

Отчет по построению филогенетических деревьев

Жерко Никита

1. При построении деревьев использовались программы:

- trimAl 1.2rev59
- ModelTest-NG v0.1.
- RAxML-NG v. 1.2.1
- IQ-TREE multicore version 2.2.2.6

2. Как вырезать плохие участки из выравнивания с помощью trimAl?

```
trimal -in SUP35_aln_prank.best.fas -out SUP35_aln_prank.trim.fas -automated1
```

3. Как подобрать модель эволюции в ModelTest (ModelTest-NG)?

```
modeltest-ng-static -i SUP35_aln_prank.trim.fas -o SUP35_trim_modeltest
# Summary:
#
#Partition 1/1:
#
```

#		Model	Score	Weight
#	BIC	TIM3+G4	18180.5614	0.3950
#	AIC	TIM3+I+G4	18041.1550	0.5376
#	AICc	TIM3+I+G4	18041.1550	0.5376

Установим, что лучшая модель TIM3+G4.

4. Строим ML-дерево в RAxML-NG, используя выбранную модель.

```
raxml-ng --msa SUP35_aln_prank.trim.fas --model TIM3+G4 --prefix SUP35_raxml --thread  
s 2 --seed 222 --outgroup SUP35_Kla_AB039749
```

5. Отрисуем полученное дерево (лучшее ML-дерево) в ggtree программе, которая вам нравится, и покажите рисунок.

Ggtree - пакет в R, код для отрисовки деревьев представлен ниже:

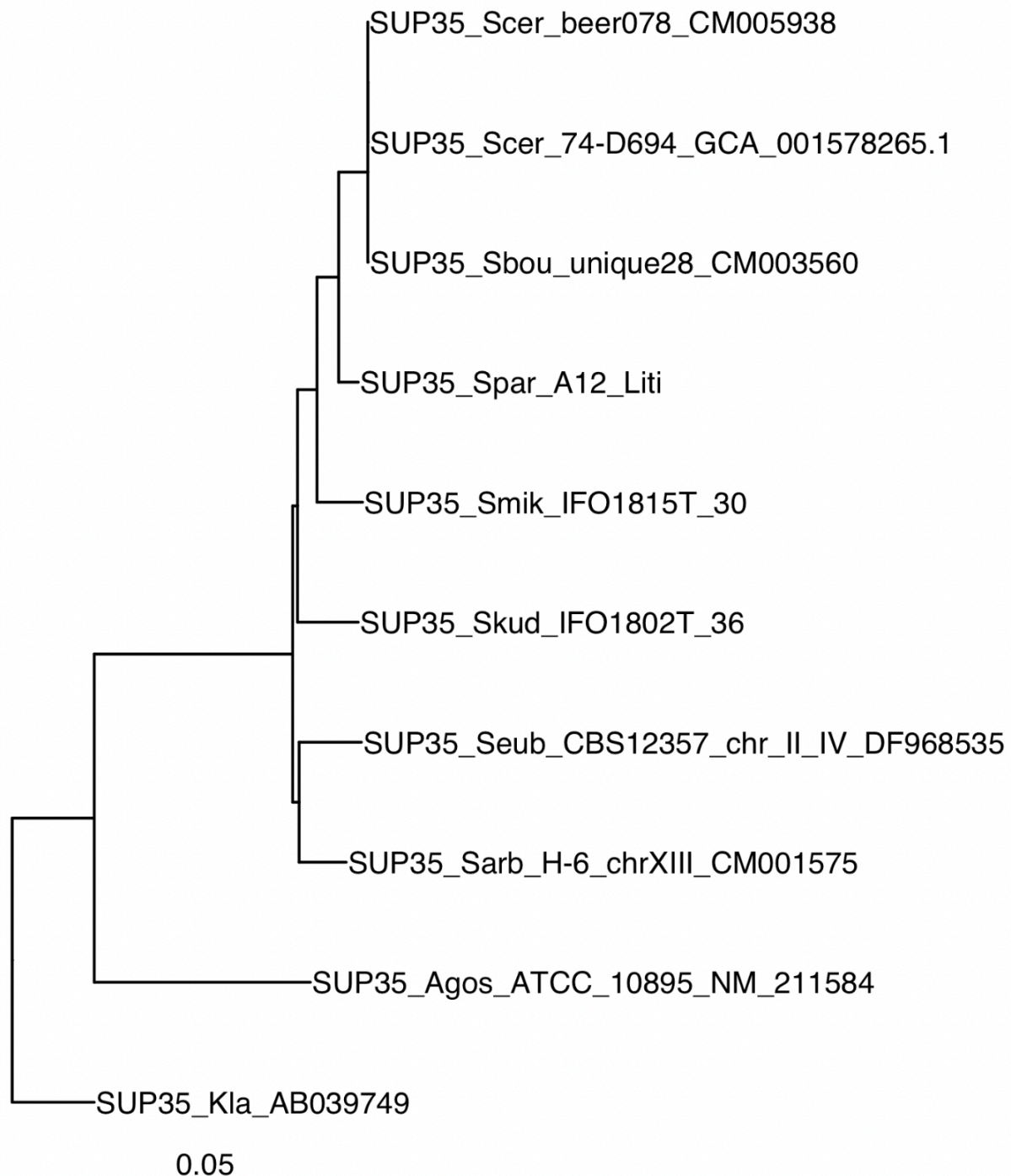
```
library(ggtree)
tr <- read.tree(args[1])
ggtree(tr) + geom_tiplab() + xlim(0,2) +
  geom_treescale()

ggsave(args[2])
```

команда для построения дерева:

```
Rscript draw_tree.R SUP35_raxml.raxml.bestTree SUP35_raxml.png
```

Полученное дерево представлено ниже:



6. Как выбрать модель в ModelFinder (можно через IQ-TREE)? Какая модель эволюции была признана наиболее подходящей для нашего выравнивания?

```
iqtree2 -m MFP -s SUP35_aln_prank.trim.fas --prefix SUP35_MF2
```

# Model	LogL	AIC	w-AIC	AICc	w-AICc	B
IC						
w-BIC						
#TIM3+F+G4	-8993.686	18035.372	0.0328	18035.972	0.0354	18170.
092 + 0.732						

Установим, что в нашем случае лучше модель TIM3+F+G4.

7. Отличаются ли модели, выбранные ModelTest и ModelFinder, и насколько сильно? Есть небольшие отличия в организации групп.

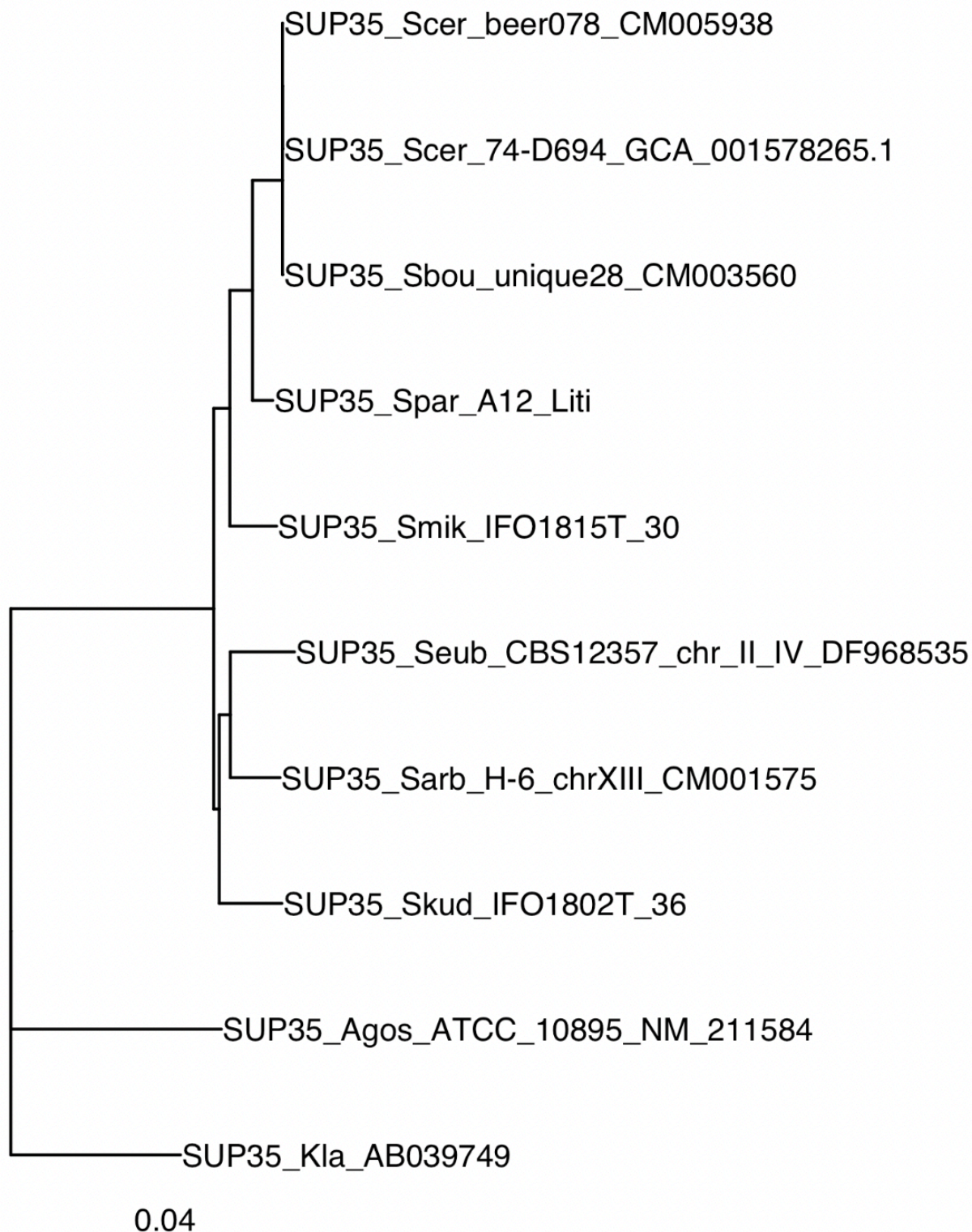
8. Постройте ML-дерево в IQ-TREE, используя выбранную модель.

```
iqtree2 -m TIM3+F+G4 -s SUP35_aln_prank.trim.fas --prefix SUP35_iqtree  
# строим дерево по указанной модели
```

9. Отрисуем полученное дерево (лучшее ML-дерево) и покажите рисунок.

```
Rscript draw_tree.R SUP35_iqtree.treefile SUP35_iqtree.png
```

Полученное дерево:



10. Сравним likelihood (log likelihood) деревьев, полученных с разными моделями, а также до и после фильтрации. Какой вывод можно из этого сделать?

Возьмем три разных дерева: 1) дерево, построенное по нефильрованному выравниванию 2) дерево, построенное по самой простой модели 3) дерево, построенное по самой лучшей модели

```
iqtree2 -s SUP35_aln_prank.best.fas -pre SUP35_iqtree_unfilt
iqtree2 -s SUP35_aln_prank.best.fas -m JC -pre SUP35_iqtree_JC
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -pre SUP35_TIM3
```

Likelihood у фильтрованного больше.

11. Как запустить базовую команду, но с генерацией 100 реплик обычного бутстрепа?

```
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -redo -pre SUP35_TIM3_b -b 100
```

12. Как запустить предыдущую команду, но с генерацией 1000 реплик ultrafast bootstrap?

```
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -redo -pre SUP35_TIM3_ufb -bb 1000
```

13. Как соотносятся время работы обычного и ultrafast bootstrap (в расчёте на одну повторность) и полученные значения?

Время работы Ultrafast bootstrap, очевидно выше в несколько раз (3 сек против 180 сек). Однако, будет хромать точность, поэтому уровень поддержки нужно ставить на 90%.

14. Как запустить предыдущую команду, но с генерацией: 1000 ultrafast bootstrap + 1000 alrt + abayes?

```
iqtree2 -s SUP35_aln_prank.trim.fas -m TIM3+F+G4 -pre SUP35_TIM3_B_alrt_abayes -bb 1000 -alrt 1000 -abayes
```

15. Отрисуйте полученное дерево с тремя поддержками.

Скрипт по отрисовке напишем на R

```
tree_alrt_abayes_ufb <- read.tree("SUP35_TIM3_B_alrt_abayes.treefile")
ggtree(tree_alrt_abayes_ufb) +
  geom_tiplab() + geom_nodelab() +
  geom_treescale() + xlim(0, 0.7)
```

в каждом узле по три поддержки (alrt, abayes, bb)

Полученное дерево:

