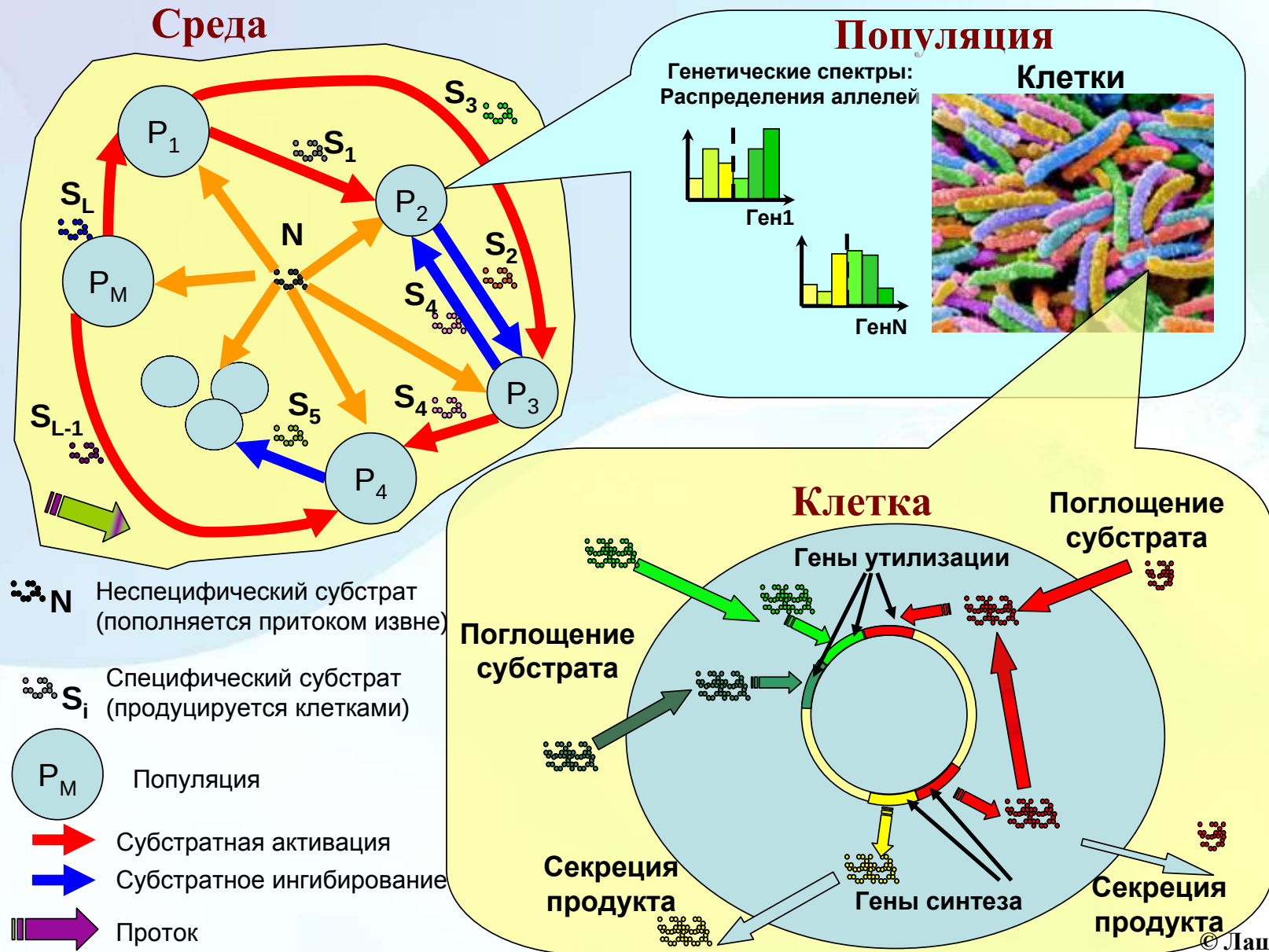
Особенности моделирования бактериальных сообществ

- Большой размер популяций ($>10^9$)
- Высокая частота возникновения мутаций
- Горизонтальный перенос генетического материала
- Полиморфизм (сложная генетическая структура популяции)
- Трофические взаимодействия
- Бактериальные системы – существенно динамические

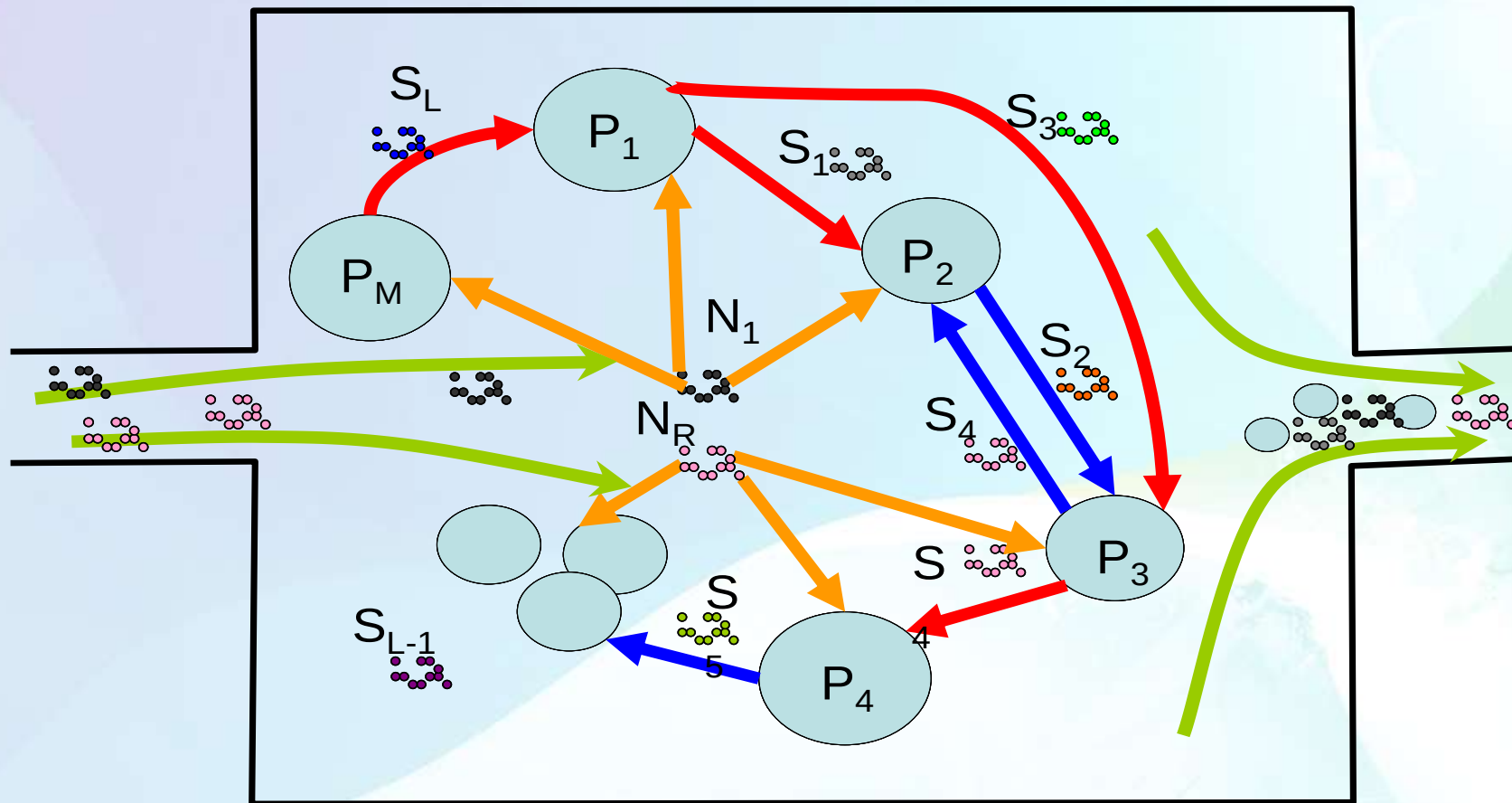
Эволюционный конструктор: постановка задачи

- Для адекватного моделирования эволюции бактериальных сообществ (или более общо — сообществ одноклеточных гаплоидных организмов) необходима методика моделирования, позволяющая строить и исследовать:
- Существенно динамические модели (*модели с изменяющимся количеством как переменных, так и уравнений*) сообществ высокой численностью (10^9 особей и более).
- Модели, **одновременно** рассматривающие следующие уровни организации:
 - Генетический
 - Популяционный
 - Экоценотический

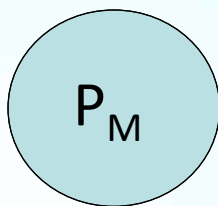
Эволюционный конструктор



Взаимодействие сообщества со средой обитания



Неспецифический
субстрат (НС)



Популяция



Специфический
субстрат (СС)

Проток



Субстратное



активирование



Субстратное
ингибирование

Мономорфные популяции

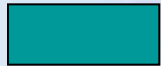
3 типа генов, определяющие процессы:



- Утилизации неспецифических субстратов (N_i)



- Утилизации специфических субстратов (S_i)



- Синтеза специфических субстратов (S_i)

- **Признак** – наличие процесса синтеза или утилизации какого-либо определённого субстрата
- **«Ген»** - единица наследования, однозначно определяющая *единственную* константу скорости реакции процесса (значение признака)
- **«Аллель»** – как вариант гена, это конкретное значение соответствующей константы скорости синтеза или утилизации субстрата
- **«Генотип» клетки** – набор констант-признаков (аллелей) трёх типов:
 - R_i – эффективность **утилизации неспецифического** субстрата
 - C_i – эффективность **утилизации специфического** субстрата
 - D_i – эффективность **синтеза** продукта
- **«Мономорфная» популяция** – множество «генетически идентичных» клеток, где у всех клеток соответствующие гены представлены одними и теми же аллелями. Общий для всех клеток такой популяции генотип называется **генотипом мономорфной популяции**
- **«Мутация»** – изменение величины признака-константы, что может интерпретироваться, как перевод гена в другое состояние (аллель)

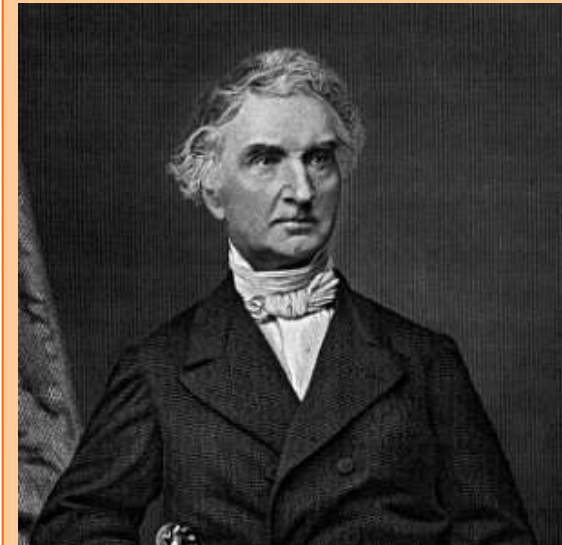
Трофические стратегии: Некомпенсаторное питание

$$\Delta P = F(\vec{N}, \vec{S}, \vec{R}, \vec{C}, P) = P \cdot \frac{\left(\frac{n_0/P}{K_{01}(r_0)} \right)^{\gamma_0}}{1 + \left(\frac{n_0/P}{K_{02}(r_0)} \right)^{\gamma_0}} \cdot \prod_{i \in I_{consumed}} \frac{1 + \left(\frac{s_i/P}{K_{i1}(c_i)} \right)^{\gamma_i}}{1 + \left(\frac{s_i/P}{K_{i2}(c_i)} \right)^{\gamma_i}} - k_{flow} \cdot P - k_{death} \cdot P^2$$

- **P** – размер популяции
- **N, S** – векторы поглощённых субстратов (соотв., неспецифических и специфических)
- **R, C** – векторы констант скоростей (аллели)
- γ_i, K_{ij} – коэффициенты нелинейного влияния субстратов на скорость роста популяции
- k_{flow} – константа скорости потока
- k_{death} – константа смертности



Закон Минимума Либиха



Ю. фон Либих

Трофические стратегии: Компенсаторное питание

$$\Delta P = F_1(\vec{N}, \vec{S}, \vec{C}, P) = \sqrt{r_0 n_0(P) \cdot \sum_{i \in I_{consumed}} c_i s_i(P) - k_{flow} \cdot P - k_{death} \cdot P^2}$$

- **P** – размер популяции
- **N, S** – векторы поглощённых субстратов (соотв., неспецифических и специфических)
- **R, C** – векторы констант скоростей (аллели)
- **k_{flow}** – константа скорости протока
- **k_{death}** – константа смертности

© Лашин С.А.

Зкон компенсации экологических факторов



Э.А. Рубель



В.В. Алехин

Трофические стратегии: «Нелинейное» питание

$$\Delta P = F_4(\vec{N}, \vec{S}, \vec{C}, P) = \frac{a_{basal}(n_0)}{1 + \left(\sum_{i \in I_{consumed}} c_i s_i(P) \right)^\gamma} - k_{death} \cdot P$$

- P – размер популяции
- N, S – векторы поглощённых субстратов (соотв., неспецифических и специфических)
- R, C – векторы констант скоростей (аллели)
- k_{flow} – константа скорости потока
- γ – коэффициент нелинейности

Биологические осцилляторы Мэки-Гласса



М. Мэки



Л. Гласс

Трофические стратегии: ингибирование специфическим субстратом

$$\Delta P = F_3(\vec{N}, \vec{S}, \vec{C}, P) = a_{basal}(n_0) \cdot P - \sqrt{\sum_{i \in I_{consumed}} c_i s_i(P)} - k_{death} \cdot P^2$$

- **P** – размер популяции
- **N, S** – векторы поглощённых субстратов (соотв., неспецифических и специфических)
- **R, C** – векторы констант скоростей (аллели)
- **k_{flow}** – константа скорости потока
- **k_{death}** – константа смертности

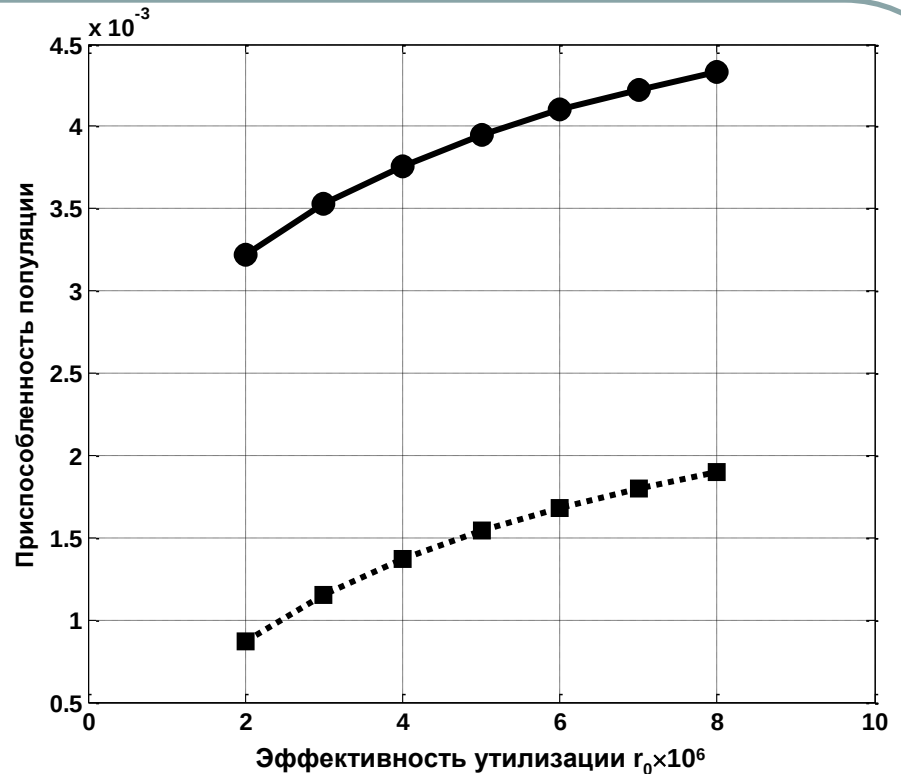
Итерационный шаг



Моделирование теоремы Фишера

Приспособленность мономорфной популяции в зависимости от эффективности утилизации неспецифического субстрата r_0 в лаг-фазе (нижняя линия) и экспоненциальной фазе (верхняя линия).

Приспособленность рассчитывалась как отношение $(P(t+1)-P(t))/P(t)$, где $P(t)$ – численность популяции в момент времени t .



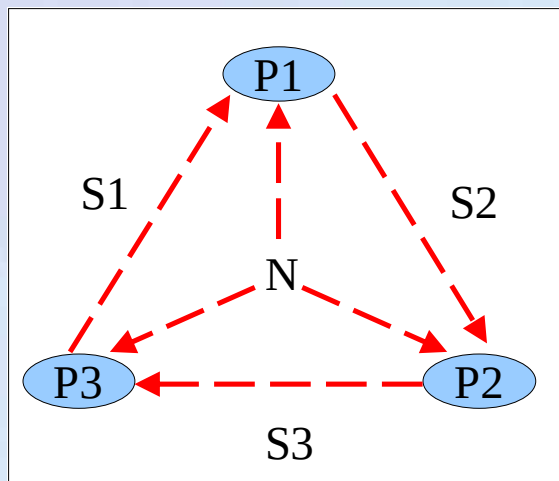
"The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time."

Fisher, R.A. The Genetical Theory of Natural Selection

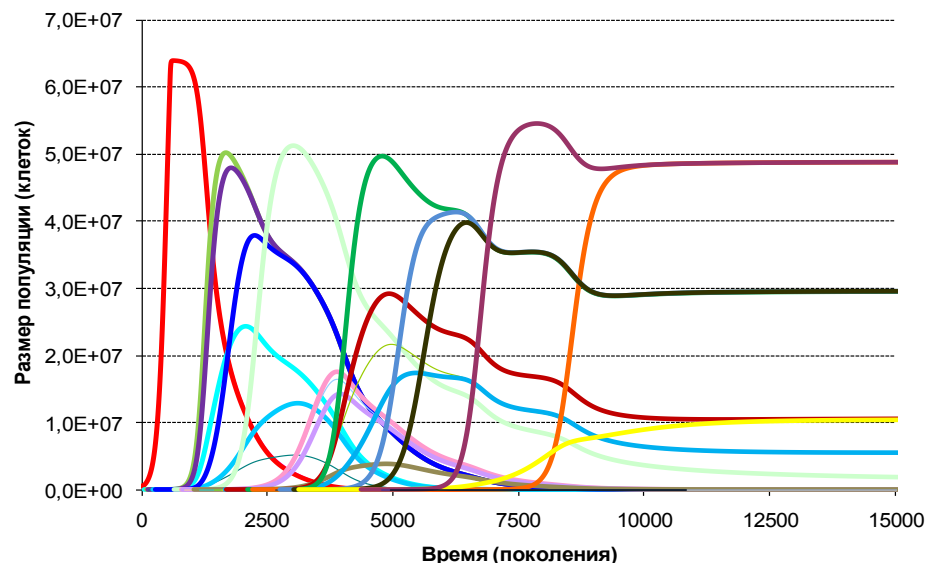
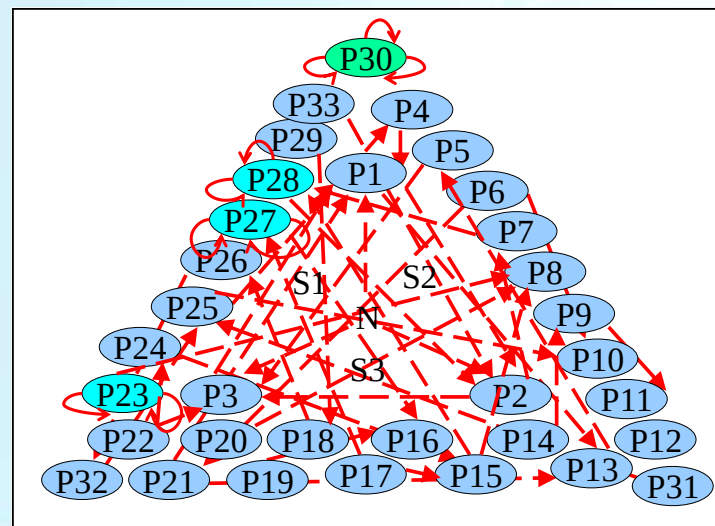
Программный комплекс «Гаплоидный эволюционный конструктор»

- Моделирование сообществ, состоящих из сотен популяций. Размер метапопуляции до 10^{20} клеток
- Моделирование «тяжёлого» полиморфизма: до 15000 аллелей в ОДНОЙ популяции (по 10-15 генам)
- Моделирование множественного видообразования (увеличение числа популяций в сообществе более чем в сто раз относительно начальных значений)

Эволюционное усложнение метаболизма

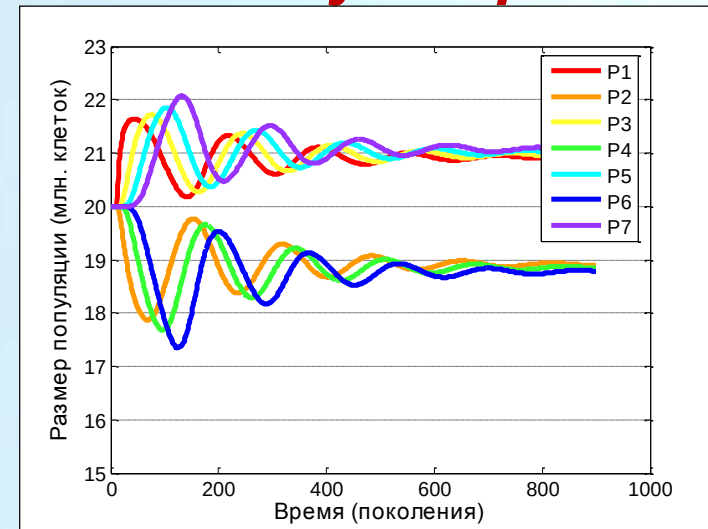
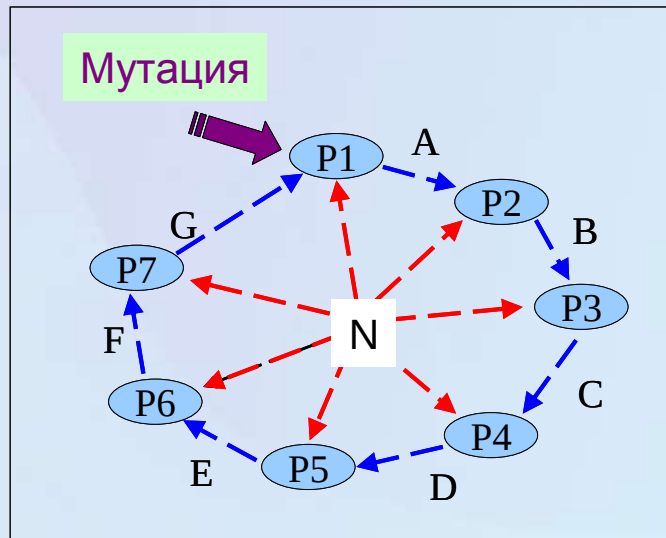


15000
поколений

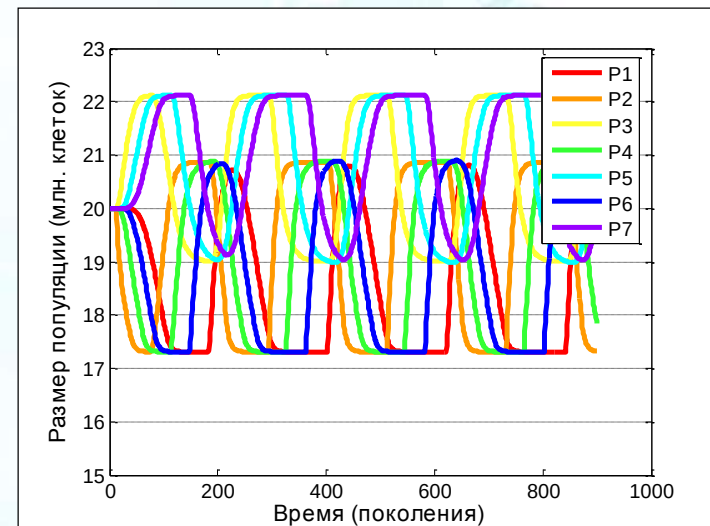
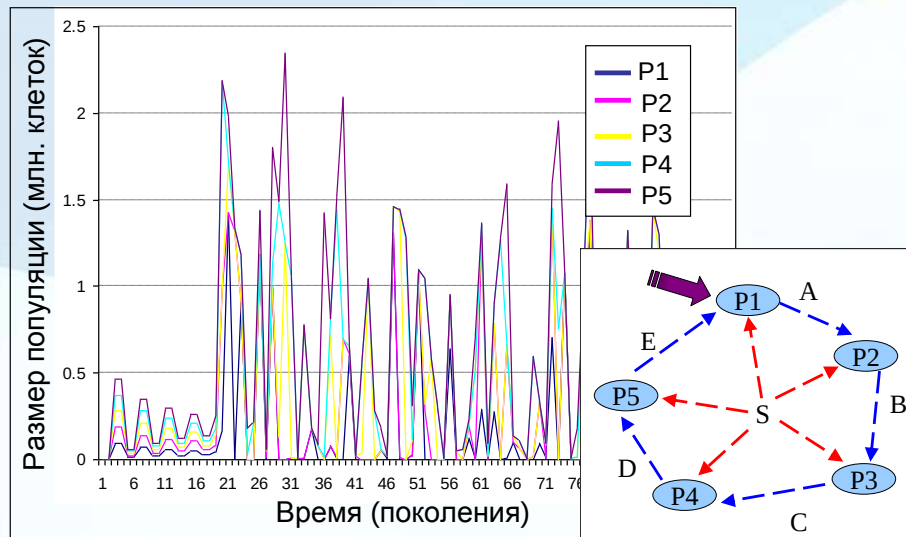


Итоговое место популяции (по численности)	Номер	Генотип (схема)	Итоговая численность (особей)
1	P30		$4.86 \cdot 10^7$
2	P24		$4.85 \cdot 10^7$
3	P28		$2.94 \cdot 10^7$
4	P23		$2.94 \cdot 10^7$
5	P27		$2.94 \cdot 10^7$
---	---	---	---
18-33	P1		0
18-33	P2		0
18-33	P3		0

Моделирование трофических колец с взаимным ингибированием специфическими субстратами



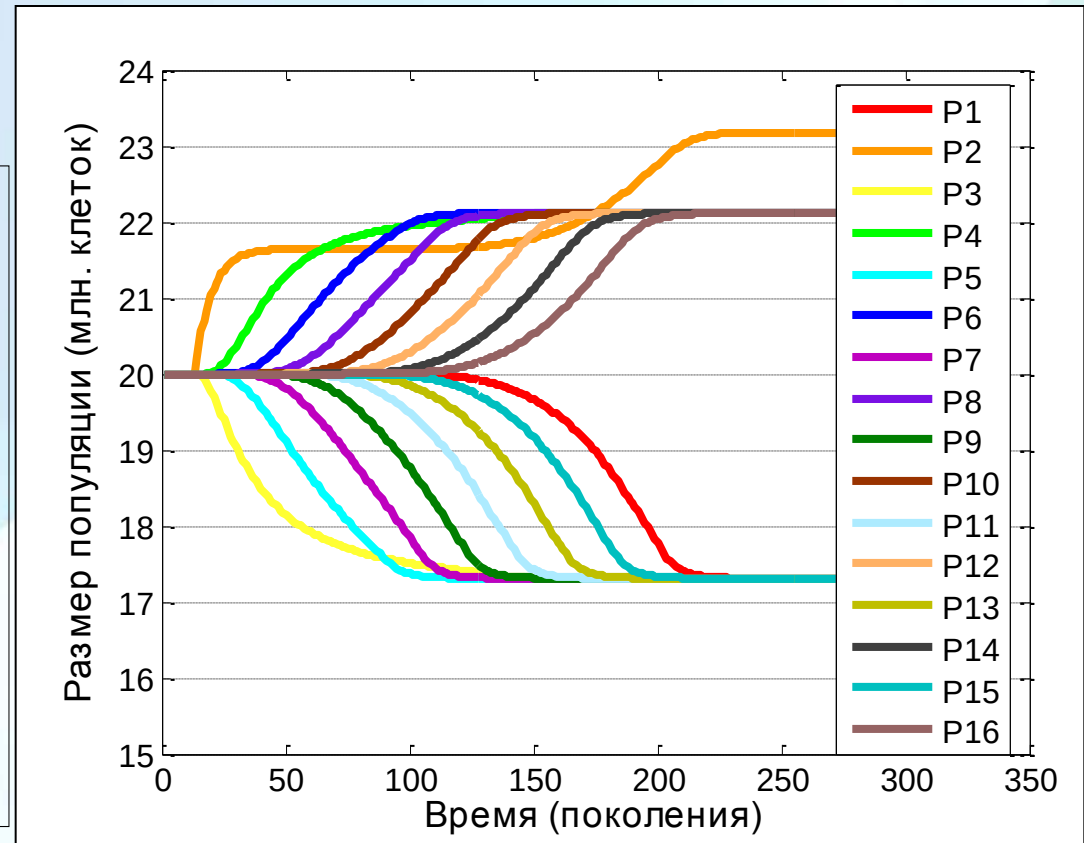
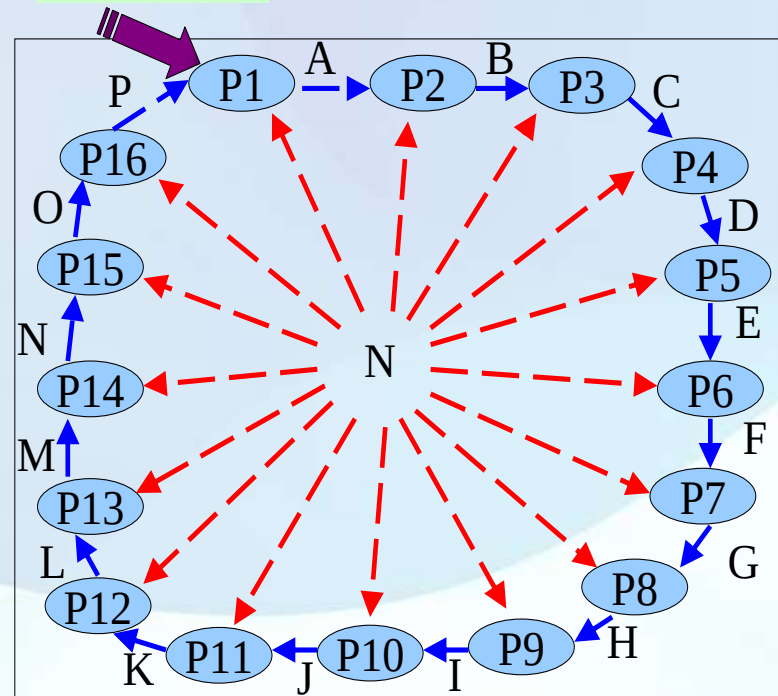
Увеличение чувствительности к ингибитору G популяцией P1



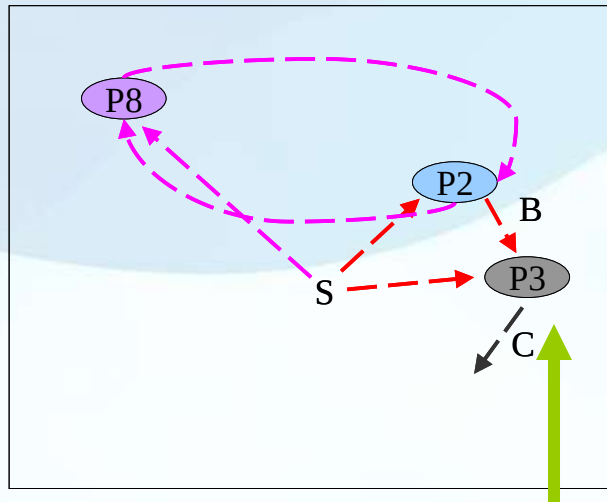
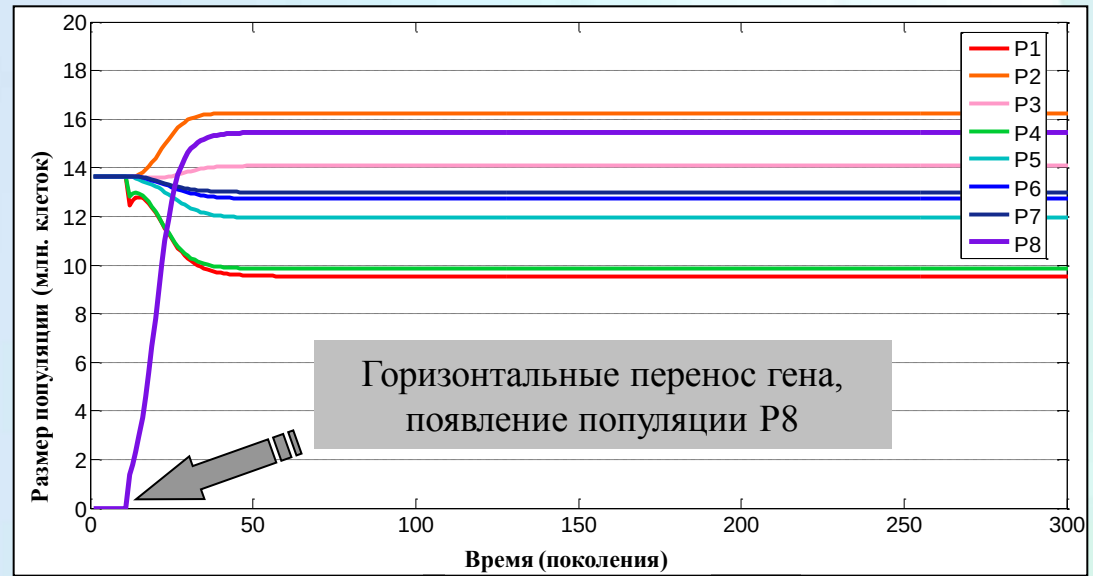
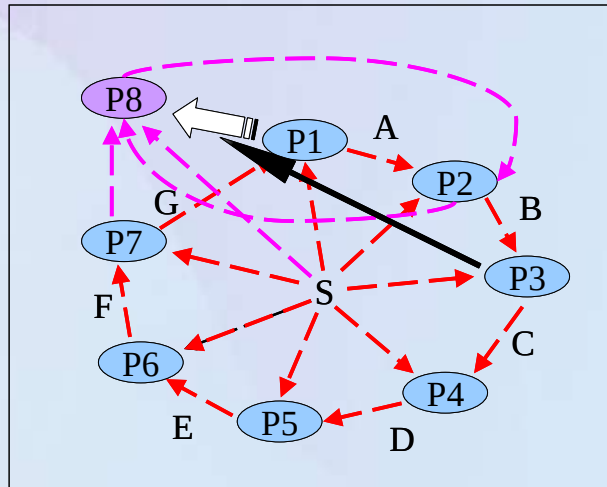
Увеличение эффективности синтеза ингибитора A популяцией P1

Моделирование трофических колец с взаимным ингибированием специфическими субстратами. Чётное число популяций в кольце

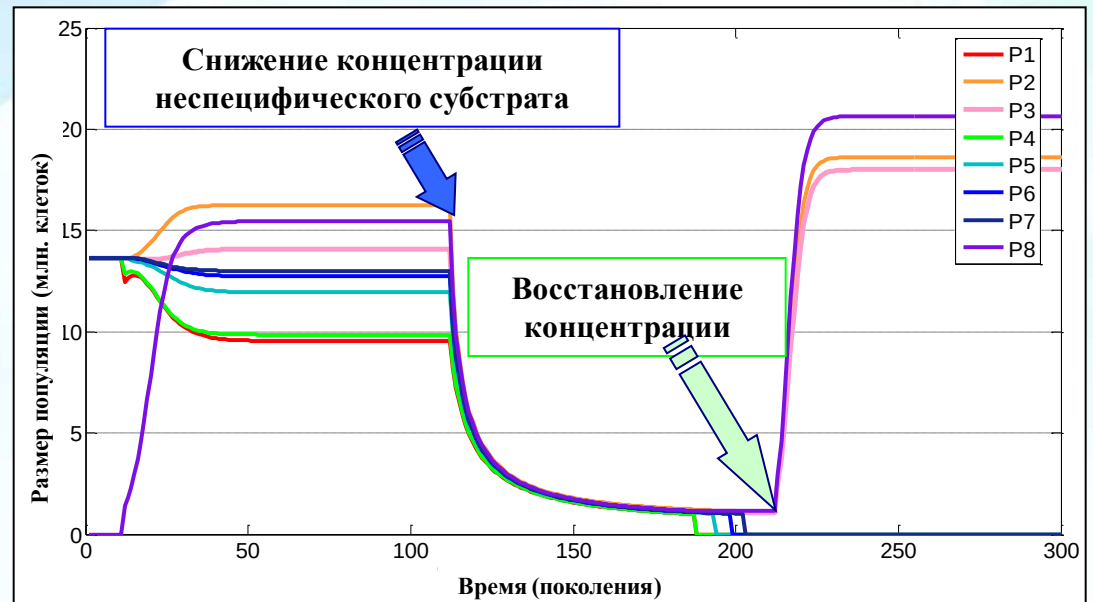
Мутация



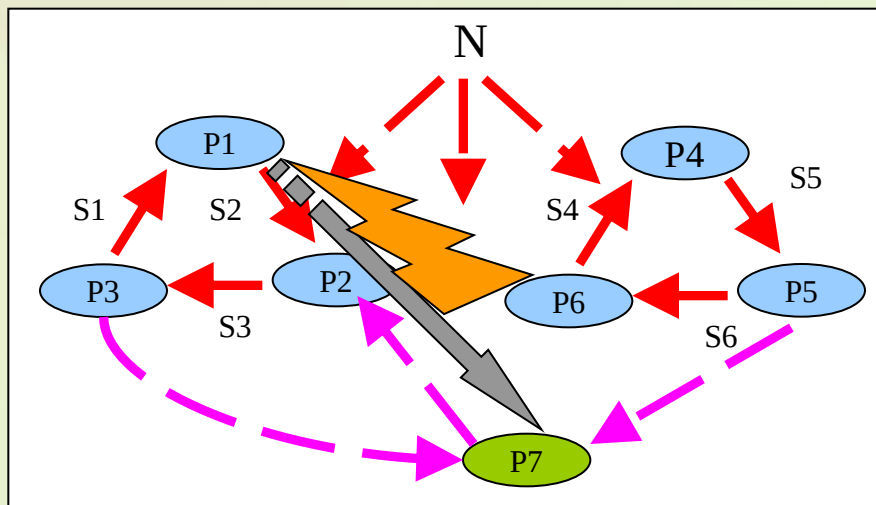
Моделирование горизонтального переноса



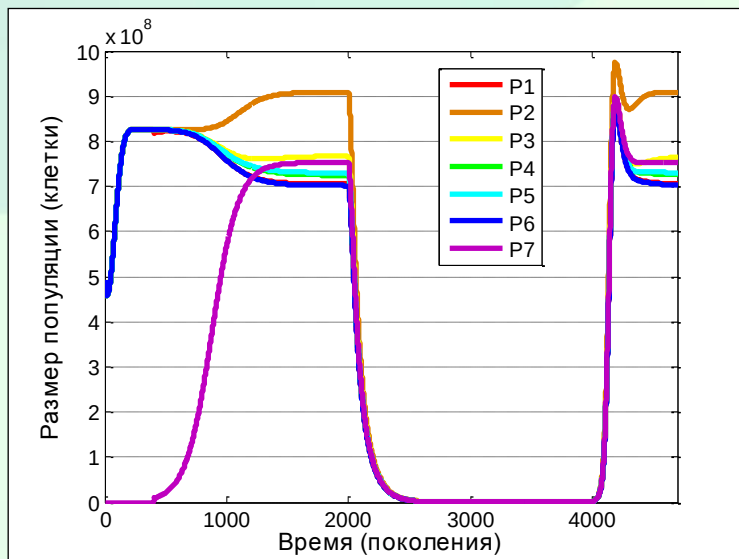
Метаболический тупик



Сравнение компенсаторного и некомпенсаторного питания



Некомпенсаторное питание: Сохранение биоразнообразия



Компенсаторное питание: Уменьшение биоразнообразия

