ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Лекция №4

Значимость выравнивания. Экспресс-сравнение последовательностей. Множественное выравнивание. Филогенетические деревья

Новоселецкий Валерий Николаевич
к.ф.-м.н., доц. каф. биоинженерии
valery.novoseletsky@yandex.ru

Сайт курса http://intbio.org/bioinf2018

Контроль посещаемости сегодня

Заполнить форму по адресу

http://intbio.org/4

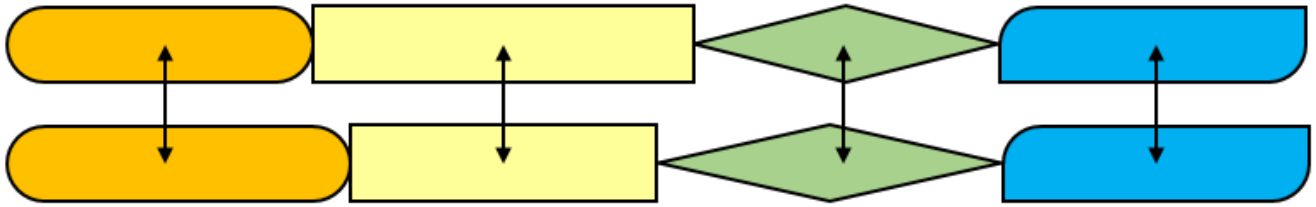
возможность закрывается сразу после перерыва

либо

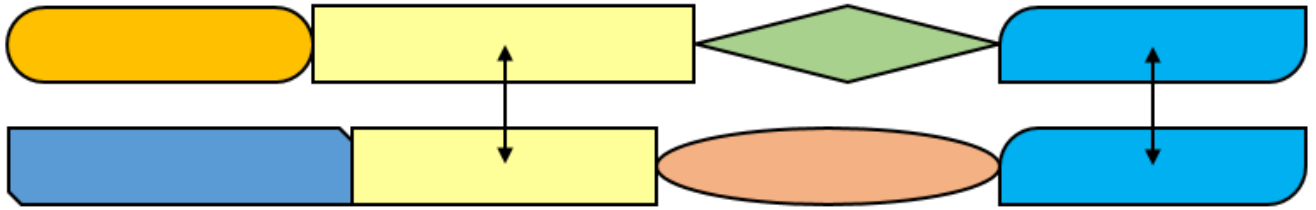
записаться в список на перерыве

Расчет выравнивания двух последовательностей. Алгоритм Смита-Уотермана (1981)

– модификация алгоритма Нидлмана-Вунша с аффинным штрафом за вставку и обнулением отрицательных значений в матрице – позволяет выявлять локальные выравнивания.



Global Alignment



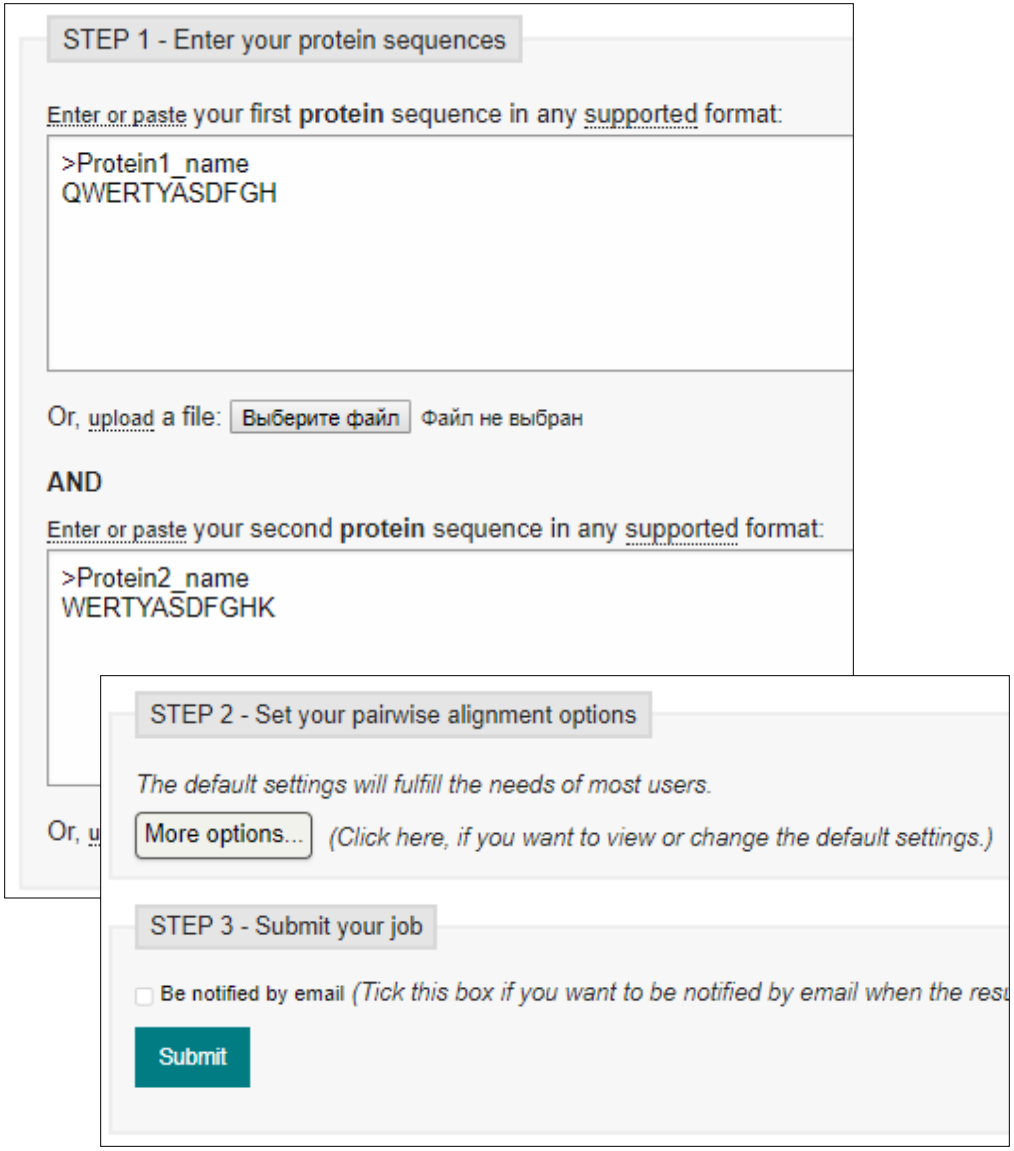
Local Alignment

Расчет выравнивания двух последовательностей. Алгоритм Смита-Уотермана

	"-"	g	c	t	g	a	a	c	g
"-"	0	0	0	0	0	0	0	0	0
c	0	0	0,5	0	0	0	0	0,5	0
t	0	0	0	1	0	0	0	0	0
a	0	0	0	0	0	0,5	0,5	0	0
t	0	0	0	0,5	0	0	0	0	0
a	0	0	0	0	0	0,5	0,5	0	0
a	0	0	0	0	0	0,5	1	0	0
t	0	0	0	0,5	0	0	0	0	0
c	0	0	0,5	0	0	0	0	0,5	0

match	mut	gap
0,5	-1	-1

Пример выравнивания двух последовательностей. EMBOSS Needle




```
# Aligned_sequences: 2
# 1: Protein1_name
# 2: Protein2_name
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 13
# Identity: 11/13 (84.6%)
# Similarity: 11/13 (84.6%)
# Gaps: 2/13 (15.4%)
# Score: 67.0
#
#=================================
Protein1_name 1 QWERTYASDFGH- 12
                  ||||||||||
Protein2_name 1 -WERTYASDFGHK 12
```

Значимость выравнивания

Насколько значимо полученное выравнивание?

Имеет ли оно биологический смысл или образовалось случайно?

```
SEQUENCE   1 VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHFDLSH   50
             |||.|||||||||||.|||.||||||||||||||.||||||||||||||
SEQUENCE   1 VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSH   50

SEQUENCE  51 GSAQVKAHGKKVADGLTLAVGHLDDLPGALSDLSNLHAHKLRVDPVNFKL  100
             ||||||.|||||||.||.||.|:||:|.|||.||:|||||||||||||||
SEQUENCE  51 GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKL  100

SEQUENCE 101 LSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTVLTSKYR  141
             ||||||.|||.|||.:||||||||||||||:||||||||||
SEQUENCE 101 LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR  141

# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 141
# Identity:     123/141 (87.2%)
# Similarity:   128/141 (90.8%)
# Gaps:           0/141 ( 0.0%)
```

Значимость выравнивания

```
SEQUENCE   1 -VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHL    49
              .|:|.:..||...|.:..|::..|.....|.:.:..|...   |.|..|
SEQUENCE   1 GALTESQAALVKSSWEEFNANIPKHTHRFFILVLEIAPAAK---DLFSFL    47

SEQUENCE  50 KTEAEM-KASEDLKKHGVTV----------LTALGAILKKKGHHEAELKP    88
             |..:|: :.:.:|:.|...|          |...|.::.     :|.||.
SEQUENCE  48 KGTSEVPQNNPELQAHAGKVFKLVYEAAIQLEVTGVVVT-----DATLKN    92

SEQUENCE  89 LAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELF   138
             |...|.:| .:...:...:.|||:.:..    ..||.....:|.|..:.
SEQUENCE  93 LGSVHVSK-GVADAHFPVVKEAILKTIKE----VVGAKWSEELNSAWTIA   137

SEQUENCE 139 RKDIAAKY-KELGYQG   153
             ..::|... ||:....
SEQUENCE 138 YDELAIVIKKEMDDAA   153

# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 166
# Identity:      35/166 (21.1%)
# Similarity:    62/166 (37.3%)
# Gaps:          26/166 (15.7%)
```

Значимость выравнивания

Миоглобин кашалота (1mbn, 1969)

Леггемоглобин люпина (1gdj, 1995) (ИК РАН)

What about Impossible Burger? ☺

https://impossiblefoods.com/

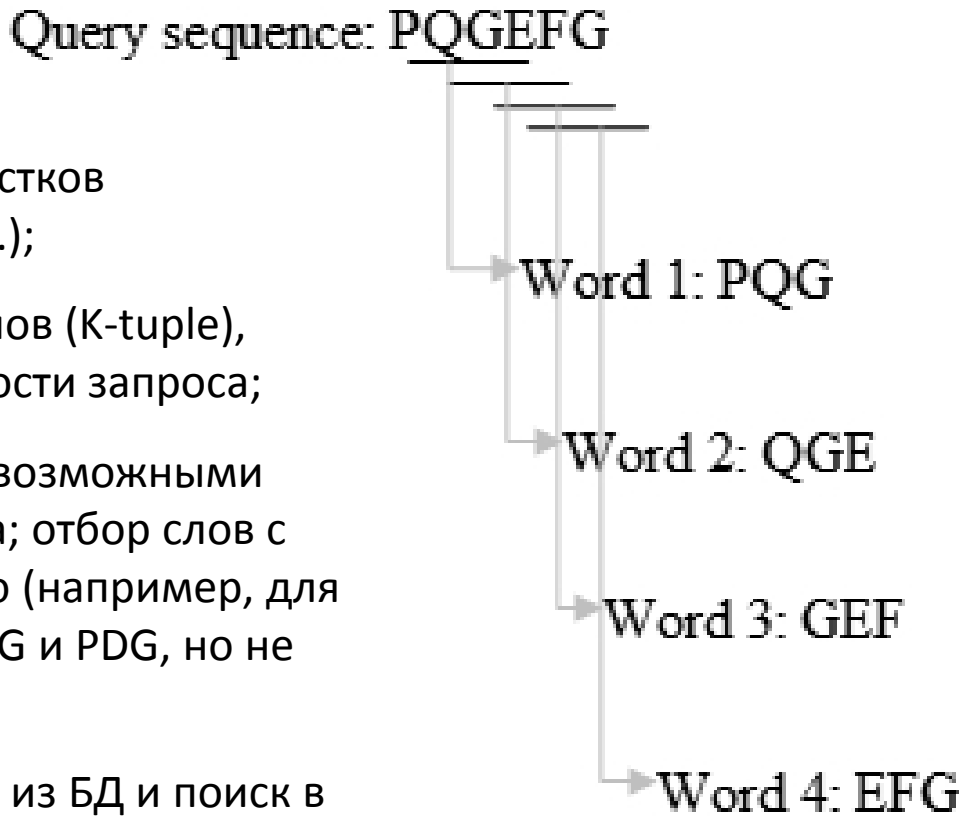
Значимость выравнивания

миоглобин кашалота и леггемоглобин люпина

Идентичность 18%*, но родство подтверждается сходством структуры и функции

Экспресс-методы сравнивания последовательностей. BLAST

BLAST - **B**asic **L**ocal **A**lignment **S**earch **T**ool (Altschul et al.,1990) предназначен для сравнения новых последовательностей с уже содержащимися в базах данных.

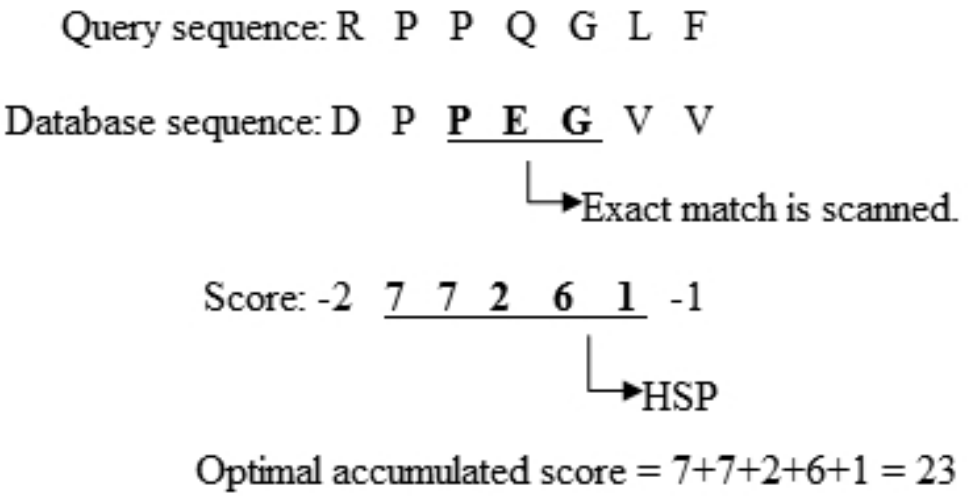


Алгоритм:

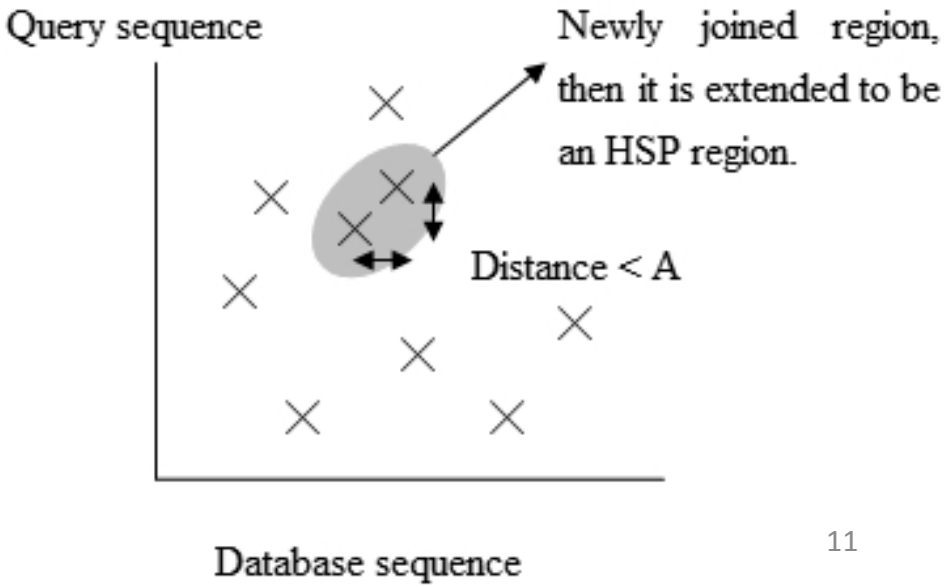
- Удаление малоинформативных участков последовательности (повторы и т.п.);
- Составление списка *к*-буквенных слов (K-tuple), присутствующих в последовательности запроса;
- Сопоставление этих слов со всеми возможными словами длины *к* и оценка сходства; отбор слов с оценкой, превышающей пороговую (например, для слова PQG сходными будут PNG, PEG и PDG, но не PQW)
- Сканирование последовательности из БД и поиск в ней слов с высокой оценкой, полученных на предыдущем шаге;

Экспресс-методы сравнивания последовательностей. BLAST

- Расширение локальных выравниваний в обе стороны до тех пор, пока суммарная оценка выравнивания не начинает уменьшаться (построение сегментных пар (high-scoring segment pair, HSP));



- Объединение сегментных пар, лежащих на удалении меньше A;
- Составление списка сегментных областей с высокой оценкой;
- **Расчет статистической значимости этих оценок.**



BLAST. Значимость выравнивания

$$\tilde{S}_{n,m} = S_{n,m} - \frac{\ln(nm)}{\lambda}$$

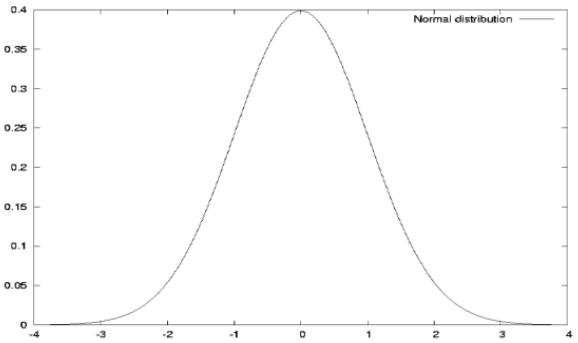
↑

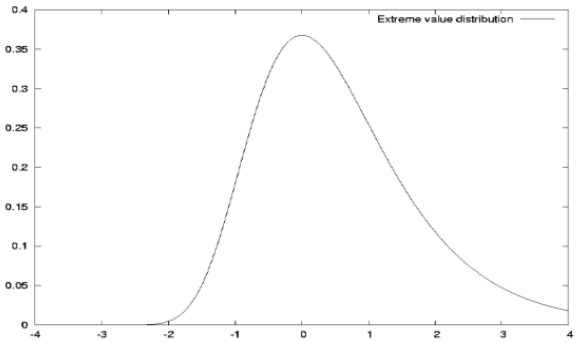
Распределение нормализованных максимальных оценок HSP подчиняется распределению Гумбеля (распределению экстремальных значений, Gumbel, 1937; Гнеденко, 1943), для которого

$$P\left(\tilde{S}_{n,m} > S\right) \approx 1 - \exp\left(-K\, m\, n\, e^{-\lambda S}\right) \approx K\, m\, n\, e^{-\lambda S}$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

$$\varphi(x) = e^{-x} \cdot e^{-e^{-x}}$$





BLAST. Значимость выравнивания

$$P(\tilde{S}_{n,m} > S) \cdot m \equiv E - value$$

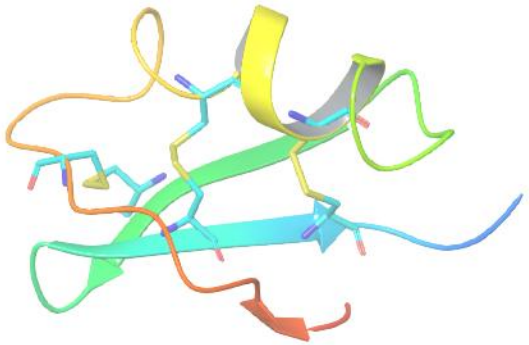
E < 0,02	высокая вероятность гомологии
0,02 < E < 1	гомология не очевидна
E > 1	сходство случайно

Для коротких последовательностей сходство может быть НЕ случайным даже при E > 1 !!

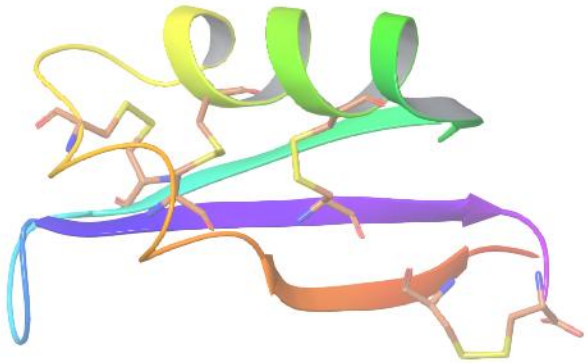
Пример: поиск в PDB по последовательности калиотоксина дает, среди прочего, неожиданный результат:

```
>lcl|PDB:1TI5_A mol:protein length:46 plant defensin Length=46

Score = 25.8 bits (74), Expect = 1.9
Identities = 12/31 (39%), Positives = 14/31 (45%), Gaps = 2/31 (6%)

Query  7 KCSGSPQCLKPCKDAGMRFGKC--MNRKCHC 35  калиотоксин
         KC     C   CK+ G G   C  M R C+C
Sbjct 12 KCLIDTTCAHSCKNRGYIGGNCKGMTRTCYC 42  дефензин
```

BLAST. Значимость выравнивания



калиотоксин



дефензин

```
3ODV_T    1 -GVEINV----KCSGSPQCLKPCKDAGMRFGKCM-N-RKCHCTPK-  38
1TI5_A    1 RTCMIKKEGWGKCLIDTTCAHSCKNRGYIGGNCKGMTRTCYCLVNC  46
                *:     **  .  * : **: *   *:*    * *:*  :
```

```
>lcl|PDB:1TI5_A mol:protein length:46 plant defensin Length=46

Score = 25.8 bits (74), Expect = 1.9
Identities = 12/31 (39%), Positives = 14/31 (45%), Gaps = 2/31 (6%)

Query  7 KCSGSPQCLKPCKDAGMRFGKC--MNRKCHC 35  калиотоксин
         KC     C   CK+ G G   C  M R C+C
Sbjct 12 KCLIDTTCAHSCKNRGYIGGNCKGMTRTCYC 42  дефензин
```

Множественное выравнивание последовательностей

Что полезного?

- Выявление удаленной гомологии
- Выявление консервативных остатков и мотивов
- Построение филогенетических деревьев
- …

Алгоритмы:

-Динамическое программирование
- Построение профилей
- Прогрессивное выравнивание
- Итеративные методы
- Скрытые марковские модели
- Квантовые компьютеры?! (2017)

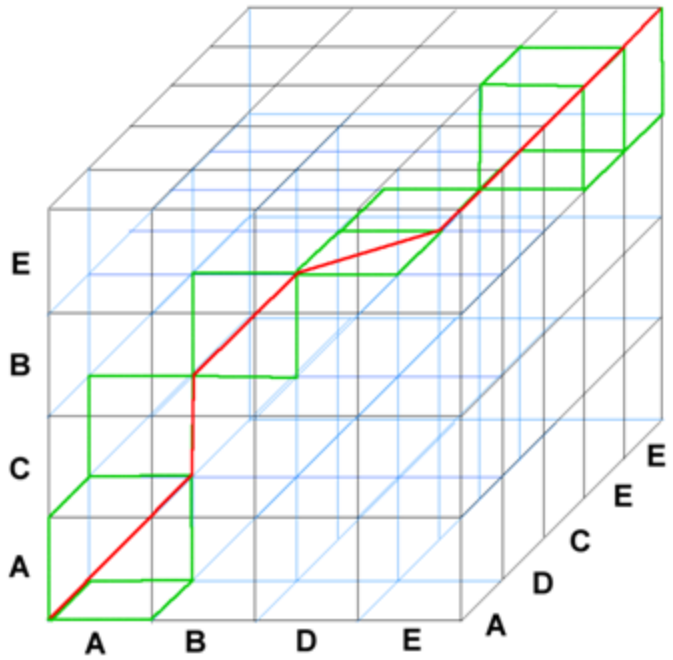
Визуализация:

Динамическое программирование

- Прямой метод выполнения множественного выравнивания, обеспечивающий нахождение глобального оптимума.

- Для выравнивания *N* последовательностей требуется построение *N*-мерной матрицы. Таким образом, пространство поиска растет экспоненциально с ростом *N* и также зависит от длины последовательностей, а время поиска может быть оценено как O(L^N).

```
A-BD-E-
ACB--E-
A--DCEE
```



Деревья: определения

Граф – структура, содержащая вершины, соединенные ребрами.

Граф называется связным, если содержит хотя бы один путь между двумя любыми вершинами.

Дерево – связный граф, содержащий только один путь между двумя вершинами. Может быть укорененным и неукорененным.

Двоичное дерево – ориентированное дерево, в котором исходящие степени вершин (число исходящих рёбер) не превосходят 2.

Длина ребра – число, соотнесенное с каждым ребром и обозначающее, в каком-то смысле, расстояние между двумя вершинами, соединенными этим ребром.

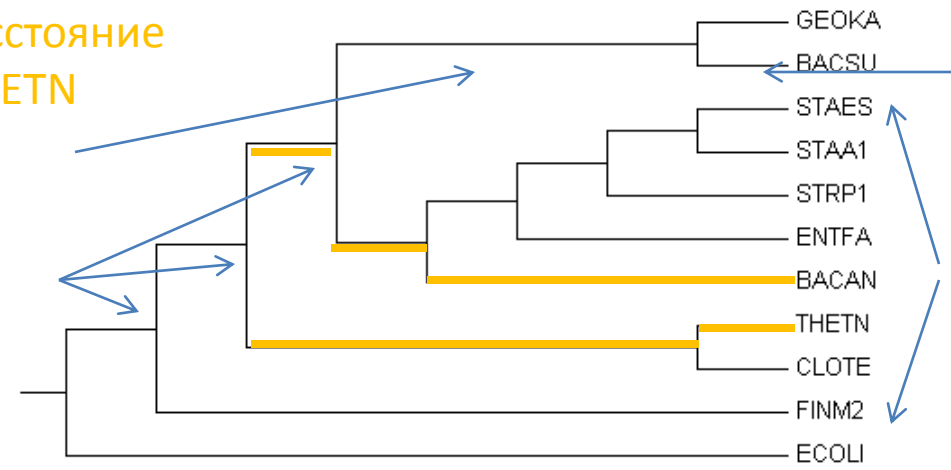
Длина пути – сумма всех длин ребер, составляющих путь.

Эволюционное расстояние между BACAN и THETN

Внутренние ветви

Внутренние узлы

Корень



Внешние ветви

Внешние узлы (листья)

Деревья: свойства

Число деревьев

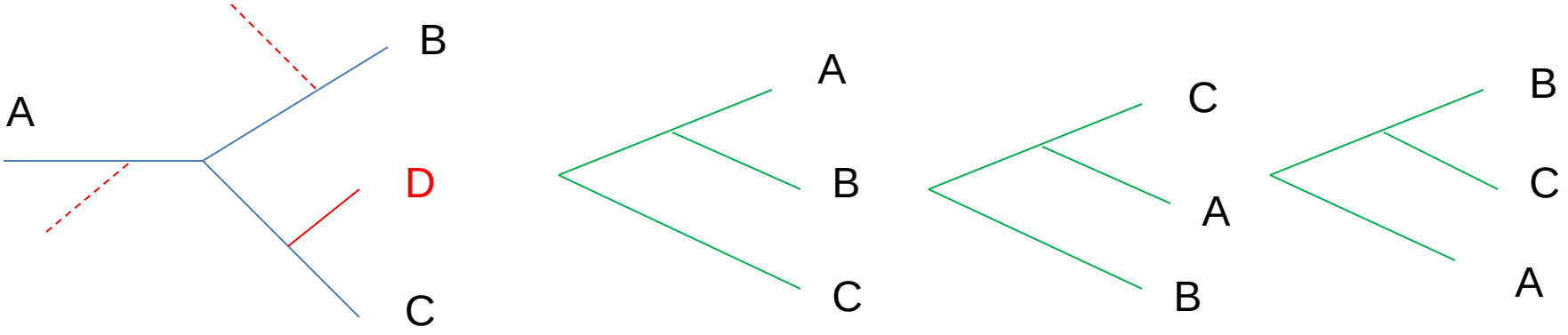
неукорененных

$$N_U = (2n-5)!!$$

укорененных

$$N_R = (2n-3)!!$$

С точки зрения филогении, правильное только одно!

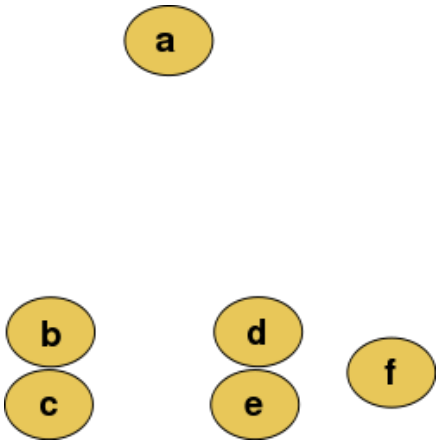


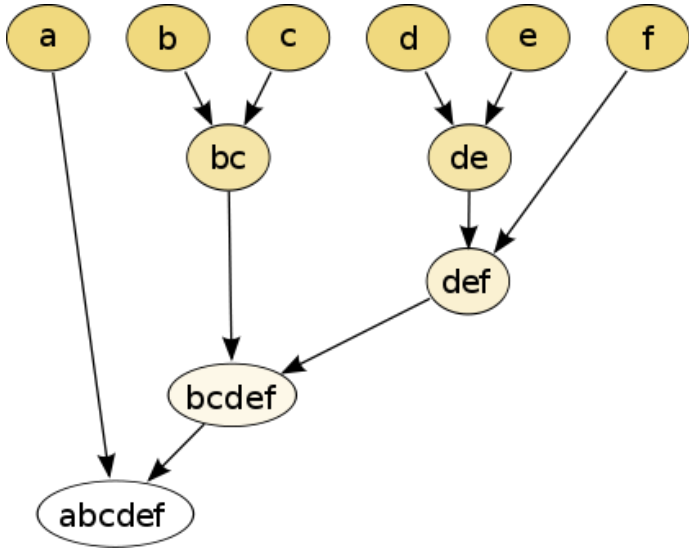
Кладограмма — филогенетическое дерево, не содержащее информации о длинах ветвей.

Филограмма — филогенетическое дерево, содержащее информацию о длинах ветвей; эти длины представляют изменение некой характеристики.

Методы кластеризации. UPGMA

UPGMA – Unweighted Pair Group Method with Arithmetic mean (1958) – метод невзвешенной группировки с арифметическим средним – пример алгоритма иерархической кластеризации





Расстояние между элементами

$$\|a-b\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$$

Расстояние между кластерами

$$\frac{1}{|\mathcal{A}| \cdot |\mathcal{B}|} \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} d(x, y).$$

Методы кластеризации. UPGMA

Дан набор объектов Sk, где для каждой пары (Si, Sj) установлена мера сходства L(Si, Sj). Для построения дерева выбирают два наиболее близких объекта (Sm, Sn) и добавляют вершину, изображающую их общего «предка» (Smn). Затем замещают эти два объекта группой, содержащий обоих, и присваивают расстояниям от этой пары до остальных объектов Sk средние значения от каждого из элементов этой группы до Sk:

$$L(S_{mn}, S_k) = \frac{L(S_m, S_k) + L(S_n, S_k)}{2}$$

В случае объединения кластеров C_i и C_j с образованием кластера C_k, содержащего *n_i + n_j = n_k* элементов, расстояние от кластера C_k до остальных кластеров C_m вычисляется как

$$L(C_k, C_m) = \frac{n_i L(C_i, C_m) + n_j L(C_j, C_m)}{n_i + n_j}$$

Методы кластеризации. UPGMA

Дано 6 последовательностей – ATTTG, AGCGT, ACCGT, CGCGA, GGCGA, CGGGC.

Используя расстояние по Хэммингу, получаем матрицу расстояний:

L	ATTTG	AGCGT	ACCGT	CGCGA	GGCGA	CGGGC
ATTTG	0	4	4	5	5	4
AGCGT		0	***1***	2	2	3
ACCGT			0	3	3	4
CGCGA				0	***1***	2
GGCGA					0	3
CGGGC						0

Методы кластеризации. UPGMA

Дано 6 последовательностей – ATTTG, AGCGT, ACCGT, CGCGA, GGCGA, CGGGC.

Используя расстояние по Хэммингу, получаем матрицу расстояний:

L	ATTTG	AGCGT	ACCGT	CGCGA	GGCGA	CGGGC
ATTTG	0	4	4	5	5	4
AGCGT		0	***1***	2	2	3
ACCGT			0	3	3	4
CGCGA				0	***1***	2
GGCGA					0	3
CGGGC						0

L	ATTTG	AGCGT, ACCGT	CGCGA	GGCGA	CGGGC
ATTTG	0	(4+4)/2=4	5	5	4
AGCGT, ACCGT		0	(2+3)/2=2,5	(2+3)/2=2,5	(3+4)/2=3,5
CGCGA			0	**1**	2
GGCGA				0	3
CGGGC					0

Методы кластеризации. UPGMA

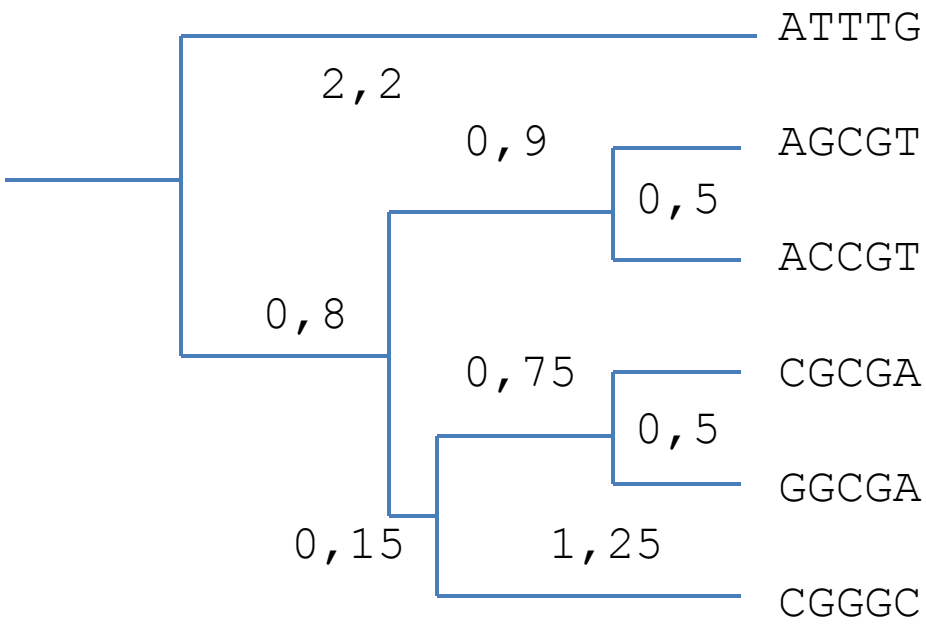
L	ATTTG	AGCGT, ACCGT	CGCGA, GGCGA	CGGGC
ATTTG	0	4	(5+5)/2=5	4
AGCGT, ACCGT		0	(2,5+2,5)/2=2,5	3,5
CGCGA, GGCGA			0	(2+3)/2=**2,5**
CGGGC				0

L	ATTTG	AGCGT, ACCGT	(CGCGA, GGCGA), CGGGC
ATTTG	0	4	(2*5+4)/3=4,7
AGCGT, ACCGT		0	(2*2,5+3,5)/3=**2,8**
(CGCGA, GGCGA), CGGGC			0

L	ATTTG	((CGCGA, GGCGA), GGGC), (AGCGT, ACCGT)
ATTTG	0	(4*2+4,7*3)/5=4,4
((CGCGA, GGCGA), CGGGC), (AGCGT, ACCGT)		0

Методы кластеризации. UPGMA

Объединяя теперь кластеры, получим дерево :

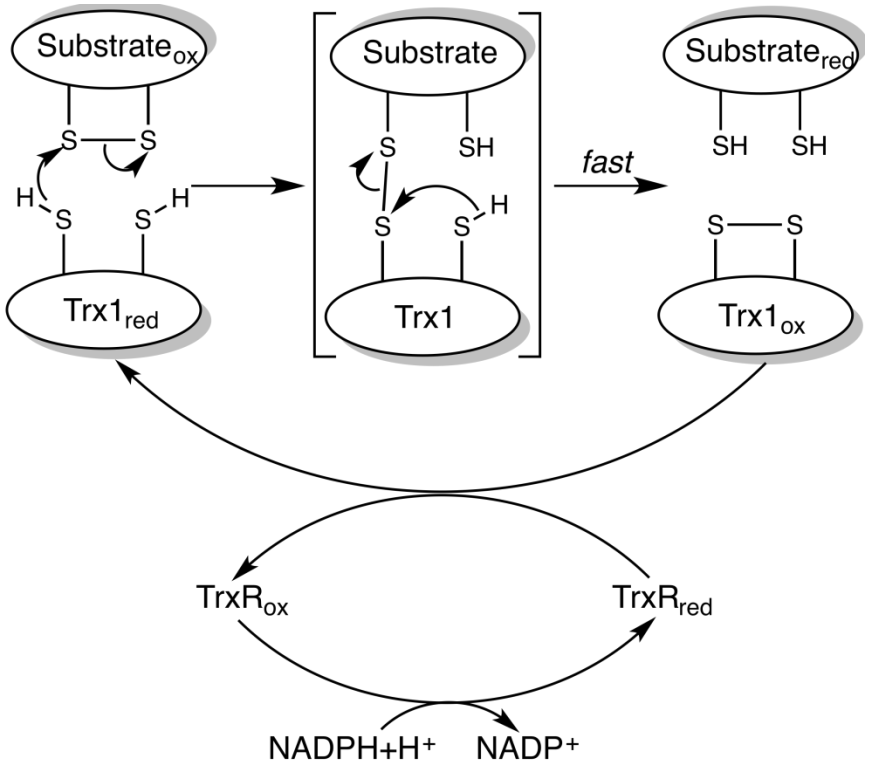


Длины ветвей установлены так, что расстояние от корня одинаково для всех листьев - ультраметричность.

Метод UPGMA подразумевает справедливость гипотезы молекулярных часов (постоянной скорости эволюции) (Э. Цукеркандль, Л. Полинг, 1962).

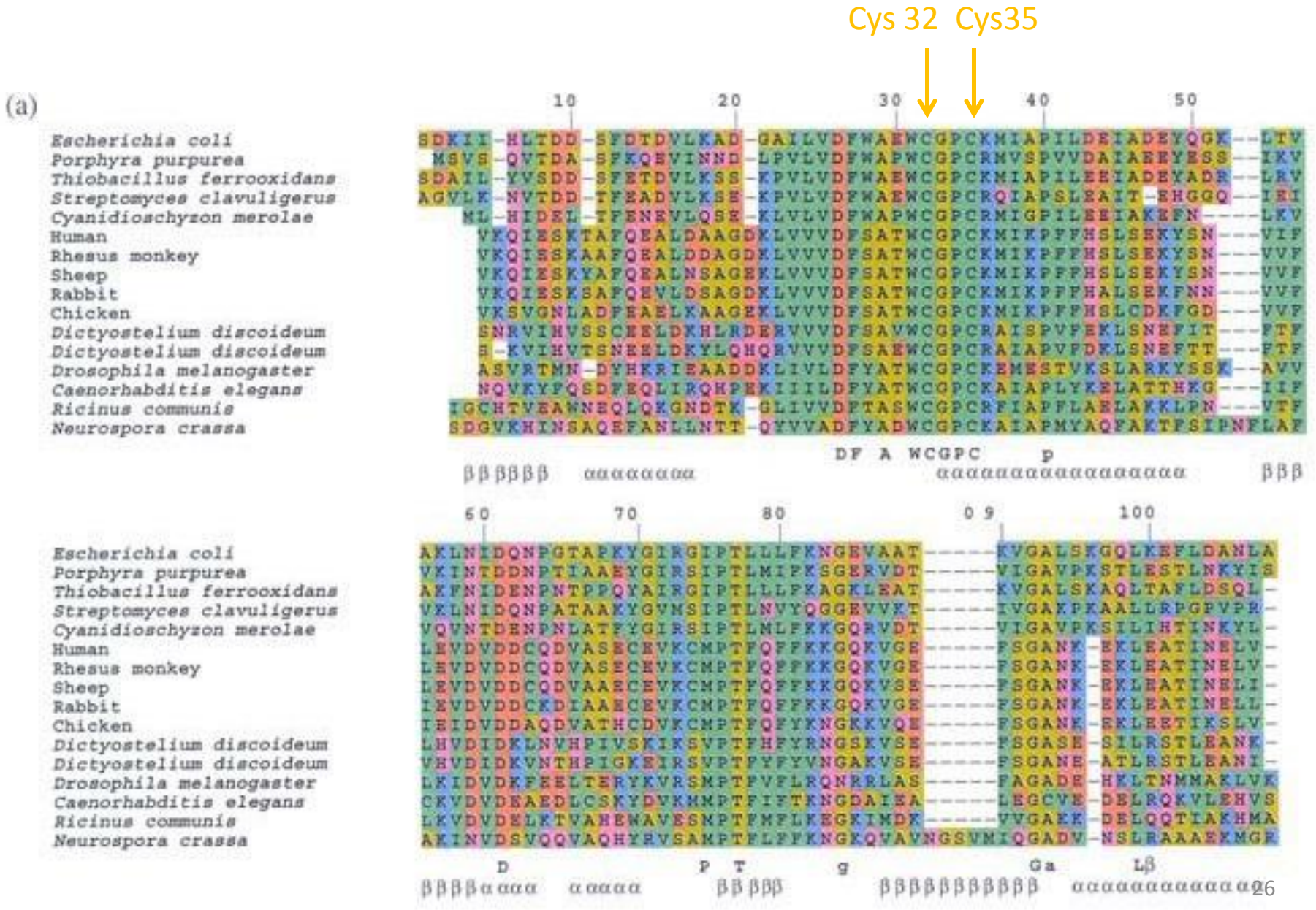
Построение профилей

Тиоредоксины – семейство белков, отвечающих за восстановление дисульфидных связей в белках и встречающихся как в животном, так и в растительном мире.

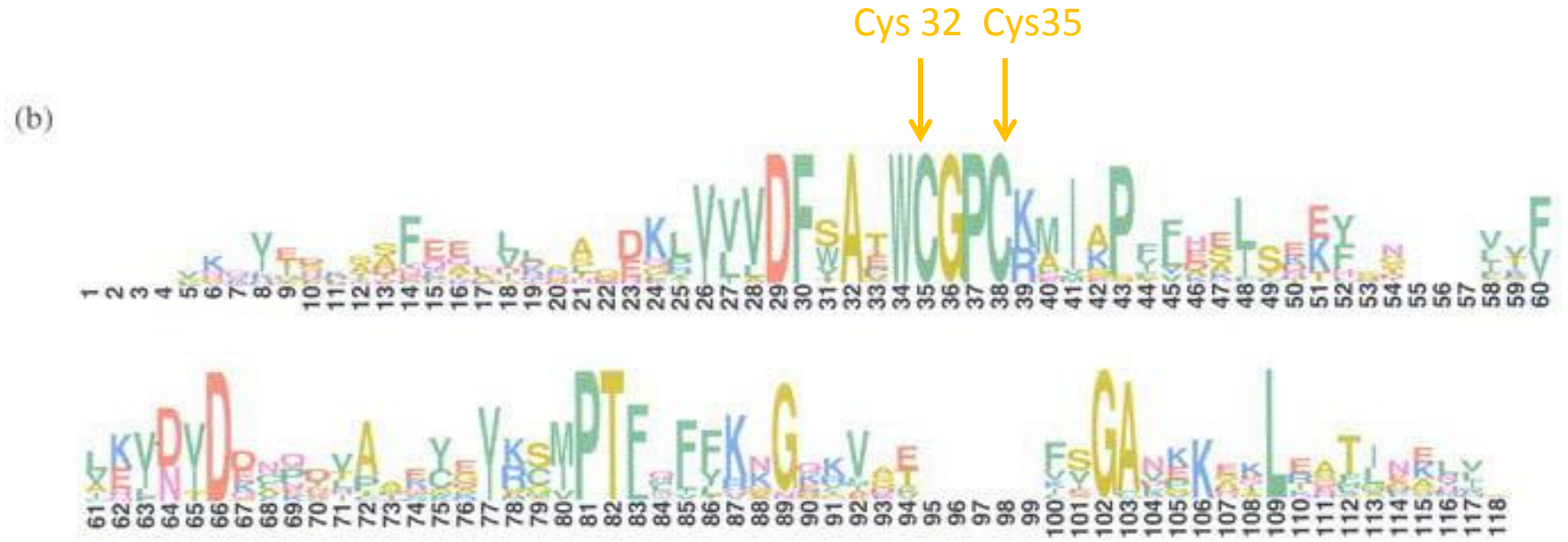


Выравнивание структур тиоредоксина человека и мушки *Drosophila melanogaster*.

Построение профилей



Построение профилей



Номер остатка	Число остатков каждой аминокислоты																		
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
28 (25)	1									2								13		
29 (26)			16																	
30 (27)					16															
31 (28)																7	1		5	3
32 (29)	16																			
33 (30)			1	4									2			1	7	1		

Построение профилей

В реальности нужно учитывать, что встречаемость аминокислот не равновероятна, и использовать матрицы замен (PSSM).

Применение:

-- Аккуратное выравнивание дальнородственных последовательностей

-- Выявление консервативных остатков и паттернов для:

-- предположения об активном центре или сайте связывания

-- идентификации и классификации родственных последовательностей

-- Выявление вариабельных остатков для потенциальных замен (петли в антителах)

Недостатком, очевидно, является то, что само выравнивание должно быть известно заранее.

Прогрессивное выравнивание. Clustal

Clustal (1988) выполняет постепенное выравнивание все новых последовательностей, начиная с наиболее <эволюционно> близких, ориентируясь на предварительно построенное на основании парных выравниваний филогенетическое дерево.

Алгоритм:

1) Экспресс-оценка выравнивания двух последовательностей вычисляется как число совпадающих остатков в словах длины K (K = 1-2 для белковых последовательностей и 2-4 для нуклеотидных) за вычетом штрафа за сделанные вставки.

2) Методом UPGMA (позже NJ) рассчитывается направляющее дерево, по которому затем рассчитываются веса последовательностей, причем более близкие последовательности получают меньшие веса.

3) Согласно направляющего дерева выбираются наиболее близкие последовательности и выполняется их выравнивание методом динамического программирования с использованием матрицы замен и штрафов за открытие/расширение вставок, с полученным выравниванием сопоставляются всё новые последовательности.

Прогрессивное выравнивание. Clustal

Особое внимание уделено значениям штрафов за вставки. Введена их зависимость от:

А) типа сопоставляемого и предшествующего остатков;

Б) степени близости последовательностей;

В) длин рассматриваемых последовательностей;

Г) наличия вставок в уже имеющемся выравнивании;

Д) характера аминокислотной последовательности.

One gap, always gap

Текущая версия - Clustal Omega - использует СММ

http://www.ebi.ac.uk/Tools/msa/clustalo/

Итеративное выравнивание. MUSCLE

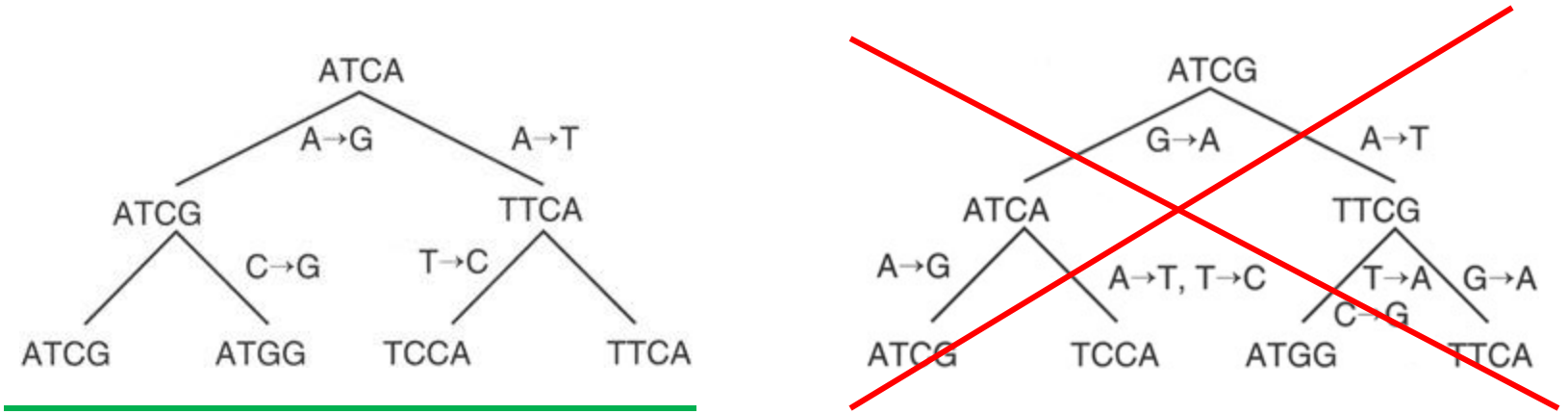
Три стадии:

1) быстрое «черновое» множественное выравнивание (попарные глобальные выравнивания; оценка сходства как доля совпадающих позиций; построение направляющего дерева методом UPGMA или NJ; прогрессивное выравнивание)
2) улучшенное множественное выравнивание (оценка сходства как доля совпадающих позиций в текущем множественном выравнивании; построение направляющего дерева через построение матрицы расстояний по Кимуре и ее кластеризацию; сравнение текущего дерева с построенным ранее; пересчет выравнивания для отличающихся узлов; повторение до сходимости)
3) уточнение выравнивания (удаление произвольного узла для разбиения дерева на два; построение профилей для каждого поддерева и их выравнивание; расчет суммы парных оценок в получающемся множественном выравнивании; перебор всех узлов от листьев к корню и выбор выравнивания с максимальной суммой)

http://www.drive5.com/muscle/ 25000+ цитирований (2018)

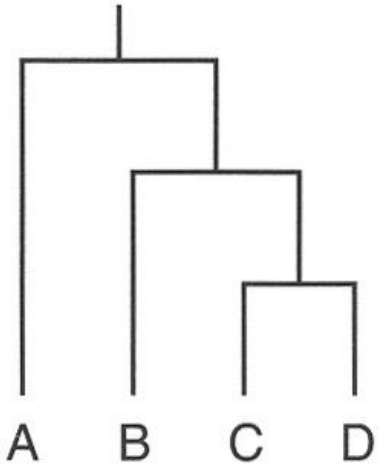
Филогенетические деревья – методы

Максимальная экономия (Fitch, 1971) (метод оценки!) – критерий оптимальности, согласно которому предпочтительнее деревья с меньшим суммарным числом мутаций. Однако алгоритма быстрого построения такого дерева не существует.



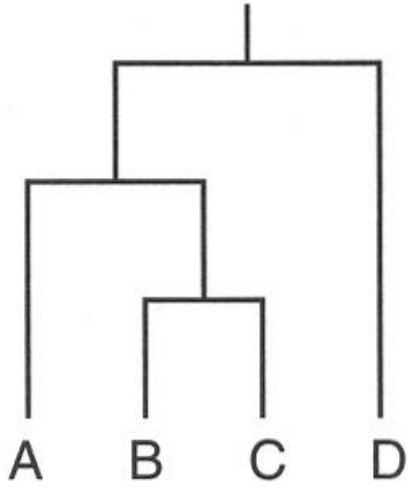
Метод максимального правдоподобия учитывает не просто число мутаций, но и их вероятность.

Филогенетические деревья – проблема переменной скорости эволюции

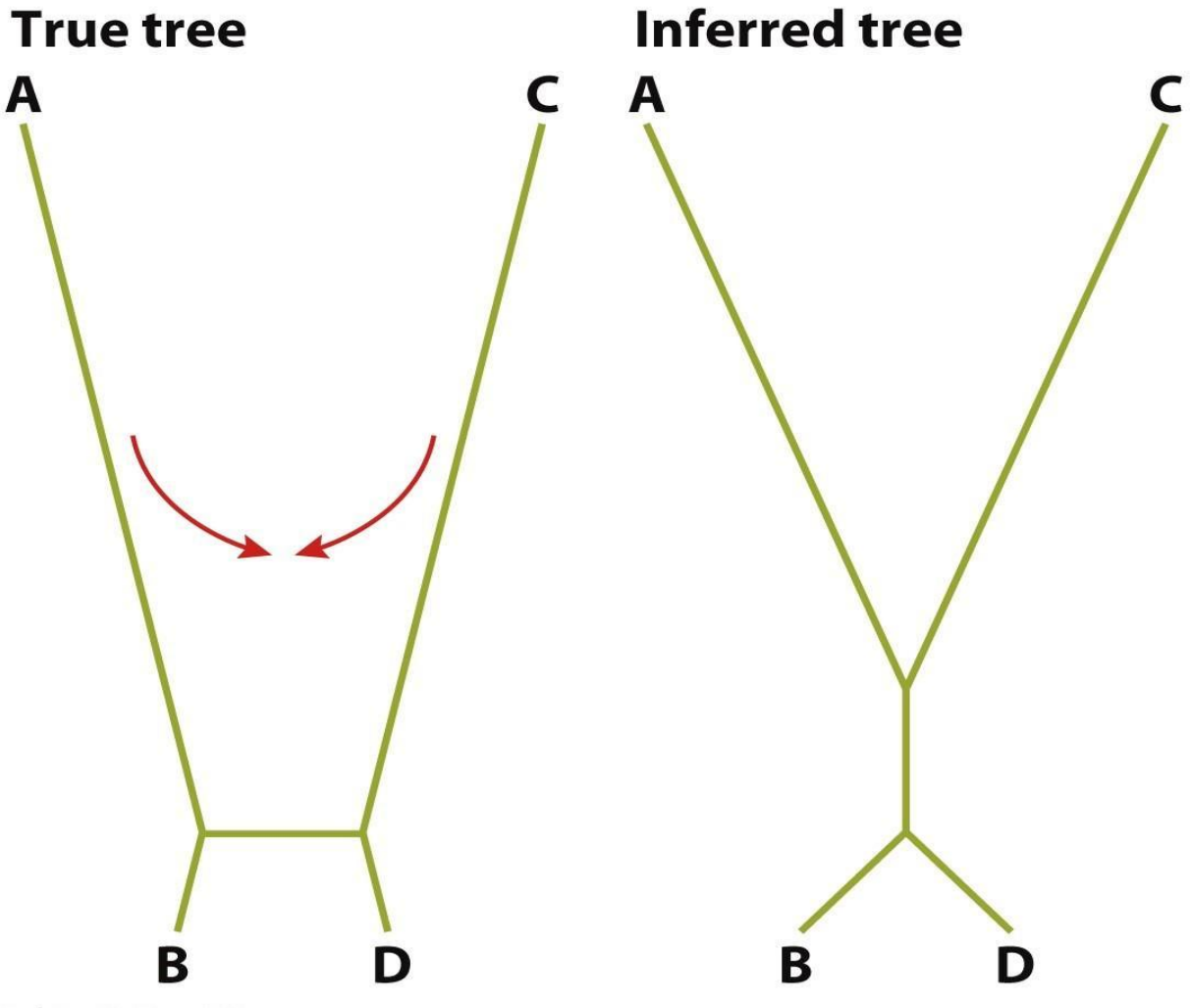


	A	B	C	D
A	0	3	3	3
B		0	2	2
C			0	1
D				0

	A	B	C	D
A	0	3	3	20
B		0	2	20
C			0	20
D				0

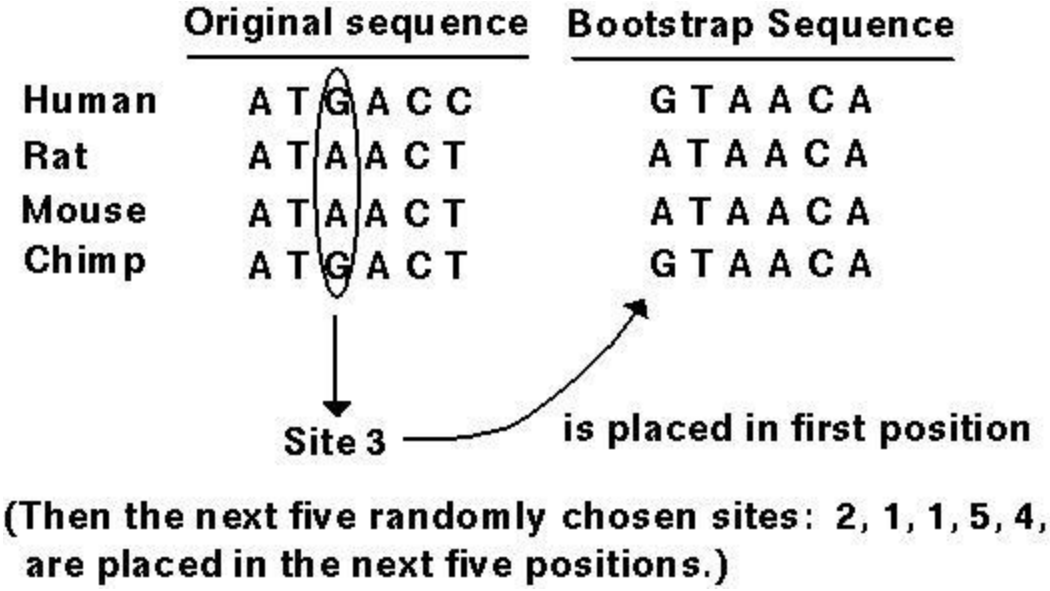


Филогенетические деревья – «притяжение длинных ветвей»

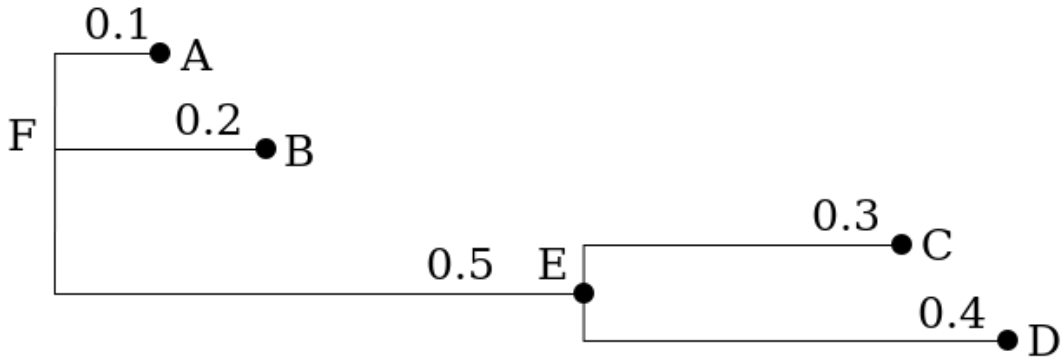


Филогенетические деревья – методы проверки

1) Использование внешней группы, т.е. видов, которые заведомо более удалены ото всех видов, для которых строится дерево (приматы и корова);
2) Сравнение деревьев, полученных на основе разных характеристик. Очевидно, они должны быть согласованными;
3) Оценка результата с помощью формальных статистических тестов. Например, построение дерева для подмножества последовательностей из исходного множественного выравнивания должно дать поддерево дерева, полученного для этого выравнивания;
4) Бутстреп (boot strap)



Филогенетические деревья.
Скобочная формула (Newick format) (1986)

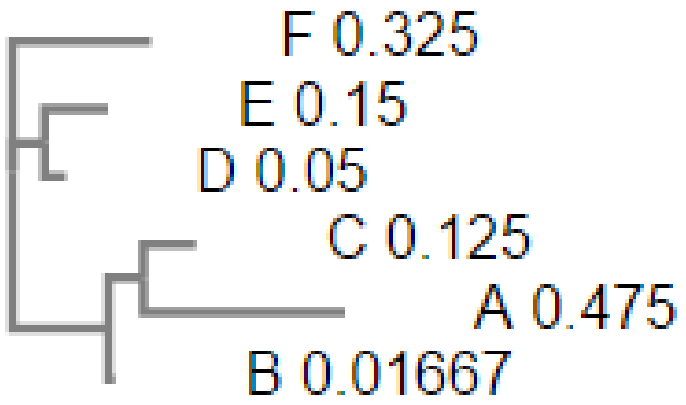


`(,,(,));`	*имена узлов не указаны*
`(A,B,(C,D));`	*указаны только имена листьев*
`(A,B,(C,D)E)F;`	*указаны имена всех узлов*
`(:0.1,:0.2,(:0.3,:0.4):0.5);`	*для всех узлов кроме корня указано расстояние до родительского узла*
`(:0.1,:0.2,(:0.3,:0.4):0.5):0.0;`	*для всех узлов указано расстояние до родительского узла*
`(A:0.1,B:0.2,(C:0.3,D:0.4):0.5);`	*указаны имена листьев и расстояния*
`(A:0.1,B:0.2,(C:0.3,D:0.4)E:0.5)F;`	*указаны все имена и расстояния*

Филогенетические деревья.
Скобочная формула (Newick format) (1986)

```
CLUSTAL O(1.2.1) multiple sequence alignment

F CGGGC
C ACCGT
B AGCGT
E GGCGA
A ATTTG
D CGCGA
```

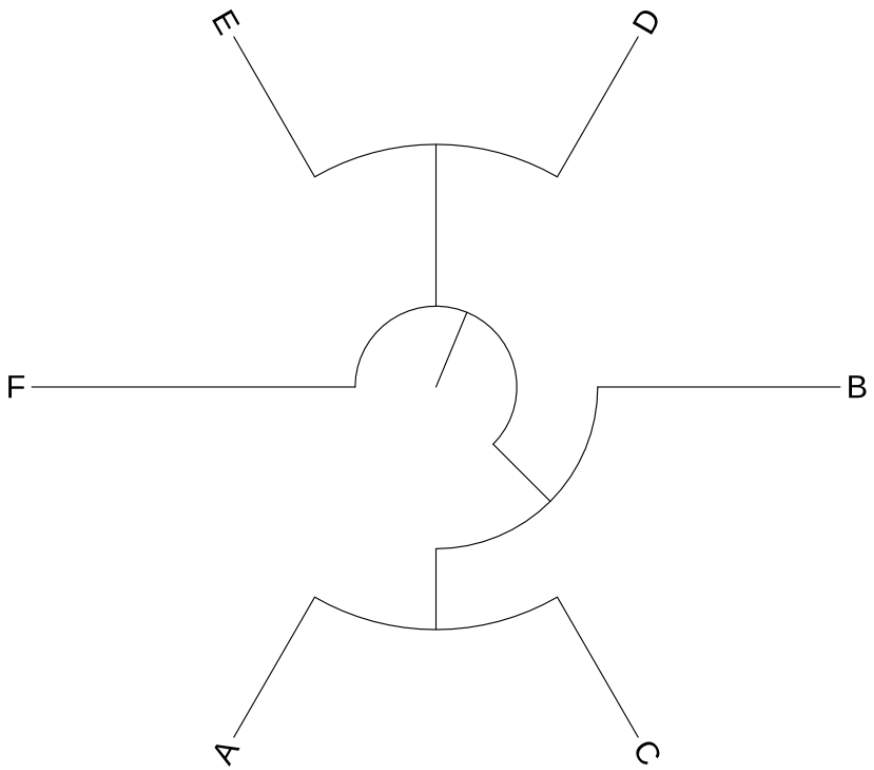
```
(
F:0.32500,
(
E:0.15000,
D:0.05000)
:0.07500,
(
(
C:0.12500,
A:0.47500)
:0.08333,
B:0.01667)
:0.22500);
```



This is a Neighbour-joining tree without distance corrections.

Филогенетические деревья. Круговое представление

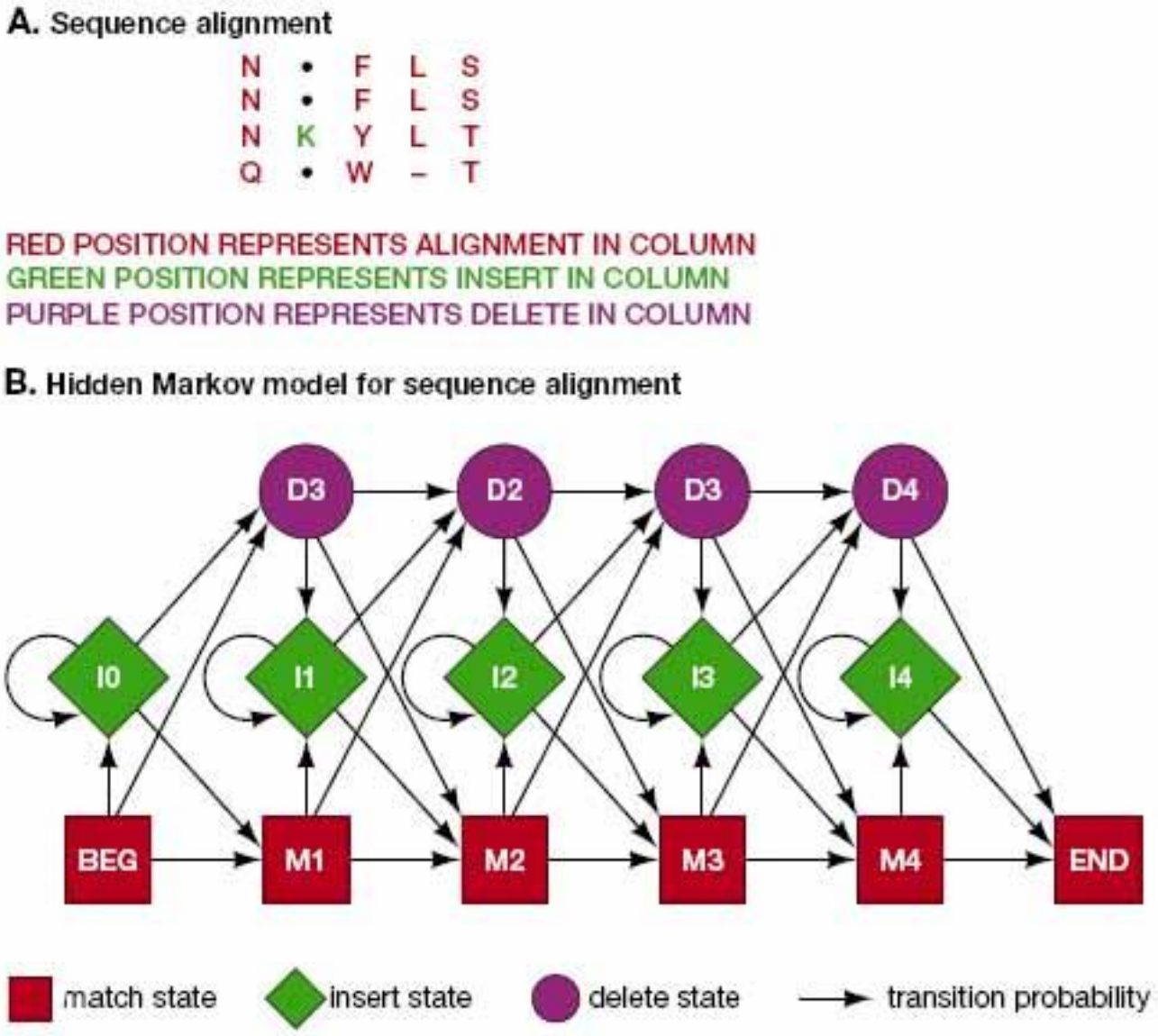
```
CLUSTAL O(1.2.1) multiple sequence alignment

F CGGGC
C ACCGT
B AGCGT
E GGCGA
A ATTTG
D CGCGA
```




```
(
F:0.32500,
(
E:0.15000,
D:0.05000)
:0.07500,
(
(
C:0.12500,
A:0.47500)
:0.08333,
B:0.01667)
:0.22500);
```

http://itol.embl.de

Скрытые марковские модели



Скрытые марковские модели

Алгоритм:

- Обучение. Имея ряд невыровненных последовательностей, можно выровнять их и подогнать вероятности переходов и порождения остатков, чтобы определить модель, описывающую данный набор последовательностей.
- Поиск гомологов. Имея модель и исследуемую последовательность, можно посчитать вероятность того, то модель могла бы сгенерировать эту последовательность. Если вероятность достаточно высока, то рассматриваемая последовательность принадлежит тому же семейству, что и обучающие.

Скрытые марковские модели

Алгоритм:

- Обучение. Имея ряд невыровненных последовательностей, можно выровнять их и подогнать вероятности переходов и порождения остатков, чтобы определить модель, описывающую данный набор последовательностей.
- Поиск гомологов. Имея модель и исследуемую последовательность, можно посчитать вероятность того, то модель могла бы сгенерировать эту последовательность. Если вероятность достаточно высока, то рассматриваемая последовательность принадлежит тому же семейству, что и обучающие.

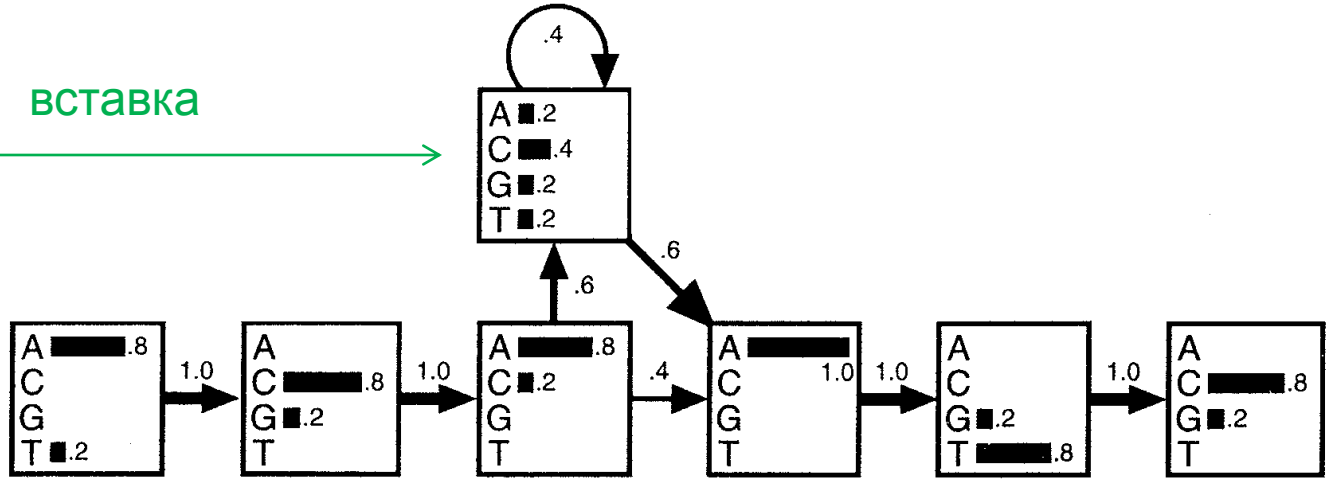
```
ACA---ATC
TCAACTATC
ACAC--AGC
AGA---ATC
ACCG--ATG
```

Построим?

Скрытые марковские модели

ACA---ATC
TCAACTATC
ACAC--AGC
AGA---ATC
ACCG--ATG

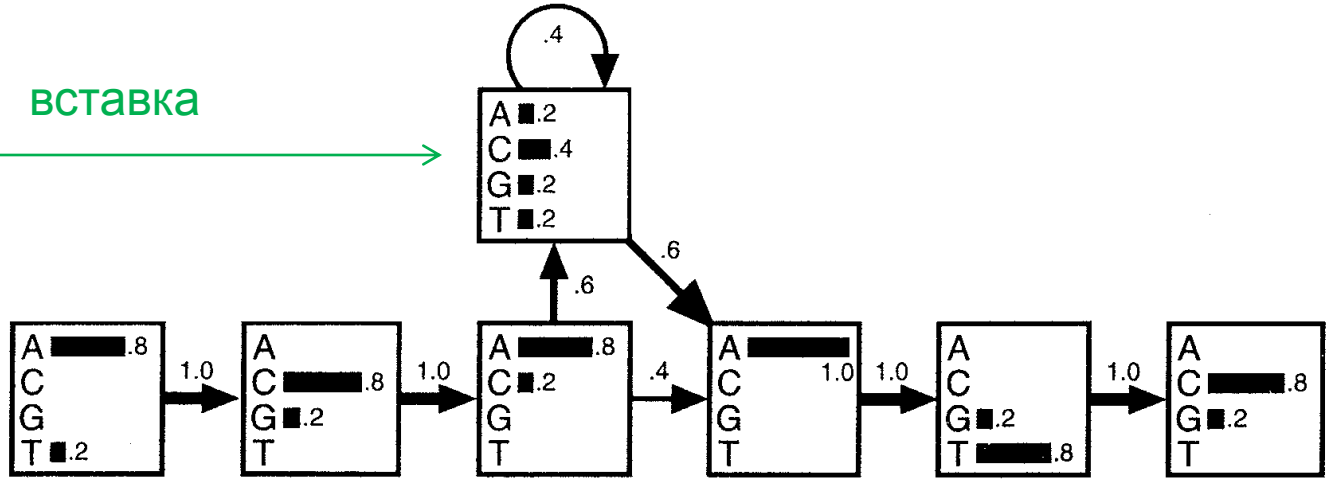
вставка



Скрытые марковские модели

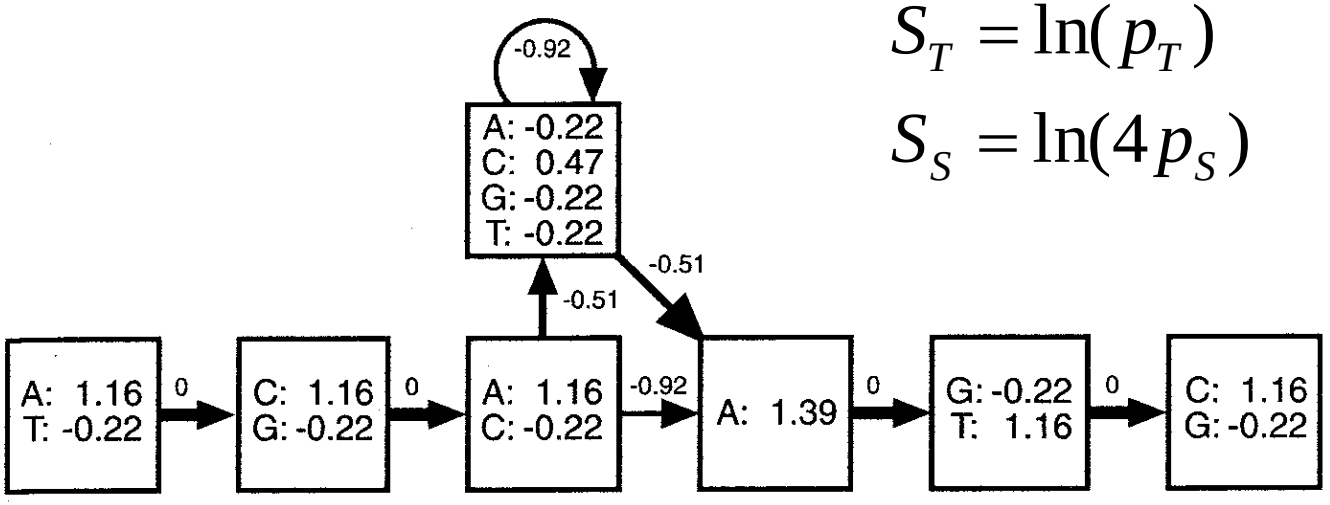
ACA---ATC
TCAACTATC
ACAC--AGC
AGA---ATC
ACCG--ATG

вставка



$$S_T = \ln(p_T)$$

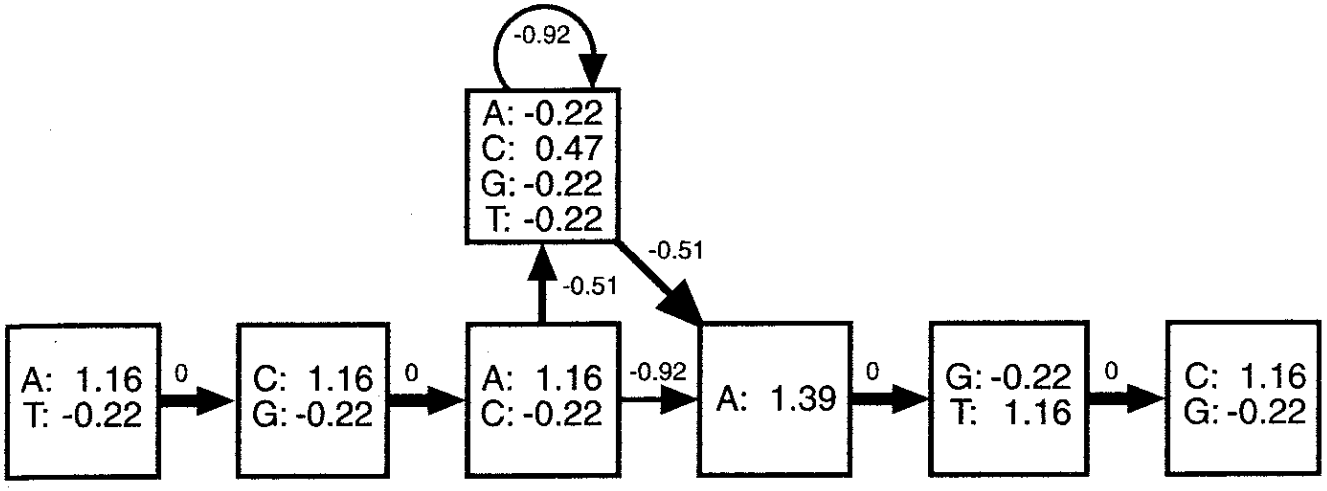
$$S_S = \ln(4p_S)$$



Скрытые марковские модели

```
ACA---ATC
TCAACTATC
ACAC--AGC
AGA---ATC
ACCG--ATG
CGCGT-CGG
```

Посчитаем: описывает ли построенная модель новую последовательность?



Скрытые марковские модели

ACA---ATC S = 1.16 + 0 + 1.16 + 0 + 1.16 - 0.92 + 1.39 + 0 + 1.16 + 0 + 1.16 = 6.29

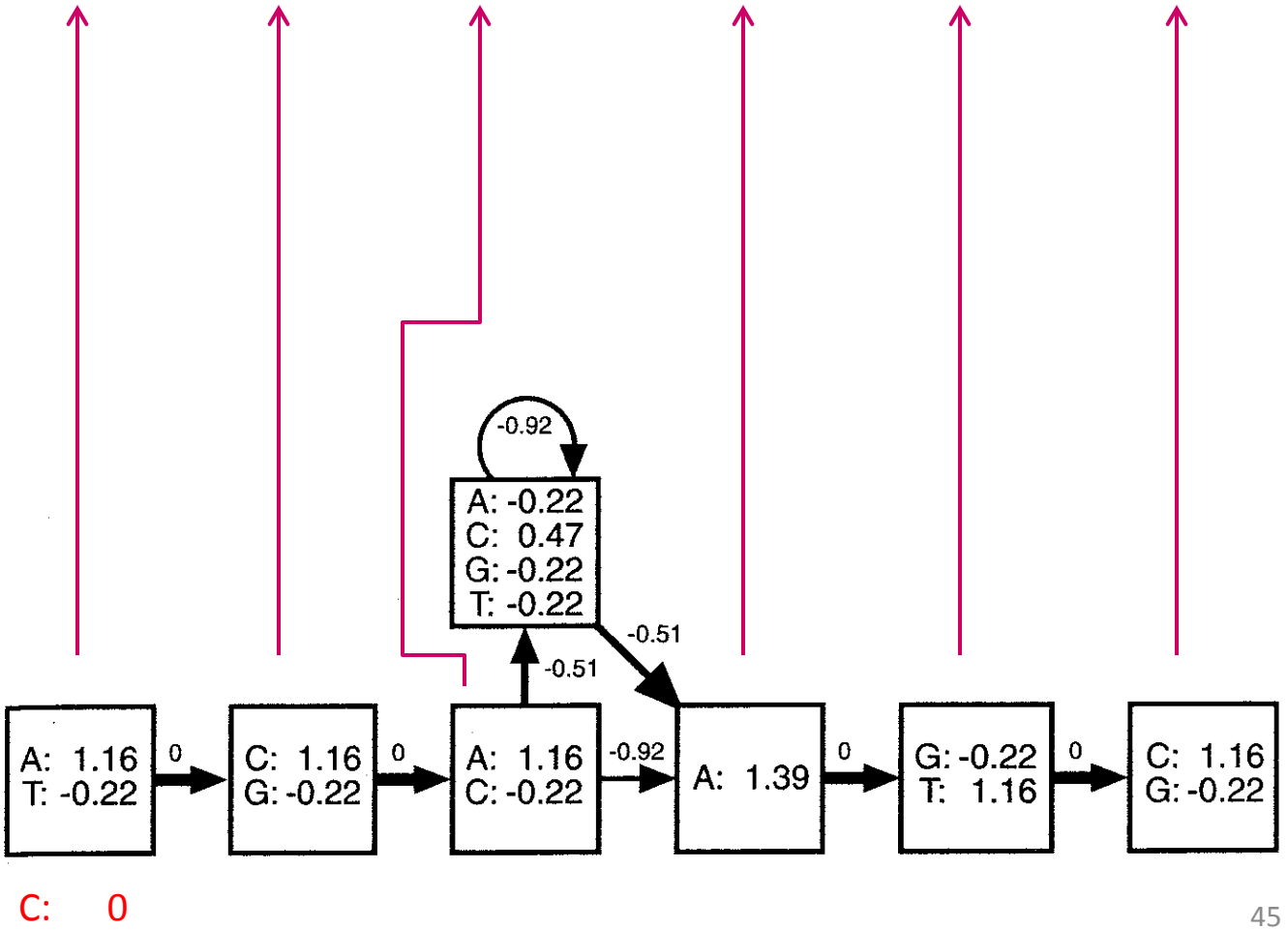
TCAACTATC ?

ACAC--AGC

AGA---ATC

ACCG--ATG

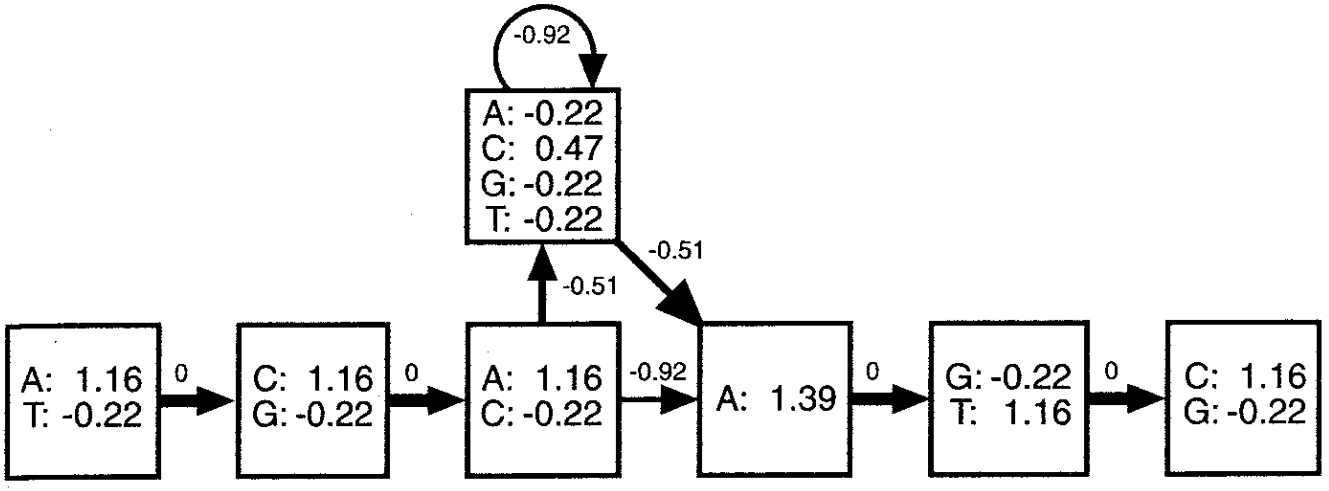
CGCGT-CGG ?



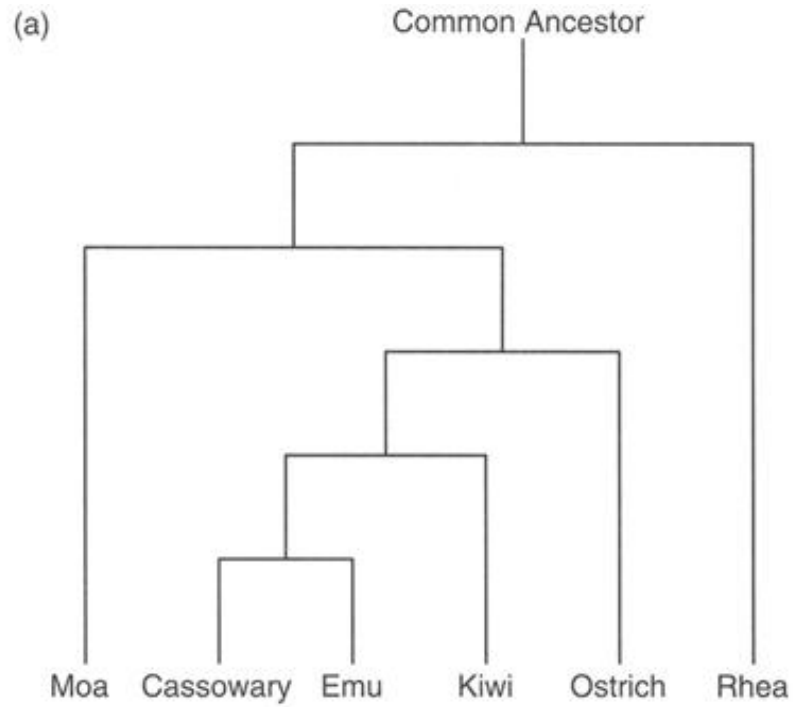
C: 0

Скрытые марковские модели

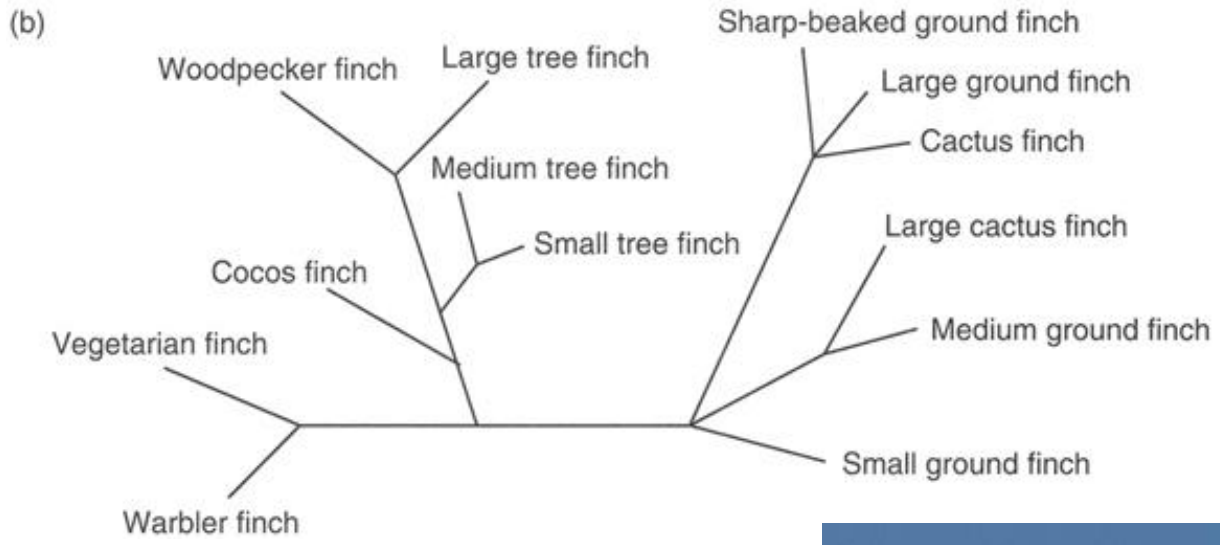
```
ACA---ATC    6,29
TCAACTATC    2,99
ACAC--AGC    5,26
AGA---ATC    4,90
ACCG--ATG    3,18
CGCGT-CGG   -3,28
```



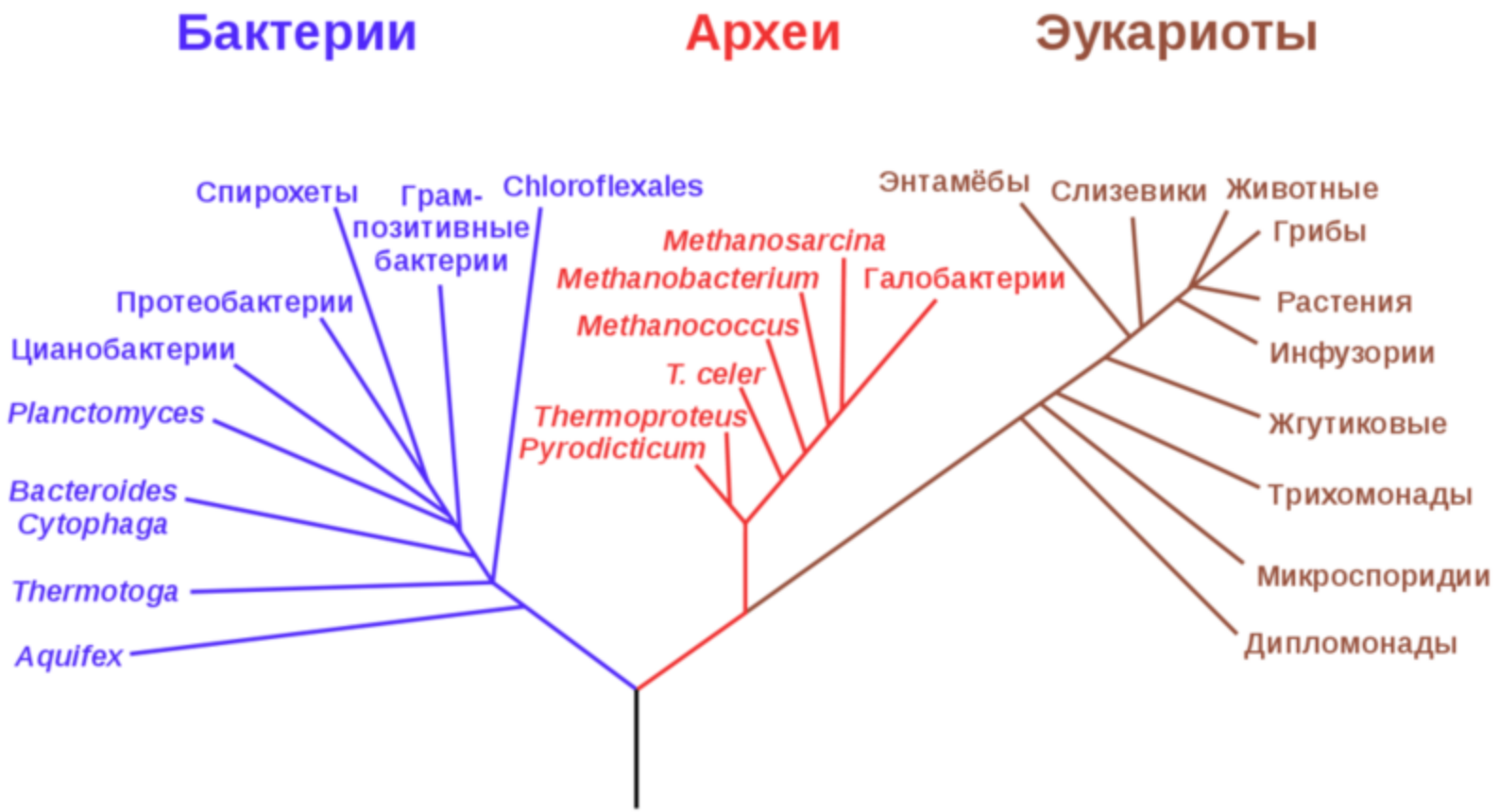
Филогенетические деревья – примеры



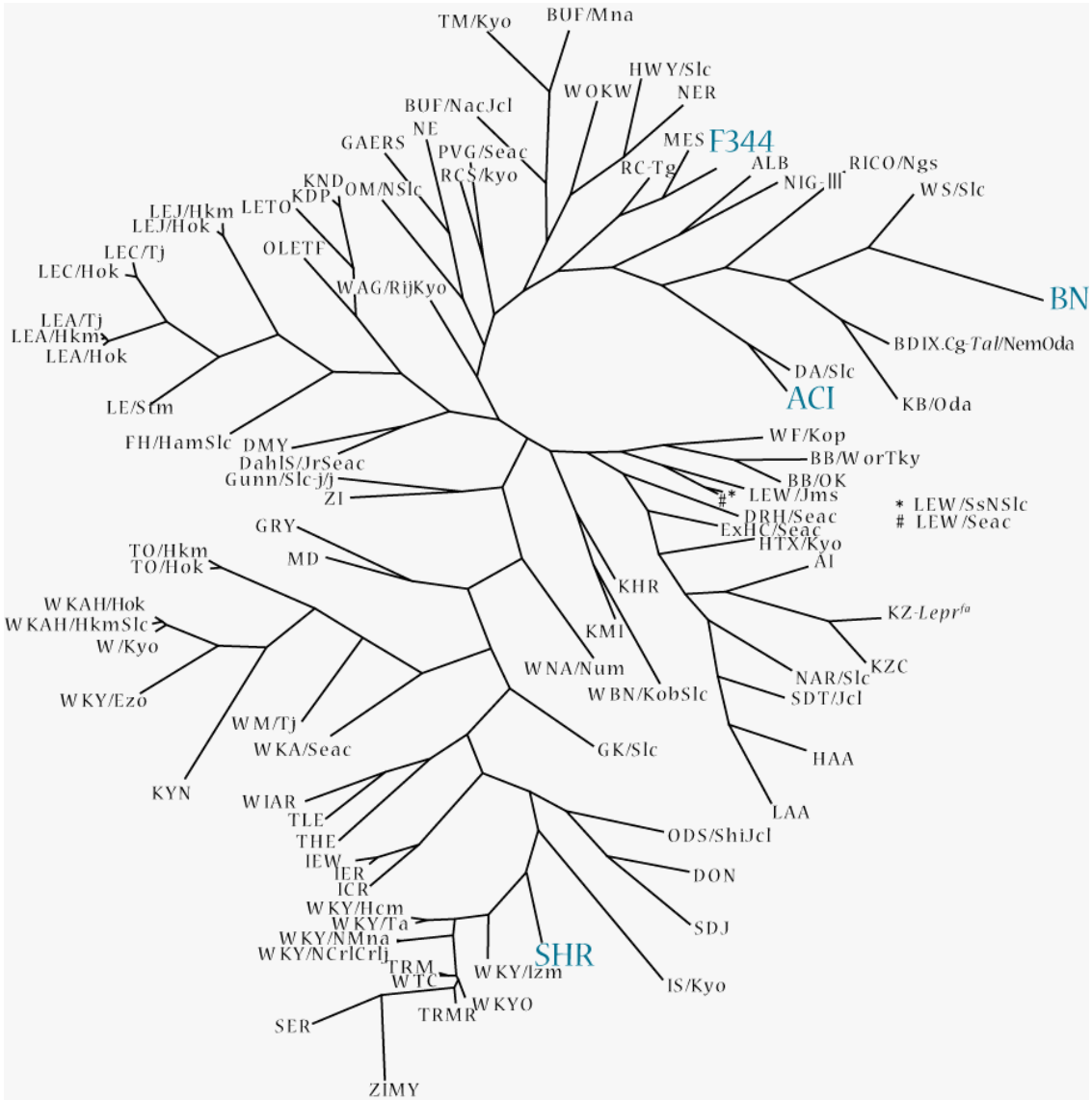
Филогенетические деревья – примери



Филогенетические деревья – такие разные



Филогенетические деревья – такие разные



Филогенетические деревья – такие разные