

Отчет по работе с укоренением и сравнением деревьев; датировкой

Жерко Никита

1. Как укоренить дерево по известной внешней клade в IQ-TREE или другой программе?

```
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -pre SUP35_TIM3_ufb -bb 1000
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -pre SUP35_TIM3_ufb_alrt_abayes -bb 1000 -alrt 1000 -abayes
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -pre SUP35_TIM3_root_outgroup -bb 1000 -alrt 1000 -abayes -o SUP35_Kla_AB039749,SUP35_Agos_ATCC_10895_NM_211584
```

2. Как укоренить дерево с помощью midpoint rooting? Приведите хотя бы один метод (можно получить бонус за разбор большего количества).

```
# команда нужна, так как скачивание происходит из интернета
chmod +x midpoint_root.py
python3 midpoint_root.py SUP35_TIM3_ufb.treefile >SUP35_TIM3_ufb_midpoint.treefile
```

А сам скрипт .py выглядит следующим образом:

```
import sys
from ete3 import Tree
intre = sys.argv[1]
tre = Tree(intre, quoted_node_names=True)
midpoint = tre.get_midpoint_outgroup()
tre.set_outgroup(midpoint)
print(tre.write())
```

Есть альтернативный вариант, с использованием скрипта, написанного на R.

```
Rscript midpoint_root.R
```

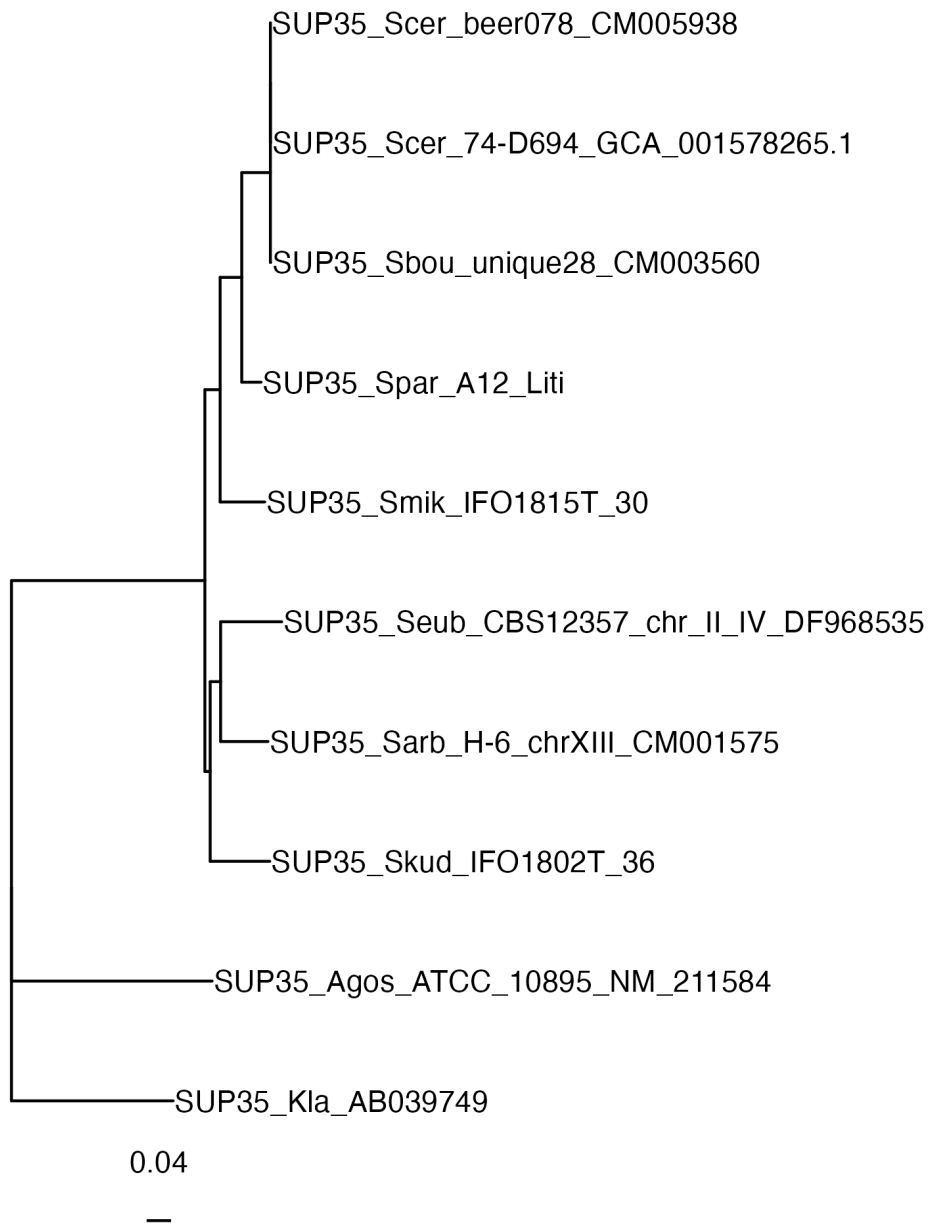
Непосредственно скрипт выглядит следующим образом:

```
library(phytools)
tree_alrt_abayes_ufb <- read.tree("SUP35_TIM3_ufb_alrt_abayes.treefile")
midpoint.root(tree_alrt_abayes_ufb)
write.tree(tree_alrt_abayes_ufb, "SUP35_TIM3_ufb_alrt_abayes_rooted.treefile")
```

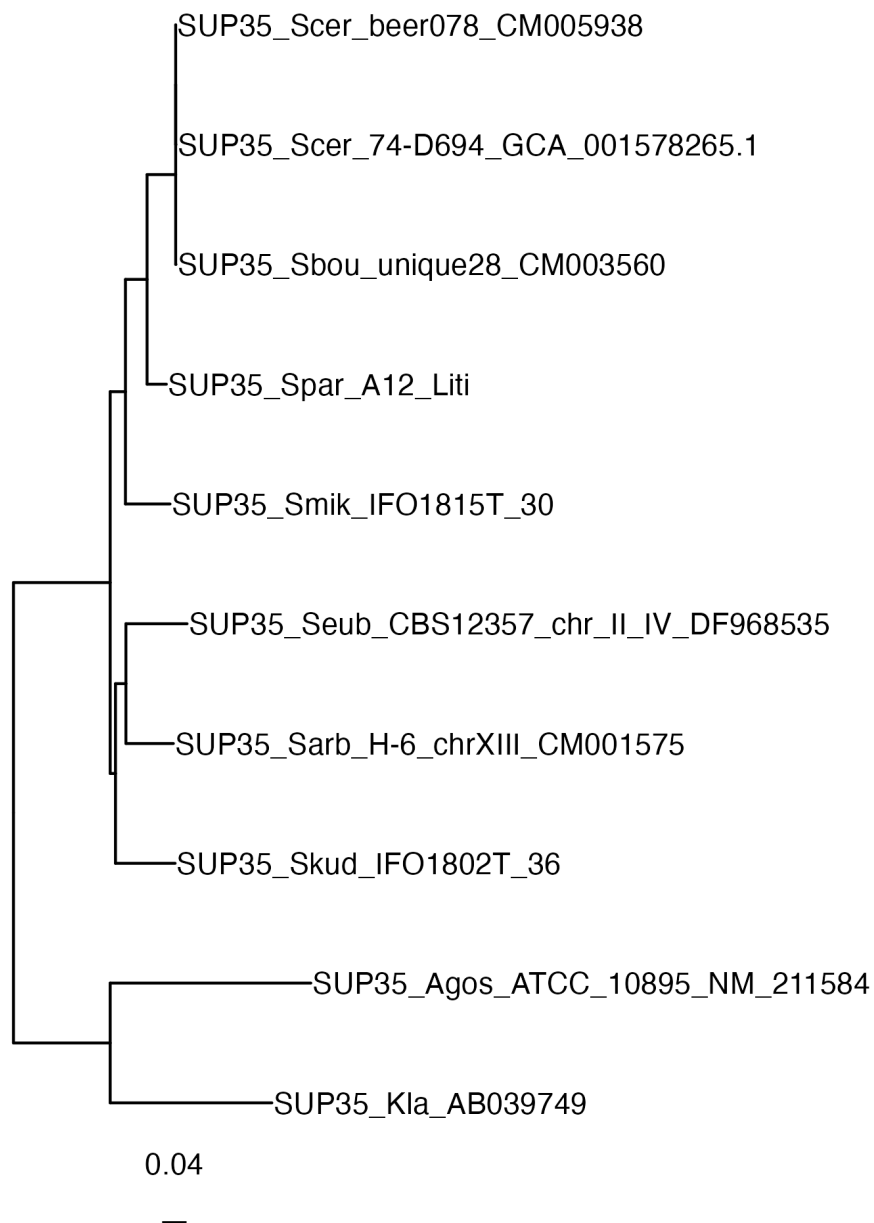
Вероятно, предварительно, нужно будет установить пакет `phytools`.

2. Визуализируйте укоренённые деревья из предыдущих пунктов и приведите их в сравнении с деревом с прошлого занятия. Чем отличаются изображения?

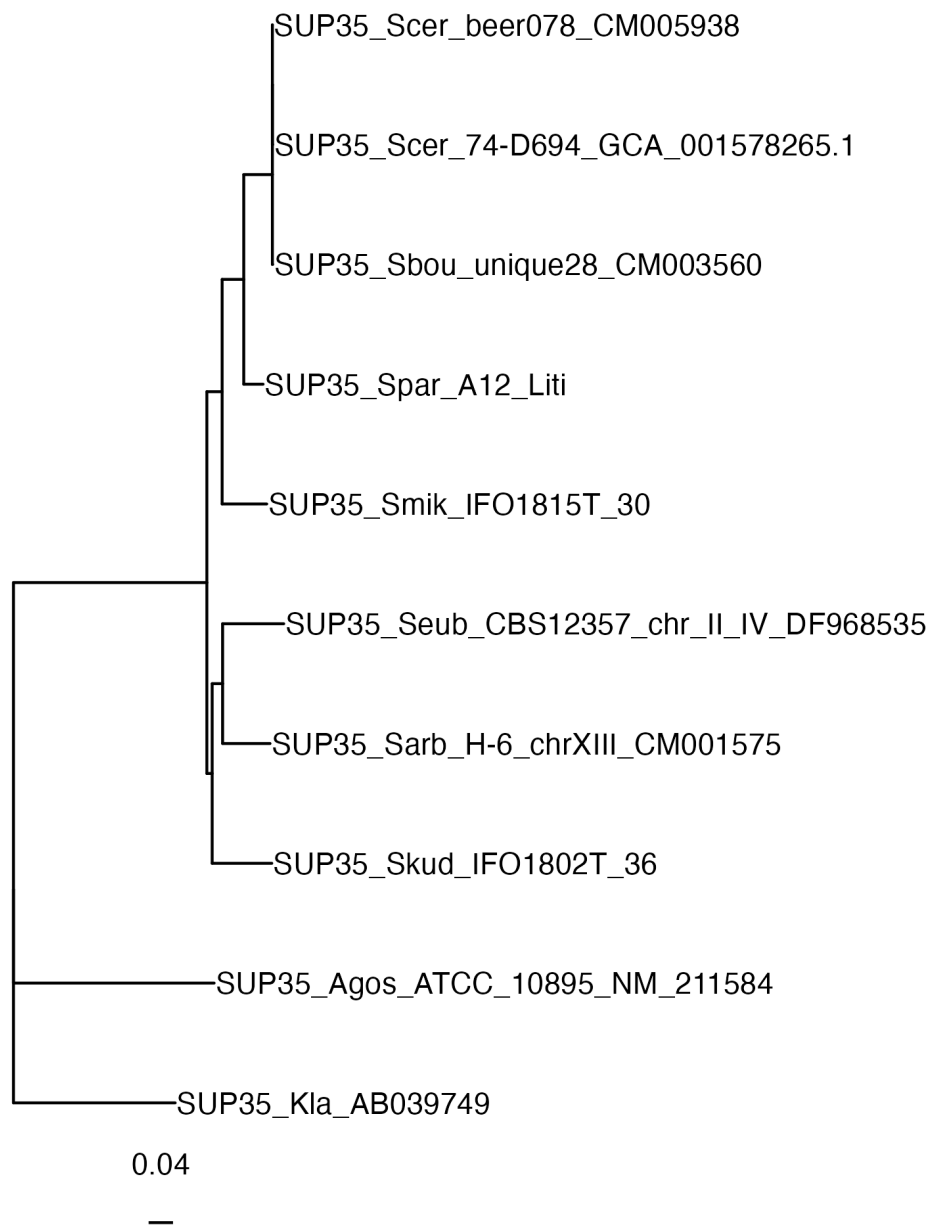
```
Rscript draw_tree.R SUP35_TIM3_ufb.treefile SUP35_TIM3_ufb.png
```



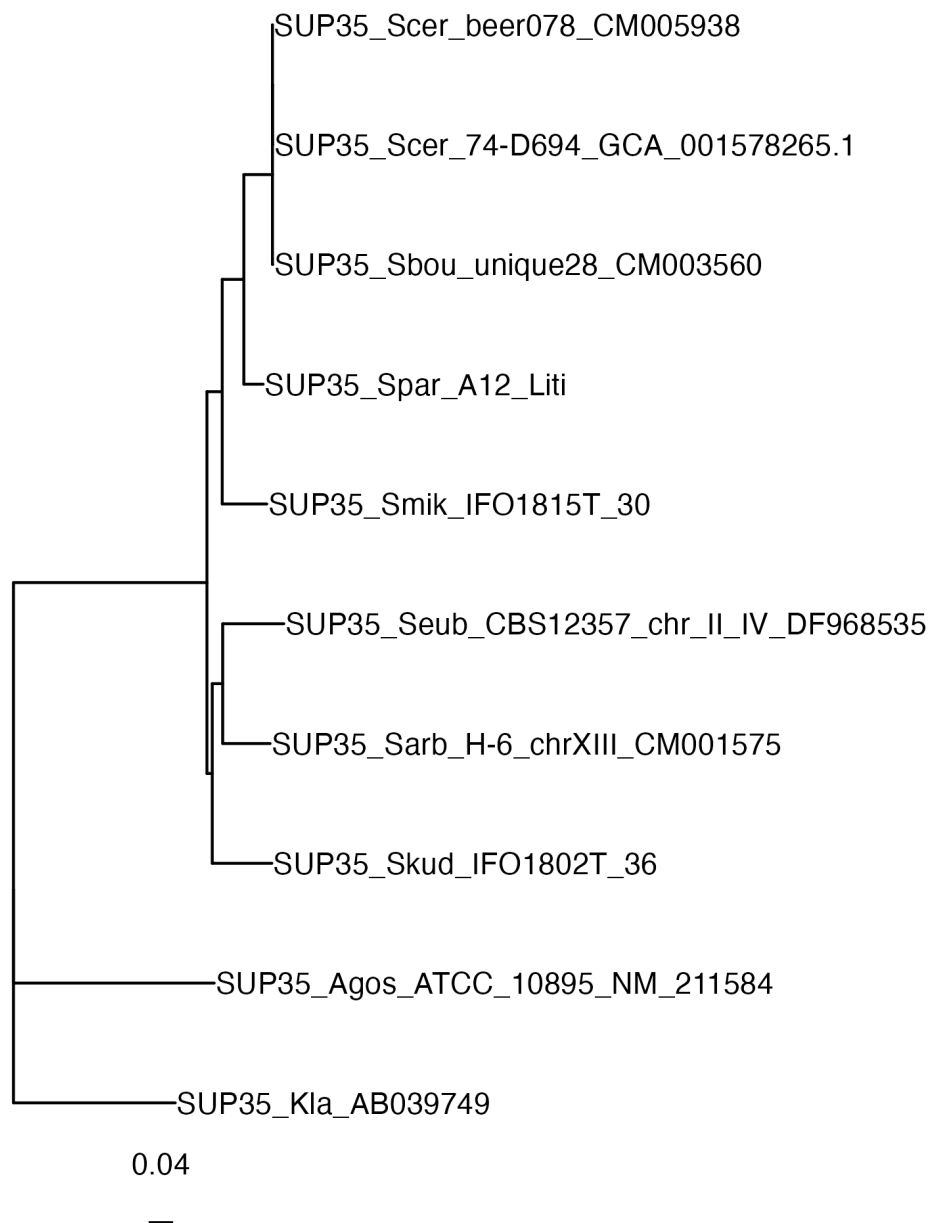
```
Rscript draw_tree.R SUP35_TIM3_ufb_midpoint.treefile SUP35_TIM3_ufb_midpoint.png
```



Rscript draw_tree.R SUP35_TIM3_root_outgroup.treefile SUP35_TIM3_root_outgroup.png



Rscript draw_tree.R SUP35_TIM3_root_outgroup.treefile SUP35_TIM3_root_outgroup.png

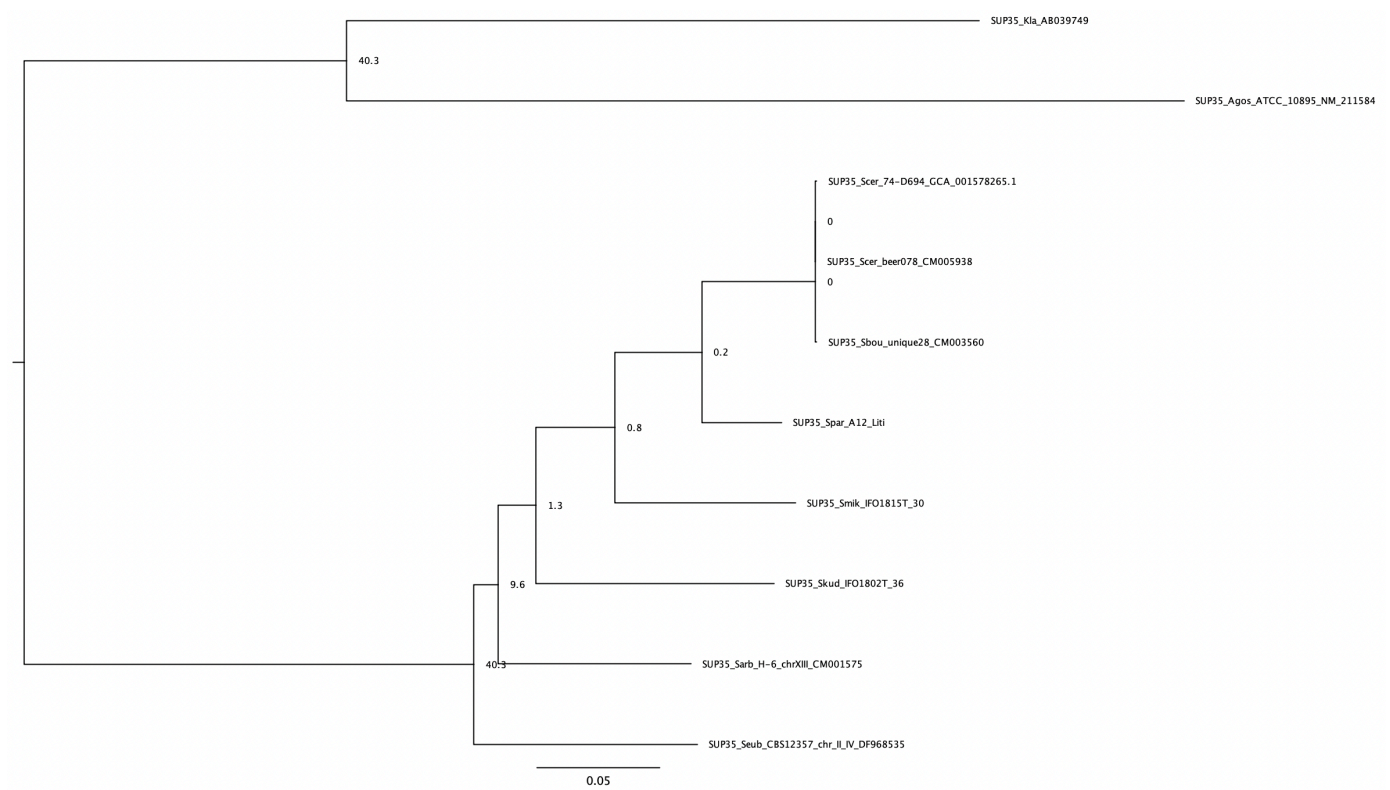


На первых двух изображениях нижние два вида объединены в одну кладу, в отличие от 3 и 4 изображения.

3. Как укоренить дерево с помощью необратимой (non-reversible) модели (iq-tree2)?

```
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -pre SUP35_TIM3_root_auto --model-joint 12.12 -B 1000
```

4. Визуализируйте дерево с поддержкой корня (rootstrap). Что мы можем сказать об уверенности алгоритма в выборе корня?



Алгоритм разделяет дерево с уверенностью в 40,3%.

5. (5 баллов) проведите анализ возраста общего предка двух видов дымчатых леопардов из статьи <https://doi.org/10.1016/j.cub.2006.08.066> (<https://doi.org/10.1016/j.cub.2006.08.066>) на данным секвенирования участка гена *atp8*, опираясь на известные данные о частоте замен в мтДНК (примерно 2% за миллион лет) в *beauti* и *beast*. Проверьте качество в *Tracer*. Объедините деревья в *treeannotator*. Отрисуйте финальное дерево (можно в *FigTree*, бонус за *ggtree*). Обязательно покажите оценки возраста общего предка на узлах!

а. Скачаем данные из NCBI, удаляем пробелы и убираем все лишнее.

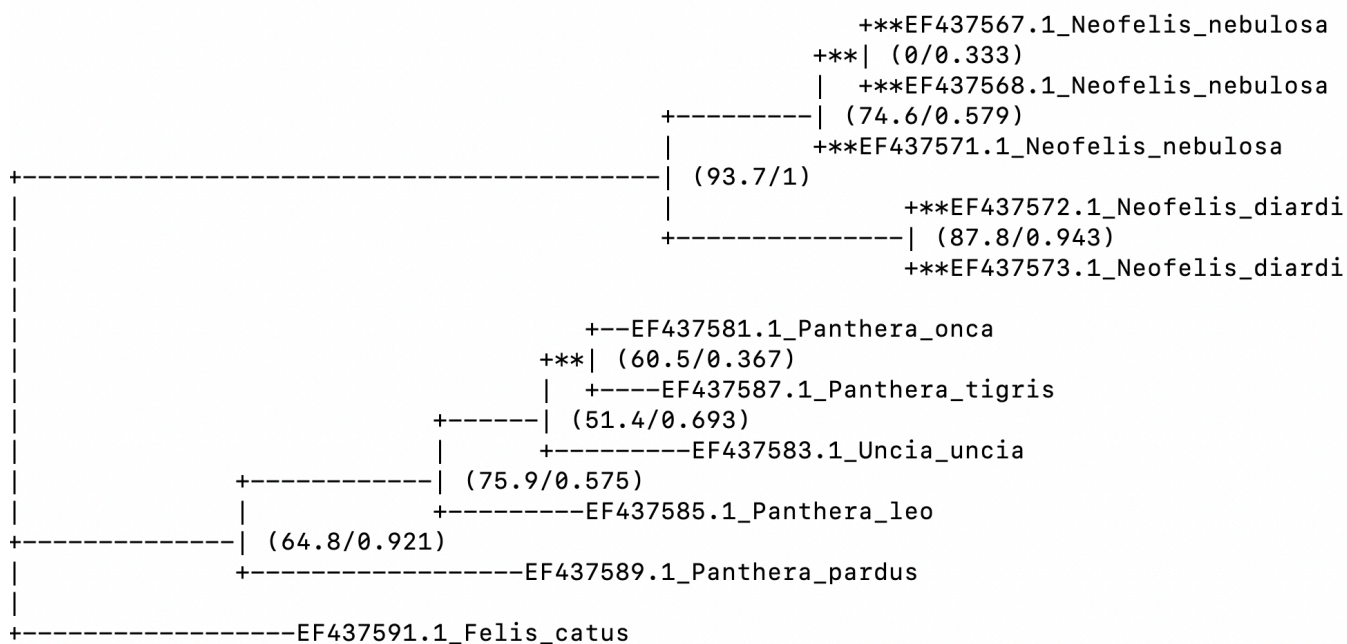
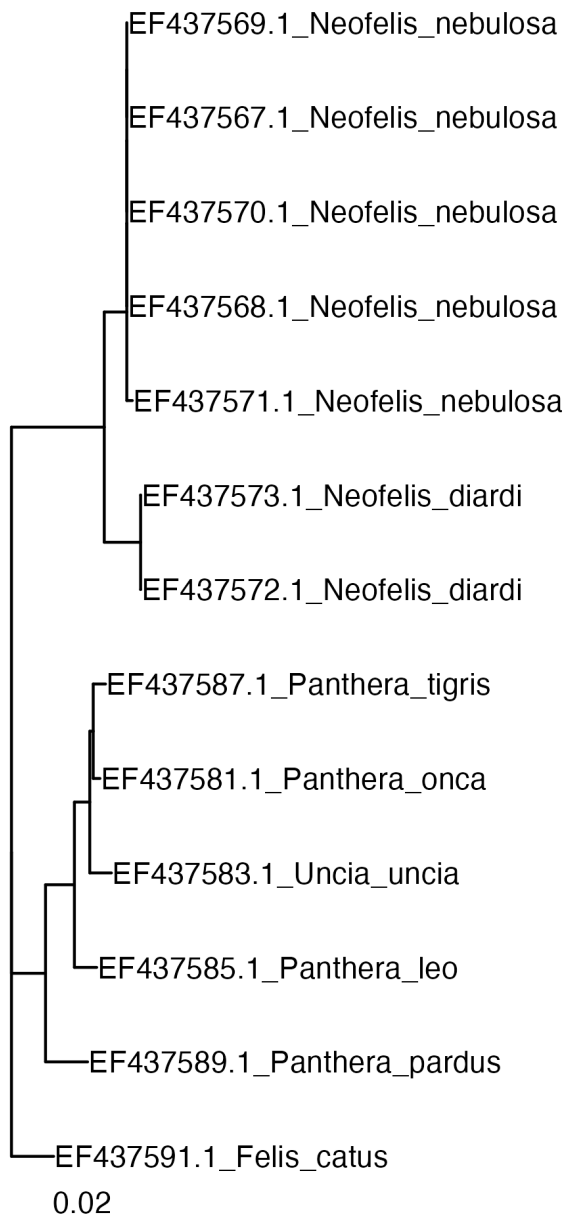
```
efetch -db popset -id 126256179 -format fasta >felidae_atp8.fa
cut -d ' ' -f 1,2,3 felidae_atp8.fa | sed -e 's/ /_/g' > felidae_atp8.renamed.fa
```

Производим выравнивание с помощью *mafft* и обрезаем гэпы с помощью *trimal*

```
mafft --auto felidae_atp8.renamed.fa >felidae_atp8.aln
trimal -in felidae_atp8.aln -out felidae_atp8.trim.fas -nogaps
```

Построим и отрисуем дерево

```
iqtree2 -s felidae_atp8.trim.fas -o EF437591.1_Felis_catus -alrt 1000 -abayes
Rscript draw_tree.R felidae_atp8.trim.fas.treefile felidae_atp8.trim.png
```



Лучшая модель для наших данных: TN+F.

b. Теперь запустим `beauti` и создадим конфигурационный файл для BEAST'a. Настраиваем параметры, как на скриншотах, остальные параметры оставляем по дефолту.

Partitions

Tip Dates

Site Model

Clock Model

Priors

MCMC

Link Site Models

Unlink Site Models

Link Clock Models

Unlink Clock Models

Link Trees

Unlink Trees

Name	File	Taxa	Sites	Data Type	Site Model	Clock Model	Tree	Ambiguities
felidae_atp8	felidae_atp8	13	103	nucleotide	felidae_atp8	felidae_atp8	felidae_atp8	<input type="checkbox"/>

Partitions

Tip Dates

Site Model ?

Clock Model

Priors

MCMC

Gamma Site Model ?

Substitution Rate1.0 ? e ☐ estimate

Gamma Category Count0 ?

Proportion Invariant0.0 ? e ☐ estimate

Subst ModelTN93 ?

Kappa12.0 ? e ☒ estimate

Kappa22.0 ? e ☒ estimate

FrequenciesEmpirical ? e

Partitions

Tip Dates

Site Model

Clock Model ?

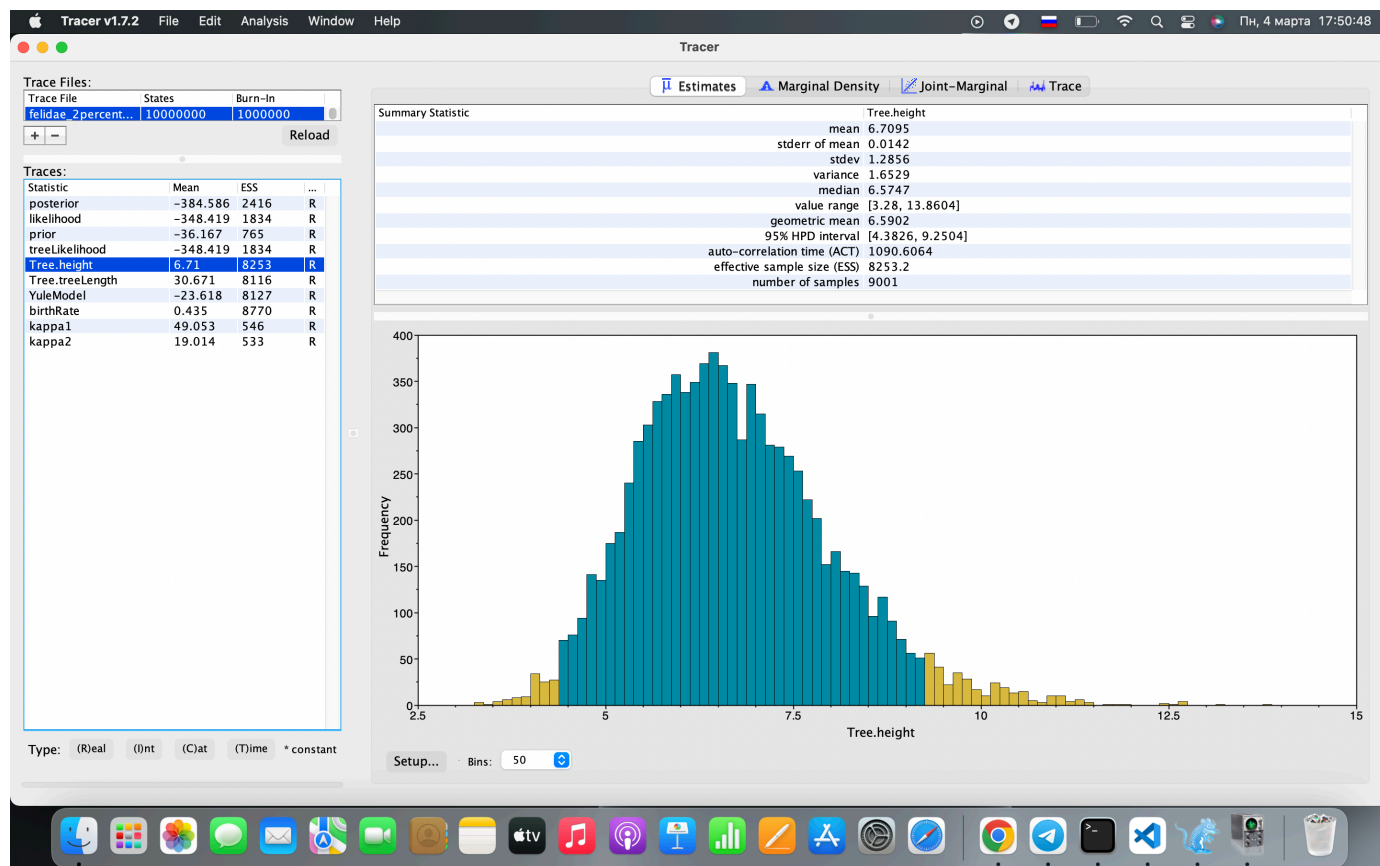
Priors

MCMC

Strict Clock ?

Mean clock rate0.02 ? e ☐ estimate

c) Запускаем `beast` через диалоговое окошко. Получаем файл `felidae_2percent.log`. Открываем `Tracer` и загружаем полученный файл. `Tree.height` и `Tree.treeLength` равны 8253 и 8116, что приемливо в нашем случае.



- d. Собираем все деревья из файла `felidae_2percent-felidae_atp8.trees` в одно дерево с помощью TreeAnnotator и сохраняем конечное дерево в файл `felidae_2percent-felidae_atp8.tree`

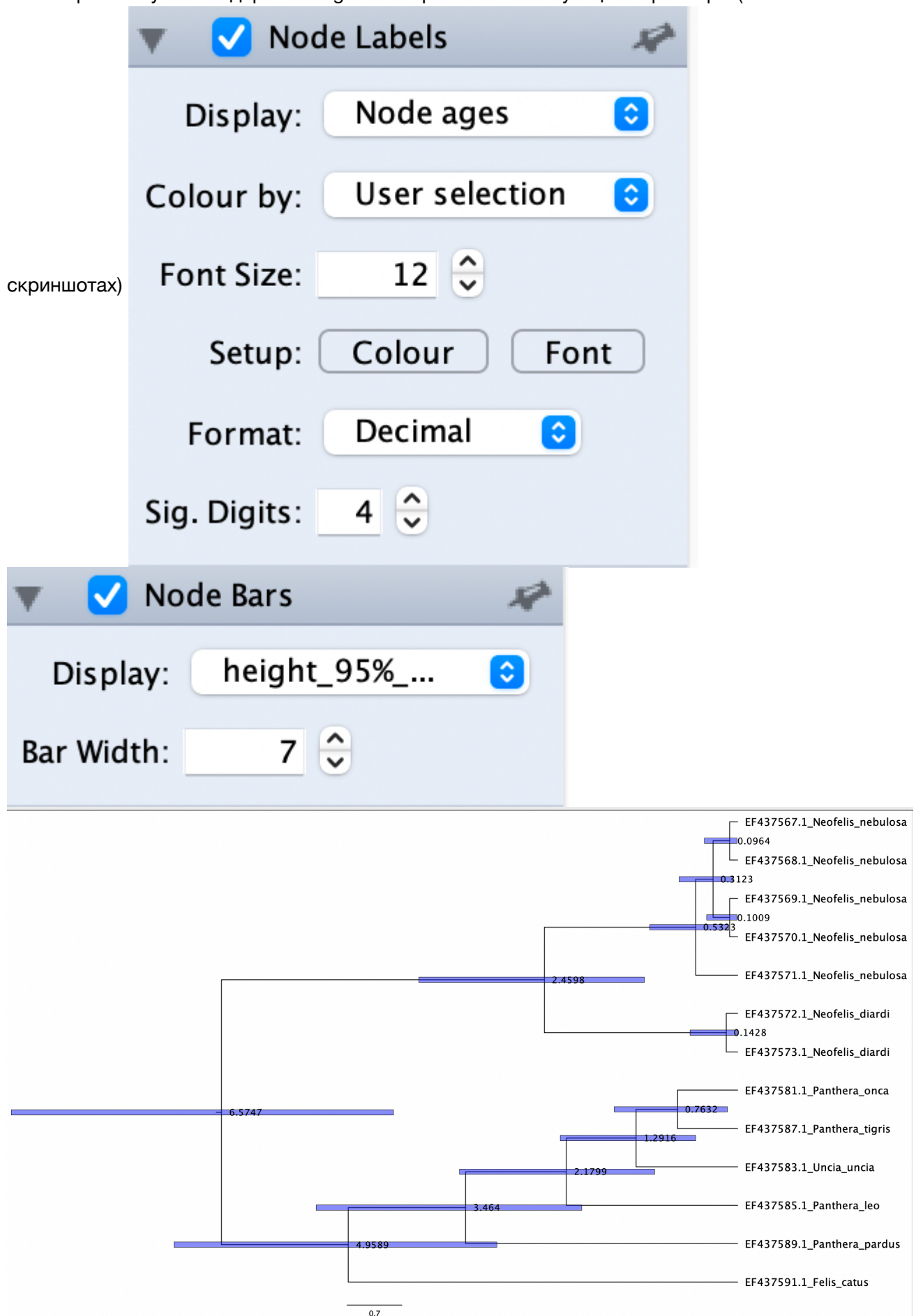
The screenshot shows the TreeAnnotator 2.7.6 interface. The configuration settings are as follows:

- Burn in percentage: 10
- Posterior probability limit: 0.0
- Target tree type: Maximum clade credibility tree
- Node heights: Median heights
- Target tree file: (empty field)
- Input Tree File: `/Users/nikitazherko/Desktop/class_5/felidae_2t`
- Output File: `/Users/nikitazherko/Desktop/class_5/felidae`
- Low memory: (unchecked checkbox)

Buttons at the bottom include 'Quit' and 'Run'.

е. Посмотрим полученное дерево в FigTree выбрав соответствующие параметры (как на

скриншотах)



6. Сравните результаты ваших анализов (возраст последнего общего предка *Neofelis*) с опубликованными статьями (<https://www.science.org/doi/10.1126/sciadv.adh9143> (<https://www.science.org/doi/10.1126/sciadv.adh9143>),

<https://www.sciencedirect.com/science/article/pii/S2589004222019198>

(<https://www.sciencedirect.com/science/article/pii/S2589004222019198>)). Какие выводы можно сделать?

Оценка по полным геномам этих двух видов равна 2.2 млн. лет. Соответственно их общий предок был столько лет назад.