



# **Введение в предмет молекулярной эволюции.**

***Д.А.Афонников***

**лаборатория теоретической генетики**

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia



# Предмет молекулярной эволюции



**Эволюция – развитие, постепенное изменение во времени.**

**Молекулярная эволюция – наука об эволюционных изменениях генетических макромолекул.**



# Основные задачи



- **Выявление закономерностей эволюции генетических макромолекул**
- **Реконструкция эволюционной истории генов и организмов**



## **Взаимосвязь с другими областями науки**



- **Палеонтология (датировка эволюционных событий)**
- **Генетика (принципы организации и передачи наследственной информации)**
- **Молекулярная биология, биофизика (функционирование генетических молекул)**
- **Математика (построение моделей эволюции)**
- **Информатика (обработка и анализ данных)**



## Исторический экскурс (XIX век)



- Креационизм (все сущее создано сверхъестественной силой и неизменно во времени)
- Ламарк, 1809 (влияние внешней среды, теория трансформации)
- Дарвин, 1859 («Происхождение видов»)



# Теория естественного отбора (Дарвин, 1859)



- **Внутривидовые вариации (случайные)**
- **Приспособленность организмов (способность давать большее потомство)**
- **«Перевоспроизводство» потомства**
- **Естественный отбор наиболее приспособленных особей (по репродуктивности)**

**Резюме :эволюция как «хаос с обратной связью»**



## Исторический экскурс (первая половина XX в.)



- Мендель, 1865 (генетическая теория наследственности)
- де Фриз, Вейсман, Морган, 1900-1920 (становление и развитие генетики)
- Фишер, Райт, Холдейн, 30-е годы (теории популяционной генетики, нео-дарвинизм)



## Исторический экскурс (вторая половина XX в.)



- Уотсон, Крик, Перутц, 50-е – 60-е годы (установление структуры генетических макромолекул)
- Методы секвенирования ДНК, 60-е годы
- Кимура, 60-е годы (теория нейтральной эволюции)
- Секвенирование полных геномов десятков организмов, конец XX в.





# Исторический экскурс (постгеномная эра)



- Развитие теории генных сетей
- Секвенирование полного генома человека, 2001-2003 гг.
- Сравнительный анализ полных геномов различных организмов
- Структурная биология



# Объект исследования молекулярной эволюции



- **ДНК (носитель генетической информации)**
- **РНК (участвуют в базовых биохимических процессах клетки – синтез аминокислот, белка, катализ)**
- **Белки (выполняют структурные, регуляторные, каталитические функции)**



# Особенности структуры и функции ДНК



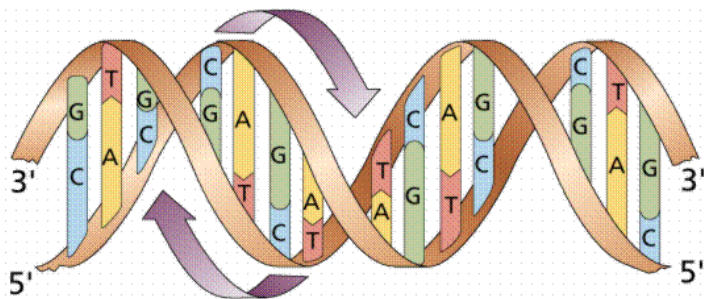
- **Основная функция – кодирование генетической информации**
- **Передается наследственным путем**
- **Подвержена случайным изменениям в результате спонтанных мутаций**



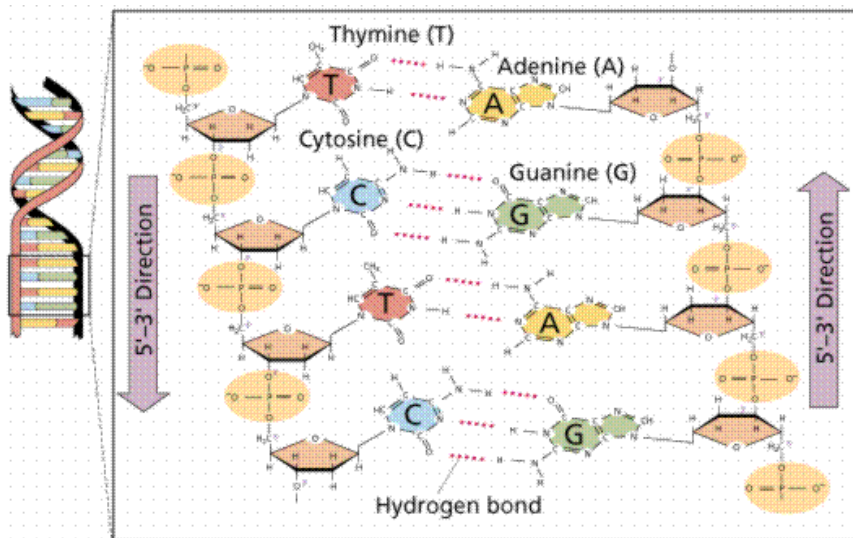
# Структура ДНК



CACCAGAAAGTCCTG



- Первичная структура – последовательность нуклеотидов
- 4 нуклеотида А,Т,Г,С, которые различаются основаниями
- Две комплиментарные нити
- Правила образования пар:
  - А-Т (слабые)
  - Г-С (сильные)
- Возможно описание в виде строк символов



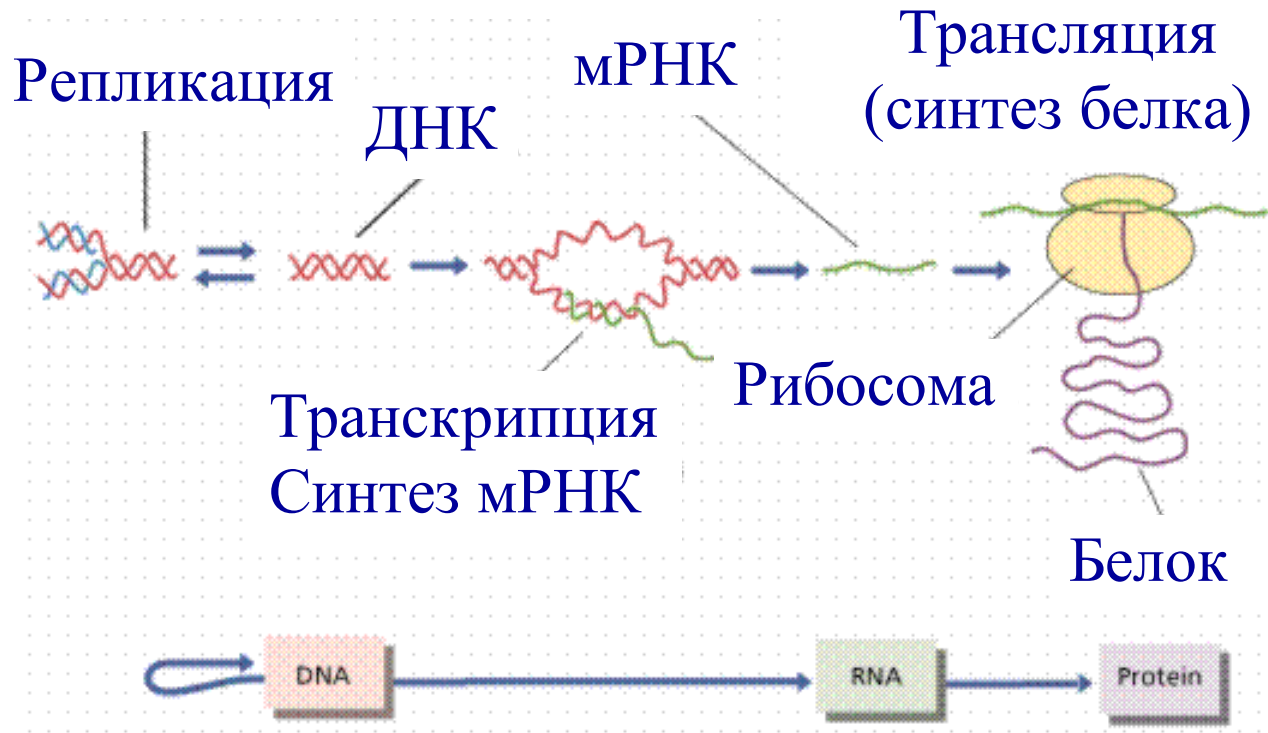
Адаптировано из **On-Line Biology Book** ©1995, 1999, 2000, 2001, by M.J. Farabee  
<http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookTOC.html>



# Реализация генетической информации



## Центральная догма биологии



Адаптировано из **On-Line Biology Book** ©1995, 1999, 2000, 2001, by M.J. Farabee  
<http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookTOC.html>



# Организация геномной ДНК



- **Кодирующая часть (функция – кодирование РНК и белков). Доля генома, кодирующая белок у человека составляет <math><2\%</math>**
- **Некодирующая часть (содержит регуляторные районы, повторы и районы, функция которых пока неизвестна). У человека составляет  $\sim 50\%$ .**
- **Соотношение кодирующей/некодирующей частей различно для разных организмов**



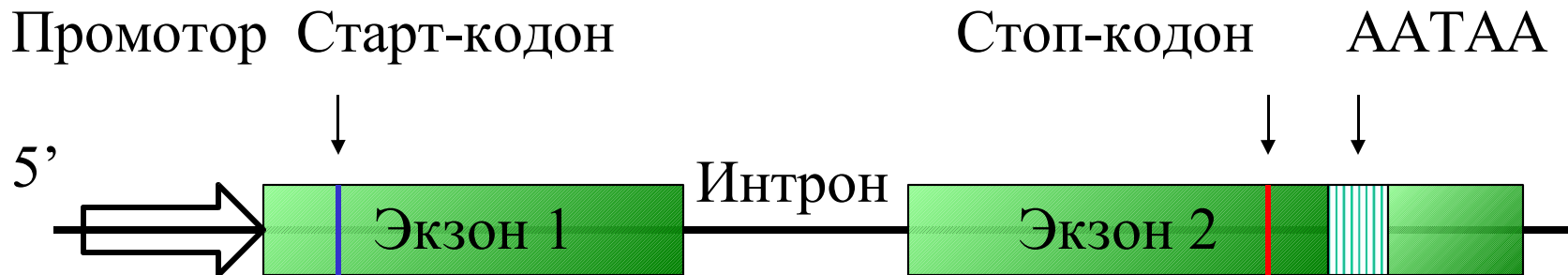
- Концепция гена изменялась на всем протяжении развития генетики и биологии
- **«Ген - это сегмент ДНК, который обуславливает фенотип или функцию. В отсутствии данных о функции ген может быть охарактеризован последовательностью, транскриптом или по гомологии»** (HUGO, J.A. White et. al. Guidelines for Human Gene Nomenclature HGNC Human Genome Nomenclature Committee, 1997) .
- Оценка числа генов ~30,000 (у человека)



# Схема структуры гена



- Гены кодируют РНК или белки и могут содержать кодирующую часть (экзон), некодирующую часть (интрон) и регуляторные зоны (промотор)



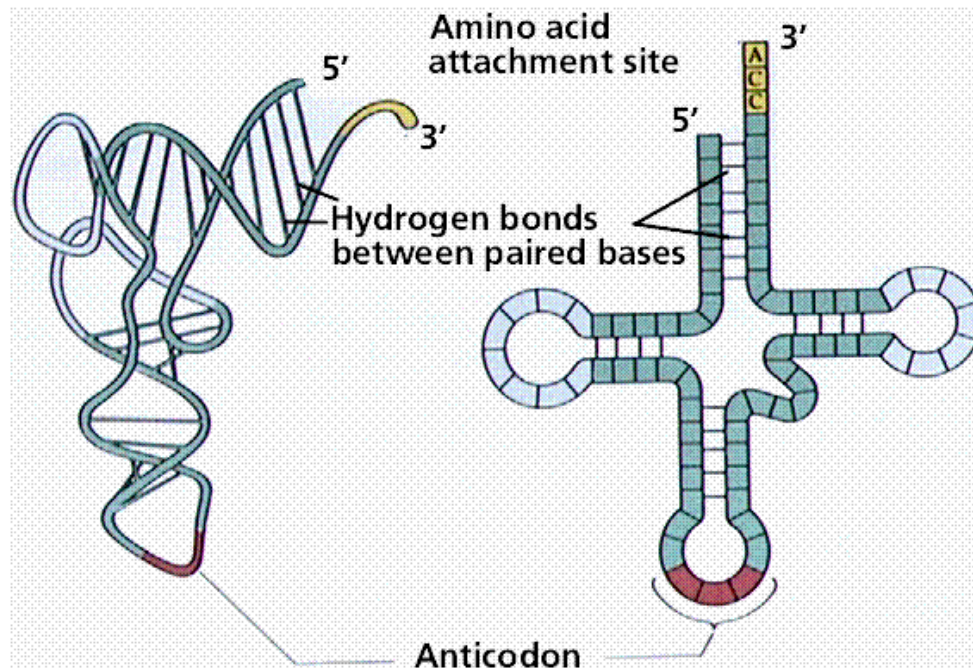




# Структура РНК



## Структура тРНК

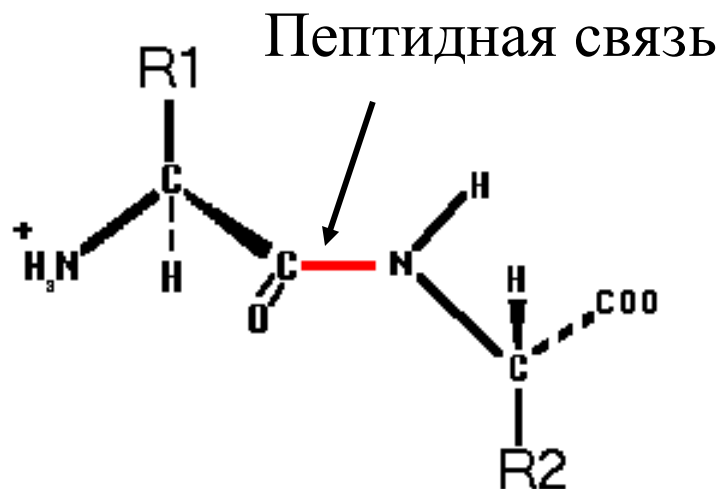
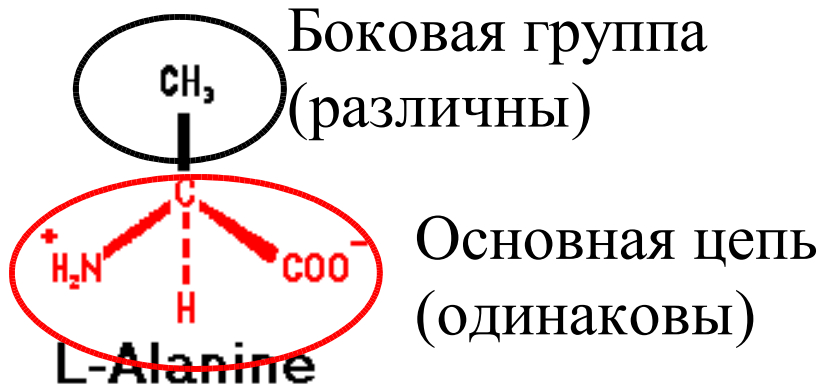


- **Относительно короткие цепи (~100 – 1000 п.о.)**
- **4 мономера А, U, G, С (замена Т↔U)**
- **Одна нить**
- **Нелинейна**
- **Вторичная структура: петли и шпильки (на участках комплиментарности )**
- **Третичная структура : дальние контакты**

Рисунок адаптирован из **On-Line Biology Book** ©1995, 1999, 2000, 2001, by M.J. Farabee  
<http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookTOC.html>



# Структура белков



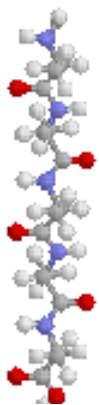
- Относительно короткие цепи (~100 – 1000 а.к.)
- 20 мономеров (различаются боковыми группами)
- Пространственная структура нелинейна
- Вторичная структура:  $\alpha, \beta, \tau$
- Третичная структура (глобула) :
  - дальние контакты
- Гетерогенность пространственной структуры



# Уровни структурной организации белков



Первичная Вторичная Третичная



Последовательность аминокислот



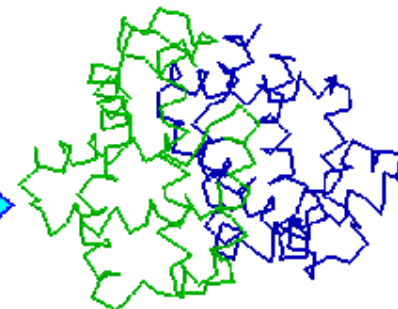
Локально упорядоченная осн. цепь



Укладка всей цепи, за счет дальних взаимодействий.



Четвертичная



Укладка нескольких цепей (мономеров)

Рисунок адаптирован из TUTORIAL ON PEPTIDE AND PROTEIN STRUCTURE

©J. E. Wampler, 1996 <http://bmbiris.bmb.uga.edu/wampler/tutorial/prot0.html>



# Генетический код



Вторая позиция кодона

	<b>U</b>	<b>C</b>	<b>A</b>	<b>G</b>		
<b>Первая позиция кодона</b>	<b>U</b>	Phe	Ser	Tyr	Cys	<b>Третья позиция кодона</b>
		Phe	Ser	Tyr	Cys	
		Leu	Ser	<b>STOP</b>	<b>STOP</b>	
		Leu	Ser	<b>STOP</b>	Trp	
<b>C</b>	Leu	Pro	His	Arg	U	
	Leu	Pro	His	Arg	C	
	Leu	Pro	Gln	Arg	A	
<b>A</b>	Ile	Thr	Asn	Ser	U	
	Ile	Thr	Asn	Ser	C	
	Ile	Thr	Lys	Arg	A	
	Met	Thr	Lys	Arg	G	
<b>G</b>	Val	Ala	Asp	Gly	U	
	Val	Ala	Asp	Gly	C	
	Val	Ala	Glu	Gly	A	
	Val	Ala	Glu	Gly	G	

Рисунок Darryl Leja  
<http://www.accessexcellence.org/AB/GG/codon.html>

Рисунок из Молекулярная биология © 2000  
Дымшиц Г.М.  
<http://www.nsu.ru/education/biology/molbio/>



# Генные мутации



- Изменения сегментов гена
  - Транслокации
  - Дупликации
  - Инверсии
  - Делеции
- Точечные мутации - замены одного нуклеотида



# Типы точечных мутаций



## Мутации, приводящие к кистозному фиброзу (пример)

Patient	Mutation	Result
A	482 C G C ↓ C A C	Arg-117 ↓ His-117
B	1609 C A G ↓ T A G	Gln-493 ↓ STOP
C	Insertion of 2 nucleotides (AT) at 2566	Frameshift
D	Deletion of one C at 3659	Frameshift
E	Deletion of 3 nucleotides at 1654-1656	Deletion of Phe-508

Миссенс-мутация  
(замена аминокислоты)

Нонсенс-мутация  
(замена на стоп-кодон)

Вставка 2 нуклеотидов  
(сдвиг рамки)

Делеция нуклеотида  
(сдвиг рамки)

Делеция 3 нуклеотидов  
(делеция аминокислоты)

Рисунок адаптирован из  
<http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/M/Mutations.html>



# Классическая теория частот генов



Ген имеет две формы (аллеля) : «дикий» тип А и мутантный В. Частоты генов А( $p$ ), В( $q$ ). ( $p+q=1$ )

Генотип	Частота
АА	$p^2$
АВ	$2pq$
ВВ	$q^2$

Закон Харди-Вайнберга:

$$p^2+2pq+q^2=1$$

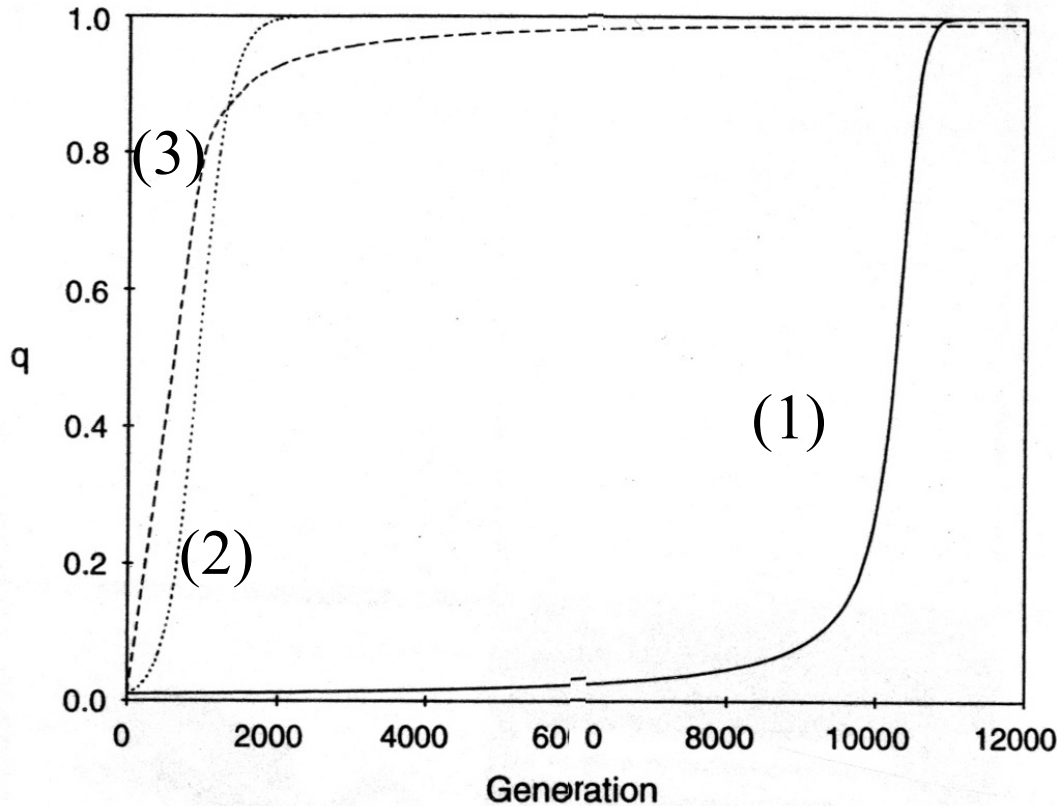
Пример - заболевание кистозным фиброзом у народностей кавказа, проживающих в США:

$$q^2=1/1700; q=0.024;$$

$$p=0.976; p^2=0.953$$



# Влияние приспособленности на фиксацию мутаций в бесконечно большой популяции



Генотип	Приспособленность.
AA	1
AB	$1 + s_{AB}$
BB	$1 + s_{BB}$

(1):  $s_{BB}=0.01$ ;  $s_{AB}=0.00$

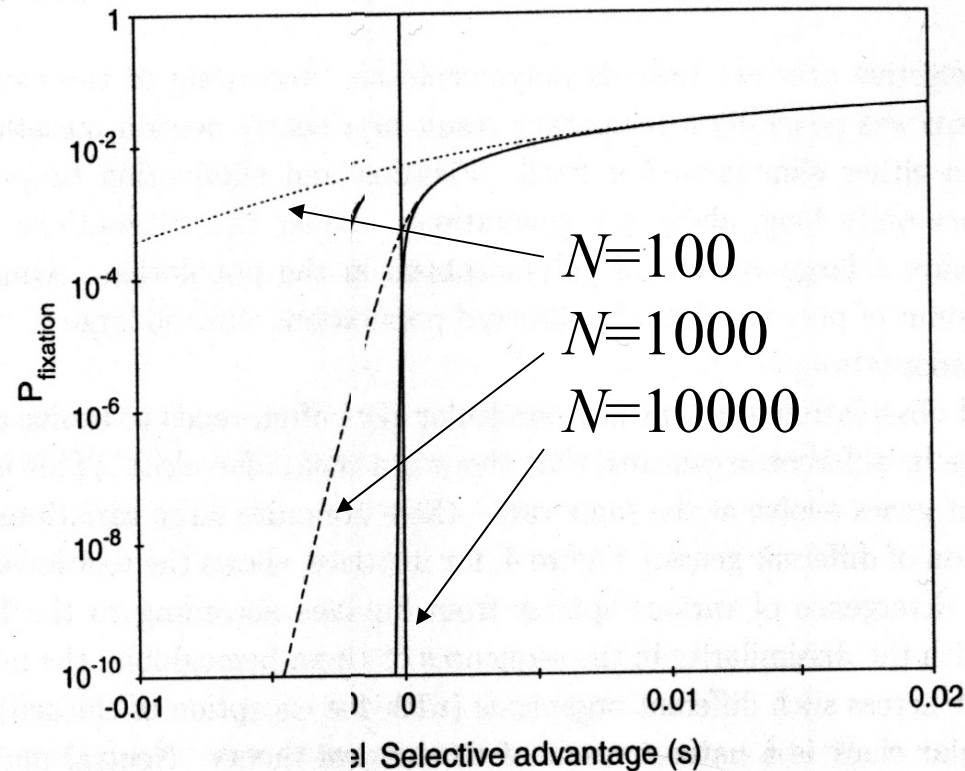
(2):  $s_{BB}=0.01$ ;  $s_{AB}=0.05$

(3):  $s_{BB}=0.01$ ;  $s_{AB}=0.01$





# Влияние приспособленности на фиксацию мутации в конечной популяции



Кимура показал, что для диплоидных популяций размера  $N$

$$P_{fix} = \frac{1 - e^{-(2N_e s)/N}}{1 - e^{-4N_e s}}$$

$N_e$  – число особей популяции, способных к воспроизводству (для человека  $N_e \sim N/3$ )

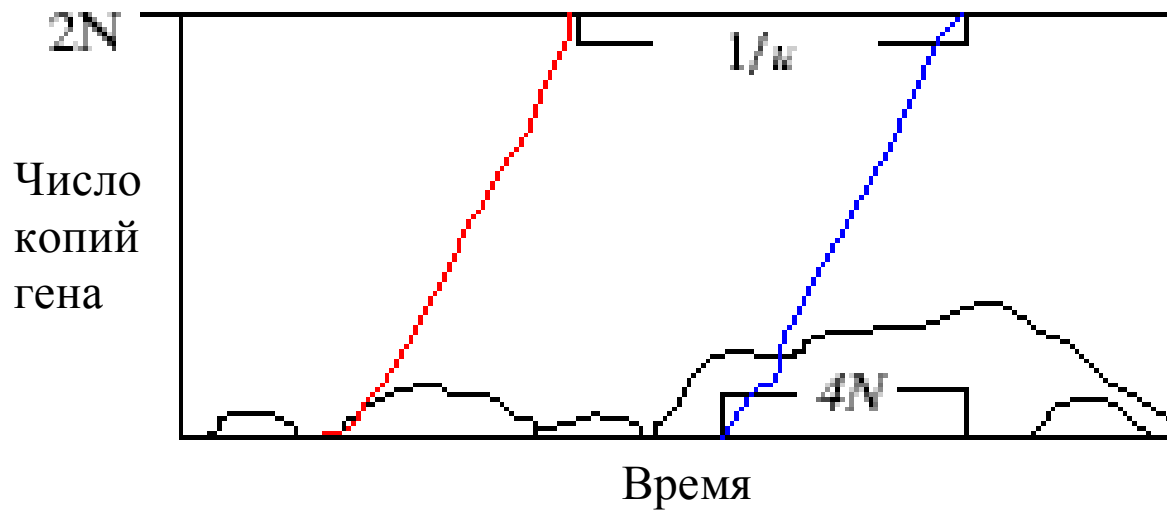
Вывод : в силу случайных флуктуаций даже вредные мутации могут фиксироваться



# Нейтральная теория Кимуры



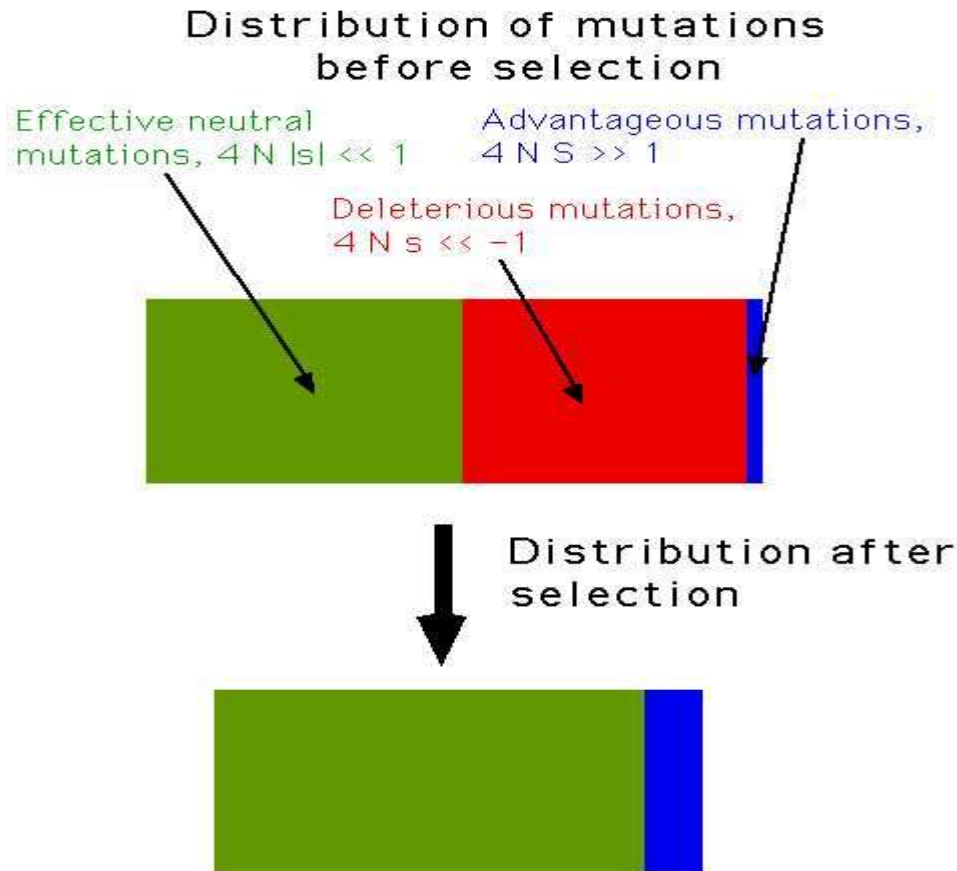
Теория Кимуры объясняет высокую степень полиморфизма белков. Пример схематически показан на диаграмме:



Время фиксации *нейтральных* мутаций  $t_{\text{fix}} = 4N$  поколений  
Скорость фиксации *нейтральных* замен (число фиксаций за единицу времени)  $K = u$ ,  $u$  - частота возникновения мутации  
Для адаптивных замен  $K = 4Nsu$



# Постулаты нейтральной теории

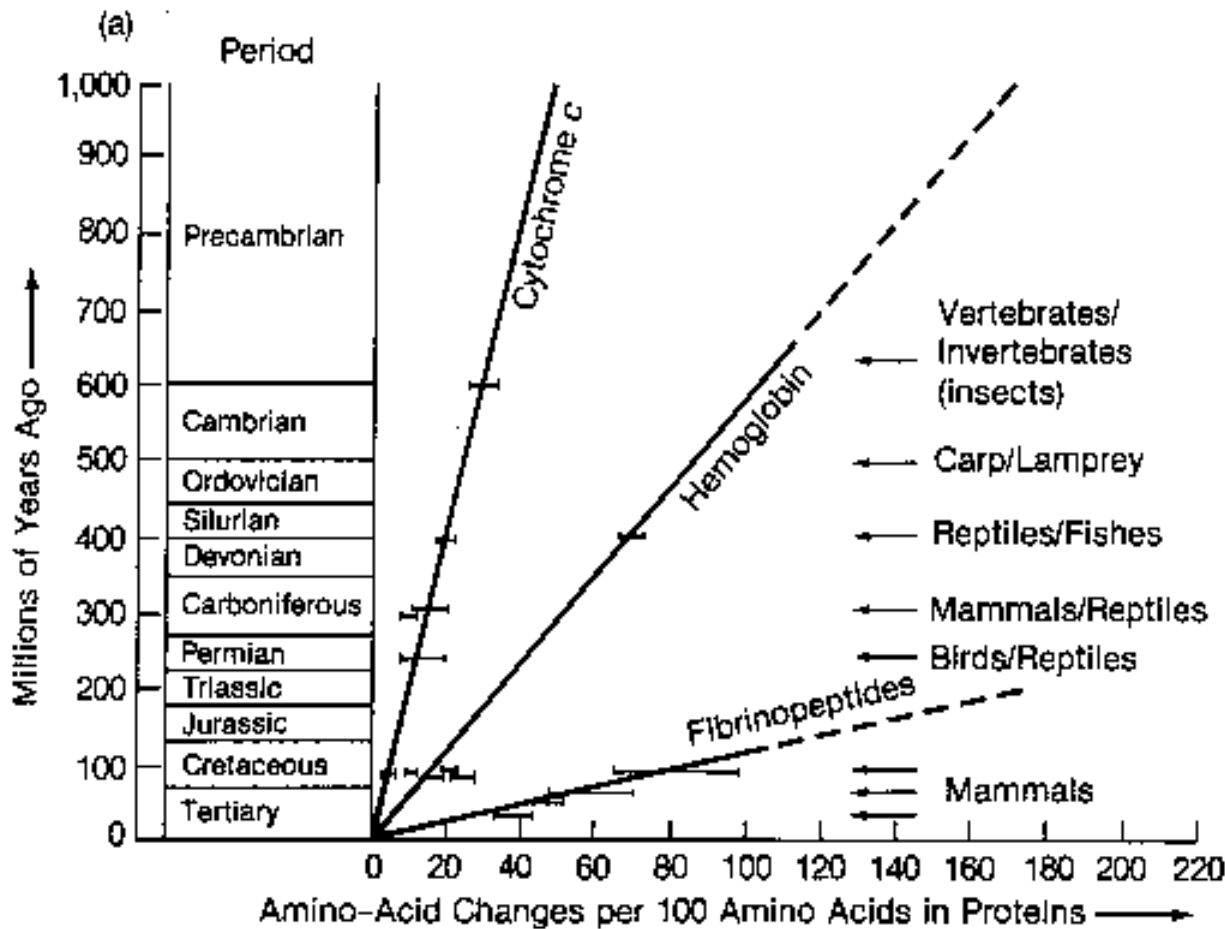


- Адаптивные мутации очень редки
- Деструктивные мутации быстро элиминируются
- Подавляющее большинство фиксированных мутаций селективно нейтральны
- Скорость фиксации замен в генах является постоянной (гипотеза молекулярных часов, Zuckercandle & Poling)
- Функционально важные районы функционируют медленнее, менее важные - быстрее

Рисунок из <http://nitro.biosci.arizona.edu/courses/EEB320/Lecture46/Lecture46.html>



# Является ли скорость замен постоянной?



Для большого числа белков наблюдаются постоянные скорости замен, однако для разных белков они различны

Рисунок из <http://nitro.biosci.arizona.edu/courses/EEB320/Lecture46/Lecture46.html>



# Сравнение скоростей синонимических/несинонимических замен



Скорости нуклеотидных замен на 1 позицию на 1000 млн. лет при сравнении некоторых белков человека и грызунов (время дивергенции ~80 млн.лет согласно ископаемым)

Ген	Число кодонов	К(син.)	К(несин.)
histone 3	135	0.00	6.38
actin beta	349	0.03	3.13
myoglobin	153	0.56	4.44
In-gamma	136	2.79	8.59



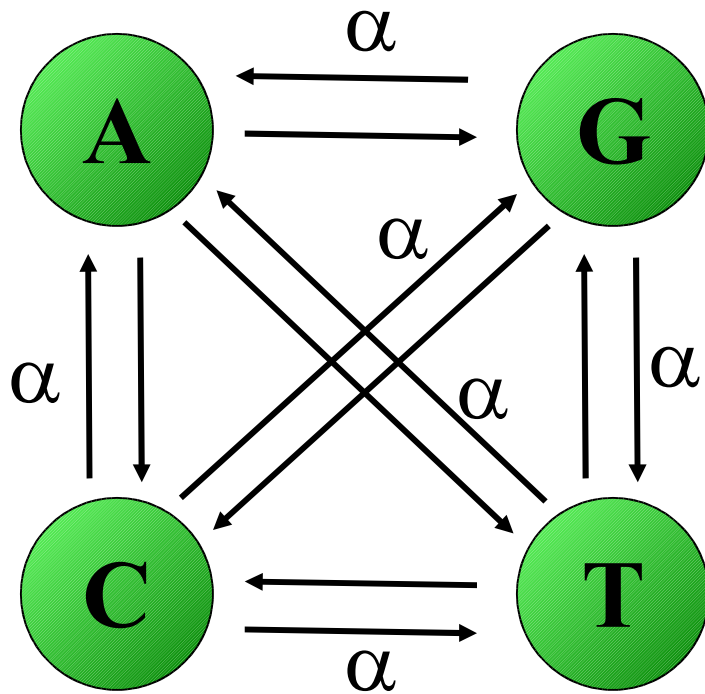
# Модели эволюции последовательностей



- Описывают временные изменения в последовательностях генетических макромолекул
- Для одиночных замен используются Марковские модели с конечным числом состояний
- Развитие от простого к сложному



# Модель Джукса-Кантора



Замены случайны,  
независимы о других  
позиций, равновероятны с  
вероятностью  $\alpha$

Дискретное приближение

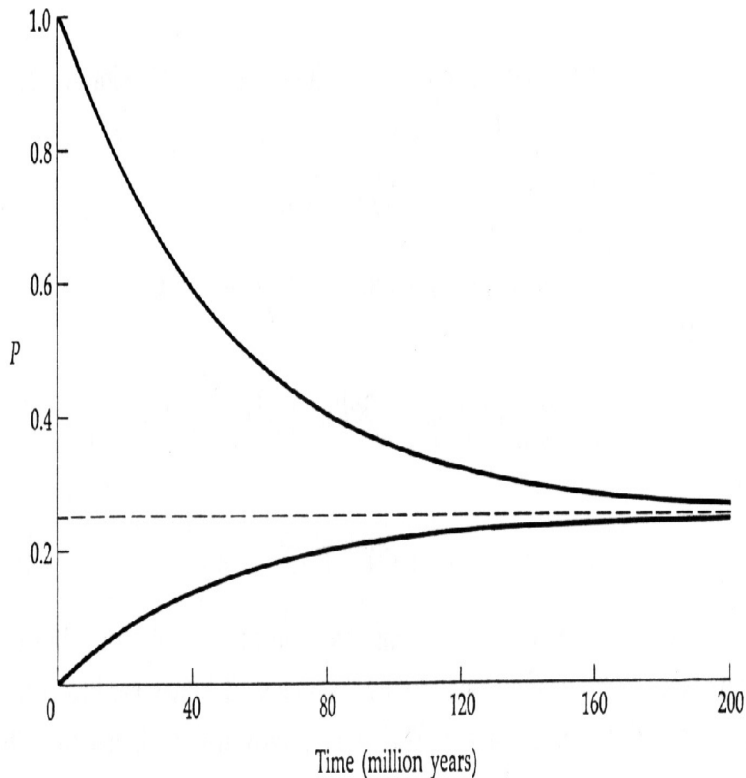
$$P_A(t+1) = (1 - 3\alpha)P_A(t) + \alpha[1 - P_A(t)] = -4\alpha P_A(t) + \alpha$$

Непрерывное приближение

$$\frac{dP_A(t)}{dt} = -4\alpha P_A(t) + \alpha$$



# Зависимость частоты нуклеотида 'А' от времени



$$P_A(t) = 0.25 + (P_A(0) - 0.25)e^{-4\alpha t}$$

$$P_A(t) = 0.25 + 0.75e^{-4\alpha t} \quad (P_A(0)=1)$$

$$P_A(t) = 0.25 - 0.25e^{-4\alpha t} \quad (P_A(0)=0)$$

Особенности:

$$P_A(\infty) = P_T(\infty) = P_G(\infty) = P_C(\infty) = 1/4$$

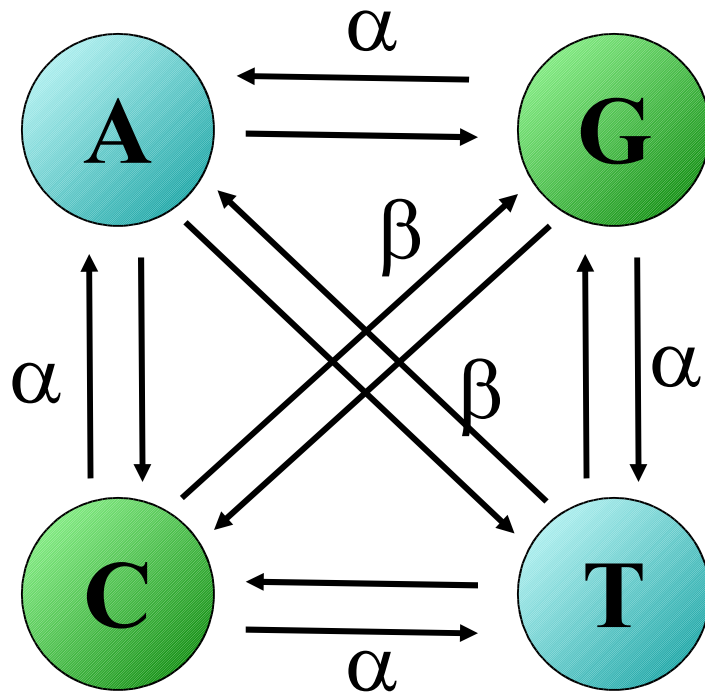
$$P_{AA}(t) = P_{ii}(t) = 0.25 + 0.75e^{-4\alpha t}$$

$$\begin{aligned} P_{AG}(t) &= P_{AT}(t) = P_{AC}(t) = \\ &= P_{ij}(t) = 0.25 - 0.25e^{-4\alpha t} \end{aligned}$$





# Модель Кимуры



Учет разных частот

Для транзиций -  $\alpha$

Для трансверсий (A-T, G-C) -  $\beta$

$$P_{ii}(t) = 0.25 + 0.25e^{-4\beta t} + \\ + 0.5e^{-2(\alpha + \beta)t}$$

Однако  $P_{ij}$  различаются по парам замен (транзиции Y, трансверсии Z):



# Матрица замен и ее свойства



$$p_A' = (1 - \lambda_1)p_A + \lambda_{21}p_T + \lambda_{31}p_G + \lambda_{41}p_C$$

$$p_T' = \lambda_{12}p_A + (1 - \lambda_2)p_T + \lambda_{32}p_G + \lambda_{42}p_C$$

$$p_G' = \lambda_{13}p_A + \lambda_{23}p_T + (1 - \lambda_3)p_G + \lambda_{43}p_C$$

$$p_C' = \lambda_{14}p_A + \lambda_{24}p_T + \lambda_{34}p_G + (1 - \lambda_4)p_C$$

$$\mathbf{p}' = \mathbf{M}\mathbf{p}$$

$$\mathbf{M} = \begin{bmatrix} (1 - \lambda_1) & \lambda_{21} & \lambda_{31} & \lambda_{41} \\ \lambda_{12} & (1 - \lambda_2) & \lambda_{32} & \lambda_{42} \\ \lambda_{13} & \lambda_{23} & (1 - \lambda_3) & \lambda_{43} \\ \lambda_{14} & \lambda_{24} & \lambda_{34} & (1 - \lambda_4) \end{bmatrix}$$

Свойства:

- Сумма по строкам равна 1
- Эволюция за  $n$  шагов эквивалентна умножению на  $\mathbf{M}^n$ .
- Если  $\mathbf{M} = \text{const}$ , то существуют равновесные частоты, которые находятся из уравнения

$$\mathbf{p} = \mathbf{M}\mathbf{p}$$



# Модель замен в белках



20 аминокислот. Замены в позициях независимы и определяются матрицей одинаковой для всех белков и всех позиций  $M(20 \times 20)$ .

Матрица замен  $M$  была определена эмпирически на основе анализа нескольких семейств гомологичных белков Дайхофф и сотр. (1978).

Свойства матрицы Дайхофф:

- Равновесные частоты равны частотам встречаемости аминокислот в последовательностях белков.
- Наиболее часты замены аминокислот на аминокислоты, сходные по физико-химическим свойствам.
- Исходная матрица нормирована на время, эквивалентное 1 замене на 100 позиций (1РАМ).
- Для оценки вероятности замен через время  $t=n$  надо матрицу 1РАМ возвести в степень  $n$ .



# Расстояние между последовательностями



AGСТА...А  $\left\{ \begin{array}{l} \text{AGTTC...A} \\ \text{AGCAG...A} \end{array} \right.$   $t$

$$p = n_d / n \quad q = e^{-2\lambda t}$$
$$q = 1 - p = n_i / n \quad d = -\log_e(q)$$

- В простейшем приближении можно использовать расстояние, основанное на частотах совпадений аминокислот (модель Пуассона).
- Зная время и  $d$  можно установить скорость замен и наоборот, зная  $\lambda$  и  $d$  можно установить  $t$ .
- Недостаток – не учитываются повторные и обратные замены, применимо на небольших временах дивергенции



# Различные меры расстояний последовательностей ДНК



Определение расстояния зависит от типа модели. Примеры:  
Модель Джукса-Кантора:

$$d = -\frac{3}{4} \log_e \left(1 - \frac{4}{3} p\right)$$

Модель Кимуры (P-частота транзиций, Q-частота трансверсий):

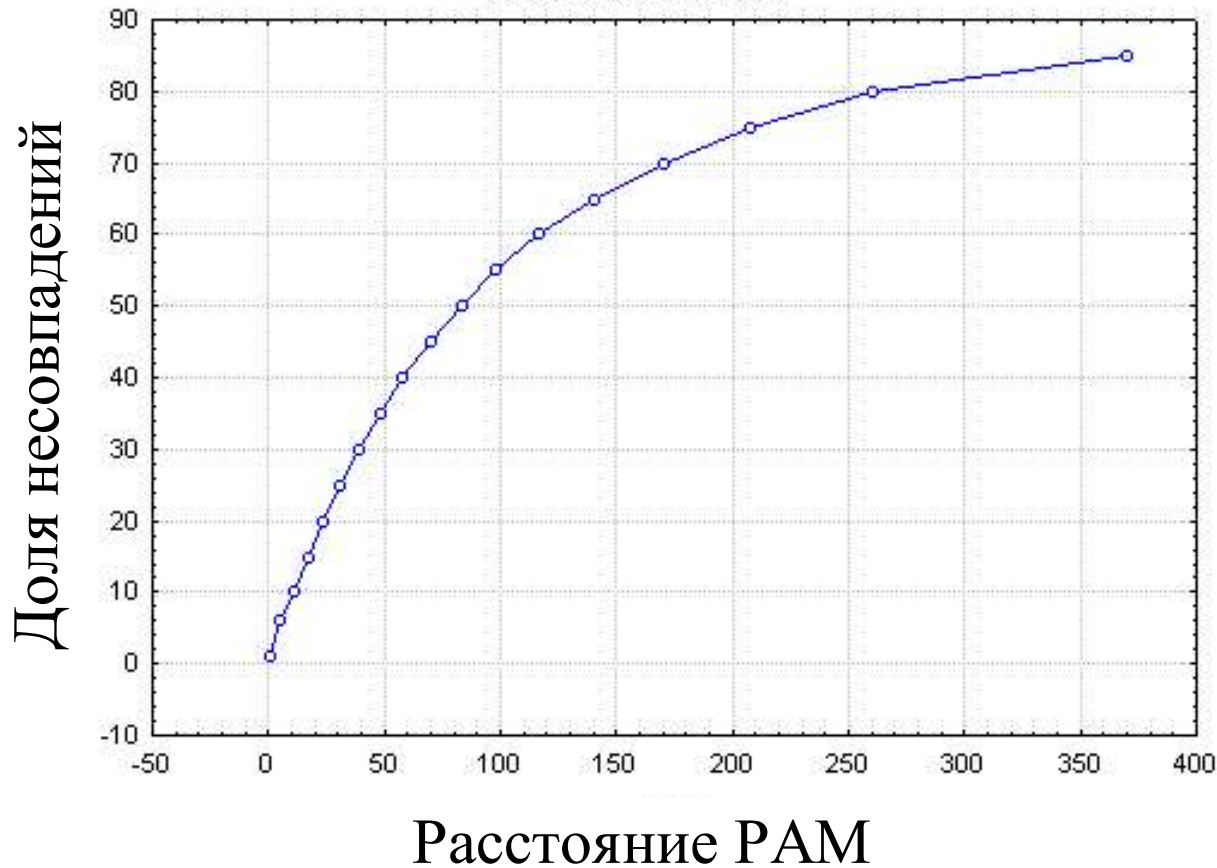
$$d = -\frac{1}{2} \log_e [(1 - 2P - Q)\sqrt{1 - 2Q}]$$

Общий случай матрицы замен:

$$d = 2 \left( \sum_1^4 p_i \lambda_i \right) t$$



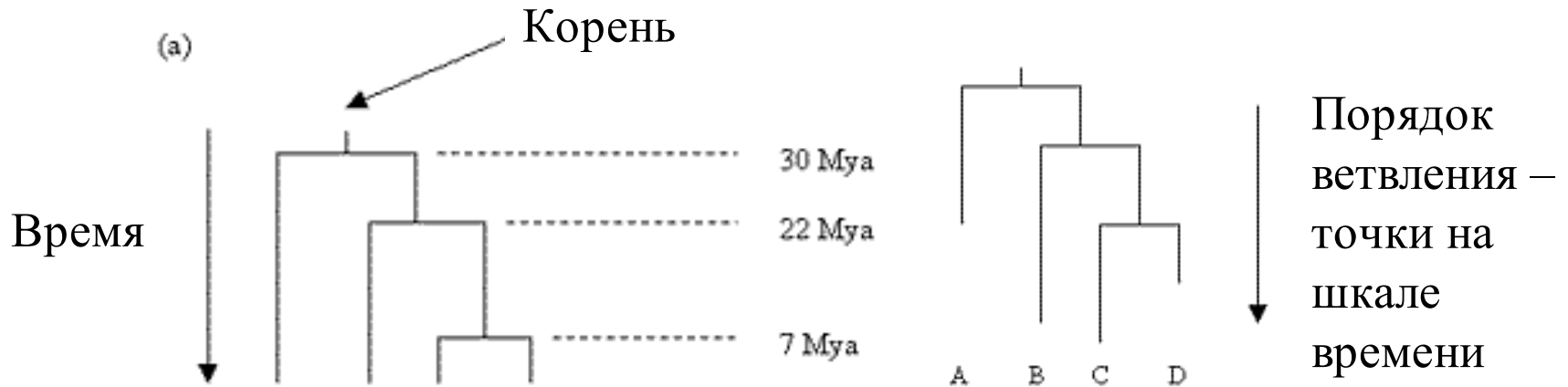
# Расстояния белковых последовательностей



- Матрицы серии RAM-N отражают вероятность замены аминокислот  $a_i$   $\rightarrow$   $a_j$  за время эволюции, эквивалентное N RAM единиц.



# Филогенетические деревья

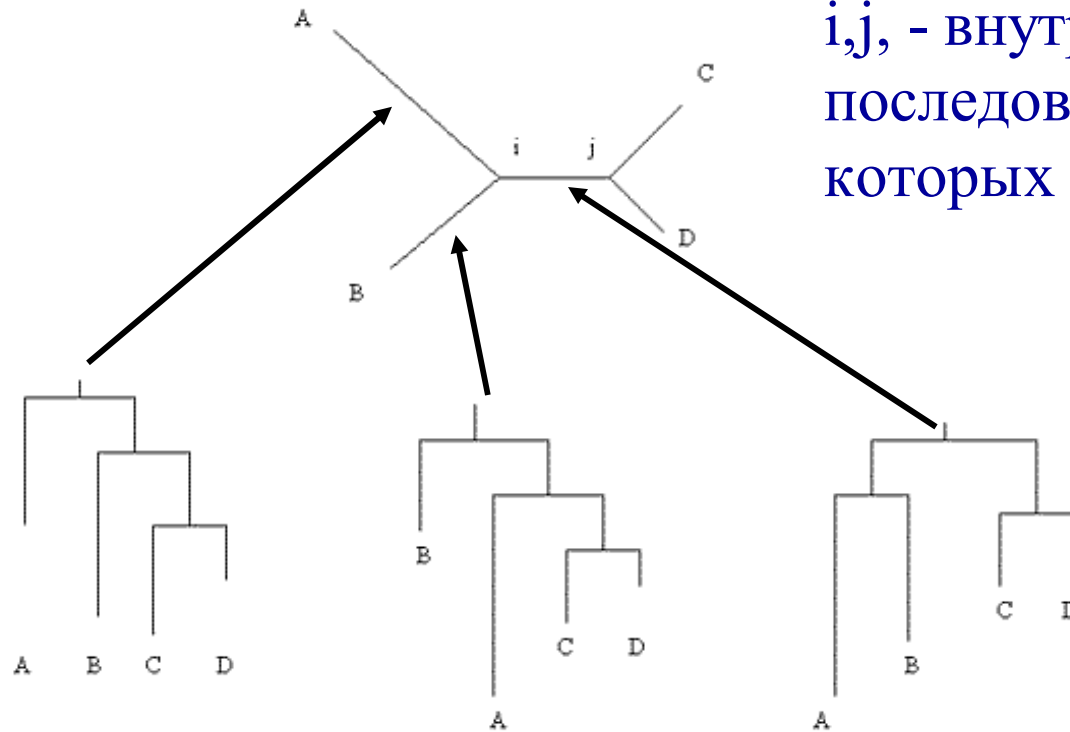


- Необходимы для восстановления эволюционной истории
- Описывают эволюционные отношения для набора из нескольких последовательностей

Рисунок из P.Higgs, Introduction to Phylogenetics Methods ,  
<http://online.itp.ucsb.edu/online/infobio01/higgs/>



# Не все деревья имеют корень



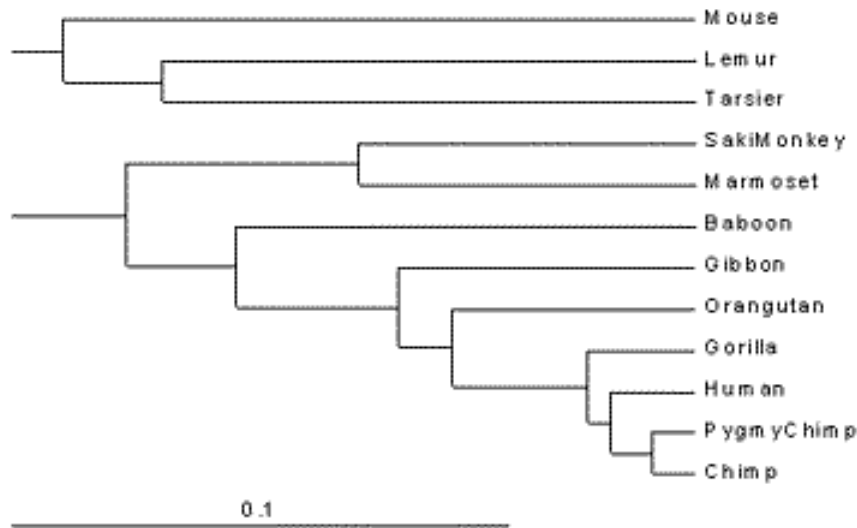
$i, j$ , - внутренние узлы,  
последовательности для  
которых неизвестны

- Не все методы построения деревьев могут давать положение корня





# Методы построения деревьев: UPGMA



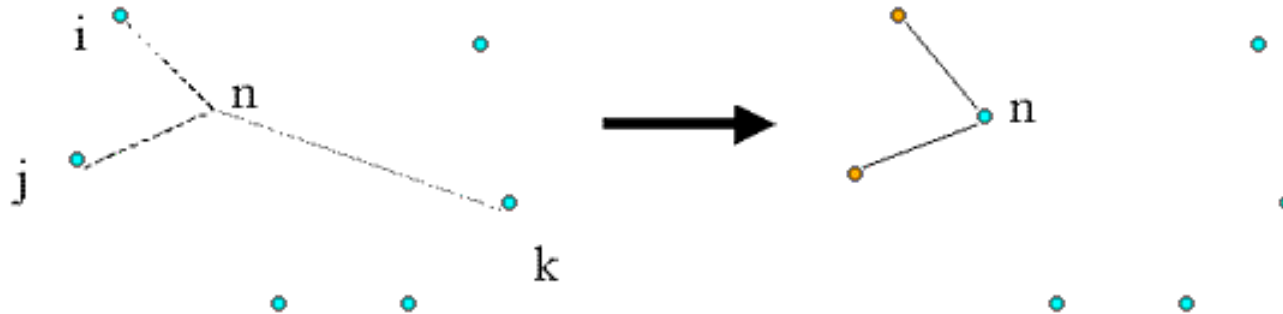
- Расстояние между кластером X и кластером Y равно среднему от парных расстояний между последовательностями этих кластеров

- Предполагает равномерность замен (молек.часы)
- Расстояние=2\*длину ветви
- Дает всегда дерево с корнем

Рисунок из P.Higgs, Introduction to Phylogenetics Methods ,  
<http://online.itp.ucsb.edu/online/infobio01/higgs/>



# Метод ближайшего соседа



Берутся два ближайших узла  $i, j$  и заменяются новым узлом  $n$

$$d_{in} + d_{nk} = d_{ik}; \quad d_{jn} + d_{nk} = d_{jk}; \quad d_{in} + d_{jn} = d_{ij};$$

$$d_{nk} = (d_{ik} + d_{jk} - d_{ij})/2; \quad \text{Для каждого}$$

$$r_i = \frac{1}{N-2} \sum_k d_{ik} \quad r_j = \frac{1}{N-2} \sum_k d_{jk}$$

$$d_{in} = (d_{ij} + r_i - r_j)/2; \quad d_{jn} = (d_{ij} + r_j - r_i)/2.$$

На следующем шаге выбирается пара  $i, j$ , для которых  $D_{ij}$  минимально ( $D_{ij} = d_{ij} - r_i - r_j$ )



# Метод ближайшего соседа - результат

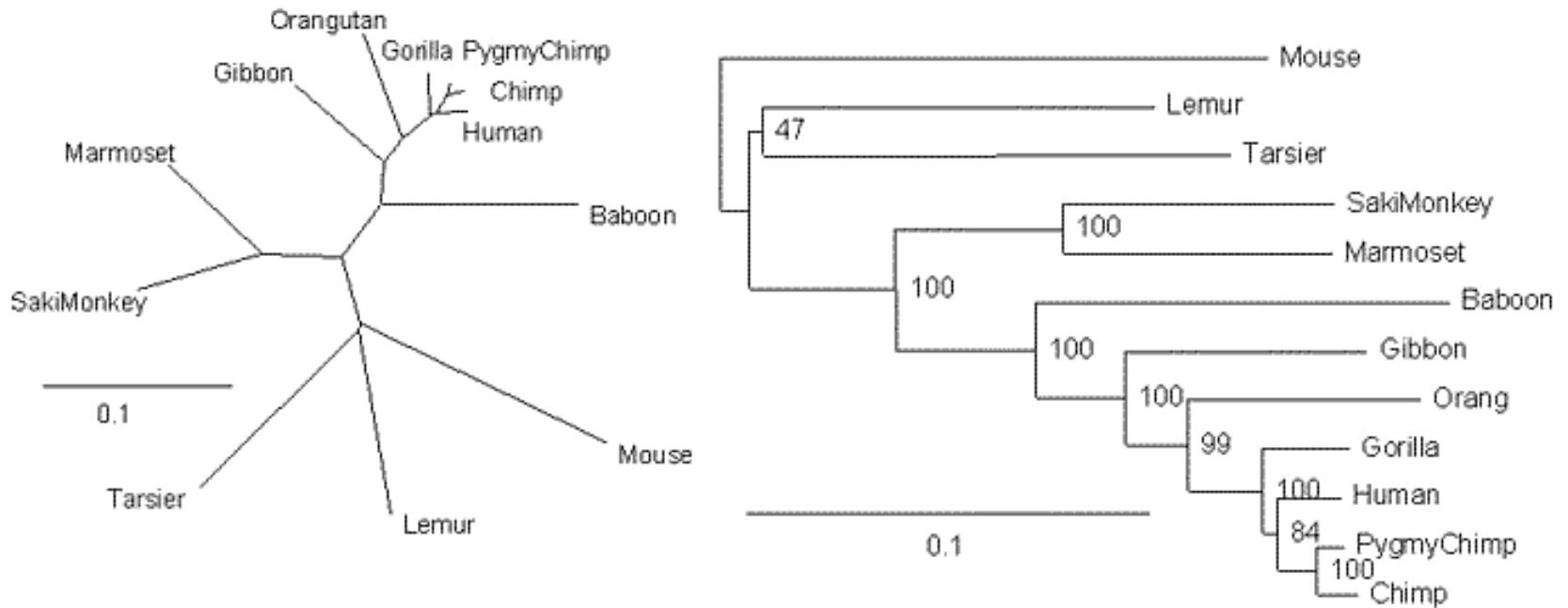


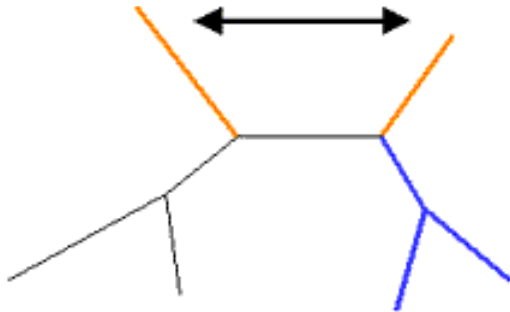
Рисунок из P.Higgs, Introduction to Phylogenetics Methods ,  
<http://online.itp.ucsb.edu/online/infobio01/higgs/>



# Дерево можно перестраивать и оценивать



Перестановка соседних узлов



Сокращение и перестройка поддеревьев

Число различных топологий дерева  
Без корня                      С корнем

N	Без корня	С корнем
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10395

$$U_N = (2N-5)U_{N-1}$$

$$R_N = (2N-3)R_{N-1}$$

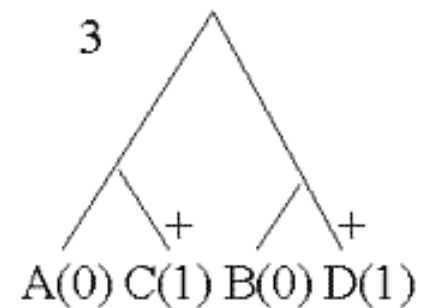
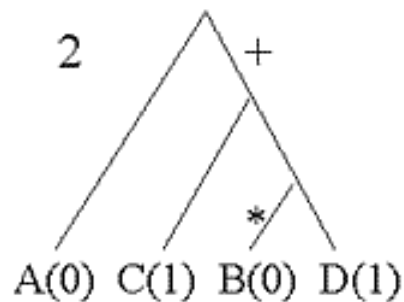
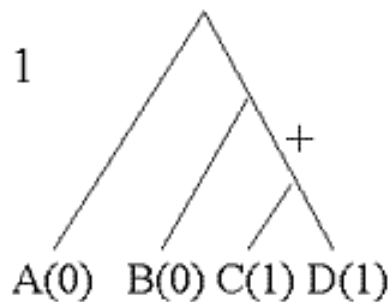
- Каждому дереву можно присвоить числовую характеристику и сравнивать их



# Метод парсимонии



Пример построения дерева для набора из четырех последовательностей, с двузнаковым алфавитом (0,1)

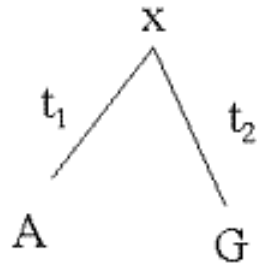


- 1) В первом дереве изменения происходят только один раз (+)
- 2) Во втором дереве 1 появляется (+) и теряется (\*)
- 3) В третьем дереве 1 появляется независимо два раза (+)

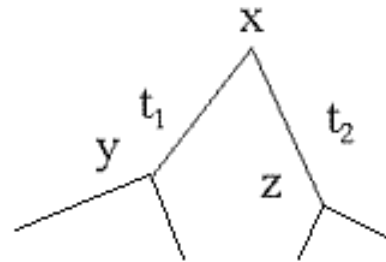
Дерево (1) содержит минимальное число эволюционных событий – его и выбираем.



# Методы максимального правдоподобия



$$L(x) = P_{xA}(t_1)P_{xG}(t_2)$$



$$L(x) = \sum_y \sum_z L(y)P_{xy}(t_1)L(z)P_{xz}(t_2)$$

- Вычисляется вероятность наблюдения данных (D), при условии, что эволюция происходила по данной топологии (T) – функция правдоподобия  $L(D|T)$
- Вычисляется рекурсивно от листьев к вершине
- Выбирается дерево, которое дает  $\max(L)$
- Решение зависит от выбора модели замен



# Методы максимального правдоподобия



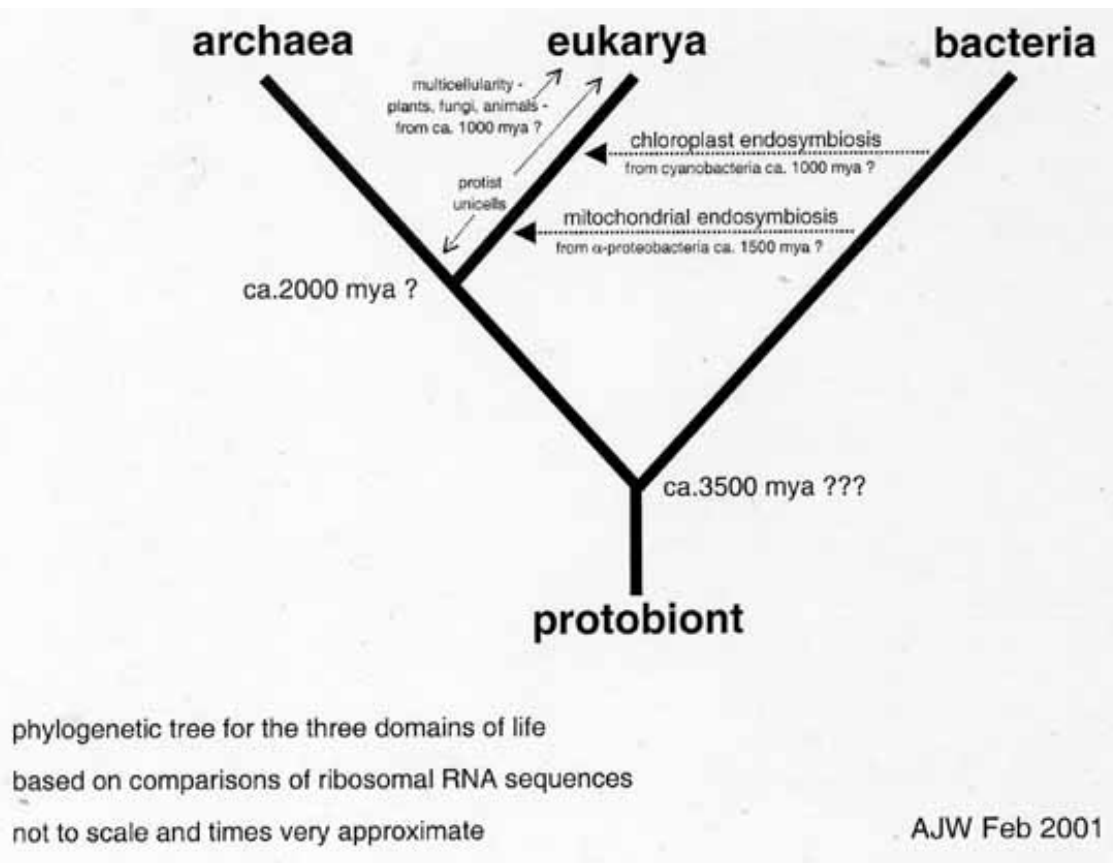
- С помощью методов МП можно оценивать и другие параметры: так как и матрица замен, и скорости замен могут быть такими параметрами.

$$L=L(T,M,t.....).$$

- Можно усложнять модель, добавляя новые параметры.
- Метод имеет статистическое обоснование
- Но требует большого количества вычислений
- Не эффективен при большом числе параметров и малом количестве данных



# Филогения: пример (открытие третьего царства)



Woese et al (1970-е) сравнивали последовательности малых субъединиц 16S rRNA и обнаружили, что для археобактерий они ближе к эукариотам, чем к бактериям. Археобактерии были выделены в отдельное царство.

Рисунок из <http://www.staffs.ac.uk/schools/sciences/biology/Handbooks/evolearly.htm>





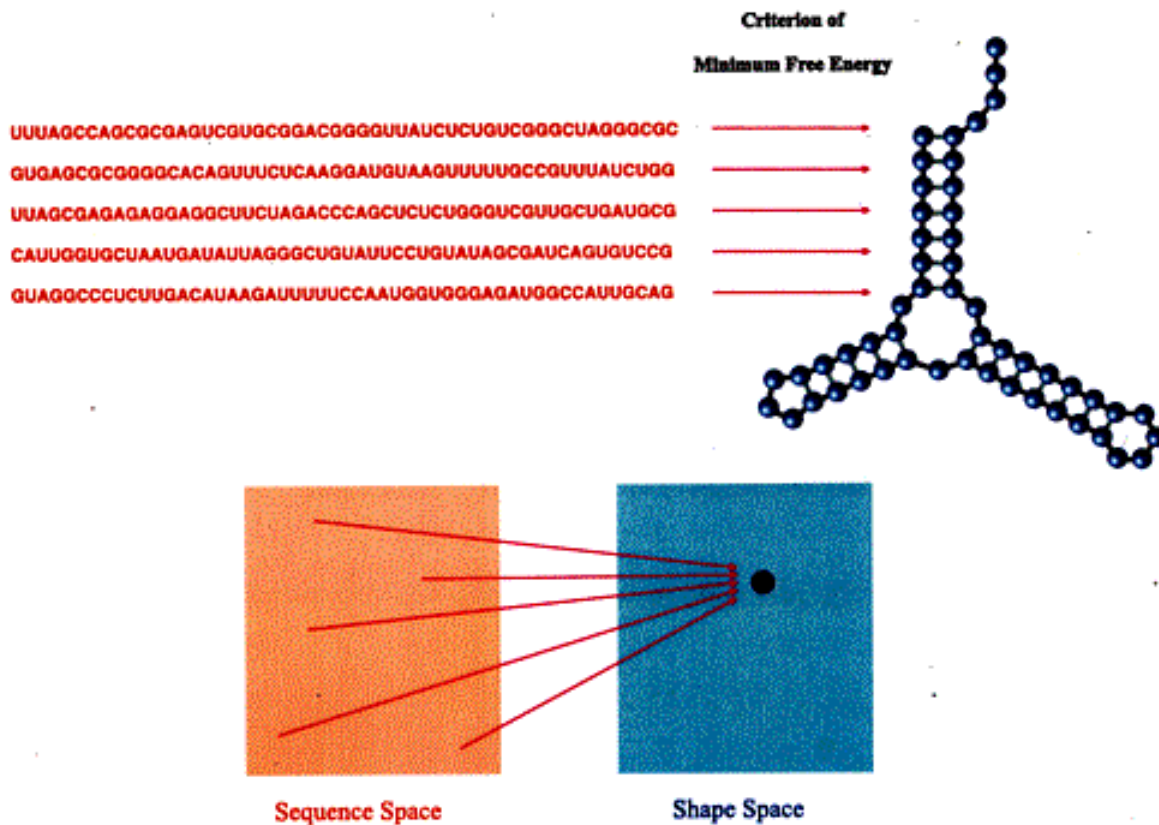
# Усложнение базовых моделей эволюции



- В природе встречаются отклонения от базовых моделей замен нуклеотидов/аминокислот.
- Сильное влияние оказывает структура макромолекул.
- Модели могут совершенствоваться с обнаружением новых фактов/явлений



# РНК – корреляции замен пар нуклеотидов



Вторичная структура РНК (шпильки) накладывает ограничения на паттерны замен

Проявление в виде корреляций

Множественность отображения последовательность-структура

Рисунок адаптирован из лекции P.Schuster, RNA Structure and Molecular Evolution, <http://online.itp.ucsb.edu/online/infobio01/schuster/>



# Модели эволюции белковых последовательностей



- Белки – гетерогенные объекты.
- Содержат различные типы вторичных структур
- Различные участки доступности растворителю
- Содержат функциональные сайты
- Давление отбора для разных участков белка различно.



# Особенности эволюции белков



скорости фиксации замен для различных ветвей эволюционного дерева могут быть различными, что не соответствует предположению о "молекулярных часах" (Ayala *et al.*, 1997; Ayala, 1997);

скорости фиксации замен аминокислот могут различаться для различных позиций белка вдоль его последовательности (Uzzel and Corbin, 1971; Morozov *et al.*, 2000);

вероятности замен аминокислот (определенные как элементы матрицы аминокислотных замен) для различных участков глобулы или даже для отдельных позиций белковой последовательности могут быть различными (Overington, 1992).

параметры эволюционного процесса (в частности элементы матрицы **M**) могут зависеть от времени (Benner *et al.*, 1994).

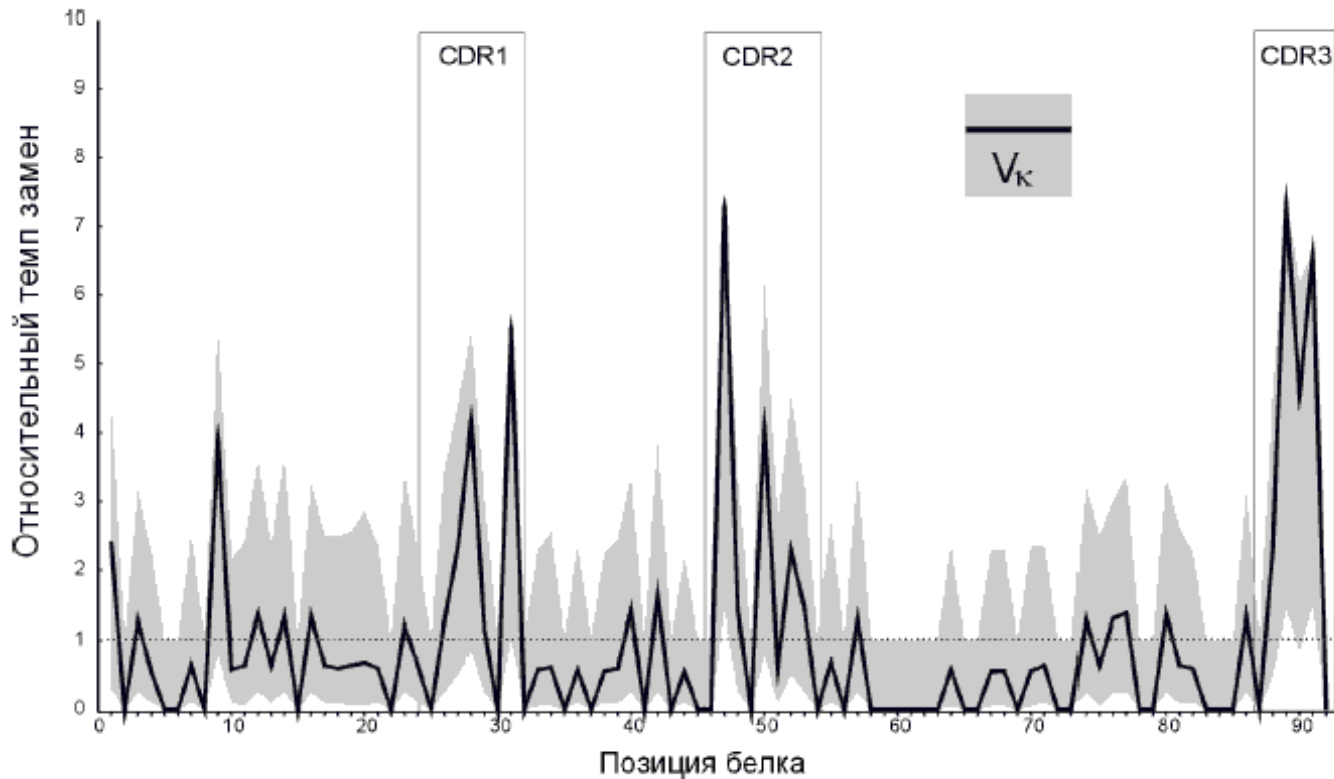
мутации в различных позициях белка могут фиксироваться зависимым образом.



# Гетерогенность скоростей замен в позициях белка



Относительные темпы замен для Ig-k (Morozov, 2000)





# Сравнение матриц замен для разных участков белка

Режимы эволюции для разных участков глобулы различны!  
Петли на поверхности                      Погруженная спираль

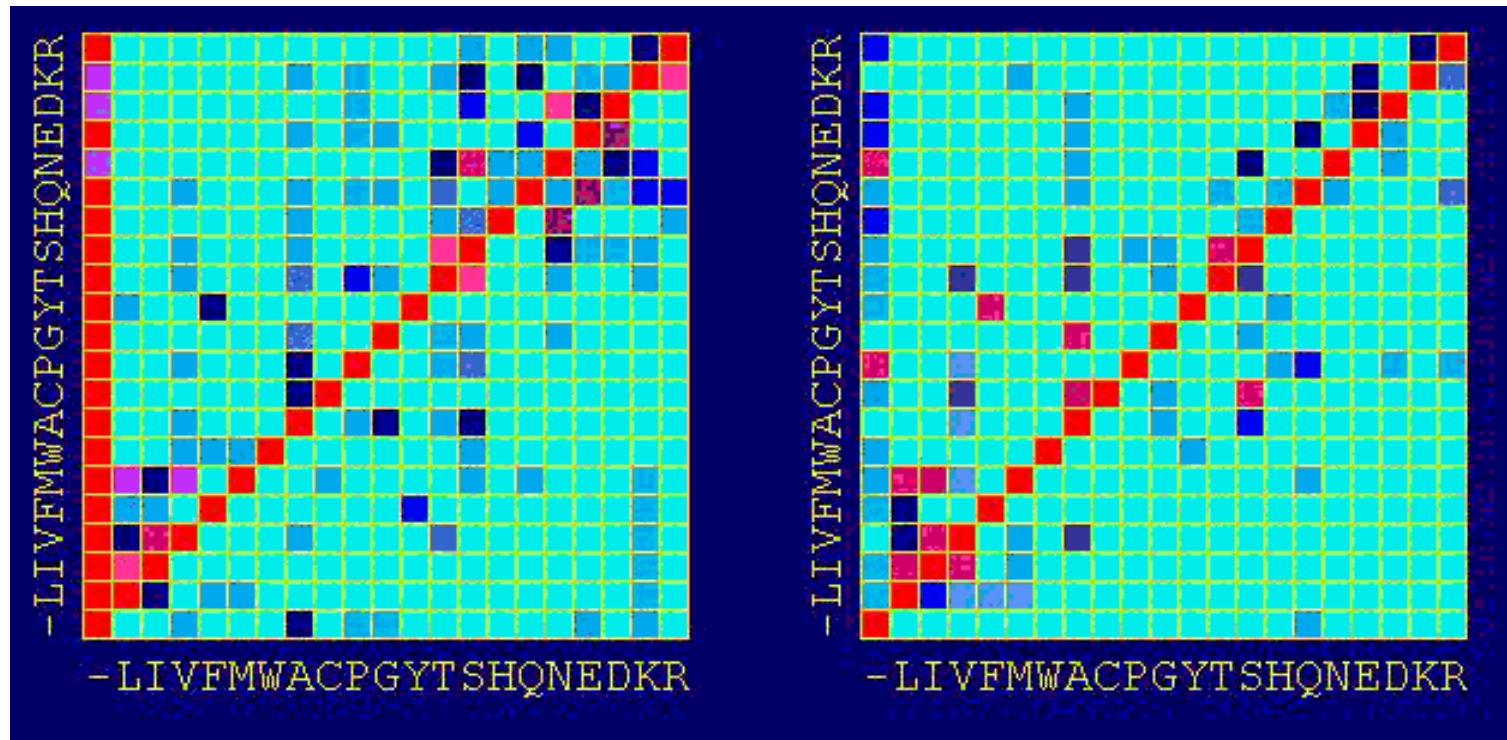


Рисунок адаптирован из лекции R.Goldstein, Modeling Evolution at the Amino Acid Level , <http://online.itp.ucsb.edu/online/infobio01/goldstein1/>



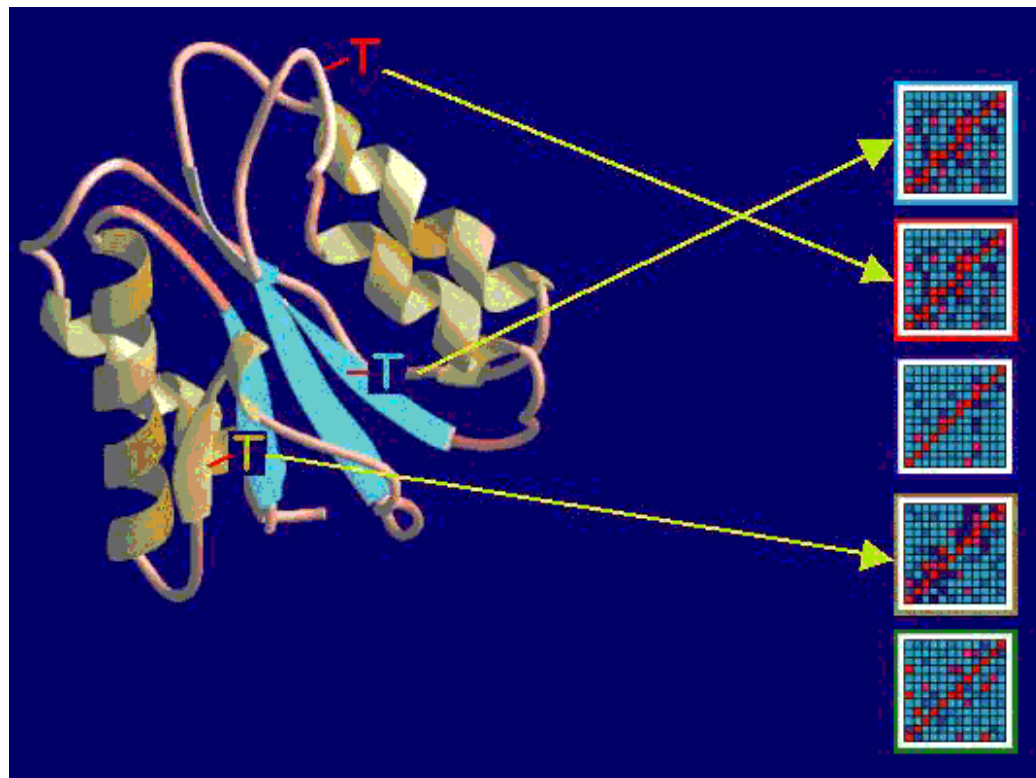
# Матрицы замен, специфичные для позиций



- Основная идея – описывать замены для разных структурных классов остатков разными матрицами замен (Koshi & Goldstein; Wako & Blundell; Overington).



# Специфичность замен аминокислот



Недостаток – слишком много параметров надо оценивать

Рисунок адаптирован из лекции R.Goldstein, Modeling Evolution at the Amino Acid Level , <http://online.itp.ucsb.edu/online/infobio01/goldstein1/>





## Учет специфичности – подход Koshi & Goldstein

Выход – параметризация матриц (вероятностей замен) на основе Больцмановского распределения по приспособленности и сходства ф-х свойств аминокислот. Подгонка – методом МП.

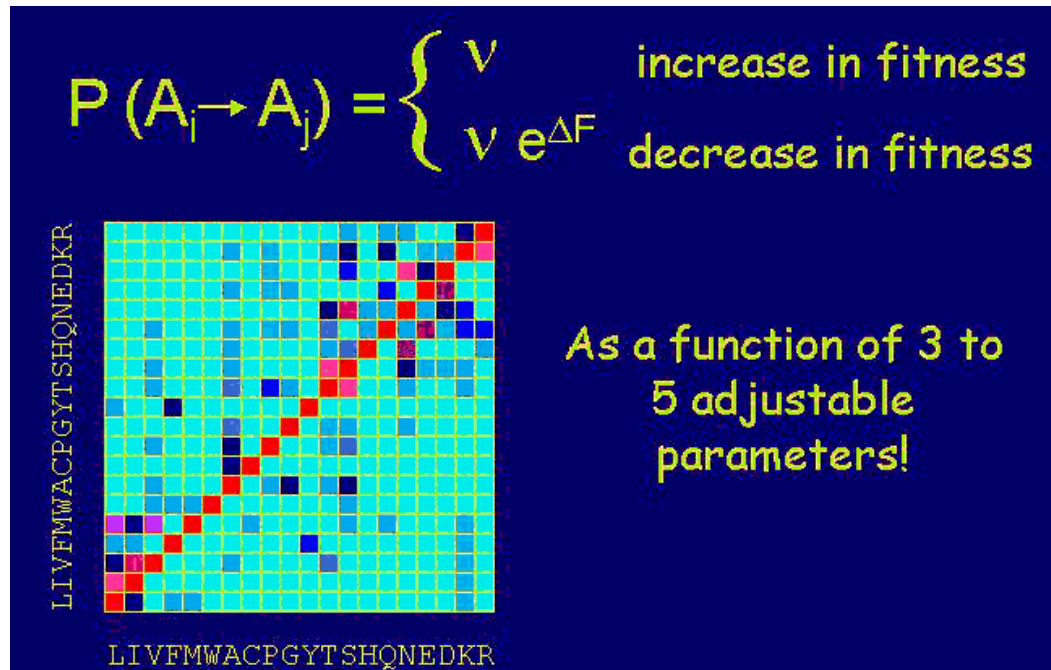


Рисунок адаптирован из лекции R. Goldstein, Modeling Evolution at the Amino Acid Level, <http://online.itp.ucsb.edu/online/infobio01/goldstein1/>

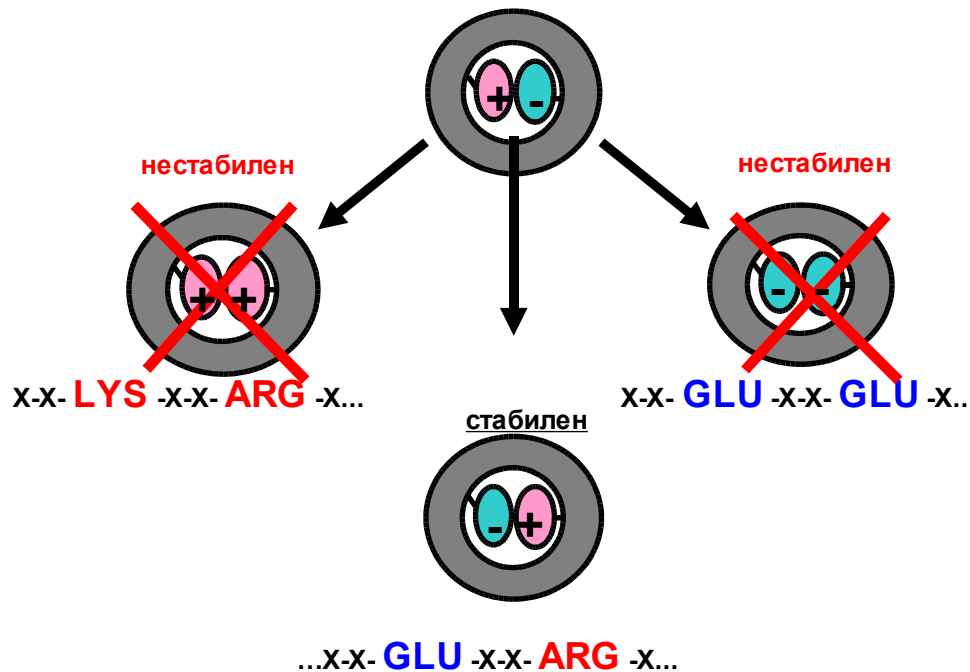


# Зависимые замены аминокислотных остатков

LIVKSM**D**GAL  
STM**E**CA**R**LIT  
GTSD**N**SH**Q**LI  
LIM**K**V**V**DGYA

} Анализ множественного  
выравнивания  
последовательностей  
белкового семейства

..X-X-X- **LYS** -X-X- **ASP** -X-X-X..

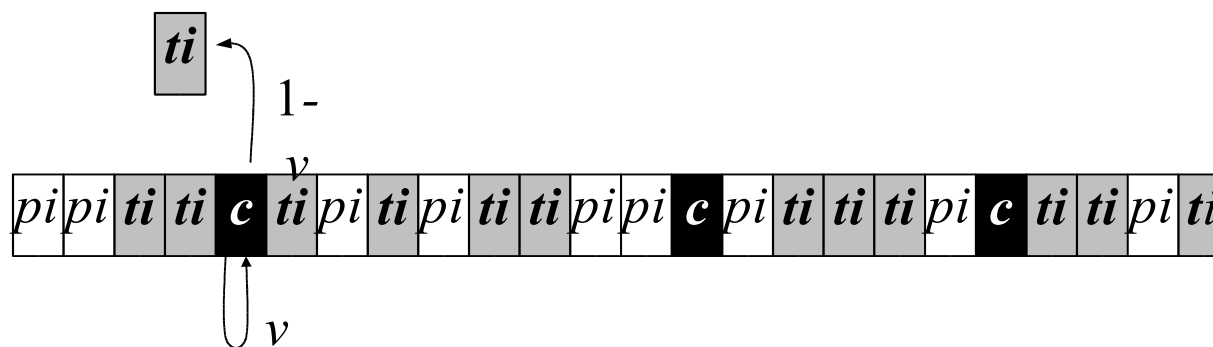




## Коварионные модели эволюции (Fitch)



- Изменение одной аминокислоты влияет на характер дальнейших мутаций.
- Три класса позиций:  $c$  – коварионы, могут заменяться в данный момент;  $ti$  – невариабельные временно позиции;  $pi$  – постоянно не изменяемые позиции





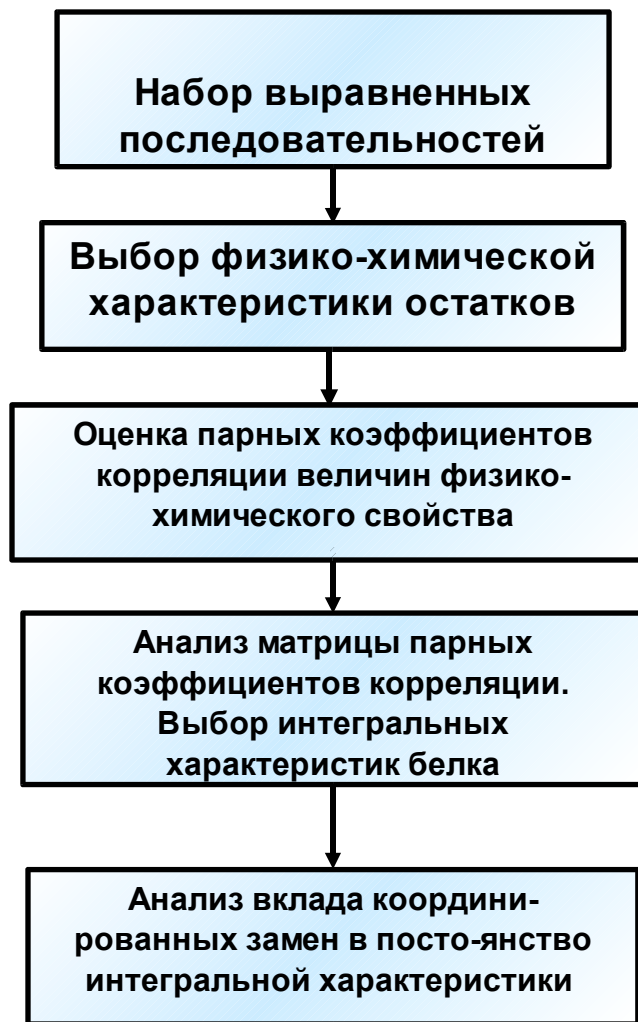
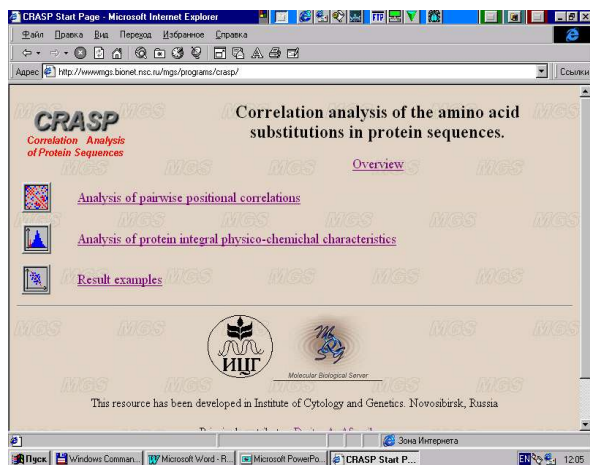
# Методы выявления зависимых замен аминокислот

- Методы теории информации
- Анализ статистики парных частот встречаемости СИМВОЛОВ
- Методы максимального правдоподобия
- Анализ корреляций физико-химических свойств



# Анализ координированных замен – пакет CRASP

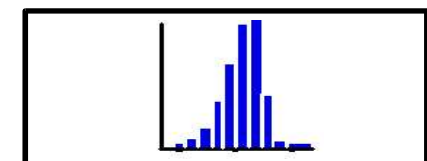
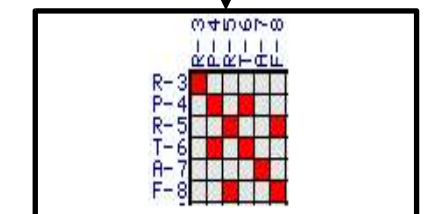
- Возможности анализа:
- Оценка парных корреляций
  - Кластерный анализ
  - Анализ интегральных характеристик



```
ASDRTFNHKLIYRRMNL-  
GAAQWCCV-RHNGTYKLS  
DFGSSR---WWRQACHMP  
DFGSSR---WWRQACHMP
```

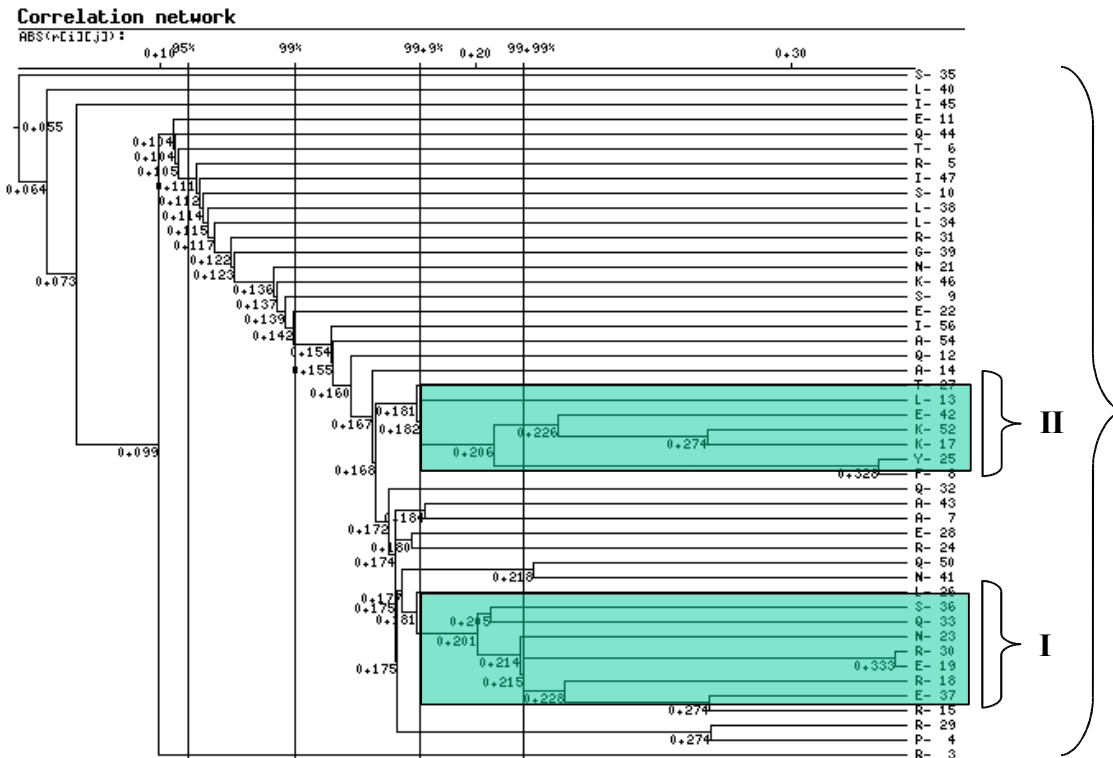
+4.0 ; +6.2 ; -1.8 ;  
+5.9 ; +3.2 ; +5.6 ;  
-3.4 ; -2.1 ; +4.0 ;  
-6.9 ; +2.3 ; +5.0 ;

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$$





# Анализ ДНК связывающих доменов класса «Гомеодомен»



Позиции белка кластеризуются по степени их взаимной коррелированности

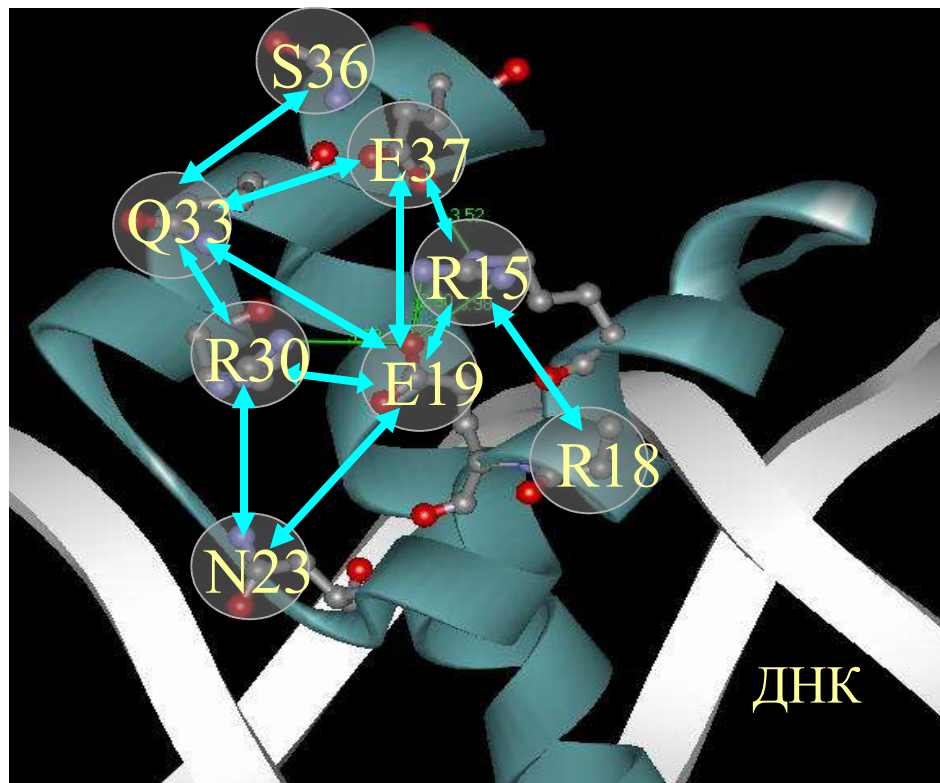
Выявлено 2 кластера коррелирующих позиций.



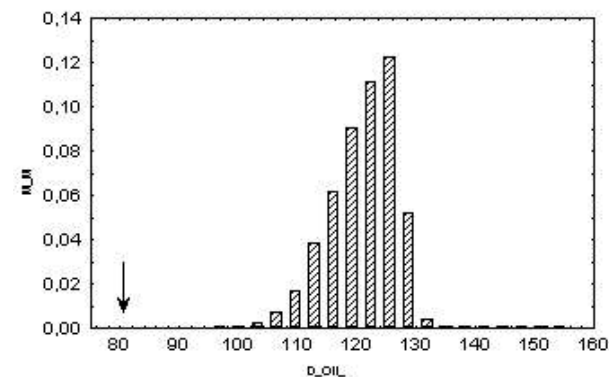
## Корреляции в «гомеодоменах»



Расположение остатков в пространственной структуре комплекса «гомеодомен»-ДНК. Отрицательные зависимости показаны голубыми стрелками.



$$Q_1 = pI_{15} + pI_{18} + pI_{19} + pI_{23} + pI_{30} + pI_{33} + pI_{36} + pI_{37}$$

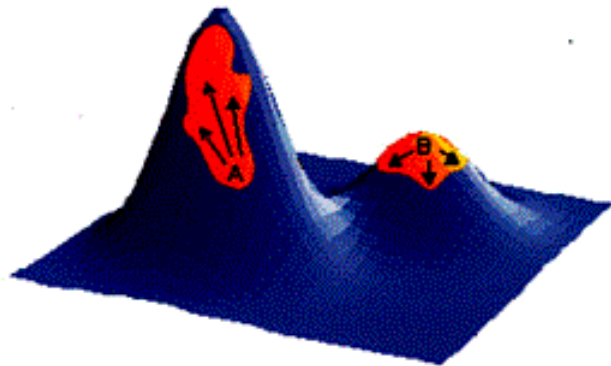


Распределение дисперсий  $Q_1$  в случайных выборках и в реальной выборке ( $P=1-10^{-5}$ )

Вывод: координированные замены вносят значимый вклад в постоянство суммарной величины изоэлектрической точки остатков кластера I



# Эволюция по ландшафту приспособленности.



Понятие ландшафта предложено Райтом. Классический подход предполагает эволюцию популяции как движение точки (отсутствие полиморфизма). Современные подходы предполагают описание вида как облака точек (высокая степень полиморфизма), которое движется по поверхности в сторону пиков (Eigen; Kauffman; Schuster).





## Ссылки в Интернете



- **Web Resources in Molecular Evolution and Systematics**  
(<http://darwin.eeb.uconn.edu/molecular-evolution.html>)
- **WebPhylip – программы филогенетического анализа последовательностей**  
(<http://sdmc.krdl.org.sg:8080/~lxzhang/phylip/>)
- **Лекции по биофизике и биоинформатике (есть лекции по эволюции) Института теоретической физики, Санта Барбара** (<http://online.itp.ucsb.edu/online/infobio01>)
- **Лекции по молекулярной эволюции, Marine Biological Laboratory**  
(<http://newfish.mbl.edu/Course/>)
- **Труды Ч.Дарвина** (<http://charles-darwin.narod.ru/>)
- **Лекция по эволюционной биологии П.М.Бородина (НГУ)**  
(<http://evolution2.narod.ru/>)
- **Обзоры книг по альтернативным эволюционным и неэволюционным теориям (G.Korthof)** (<http://home.wxs.nl/~gkorthof/index.htm>)