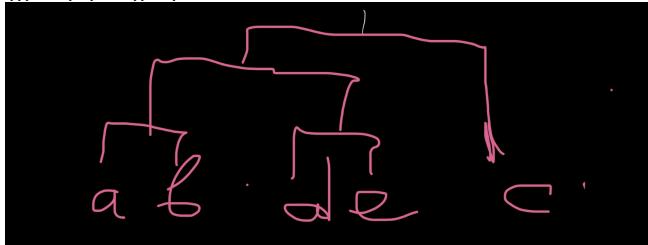


# БИ ДЗ-3. Филогенетика

## ★ Задание 1. Нарисовать дерево

((a,b),(d,e)),c;



## ★ Задание 2. Анализ существующих деревьев

🍓 Ссылка на статью

💬 <https://pubmed.ncbi.nlm.nih.gov/35689765/>

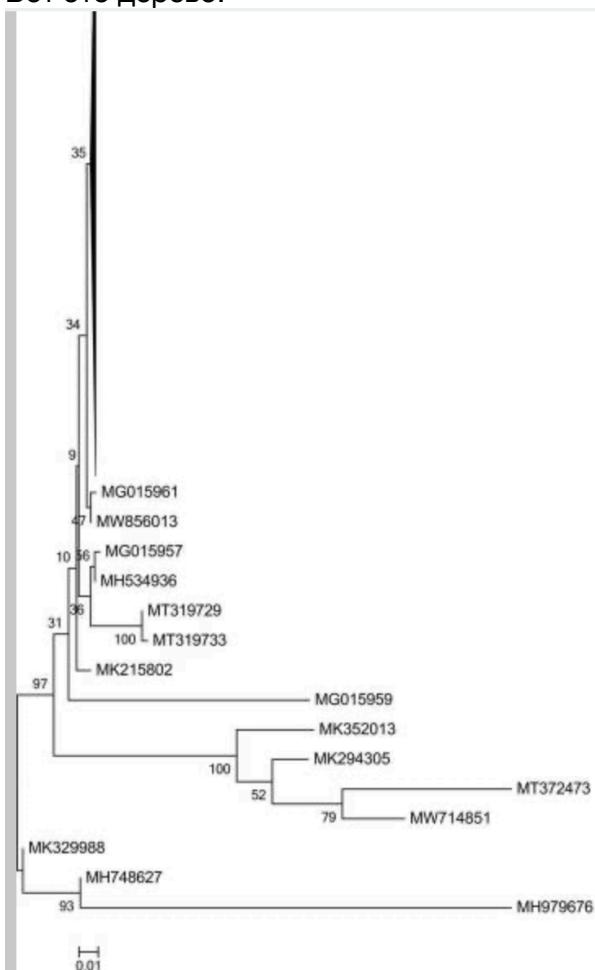
🍓 Таксономическая группа, анализируемая в работе, и метод, использованный для реконструкции дерева

💬 *Candida krusei* ITS последовательности, median-joining метод

🍓 Краткое описание дерева (масштабированное или нет, укорененное или нет)

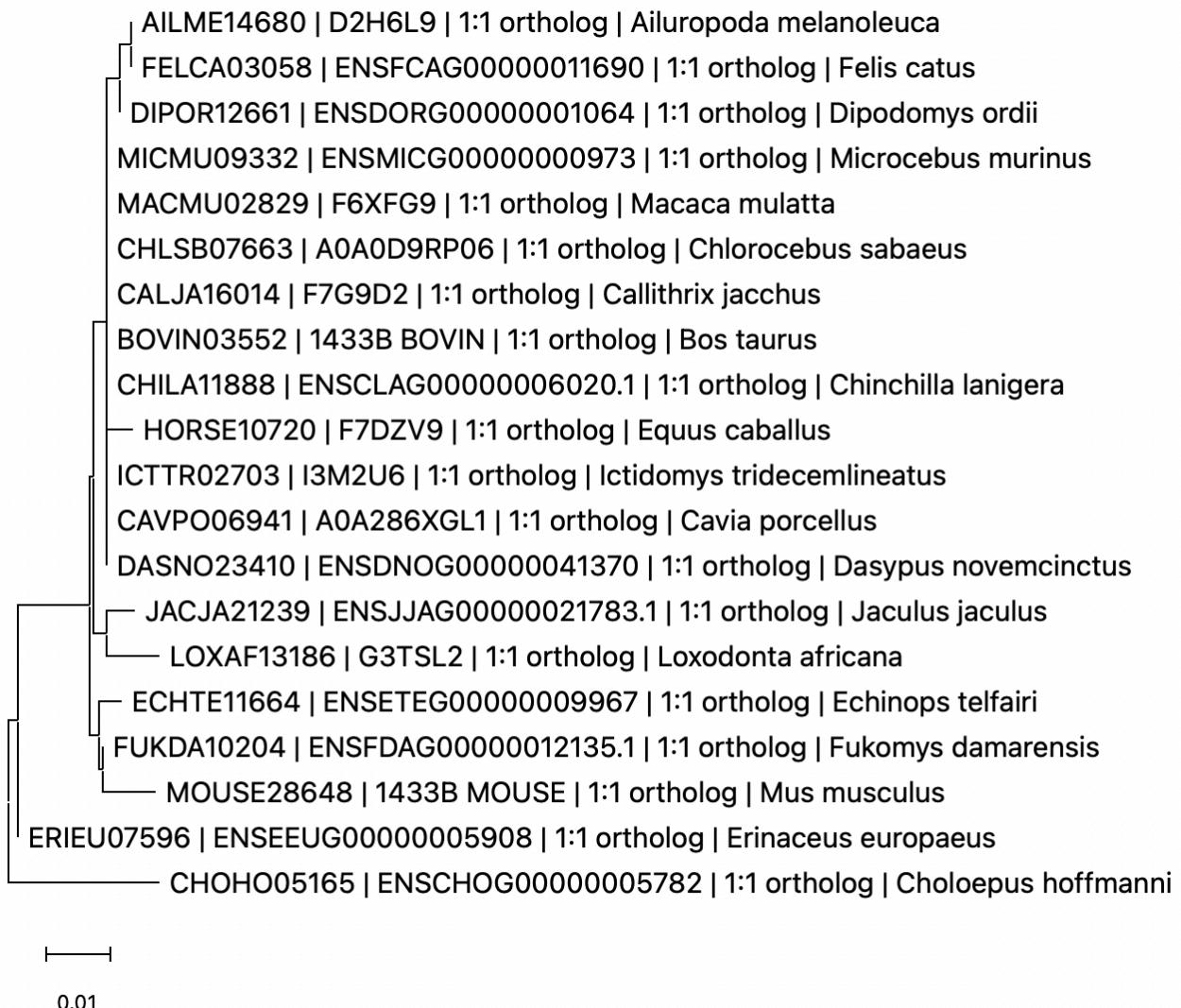
💬 Масштабированное, укорененное

Вот это дерево:



### ★ Задание 3. Генерация филогенетического дерева белка в программе MEGA

Построила дерево в MEGA с помощью алгоритма NJ (файл с гитхаба [data/sample\\_prot.fasta](#), [https://github.com/michetrofimov/hse\\_data\\_analysis\\_phylogenetics/tree/main/data](https://github.com/michetrofimov/hse_data_analysis_phylogenetics/tree/main/data)).



Дерево в формате Newick:

```
((((((((('AILME14680 | D2H6L9 | 1:1 ortholog | Ailuropoda  
melanoleuca':0.0000000, 'FELCA03058 | ENSFCAG00000011690 | 1:1 ortholog |  
Felis catus':0.0000000):0.00173821, 'DIPOR12661 | ENSDORG00000001064 | 1:1  
ortholog | Dipodomys ordii':0.0000000):0.00210547, 'MICMU09332 |  
ENSMICG00000000973 | 1:1 ortholog | Microcebus  
murinus':0.0000000):0.0003822, 'MACMU02829 | F6XFG9 | 1:1 ortholog | Macaca  
mulatta':0.0000000):0.0001433, 'CHLSB07663 | A0A0D9RP06 | 1:1 ortholog |  
Chlorocebus sabaeus':0.0000000):0.0000717, 'CALJA16014 | F7G9D2 | 1:1  
ortholog | Callithrix jacchus':0.0000000):0.0000407, ('BOVIN03552 | 1433B  
BOVIN | 1:1 ortholog | Bos taurus':0.0000000, ('CHILA11888 |  
ENSCLAG00000006020.1 | 1:1 ortholog | Chinchilla lanigera':0.0000000,  
('HORSE10720 | F7DZV9 | 1:1 ortholog | Equus caballus':0.0041006, 'ICTTR02703  
| I3M2U6 | 1:1 ortholog | Ictidomys  
tridecemlineatus':0.0000000):0.0001069):0.0000668):0.0000197):0.0000536,  
'CAVPO06941 | A0A286XGL1 | 1:1 ortholog | Cavia  
porcellus':0.0000000):0.0000838, 'DASNO23410 | ENSDNOG00000041370 | 1:1  
ortholog | Dasyurus novemcinctus':0.0000000):0.00206214, ('JACJA21239 |
```

```
ENSJJAG00000021783.1 | 1:1 ortholog | Jaculus jaculus':0.00423489, 'LOXAF13186  
| G3TSL2 | 1:1 ortholog | Loxodonta  
africana':0.00818763):0.00192410):0.00055275, ('ECHTE11664 |  
ENSETEG00000009967 | 1:1 ortholog | Echinops telfairi':0.00345684,  
('FUKDA10204 | ENSFDAG0000012135.1 | 1:1 ortholog | Fukomys  
damarensis':0.00000000, 'MOUSE28648 | 1433B MOUSE | 1:1 ortholog | Mus  
musculus':0.00816566):0.00063322):0.00155596):0.01091987, 'ERIEU07596 |  
ENSEEUG00000005908 | 1:1 ortholog | Erinaceus  
europaeus':0.00000000, 'CH0H005165 | ENSCHOG00000005782 | 1:1 ortholog |  
Choloepus hoffmanni':0.02477873);
```

#### ★ Задание 4. Полный процесс получения дерева от .fasta --> png/newick

★ 4.1. Скачала нуклеотидный FASTA-файл (data/sample\_nucl.fasta): [https://github.com/michetrofimov/hse\\_data\\_analysis\\_phylogenetics/blob/main/data/sample\\_nucl.fasta](https://github.com/michetrofimov/hse_data_analysis_phylogenetics/blob/main/data/sample_nucl.fasta)

★ 4.2. Построила выравнивание в CLUSTALW (<https://www.genome.jp/tools-bin/clustalw>) и сохранила множественное выравнивание в формате .fasta

##### General Setting Parameters:

Output Format: CLUSTAL

Pairwise Alignment:  FAST/APPROXIMATE  SLOW/ACCURATE

Enter your sequences (with labels) below (copy & paste):  PROTEIN  DNA

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

Or give the file name containing your query

Choose File  sample\_nucl.fasta

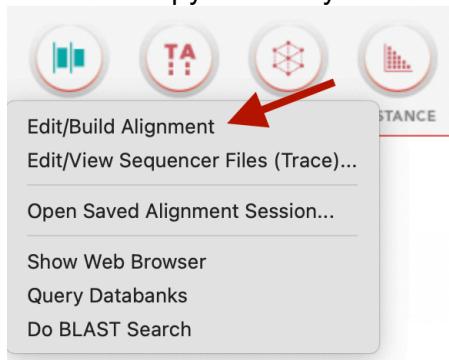
Execute Multiple Alignment  Reset

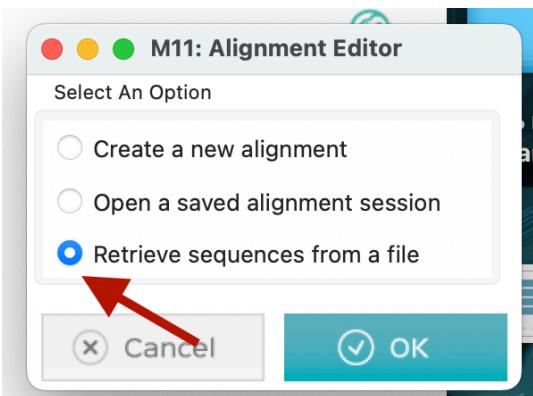
Жму сюда (красная стрелка) и скачиваю файл:

clustalw.aln

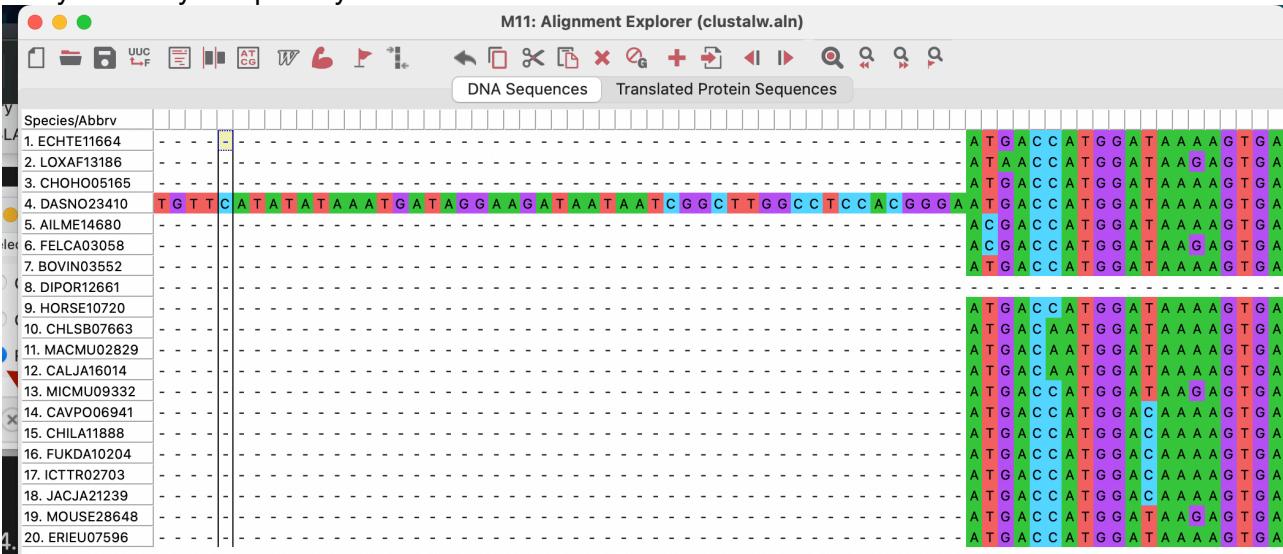
★ 4.3. Построила дерево в MEGA с помощью алгоритма ML со 100 бутстрэпами.

В MEGA загружаю полученное в предыдущем пункте выравнивание:





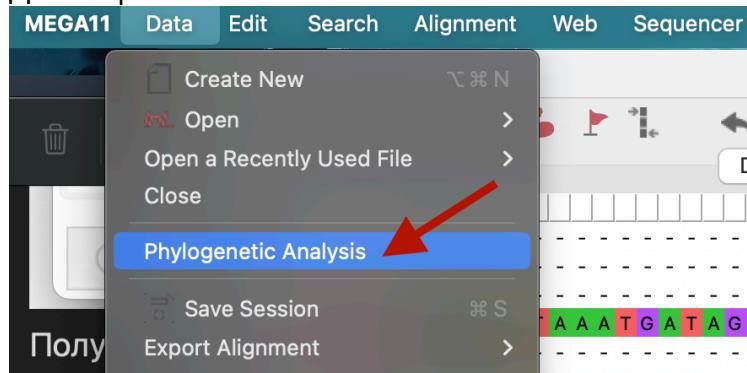
Получаю такую картинку:



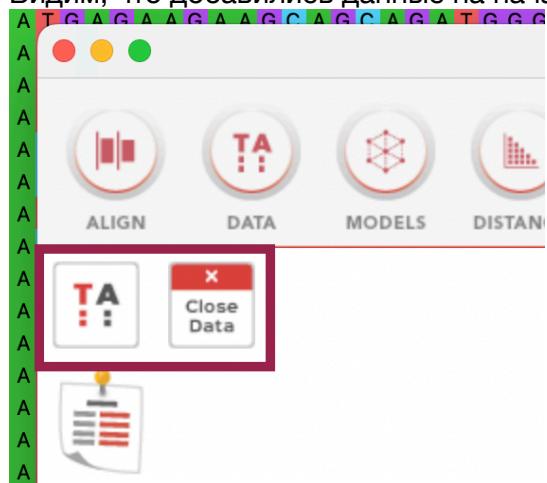
Далее необходимо сделать филогенетический анализ перед построением дерева, но есть неизвестные программы нуклеотиды X, заменю их на N:

Species/Abbrv	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1. ECHTE11664	G	A	C	A	A	C	C	T	C	A	T	C	T	T	G	G	A	A	A
2. LOXAF13186	G	A	C	A	A	T	C	T	G	A	C	T	C	T	G	T	G	A	C
3. CHOHO05165	G	A	C	A	A	T	C	T	G	A	C	T	X	X	X	G	A	C	G
4. DASNO23410	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
5. AILME14680	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
6. FELCA03058	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
7. BOVIN03552	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
8. DIPOR12661	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
9. HORSE10720	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
10. CHLSB07663	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
11. MACMU02829	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
12. CALJA16014	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
13. MICMU09332	G	A	T	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
14. CAVPO06941	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
15. CHILA11888	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
16. FUKDA10204	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
17. ICTTR02703	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
18. JACJA21239	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
19. MOUSE28648	G	A	C	A	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C
20. ERIEU07596	G	A	T	X	A	T	C	T	C	A	C	T	C	T	G	T	G	A	C

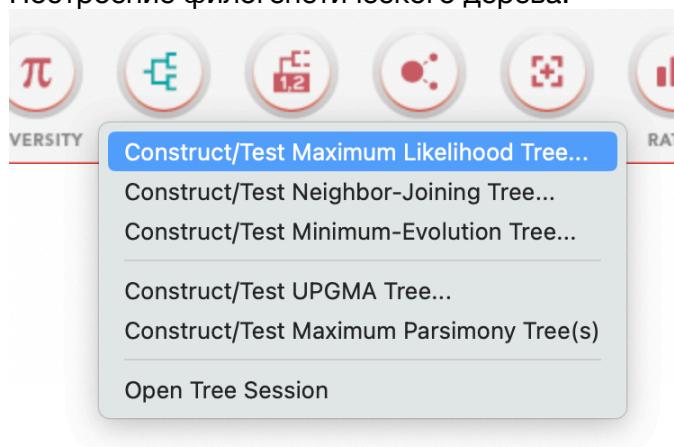
Делаю филогенетический анализ:



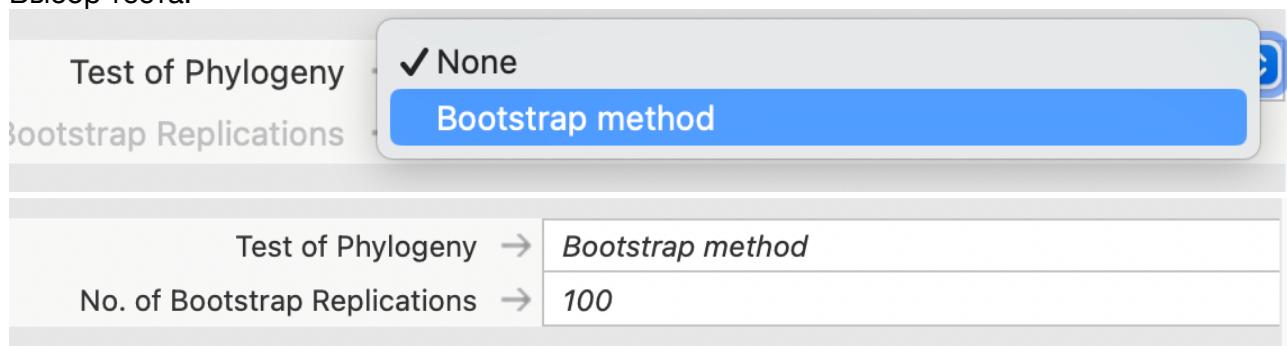
Видим, что добавились данные на начальный экран MEGA:



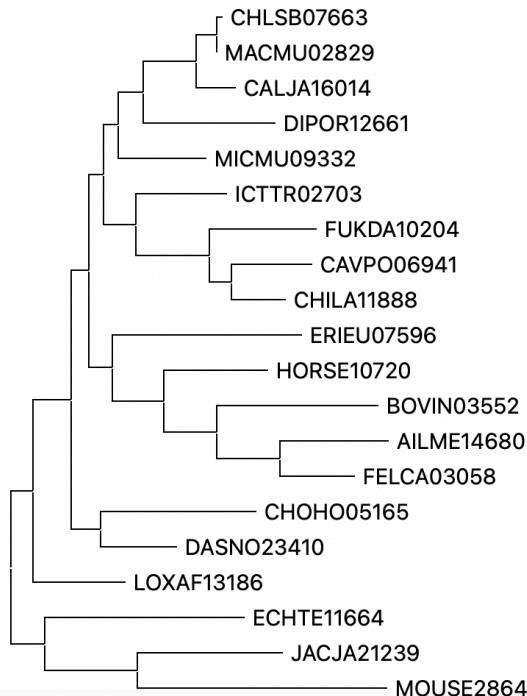
Построение филогенетического дерева:



Выбор теста:



☆ 4.4. Сохранила рисунок ML-дерева (в PNG) и дерево (в Newick).



```
((((((((CHLSB07663,MACMU02829),CALJA16014),DIPOR12661),MICMU09332),(ICTTR02703,(FUKDA10204,(CAVPO06941,CHILA11888)))),(ERIEU07596,(HORSE10720,(BOVIN03552,(AILME14680,FELCA03058)))),(CHOHO05165,DASNO23410)),LOXAF13186,(ECHTE11664,(JACJA21239,MOUSE28648)));
```

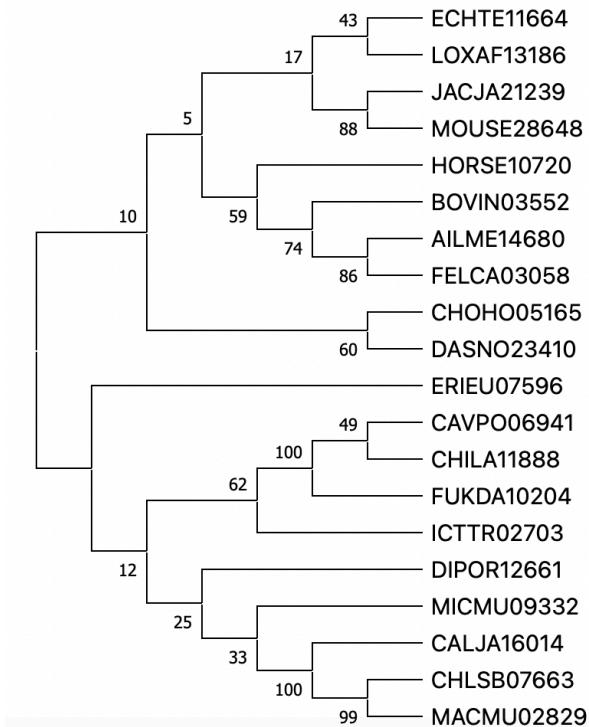
Или с дополнительной информацией:

Branch Lengths

Bootstrap Values

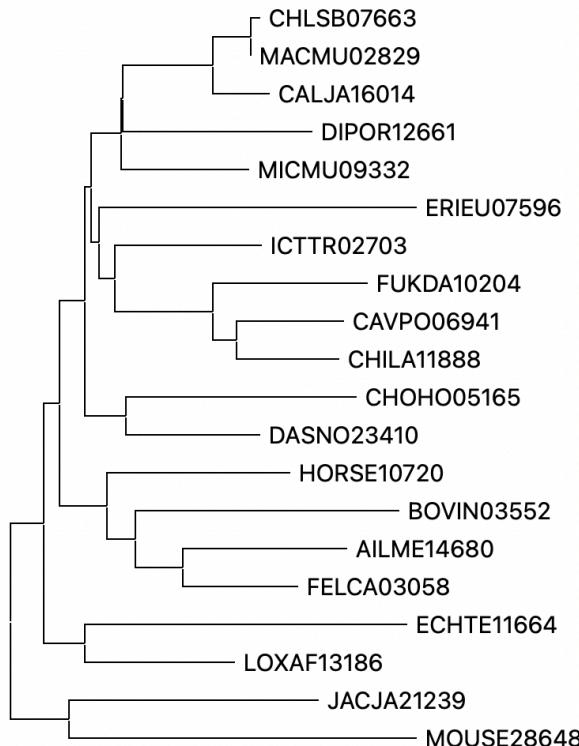
```
((((((((CHLSB07663:0.00135288,MACMU02829:0.00012309)0.9900:0.00474933,CALJA16014:0.00893784)1.0000:0.01192997,DIPOR12661:0.02982636)0.1900:0.00550624,MICMU09332:0.01990756)0.2500:0.00342900,(ICTTR02703:0.02065664,(FUKDA10204:0.02410707,(CAVPO06941:0.01834924,CHILA11888:0.01225204)0.4900:0.00495003)1.0000:0.01644940)0.6200:0.00740248)0.1200:0.00339524,(ERIEU07596:0.04274345,(HORSE10720:0.02348639,(BOVIN03552:0.03634106,(AILME14680:0.02432694,FELCA03058:0.01692798)0.8600:0.01431894)0.7400:0.01182617)0.5900:0.01171382)0.0600:0.00542751)0.0200:0.00382781,(CHOHO05165:0.03507047,DASNO23410:0.01723908)0.6000:0.00660210)0.1700:0.00860074,LOXAF13186:0.02099886,(ECHTE11664:0.04505433,(JACJA21239:0.03264869,MOUSE28648:0.05629350)0.8800:0.02097390)0.3500:0.01275415);
```

☆ 4.5. Сохранила рисунок бутстрепного дерева (в PNG).



☆ 4.6. Построила дерево в MEGA, используя алгоритм NJ.

☆ 4.7. Сохранила рисунок в формате PNG.



☆ 4.8. Сохранила дерево в формате Newick.

С длинами ветвей:

(((((CHLSB07663:0.00126888,MACMU02829:0.00008143):0.00551399,CALJA16014:0.00816000):0.01299603,DIPOR12661:0.02747941):0.00033079,MICMU09332:0.01853018):0.00441402,(FRTFU07596:0.04595097,(TCTTR02703:0.02128984,

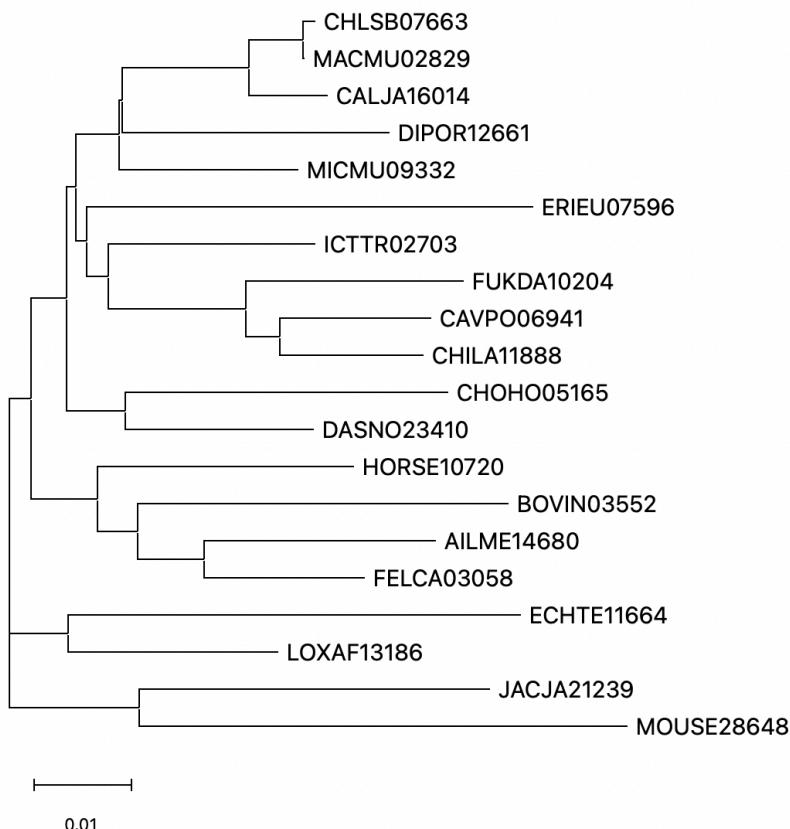
```
(FUKDA10204:0.02250799,
(CAVP006941:0.01557096,CHILA11888:0.01485295):0.00347899):0.01419672):0.00226
917):0.00111580):0.00091787,
(CHOH005165:0.03329191,DASN023410:0.01943316):0.00596795):0.00361701,
(HORSE10720:0.02640968,(BOVIN03552:0.03824392,
(AILME14680:0.02388295,FELCA03058:0.01662082):0.00681531):0.00412218):0.00676
606):0.00226390,(ECHTE11664:0.04656558,LOXAF13186:0.02162477):0.00605068,
(JACJA21239:0.03615684,MOUSE28648:0.05020310):0.01343368);
```

Без длин ветвей:

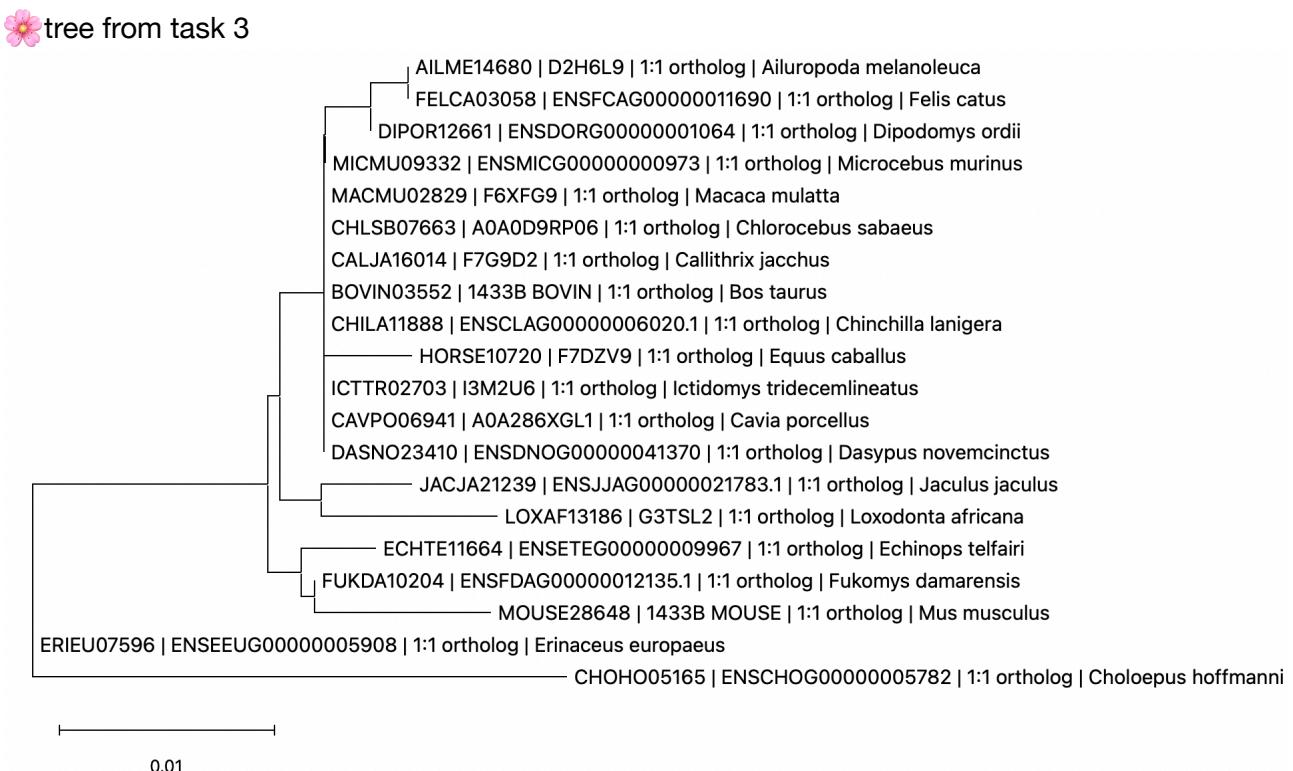
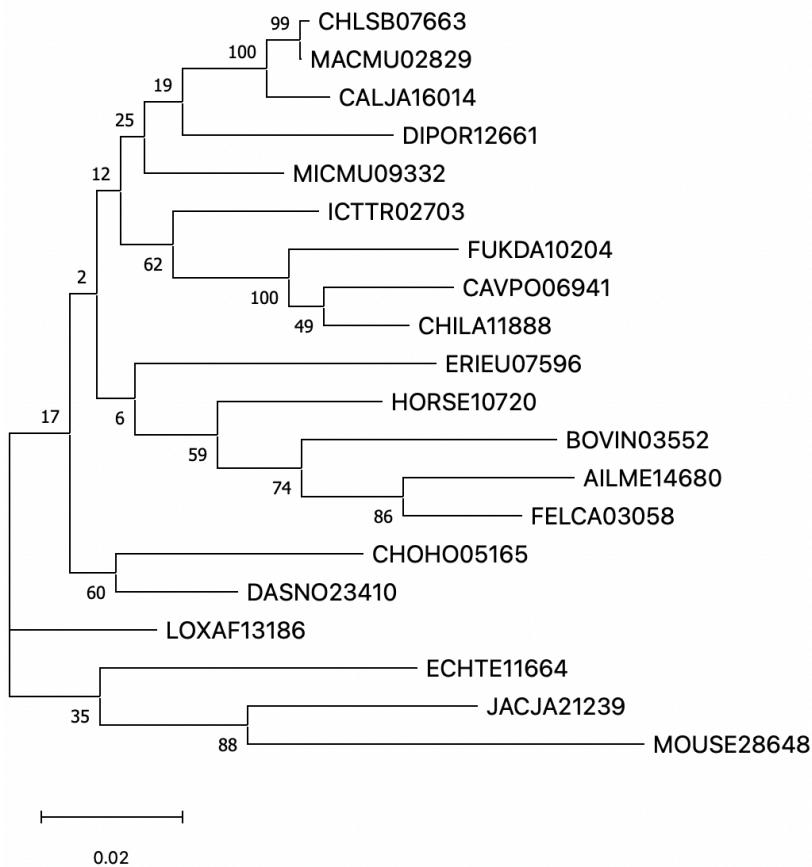
```
(((((CHLSB07663,MACMU02829),(CALJA16014),(DIPOR12661),(MICMU09332),
(ERIEU07596,(ICTTR02703,(FUKDA10204,(CAVP006941,CHILA11888)))),,
(CHOH005165,DASN023410)),(HORSE10720,(BOVIN03552,(AILME14680,FELCA03058))),,
(ECHTE11664,LOXAF13186),(JACJA21239,MOUSE28648));
```

★ 4.9. Сравните полученное дерево NJ с ранее полученным деревом ML и с деревом полученным в задании 3.

NJ



ML



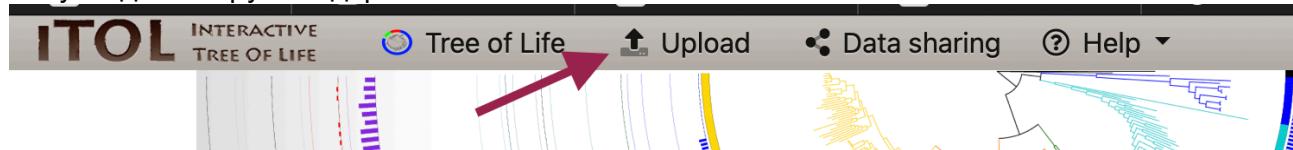
Кажется, что первые два дерева получились получше, чем дерево из третьего задания. Я как-то удалила названия таксонов - это было фатальной ошибкой, конечно, для сравнения. В идеале сделать одинаковые картинки в плане содержания.

## ★ Задание 5. Альтернативная визуализация деревьев

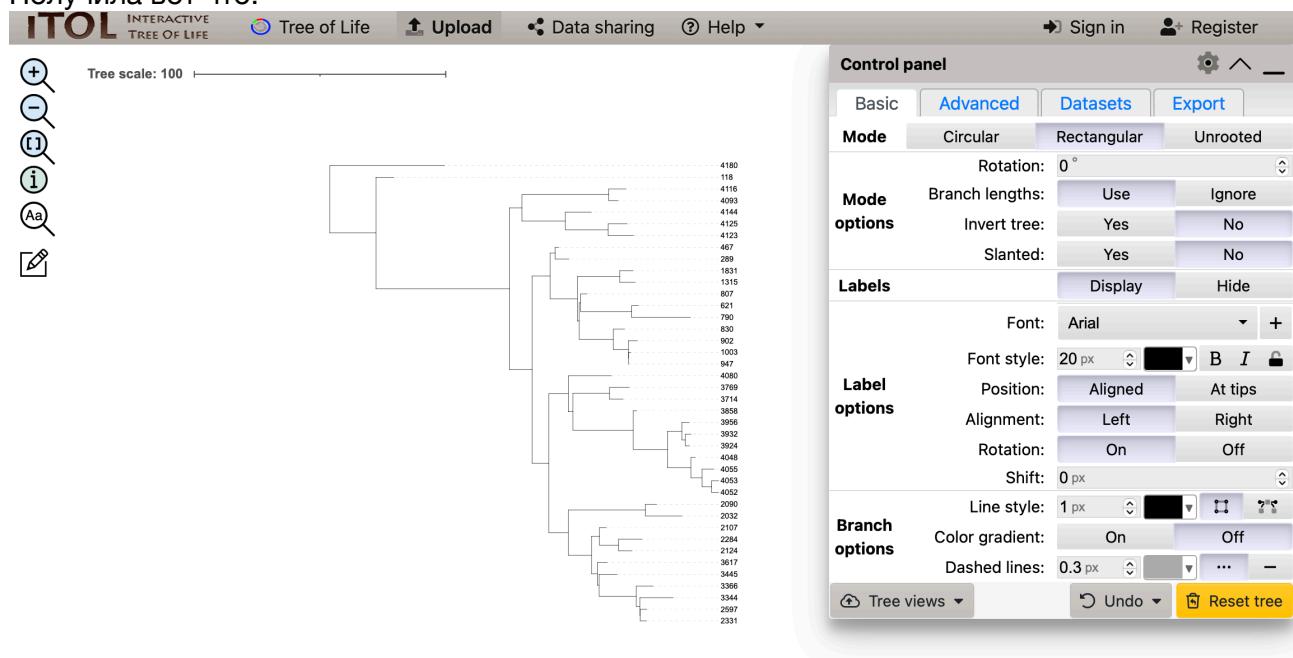
★ 5.1. Зашла на сервер iTOL: <https://itol.embl.de/>.

- ★ 5.2. Загрузила дерево (data/sample\_tree.nwk, [https://github.com/michetrofimov/hse\\_data\\_analysis\\_phylogenetics/blob/main/data/sample\\_tree.nwk](https://github.com/michetrofimov/hse_data_analysis_phylogenetics/blob/main/data/sample_tree.nwk)).

Жму сюда и загружаю дерево:



Получила вот что:

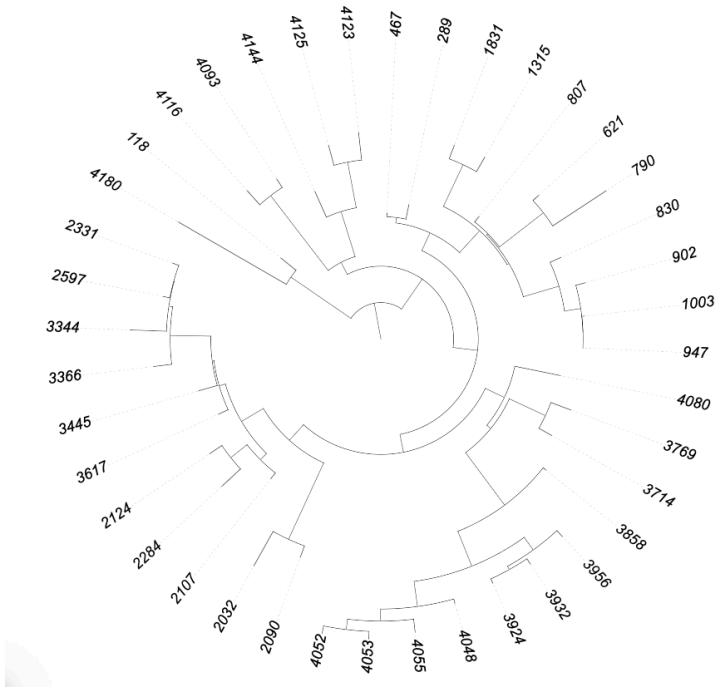


Citation: Letunic and Bork (2024) Nucleic Acids Res doi: 10.1093/nar/gkae268 | Privacy Policy

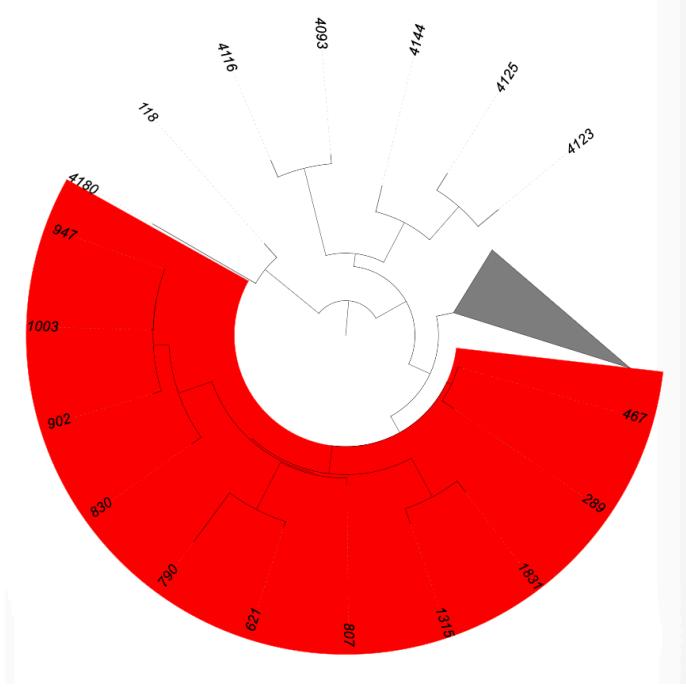
design & development: biobyte solutions

- ★ 5.3. Внесла следующие изменения (по заданию):

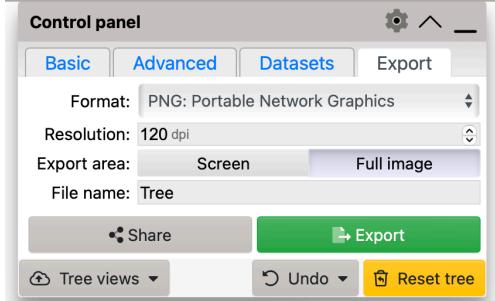
1. Круговое дерево (Circular tree)
2. Курсив для имен листьев (Italic for leaf names)
3. Укорените дерево в средней точке (Root the tree to the midpoint)



4. Свернуть любой узел (Collapse any node)
  5. Окрасьте любой клайд в красный цвет (Colorize any clade by red)



6. Загрузите рисунок в формате PNG (Download the figure in PNG format)



## ★ Задание 6.

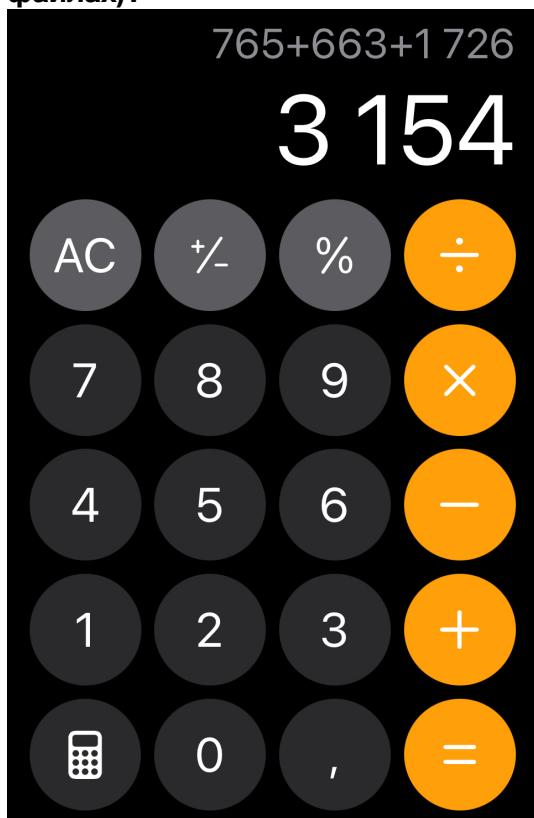
- ☆ 1. После выполнения заданий 3 и 4 у вас должно быть 3 файла Newick. Поместите их вместе в какую-нибудь папку. В командной строке введите команду, которая подсчитает количество символов во всех файлах Newick в папке. Включите в отчет скриншот с командой и ее выводом.

```
(base) jupyter-gorovenko-e@hse-students:~/common/K$ cat M  
L.nwk | wc -c  
765  
(base) jupyter-gorovenko-e@hse-students:~/common/K$ cat N  
J.nwk | wc -c  
663  
(base) jupyter-gorovenko-e@hse-students:~/common/K$ cat t  
ree3.nwk | wc -c  
1726
```

- ☆ 2. Теперь введите команду, которая объединяет эти 3 файла Newick, а затем подсчитывает количество символов в конкатенированном файле (без сохранения конкатенированного файла!). Включите скриншот экрана с командой и ее выводом в свой отчет.

```
(base) jupyter-gorovenko-e@hse-students:~/common/K$ cat M  
L.nwk NJ.nwk tree3.nwk > all_trees.nwk  
(base) jupyter-gorovenko-e@hse-students:~/common/K$ cat a  
ll_trees.nwk | wc -c  
3154
```

- ☆ 3. Ответьте на вопрос: равно ли количество символов в конкатенированном файле по сравнению с первым подзаданием (подсчетом всех символов во всех Newick файлах)?



… Калькулятор посчитал, что все равно.