

Построение и бутстреп- анализ филогенетических деревьев

Фролова Людмила Леонидовна
к.т.н., доц. каф. генетики КФУ

Филогенетическое дерево

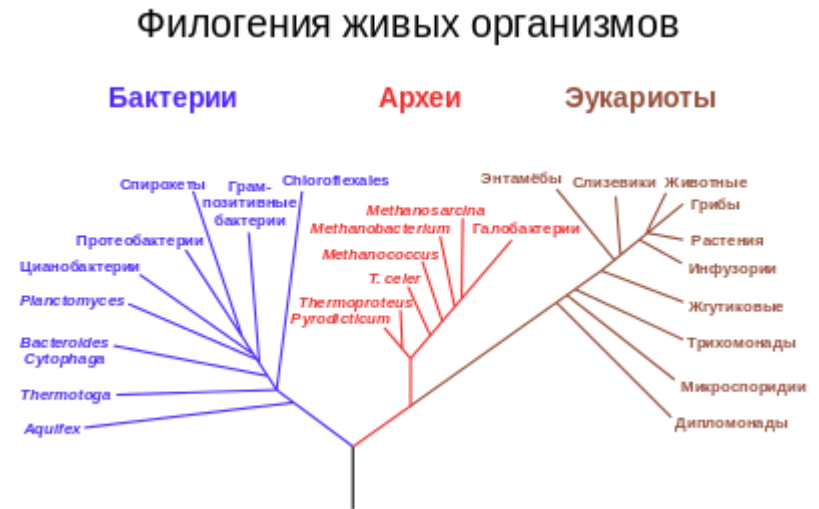
<http://ru.wikipedia.org/wiki/>

- **Филогенетическое дерево** (эволюционное дерево, дерево жизни) — дерево, отражающее эволюционные взаимосвязи между различными видами или другими сущностями, имеющими общего предка

Типы филогенетических деревьев

<http://ru.wikipedia.org/wiki/>

Укорененное дерево — дерево, содержащее выделенную вершину — корень. Корневое дерево можно считать ориентированным графом, поскольку на нем имеется естественная ориентация — от корня к листьям. Каждый узел корневого дерева отвечает последнему общему предку нижележащих листьев дерева. На рисунке представлено корневое филогенетическое дерево, окрашенное в соответствии с трёхдоменной системой живых организмов



Филогенетическое дерево

<http://ru.wikipedia.org/wiki/>

Филогенетическое дерево состоит из ветвей, узлов, листьев и (максимум один) корень:

Листья — это конечные вершины, то есть те, в которые входят ровно по одному ребру; каждый лист отображает некоторый вид живых организмов (или иной объект, подверженный эволюции, например, домен белка).

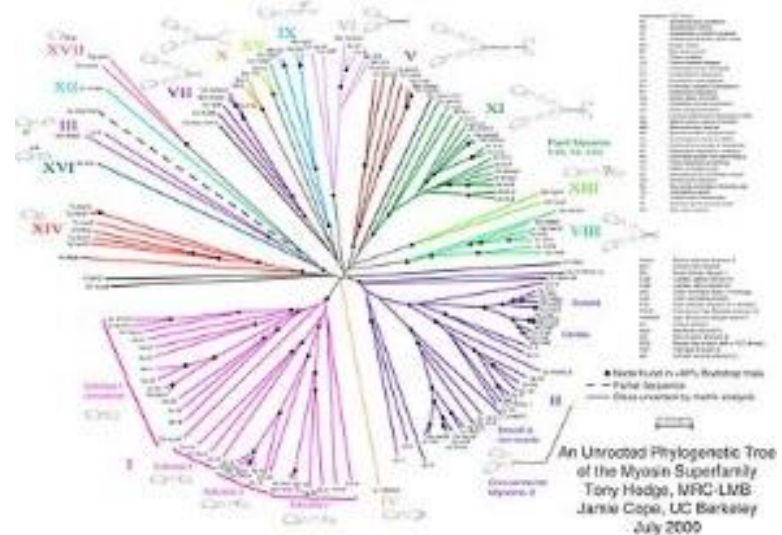
Каждый **узел** представляет эволюционное событие: разделение предкового вида на два или более, которые в дальнейшем эволюционировали независимо.

Корень представляет общего предка всех рассматриваемых объектов. Ребра филогенетического дерева принято называть «ветвями».

Типы филогенетических деревьев

<http://ru.wikipedia.org/wiki/>

Неукорененное дерево не содержит корня и отражает связь листьев без предполагаемого положения общего предка. Необходимость рассматривать некорневые деревья возникает из-за того, что часто связи между узлами восстановить легче, чем направление эволюции.



Графическое представление деревьев

- **Дендрограмма** — общий термин, обозначающий схематическое представление филогенетического дерева.
- **Кладограмма** — филогенетическое дерево, не содержащее информации о длинах ветвей.
- **Филограмма** (или **фенограмма**) — филогенетическое дерево, содержащее информацию о длинах ветвей; эти длины представляют изменение некой характеристики.
- **Хронограмма** — филограмма, длины ветвей в которой представляют эволюционное время.

Принципы дистанционных методов

1. Установление попарных эволюционных дистанций между анализируемыми последовательностями, представленных в виде матрицы дистанций. Эволюционная дистанция между двумя последовательностями измеряется их различиями, числом и качеством нуклеотидных замен или аминокислотных замещений.
2. При построении дерева с помощью дистанционного метода важен выбор эволюционной модели.

Эволюционные модели

Эволюционные модели и дистанции между нуклеотидными последовательностями построены на основе частоты замен нуклеотидов:

Модель Джукса – Кантора (1969)

Модель Кимуры (1980)

Модель Таджимы – Неи (1984)

Модель Хезегавы (1985)

Модель Тамуры (1992)

Модель Тамуры – Неи (1993)

Различные модели, рассматривающие эволюцию нуклеотидных последовательностей, учитывают множественные замены в одной позиции, а различие между моделями определяют тем, что они исходят из различных частот, вероятностей разных замен.

Простые эволюционные модели

изначально	Нуклеотид			
	после замены			
	A	T	C	G
1. Модель Джукса—Кантора				
A	—	α	α	α
T	α	—	α	α
C	α	α	—	α
G	α	α	α	—
2. Модель Кимуры				
A	—	β	β	α
T	β	—	α	β
C	β	α	—	β
G	α	β	β	—
3. Модель Таджимы—Неи				
A	—	β	γ	δ
T	α	—	γ	δ
C	α	β	—	δ
G	α	β	γ	—

α и β — частоты соответственно транзиций и трансверсий,

Модель **Джукса-Кантора** исходит из того, что вероятность замены любого нуклеотида на любой другой нуклеотид за единицу времени одинакова.

В модели **Кимуры** учитывается различная частота транзиций и трансверсий.

В модели **Таджимы – Неи** частоты замен на любой из четырех нуклеотидов отличаются друг от друга, но не зависят от того, какой другой из трех нуклеотидов находился в этой позиции раньше.

Сложные эволюционные модели

Нуклеотид				
Изначально	после замены			
	A	T	C	G
4. Модель Тамуры				
A	—	$(1-\theta)\beta$	$\theta\beta$	$\theta\beta$
T	$(1-\theta)\beta$	—	$\theta\alpha$	$\theta\beta$
C	$(1-\theta)\beta$	$(1-\theta)\alpha$	—	$\theta\alpha$
G	$(1-\theta)\alpha$	$(1-\theta)\beta$	$\theta\beta$	—
α и β — частоты соответственно транзиций и трансверсий, а θ — GC-содержание				
5. Модель Хазегавы и др.				
A	—	$g_T\beta$	$g_C\beta$	$g_G\alpha$
T	$g_A\beta$	—	$g_C\alpha$	$g_G\beta$
C	$g_A\beta$	$g_T\alpha$	—	$g_G\beta$
G	$g_A\alpha$	$g_T\beta$	$g_C\beta$	—
α и β — частоты соответственно транзиций и трансверсий, g_i — содержание нуклеотида i ($i = A, T, C, G$)				
6. Модель Тамуры—Неи				
A	—	$g_T\beta$	$g_C\beta$	$g_G\alpha_1$
T	$g_A\beta$	—	$g_C\alpha_2$	$g_G\beta$
C	$g_A\beta$	$g_T\alpha_2$	—	$g_G\beta$
G	$g_A\alpha_1$	$g_T\beta$	$g_C\beta$	—
α_1 и α_2 — частоты транзиций между соответственно пуринами и пиримидинами, β — частота трансверсий, g_i — содержание нуклеотида i ($i = A, T, C, G$)				

Модель **Тамура** учитывает влияние суммарного содержания нуклеотидов G и C (GC- содержание) на частоту транзиций и трансверсий.

Модель **Хазегавы** учитывает возможное влияние содержания каждого из 4-х нуклеотидов на частоту транзиций и трансверсий

в модели **Тамуры – Неи** учитывает возможность неодинаковой частоты транзиций для пуринов и пиримидинов.

Метод присоединения соседей

(Neighbor - Joining - NJ)

(Метод присоединения соседей.

Лукашов, 2009)

Наиболее распространенный
дистанционный метод построения
деревьев

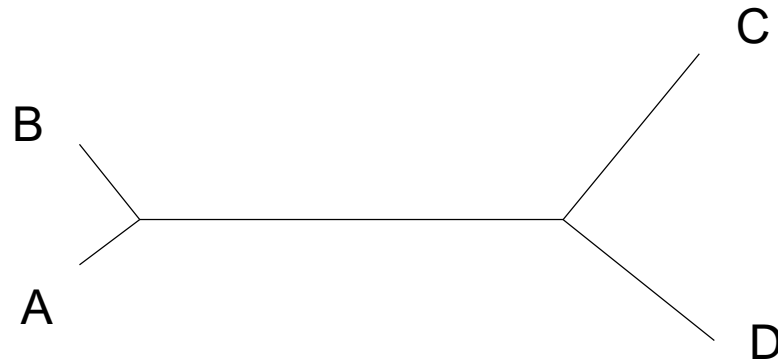
(**distance methods**, дистанционные
методы, называемые матричными)

Метод присоединения соседей (Neighbor-joining, NJ)

Соседями в филогенетическом анализе называют две последовательности, соединенные через один узел.

Последовательности А и В – соседи,
А и С – не соседи, но

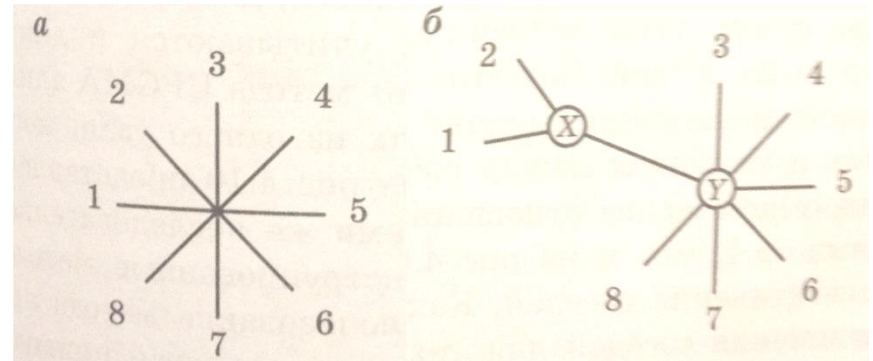
Композитная последовательность АВ является
соседом композитной последовательности CD



Метод присоединения соседей (Neighbor-joining, NJ)

Метод NJ последовательно объединяет данные до тех пор, пока все виды не объединятся в некорневое дерево.

1. Построение дерева со звездообразной топологией, т.е. отсутствием кластеров
2. Отделение 2-х последовательностей
3. Перебор каждой пары
4. Выбор пары, дающей наименьшую сумму длин ветвей
5. Расчет новой матрицы дистанций

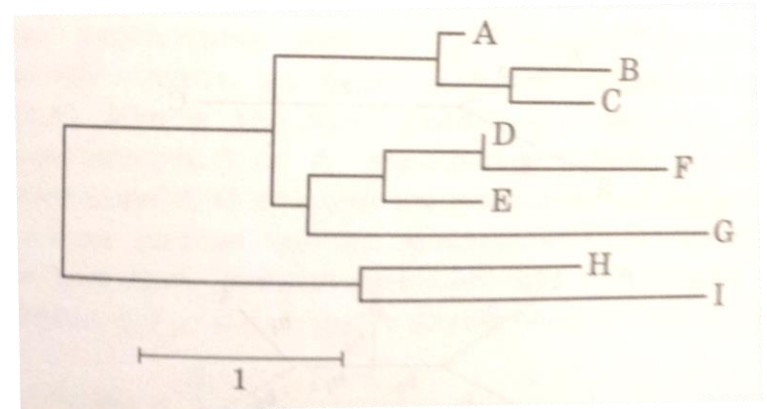


Метод присоединения соседей (Neighbor-joining, NJ)

При использовании метода присоединения соседей, длины ветвей, выходящих из каждого внутреннего узла, в общем случае не будут равны между собой

В масштабированных деревьях длина ветвей отражает число эволюционных событий.

В немасштабированном дереве длина всех ветвей одинакова, такие деревья показывают только порядок бифуркации.



Метод присоединения соседей (Neighbor-joining, NJ)

Быстрый метод, подходит для работы с большим количеством последовательностей.

Принципы методов анализа дискретных признаков

- Рассматривают различия между последовательностями в конкретных позициях.
- Целью методов анализа дискретных признаков является реконструкция сценария , наиболее вероятно объясняющего порядок конкретных нуклеотидных замен (аминокислотных замещений) в эволюционной истории анализируемой группы последовательностей.

Метод максимальной экономии

(parsimony, MP)

(Лукашов, 2009)

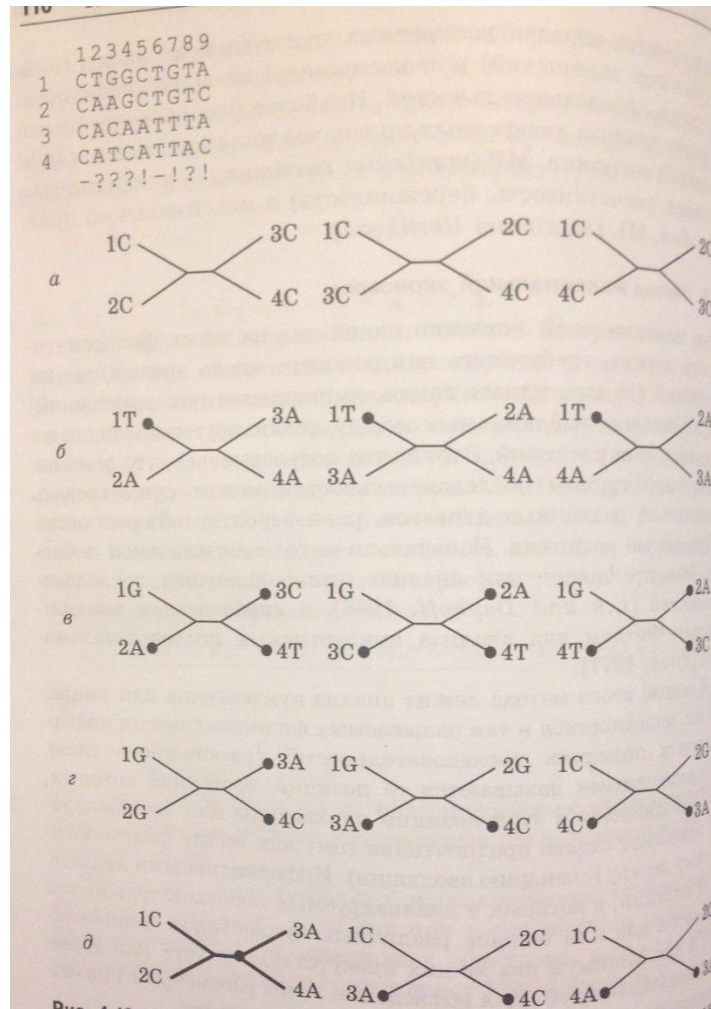
Направлен на поиск филогенетического дерева, требующего наименьшего числа эволюционных изменений (нуклеотидных замен, аминокислотных замещений) для объяснения наблюдаемых между анализируемыми последовательностями различий

(**discrete characters** - методы анализа дискретных признаков)

Метод максимальной экономии Maximum parsimony, MP

- Основан на символах, а не на расстояниях, поэтому не нужно измерять расстояния
- Деревья строятся на основе информации об изменении символов (или признаков)
- Подходит для небольшого множества последовательностей с высокой степенью схожести.

Метод максимальной экономии Maximum parsimony, MP



Статистическая оценка филогенетического дерева

- Аккуратность зависит от множественного выравнивания;
- В настоящее время только один метод выборок позволяет протестировать топологию филогенетического дерева:

Бутстрэп (Bootstrap)

Суть метода состоит в выборке с замещением колонок в используемом для построения дерева множественном выравнивании, пока размер выборки не достигнет размера изначального выравнивания (обычно получается, что какие-то колонки выбраны несколько раз, а какие-то вообще не выбраны).

Бутстрэп

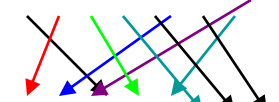
ATAGCCATA

ATACCCATG

ATACCCATA

ATAGCCATA

ATCCCCAT



TCAAATGCA

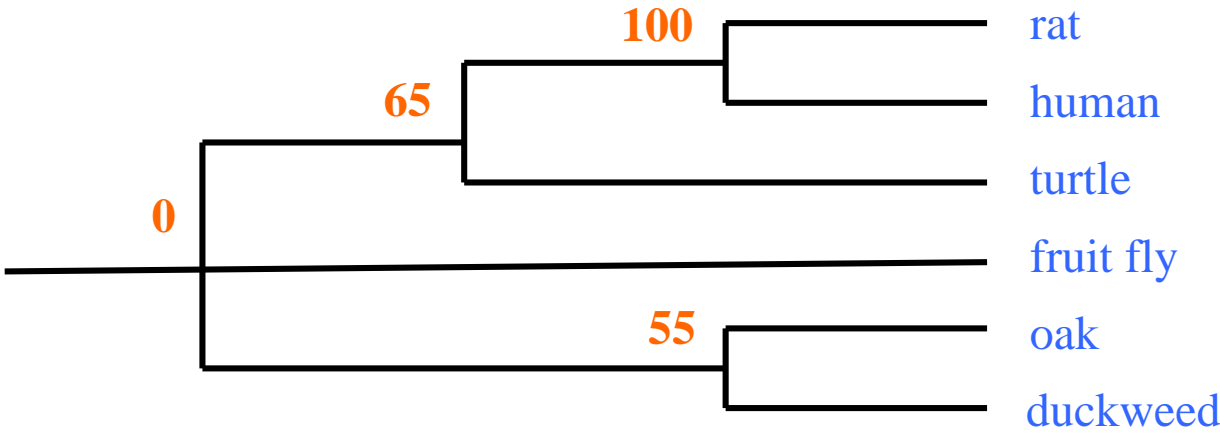
TCGAATCCA

TCAAATCCA

TCAAATGCA

TCAACACCC

100 раз



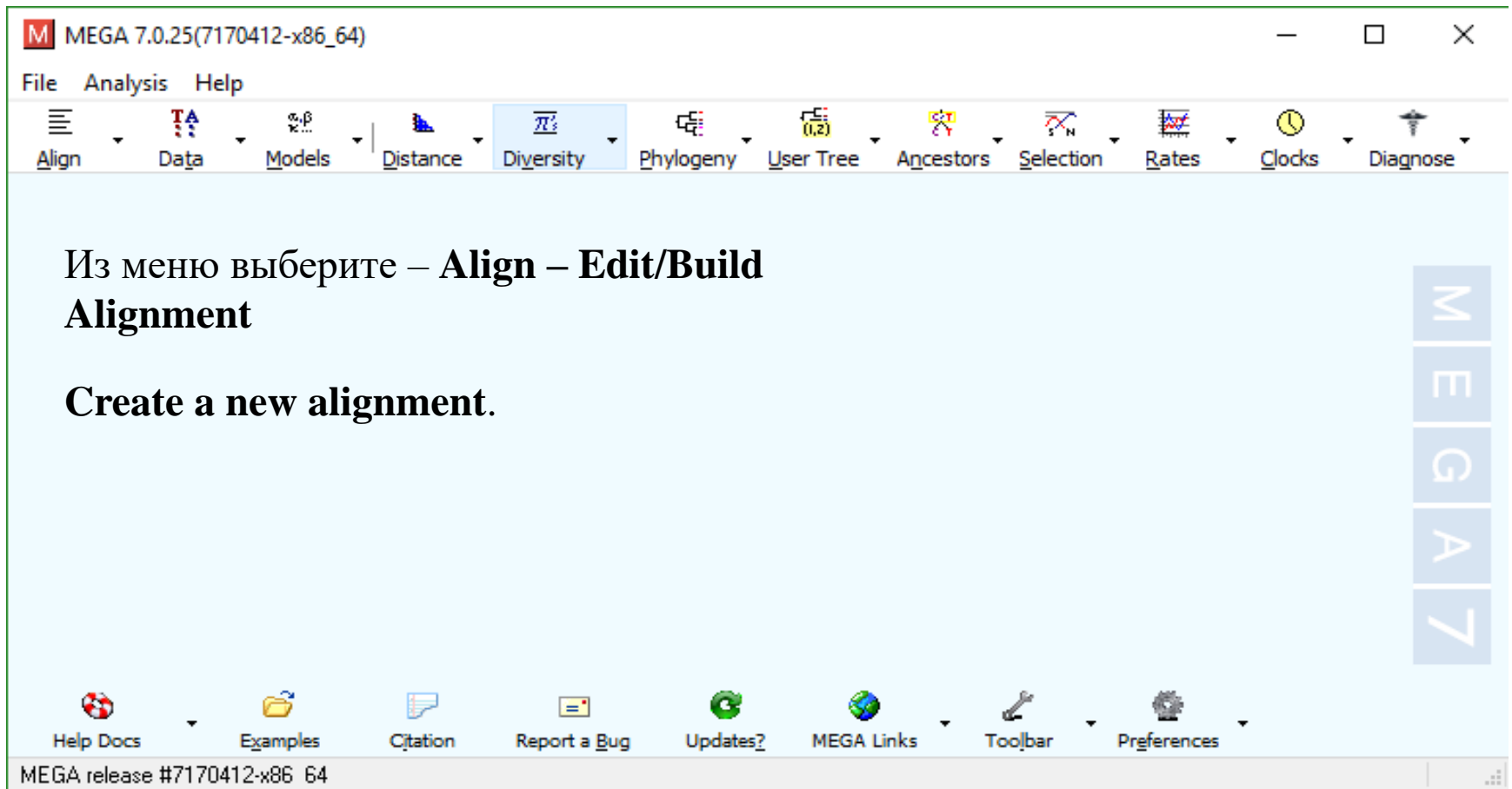
Этапы построения и анализа филогенетического дерева:

- 1. Выбрать последовательности генов/белков для анализа;**
- 2. Выравнить эти последовательности;**
- 3. Применить методы построения филогенетических деревьев (NJ, MP);**
- 4. Выполнить статистическую оценку полученных деревьев.**

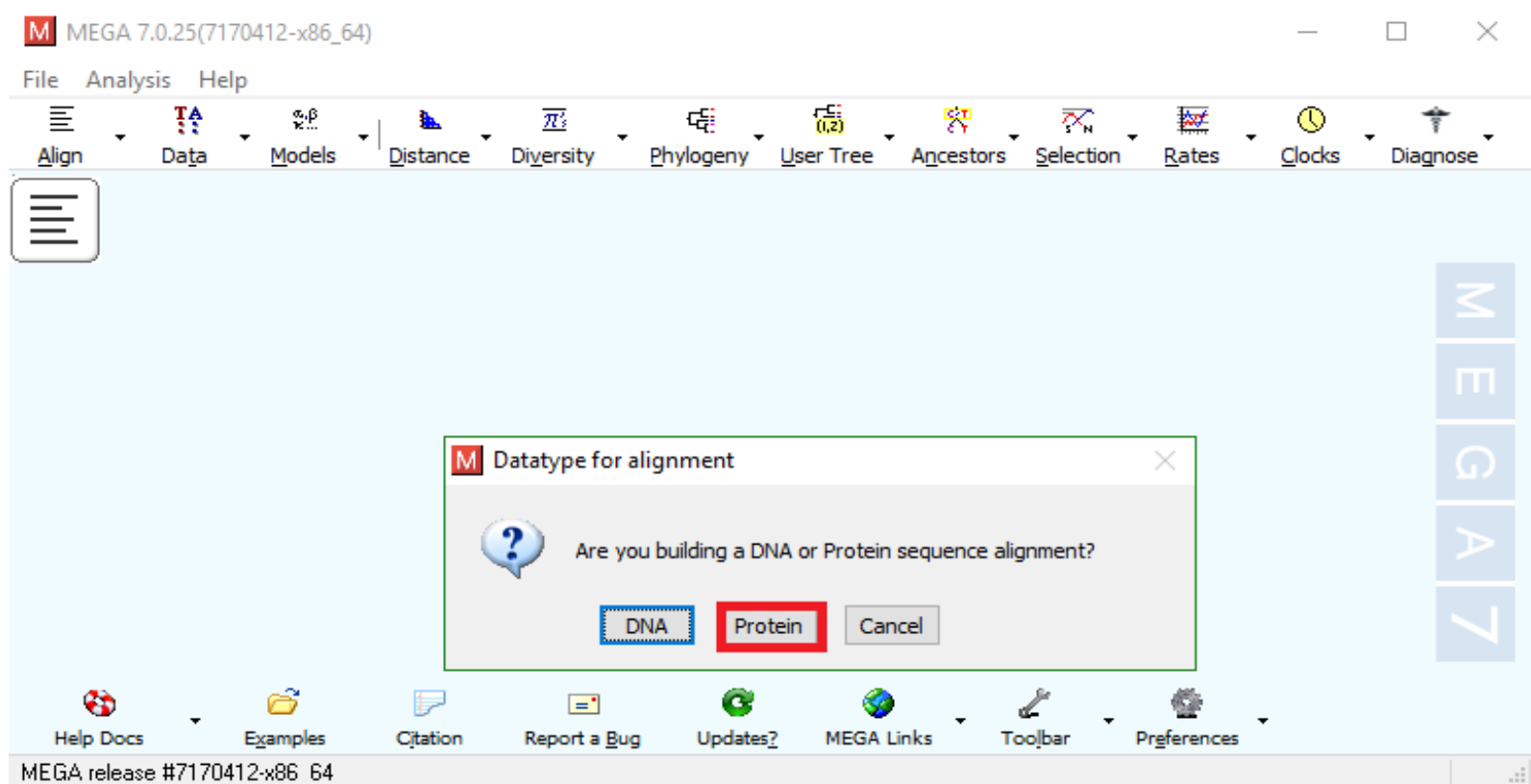
Построение филогенетического дерева

- Программа для построения молекулярных филогенетических деревьев – Mega (**M**olecular **E**volutionary **G**enetics **A**nalysis).
- Бесплатную версию программы Mega 7 можно скачать в интернете по ссылке: <http://www.megasoftware.net/>.

Рабочее окно МЕГА7



Окно выбора DNA or Protein



Для сохранения результатов выравнивания нажмите кнопку **Download Alignment File** или используйте **буфер обмена (ВЫДЕЛИТЬ – СОПЫ – PASTE)** в текстовый файл, сохраните файл, введите с клавиатуры тип файла .fas

Clustal Omega

[Input form](#)[Web services](#)[Help & Documentation](#)

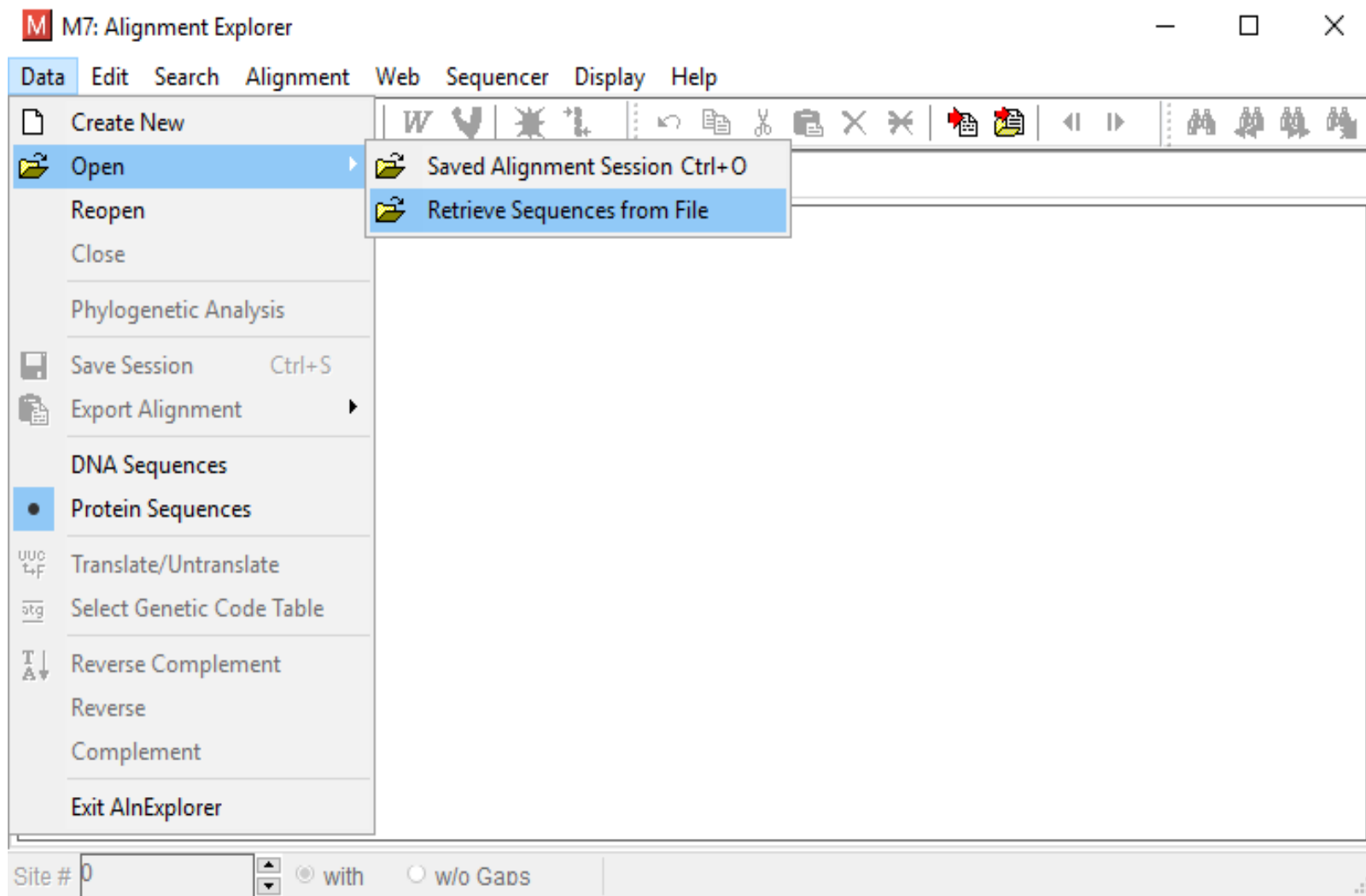
Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-l20170421-120055-0033-85848372-pg

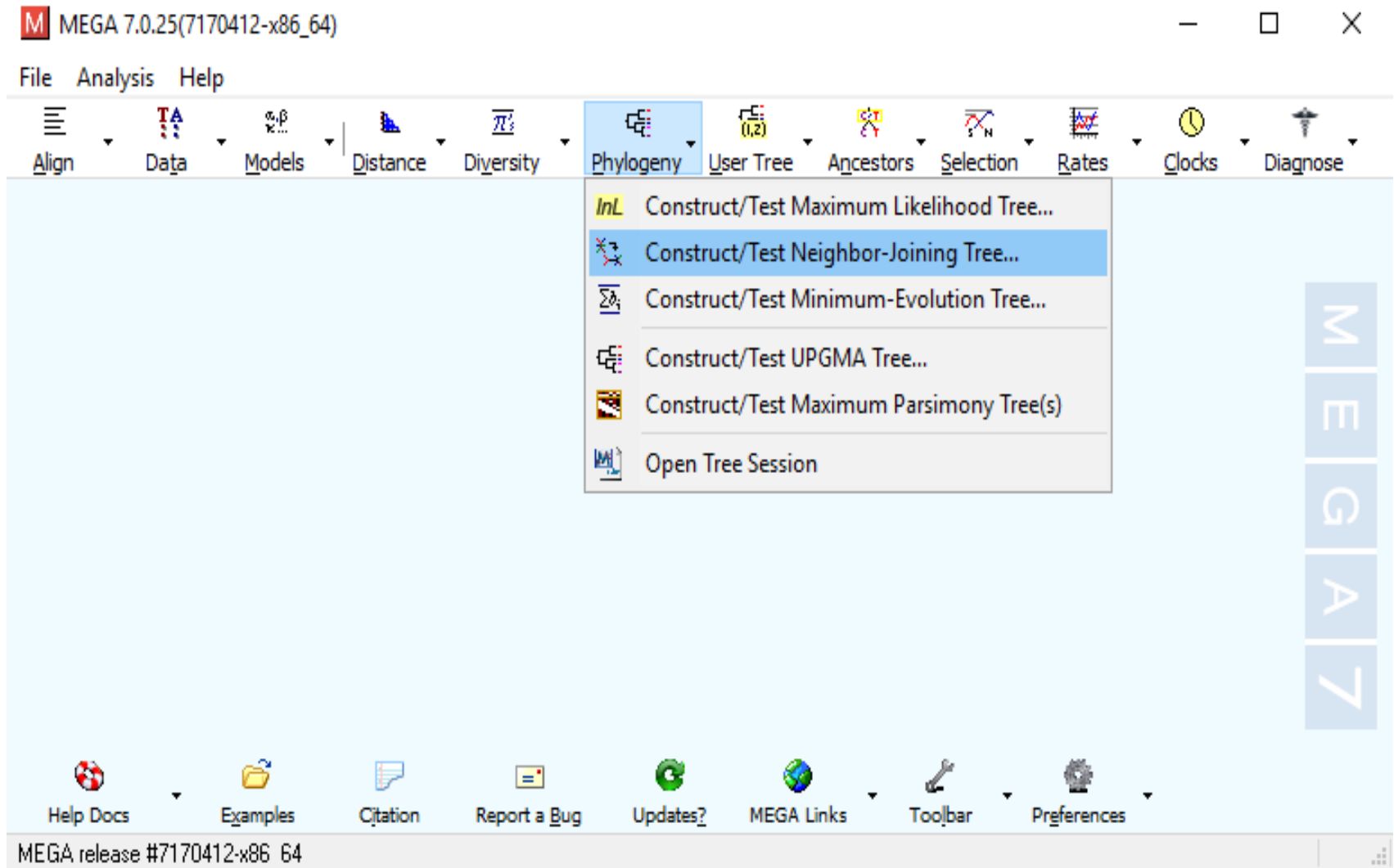
[Alignments](#)[Result Summary](#)[Phylogenetic Tree](#)[Submission Details](#)[Download Alignment File](#)[Send to Simple_Phylogeny](#)

```
>gi|697736615|gb|AIT57635.1| cytochrome oxidase subunit 1, partial (mitochondrion) [Anguilla anguilla]
LYLVFGAWAGMVGTAALSLIRAEALSQPGALLGDDQIYNVIVTAHAFVMIFFMVMFVMIGG
FGNWLIPLMIGAPDMAFPRMNNMSFWLLPPSFLLLLASSGVEAGAGTGWTVPPLAGNLA
HAGASVDLTI FSLHLAGISSILGAINFITTI INMKPPAITQYQTPLFVWAVLVTAVLLLL
SLPVLAAGITMLLTDRNLNTTFDPAGGGDPILYQHL-----
>gi|690195218|gb|AIR12026.1| cytochrome oxidase subunit 1, partial (mitochondrion) [Esox lucius]
LYLVFGAWAGMVGTAALSLIRAEALSQPGALLGDDQIYNVIVTAHAFVMIFFMVMFVMIGG
FGNWLIPLMIGAPDMAFPRMNNMSFWLLPPSFLLLLASSGVEAGAGTGWTVPPLAGNLA
HAGASVDLTI FSLHLAGISSILGAINFITTI INMKPPAISQYQTPLFVWAVLVTAVLLLL
SLPVLAAGITMLLTDRNLNTTFDPAGGGDPILYQHL-----
>gi|237651897|gb|ACR08638.1| cytochrome oxidase subunit I, partial (mitochondrion) [Salmo trutta fario]
-----GAWAGMVGTAALSLIRAEALSQPGALLGDDQIYNVIVTAHAFVMIFFMVMFVMIGG
FGNWLIPLMIGAPDMAFPRMNNMSFWLLPPSFLLLLASSGVEAGAGTGWTVPPLAGNLA
HAGASVDLTI FSLHLAGISSILGAINFITTI INMKPPAISQYQTPLFVWAVLVTAVLLLL
SLPVLAAGITMLLTDRNLNTTFDPAGGGDPILYQHLFWFFGH----
>gi|868612358|gb|AKO90570.1| cytochrome oxidase subunit 1, partial (mitochondrion) [Gymnocephalus cernua]
-----MVGTAALSLIRAEALSQPGALLGDDQIYNVIVTAHAFVMIFFMVMFVMIGG
FGNWLIPLMIGAPDMAFPRMNNMSFWLLPPSFLLLLASSGVEAGAGTGWTVPPLAGNLA
HAGASVDLTI FSLHLAGISSILGAINFITTI INMKPPAISQYQTPLFVWAVLVTAVLLLL
SLPVLAAGITMLLTDRNLNTTFDPAGGGDP-----
>gi|868612582|gb|AKO90637.1| cytochrome oxidase subunit 1, partial (mitochondrion) [Sander lucioperca]
-----MVGTAALSLIRAEALSQPGALLGDDQIYNVIVTAHAFVMIFFMVMFVMIGG
FGNWLIPLMIGAPDMAFPRMNNMSFWLLPPSFLLLLASSGVEAGAGTGWTVPPLAGNLA
HAGASVDLTI FSLHLAGISSILGAINFITTI INMKPPAISQYQTPLFVWAVLVTAVLLLL
SLPVLAAGITMLLTDRNLNTTFDPAGGGDP-----
>gi|868612651|gb|AKO90656.1| cytochrome oxidase subunit 1, partial (mitochondrion) [Thymallus thymallus]
-----GAWAGMVGTAALSLIRAEALSQPGALLGDDQIYNVIVTAHAFVMIFFMVMFVMIGG
FGNWLIPLMIGAPDMAFPRMNNMSFWLLPPSFLLLLASSGVEAGAGTGWTVPPLAGNLA
HAGASVDLTI FSLHLAGISSILGAINFITTI INMKPPAISQYQTPLFVWAVLVTAVLLLL
SLPVLAAGITMLLTDRNLNTTFDPAGGGDPILYQHLF-----
```

Загрузка файла с последовательностями с локального диска



Методы построения филогенетических деревьев



Bootstrap method

M7: Analysis Preferences

Options Summary

Option	Selection
Analysis	Phylogeny Reconstruction
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
Phylogeny Test	
Test of Phylogeny	Bootstrap method
<i>No. of Bootstrap Replications</i>	100
Substitution Model	
Substitutions Type	Amino acid
Model/Method	Poisson model
Rates and Patterns	
Rates among Sites	Uniform rates
<i>Gamma Parameter</i>	<i>Not Applicable</i>
Pattern among Lineages	Same (Homogeneous)
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
<i>Site Coverage Cutoff (%)</i>	<i>Not Applicable</i>

? Help

✓ Compute

✗ Cancel

Филогенетическое дерево

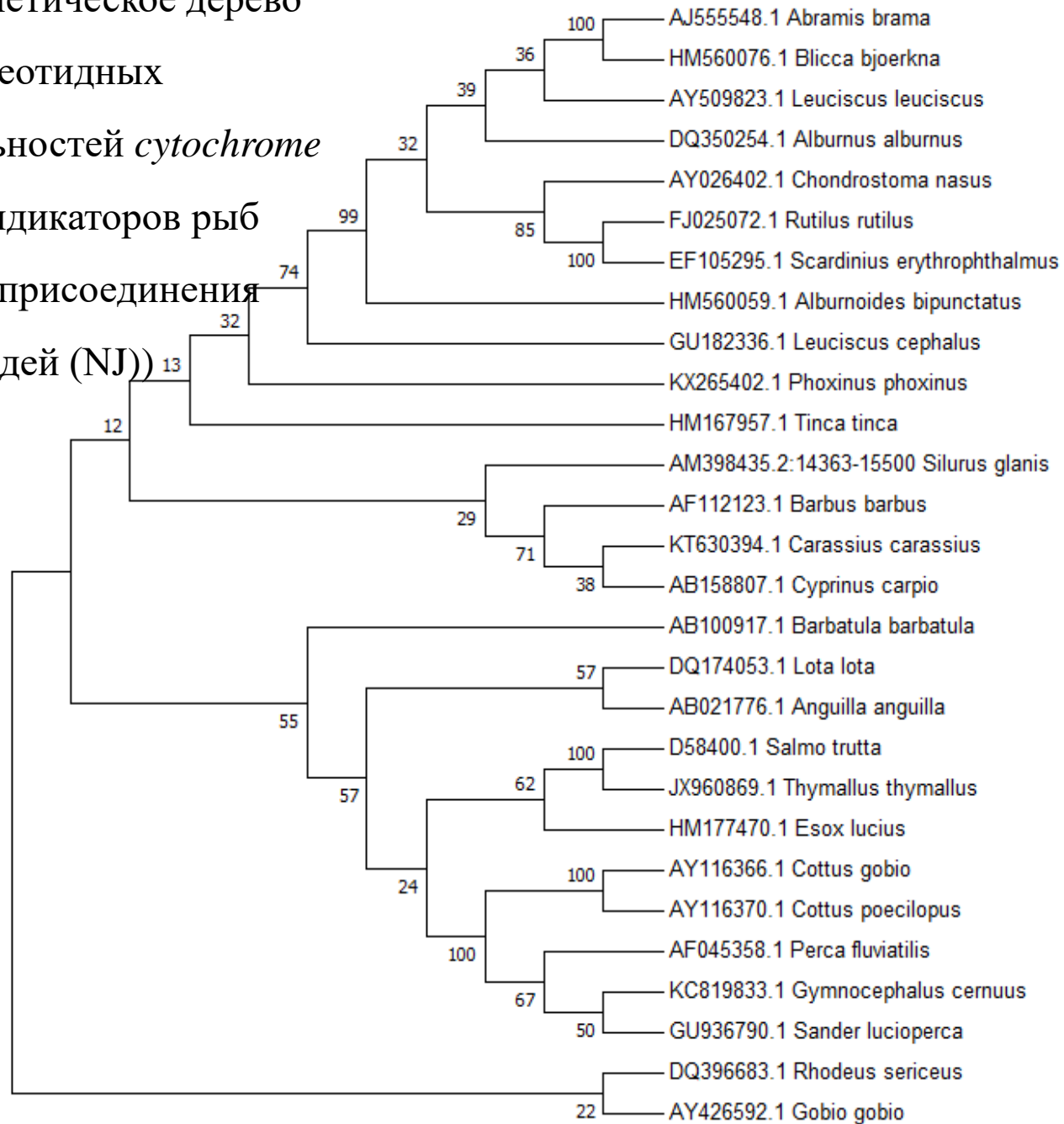
нуклеотидных

последовательностей *cytochrome*

b видов-индикаторов рыб

(Методом присоединения

соседей (NJ))



СПАСИБО ЗА ВНИМАНИЕ