

Реконструкция филогенетических деревьев, ч. 3

С.А.Спирин

3 марта 2023

ФББ МГУ

1.

СРАВНЕНИЕ ДЕРЕВЬЕВ

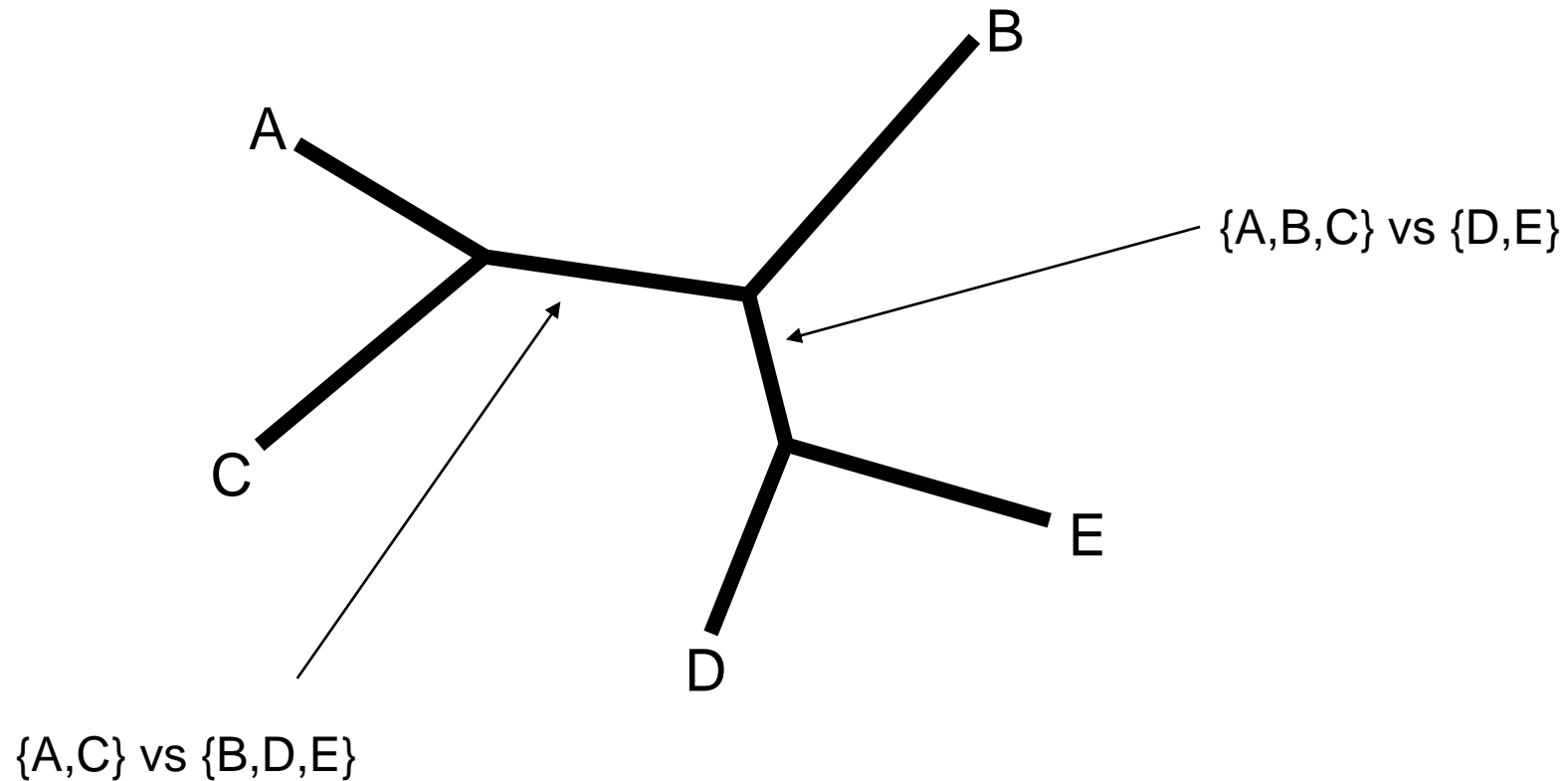
Сравнение деревьев

Программы реконструкции филогении так же ненадёжны, как и любые другие компьютерные программы предсказания биологических фактов.

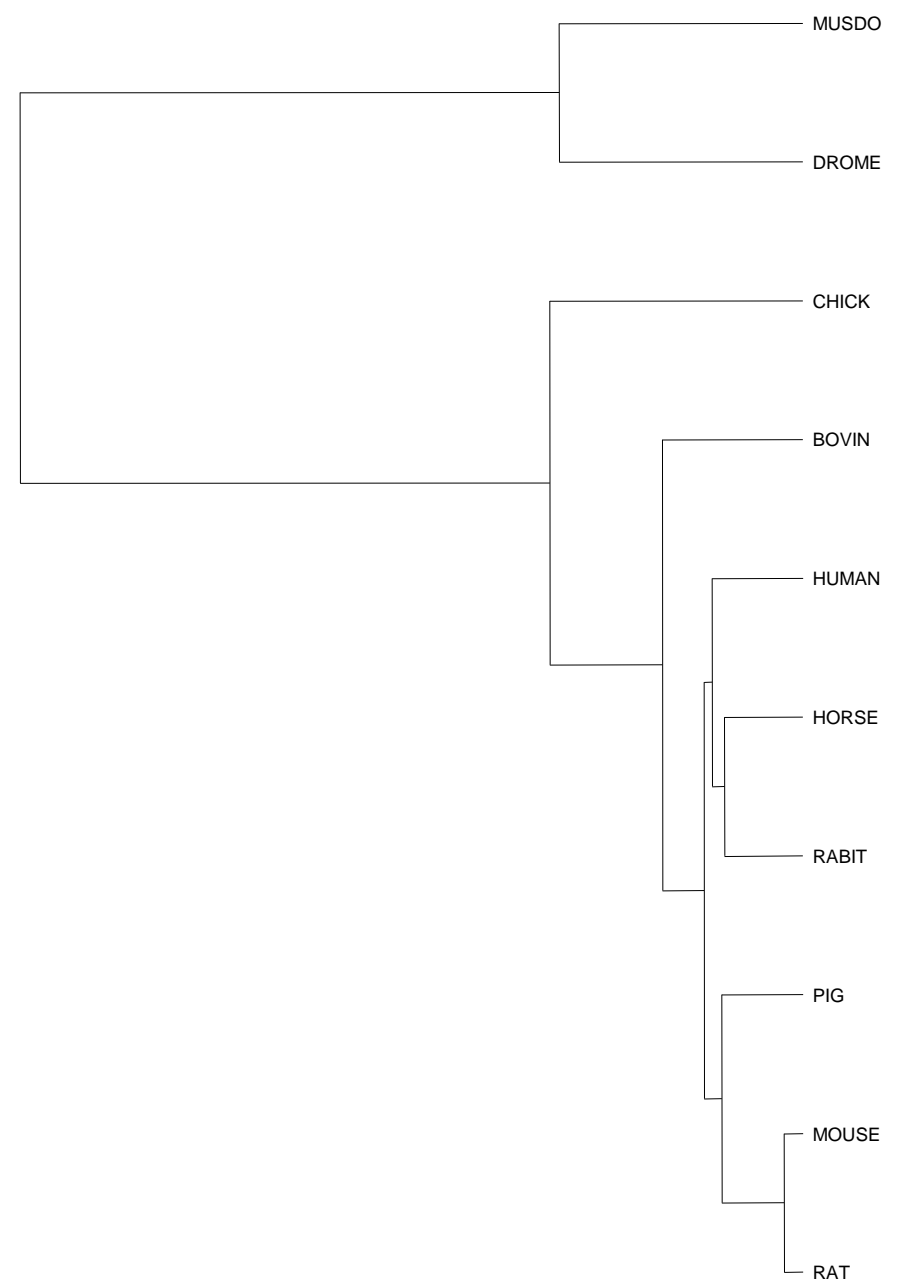
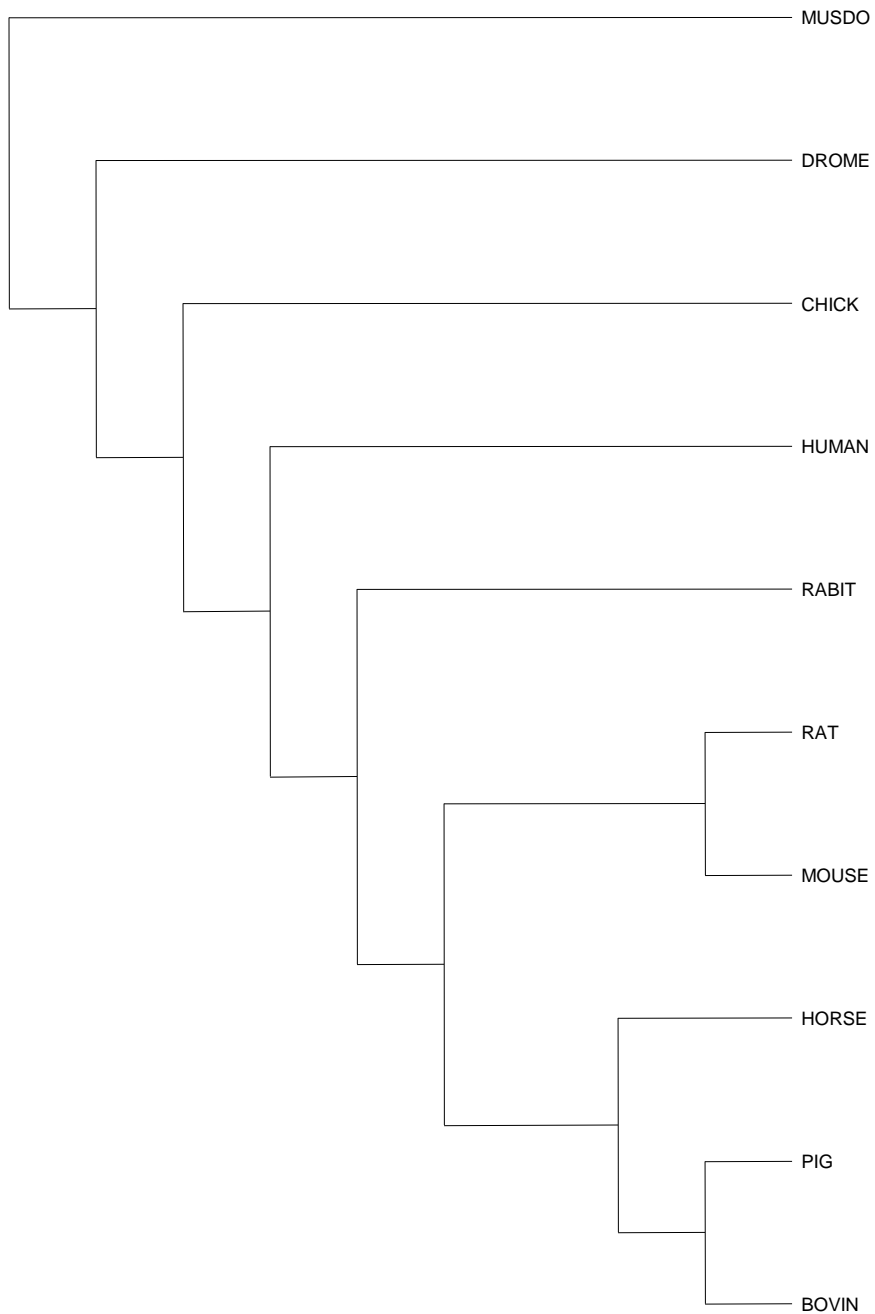
Поэтому (в частности) возможны различные варианты реконструкции по одним и тем же данным.

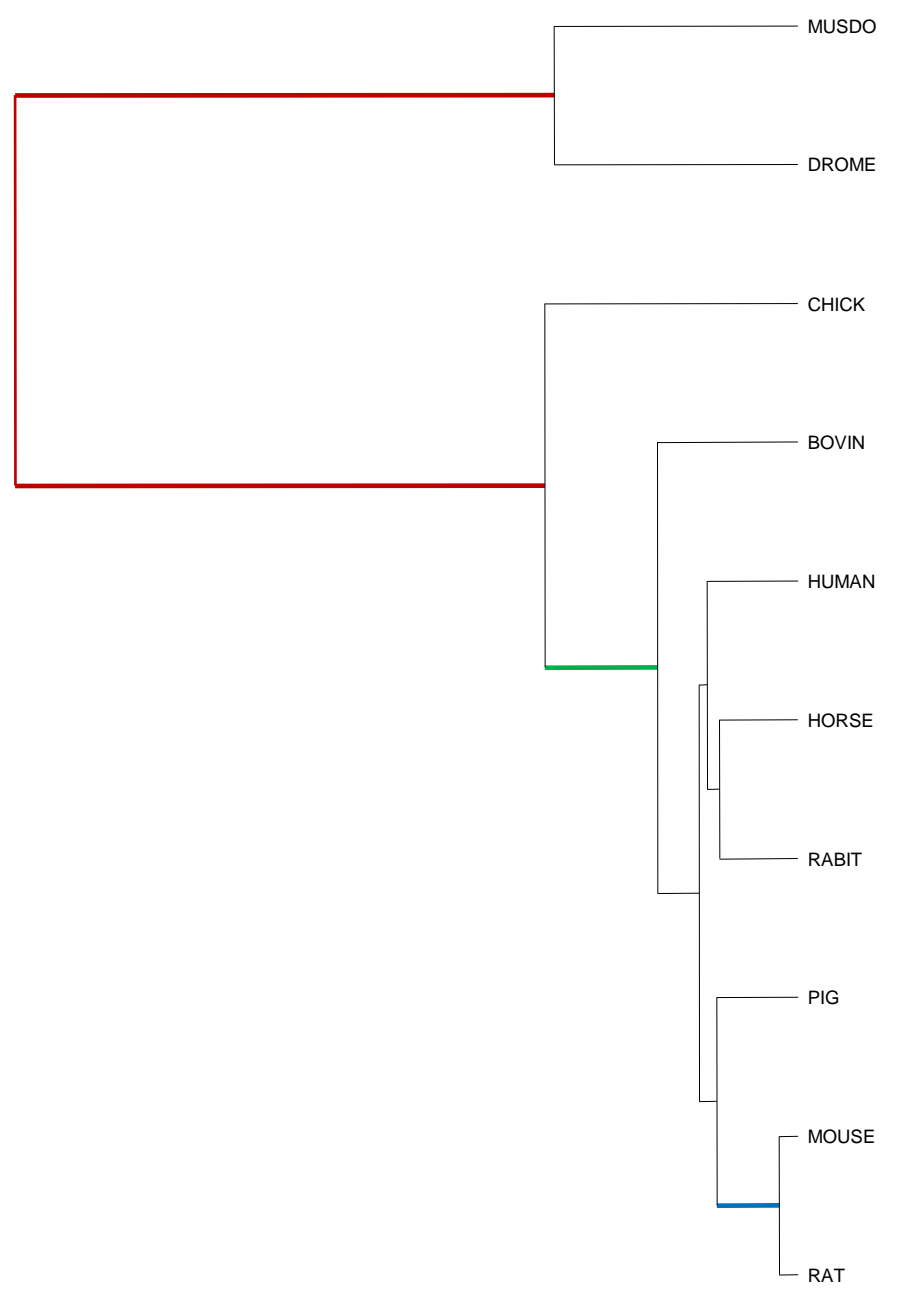
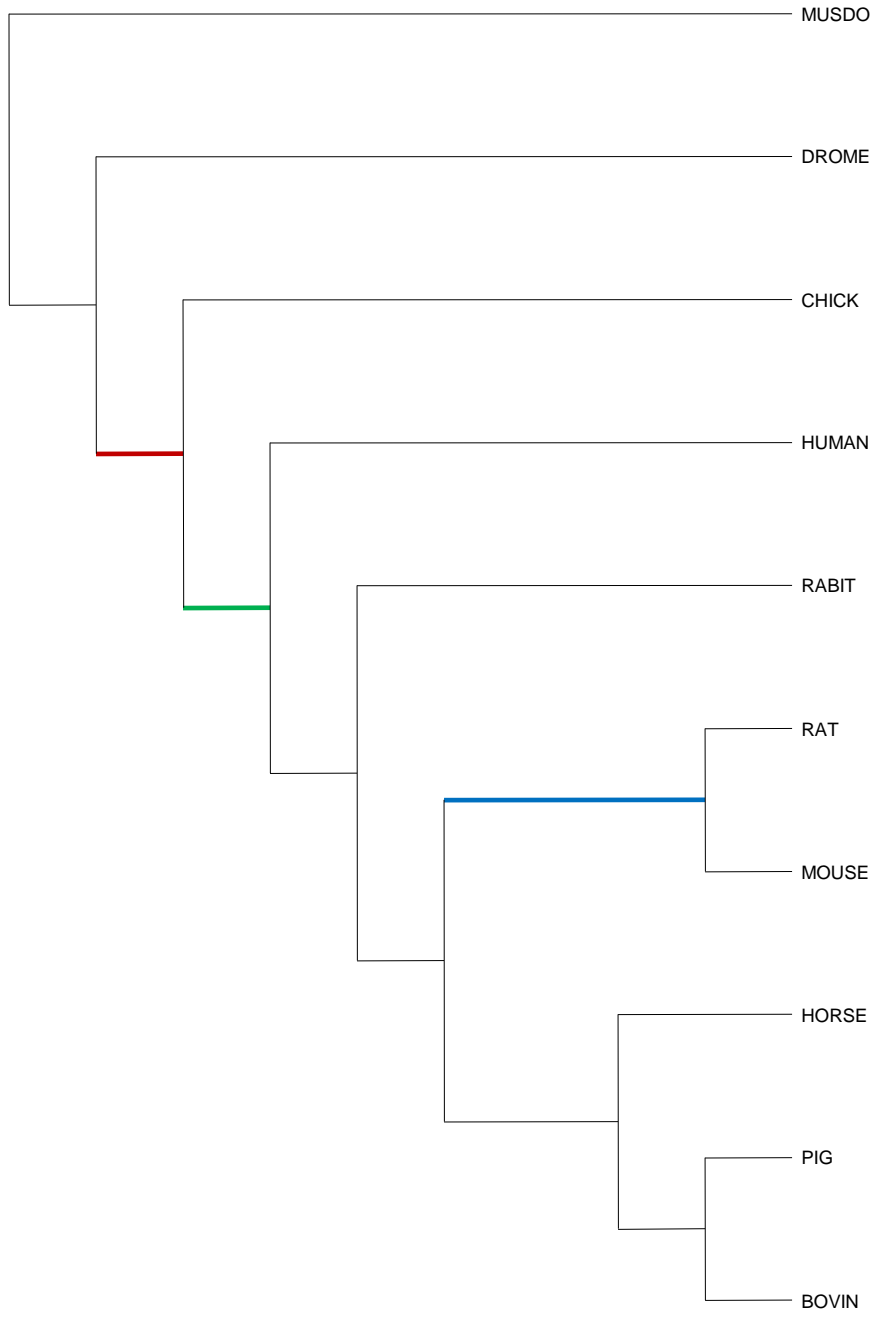
Встаёт задача сравнения различных деревьев с одним и тем же множеством листьев.

Напоминание: ветвь дерева как разбиение множества листьев

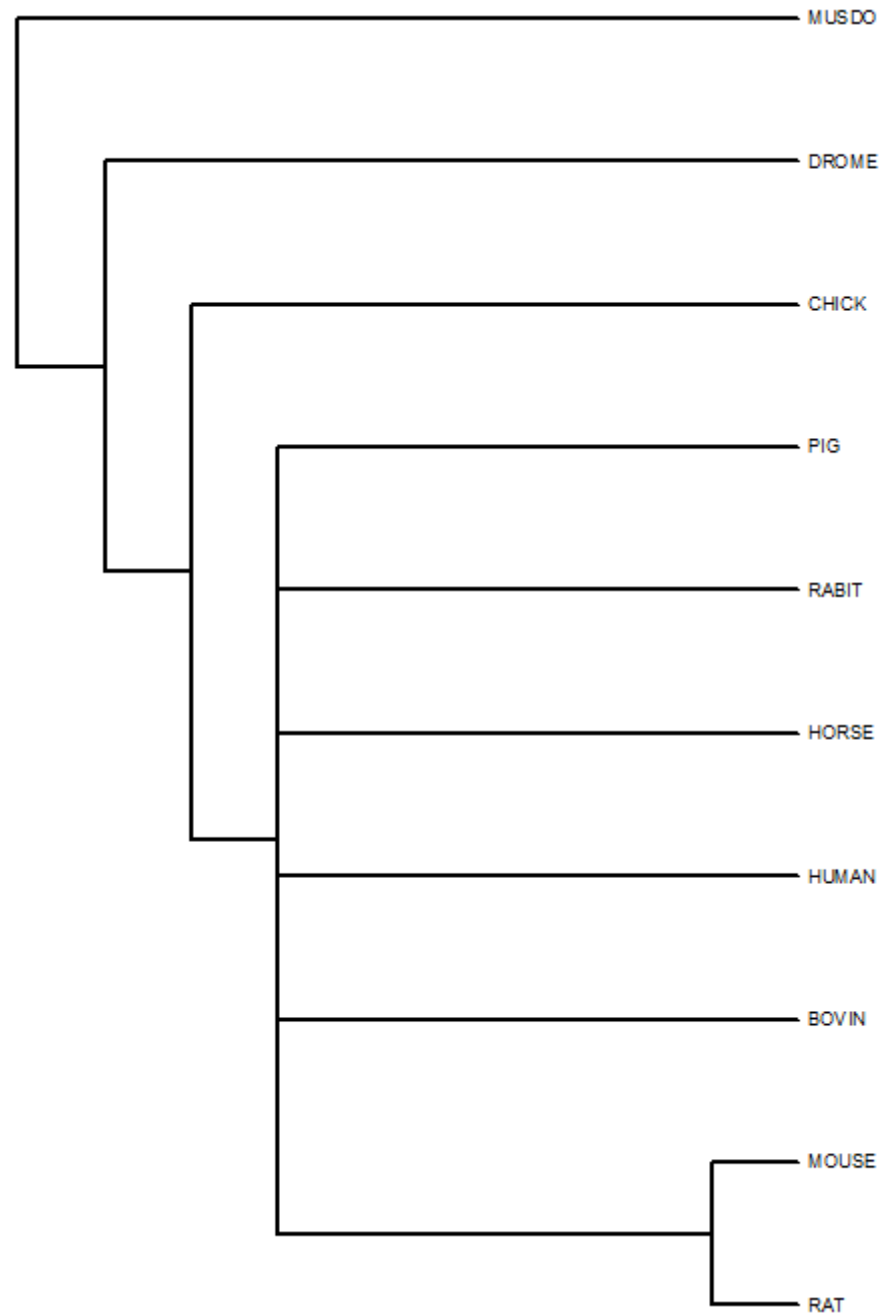


Что общего у этих двух деревьев?





Консенсусное дерево



Совмещение данных от многих деревьев

По набору деревьев с **одинаковым** множеством листьев:

- 1** Консенсус (consensus)
Включает только те ветви, которые встретились **во всех** деревьях исходного набора
- 2** Дерево большинства (majority-rule tree)
Включает только те ветви, которые встретились **в большинстве** деревьев исходного набора
- 3** Дерево расширенного большинства (extended majority-rule tree)
К дереву большинства добавляются ветви, не противоречащие уже имеющимся, начиная с наиболее «поддержанных»

Задача на дом

Даны две ветви, то есть два разбиения одного и того же множества листьев.
Как понять, могут ли эти две ветви встретиться в одном и том же дереве?

ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees

Chao Zhang, Maryam Rabiee, Erfan Sayyari & Siavash Mirarab 

BMC Bioinformatics **19**, Article number: 153 (2018) | [Cite this article](#)

16k Accesses | 507 Citations | 8 Altmetric | [Metrics](#)

Abstract

Background

Evolutionary histories can be discordant across the genome, and such discordances need to be considered in reconstructing the species phylogeny. ASTRAL is one of the leading methods for inferring species trees from gene trees while accounting for gene tree discordance. ASTRAL uses dynamic programming to search for the tree that shares the maximum number of quartet topologies with input gene trees, restricting itself to a predefined set of bipartitions.

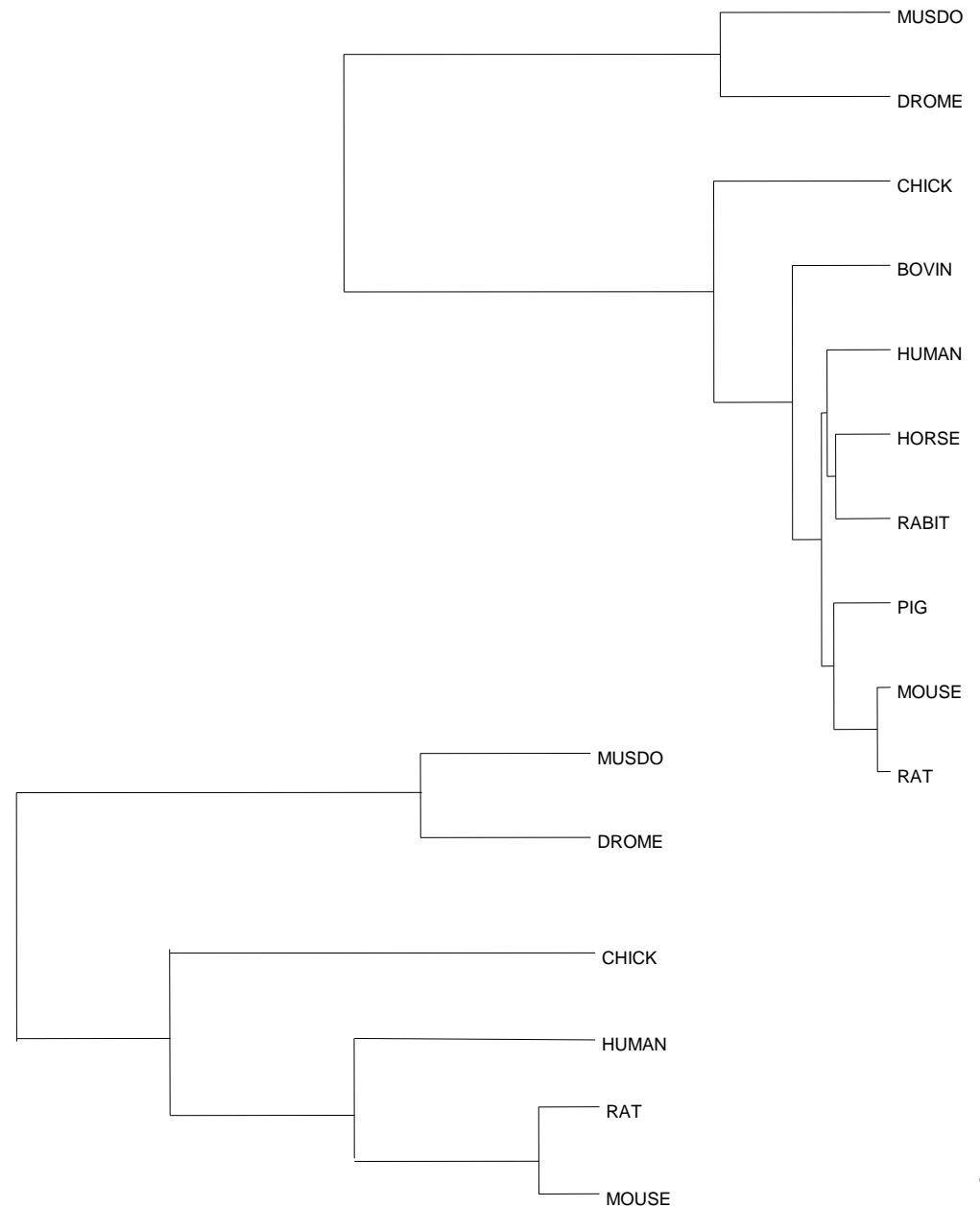
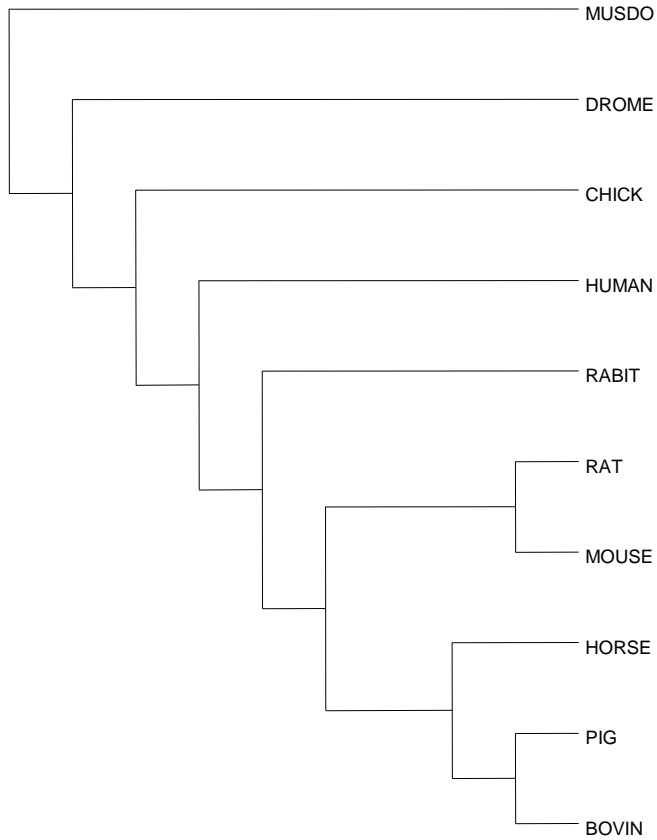
Results

We introduce ASTRAL-III, which substantially improves the running time of ASTRAL-II and guarantees polynomial running time as a function of both the number of species

[Download PDF](#)[Sections](#)[Figures](#)[References](#)[Abstract](#)[Background](#)[Methods](#)[Results](#)[Discussion](#)[Conclusions](#)[References](#)[Acknowledgements](#)[Author information](#)[Ethics declarations](#)[Additional information](#)[Additional file](#)[Rights and permissions](#)[About this article](#)

Advertisement

Наибольшее общее поддерево

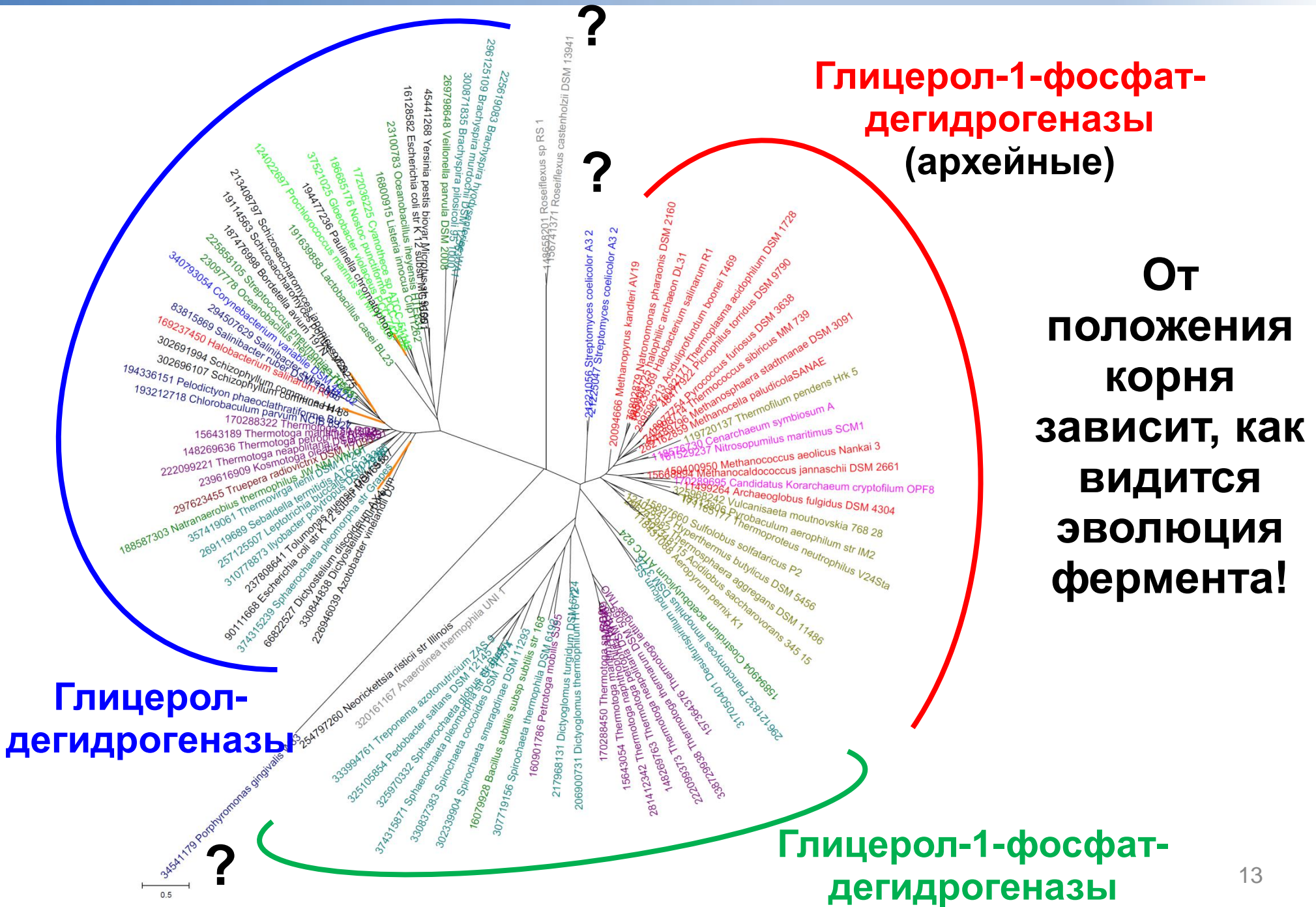


Н.О.П. определено не однозначно!

2.

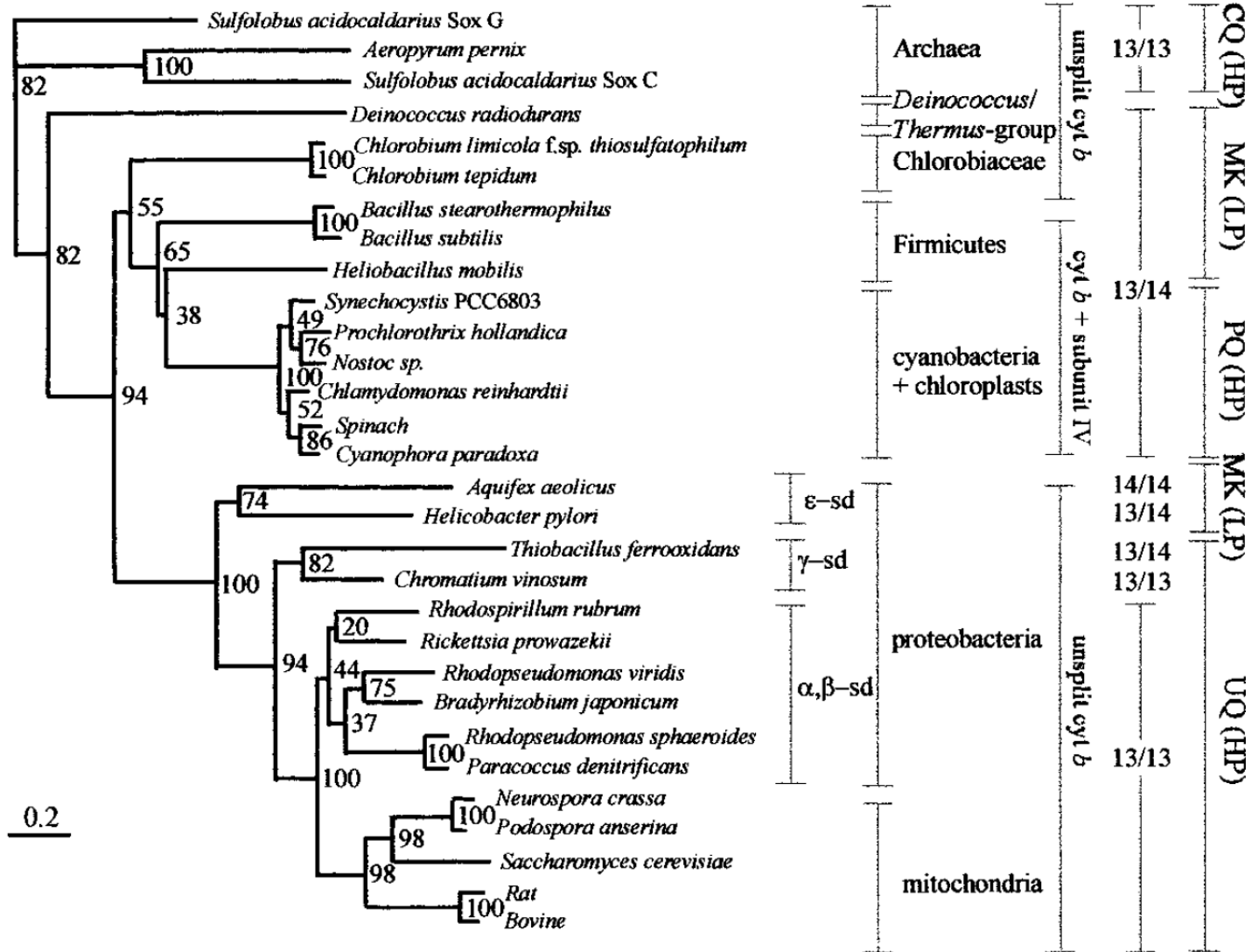
УКОРЕНЕНИЕ

Укоренение дерева



Укоренение

(Schütz et al., 2000)



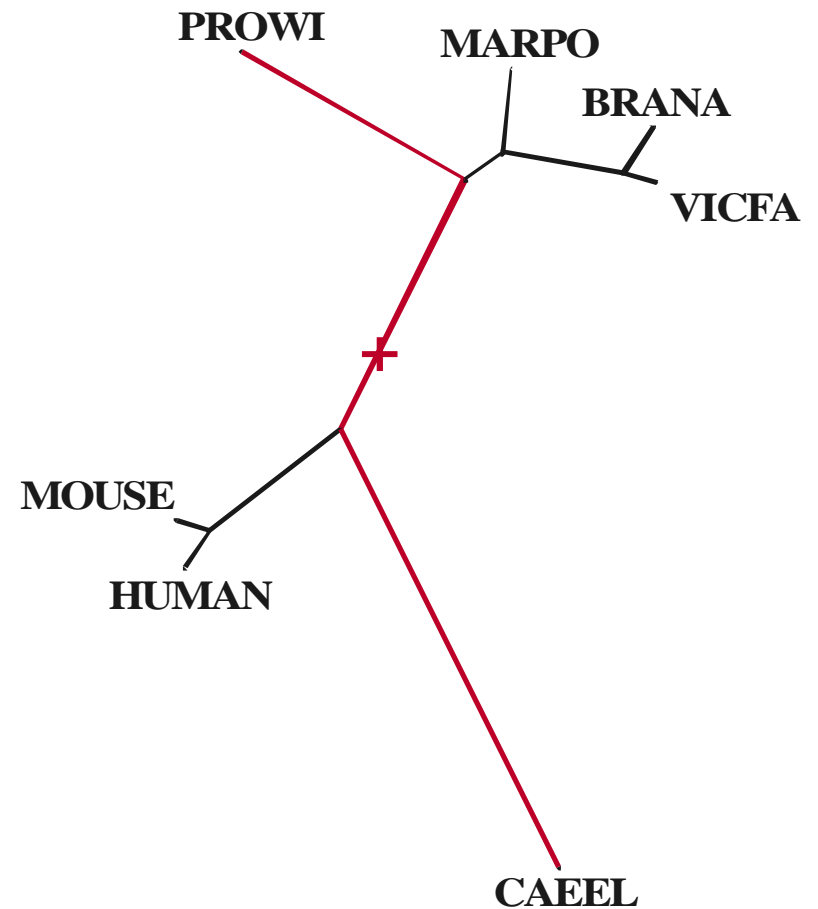
Для подтверждения своей гипотезы о древности некоторого фермента авторы работы укоренили дерево так, чтобы корневая ветвь разделяла археи и эубактерии.

Однако такое укоренение ничем не обосновано!

Укоренение

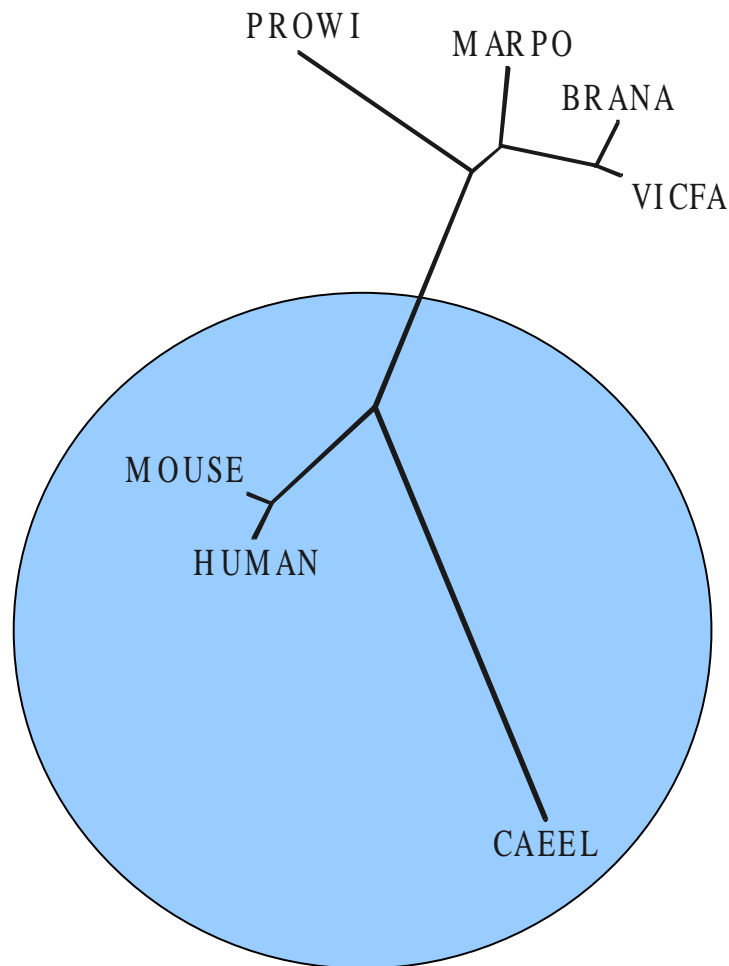
- В среднюю точку:

Находим на дереве самый длинный путь от листа к листу и за корень принимаем середину этого пути



Укоренение

- Используя внешнюю группу («аутгруппу», outgroup):



В данном случае укоренено дерево четырёх растений, для чего пришлось построить дерево с участием внешней группы — трёх животных (в синем круге)

3.

BOOTSTRAP

Bootstrap



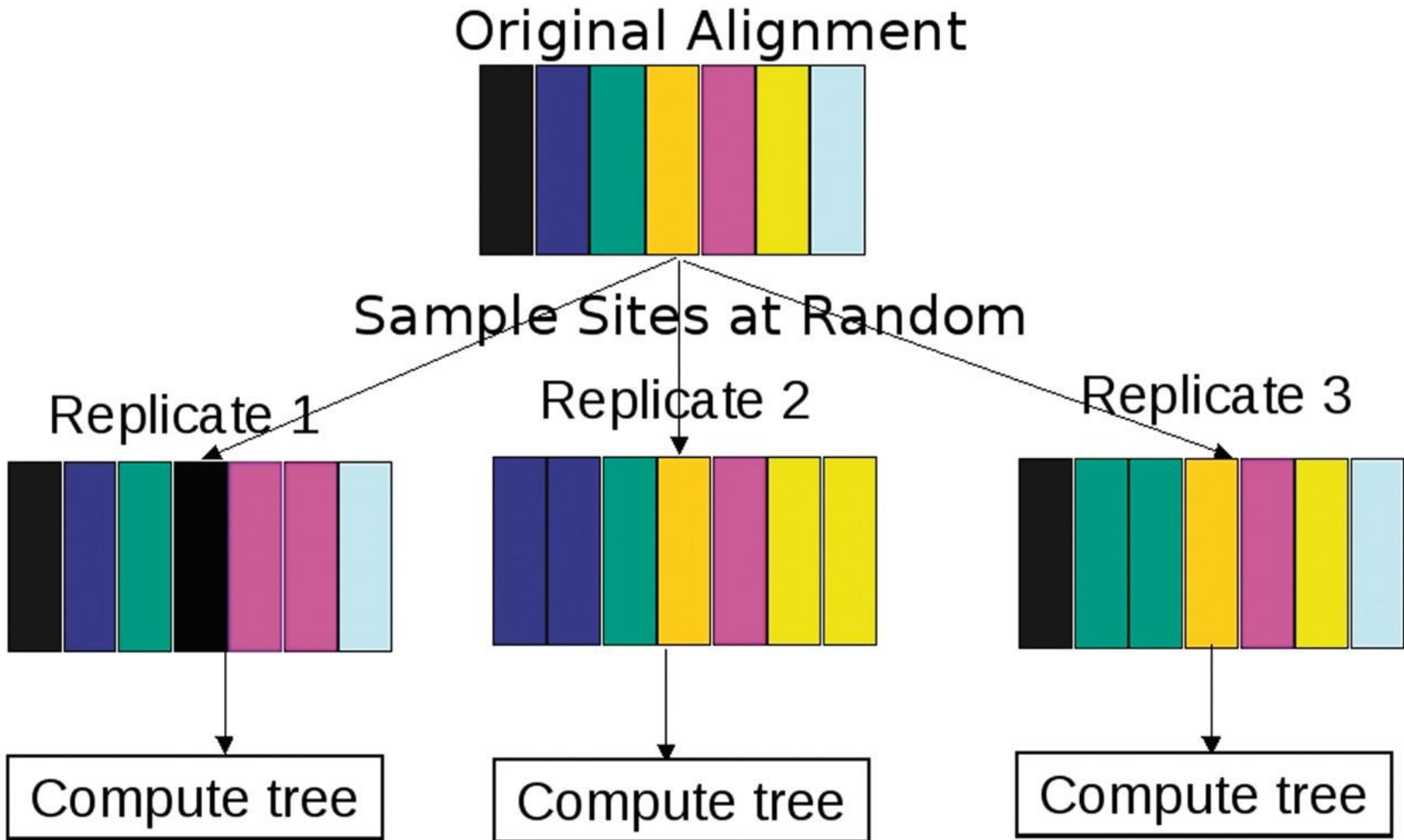
Бутстрэп-анализ

Из входного выравнивания делается много (например, 100) так наз. «бутстрэп-реплик».

Каждая бутстрэп-реплика получается в результате случайного выбора столбцов из выравнивания в том же количестве, что в исходном выравнивании. В результате некоторые столбцы выбираются один раз, некоторые дважды, некоторые ни разу.

Смысл в том, чтобы построить дерево по части данных и затем сравнить результаты от по разному выбранных частей.

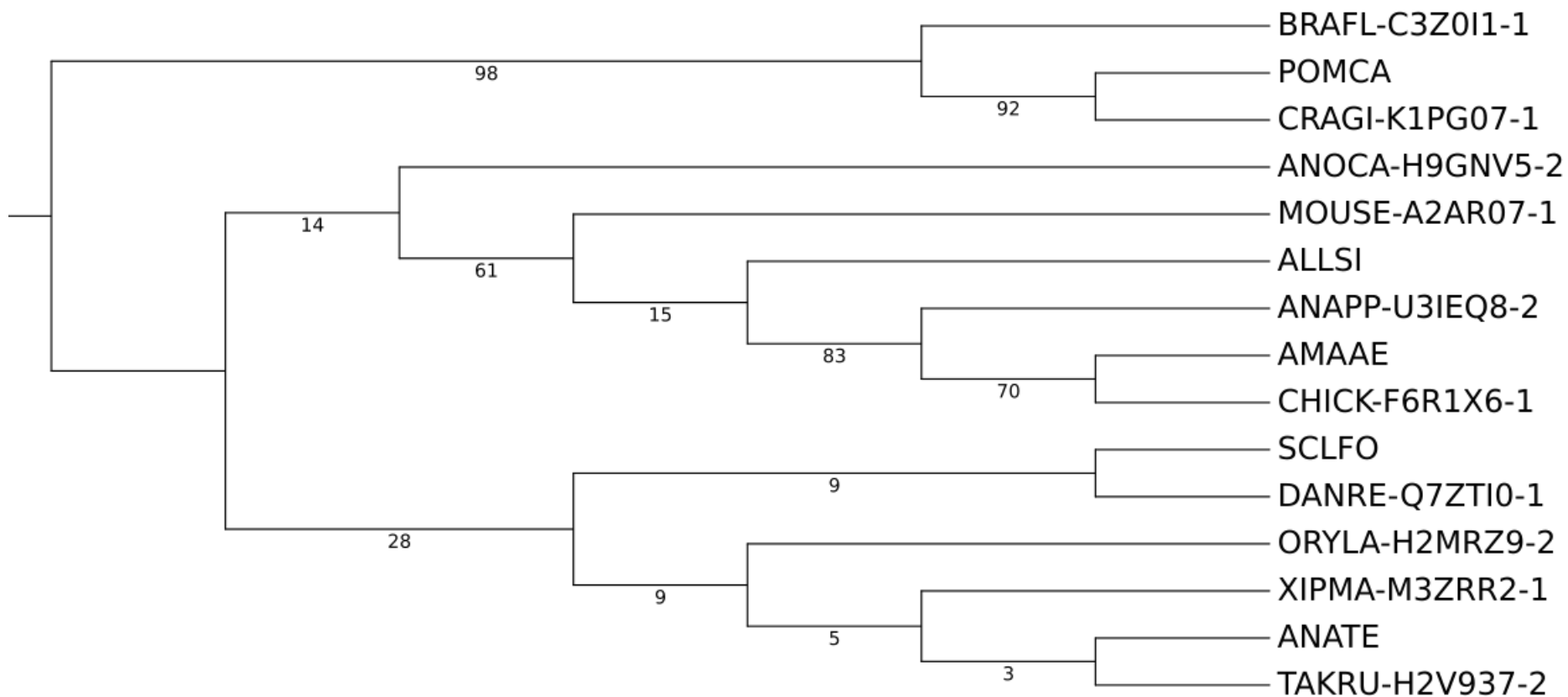
Outline of the phylogenetic bootstrap procedure.



Бутстрэп-анализ

- создаём из входного выравнивания 100 бутстрэп-реплик;
- для каждой из реплик строим по дереву;
- два варианта:
 - строим дерево по входному выравниванию, и для каждой ветви отмечаем, в каком проценте деревьев, построенных по репликам, встретилась эта ветвь.
 - из 100 деревьев строим дерево по методу расширенного большинства («Extended majority-rule tree»).
- по числам на ветвях («бутстрэп-поддержке») судим о достоверности ветви

Бутстрэп-анализ (пример результата)



4.

АНАЛИЗ ДЕРЕВА, СОДЕРЖАЩЕГО ПАРАЛОГИ

Ортологи и паралоги

Белки-гомологи: белки, имеющие общее происхождение.

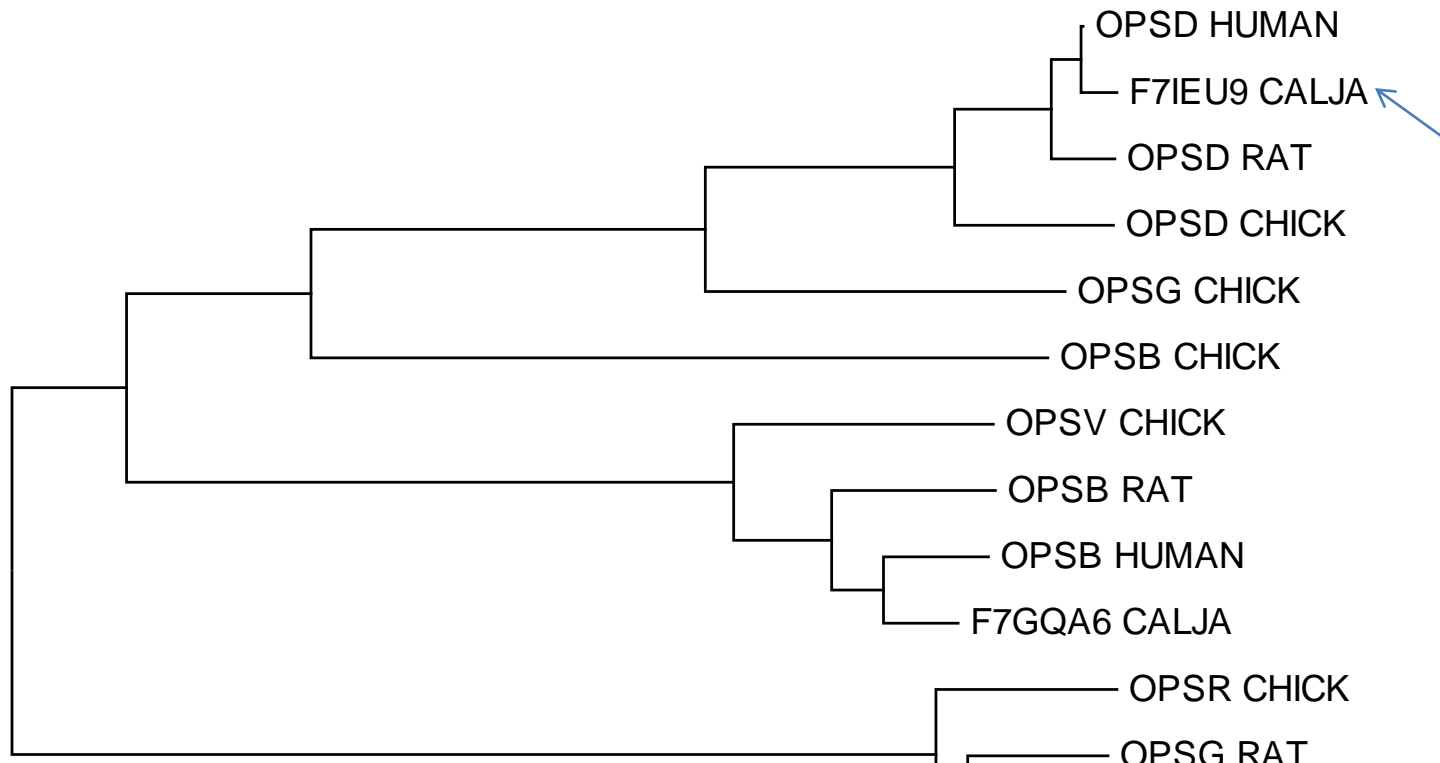
Среди гомологов выделяют **ортологи** и **паралоги**

- Белки-ортологи разошлись в результате видообразования.
- Белки-паралоги:
 - а) разошлись в результате геной дупликации;
 - б) присутствуют в одном и том же организме.

Вообще-то из б) следует а), ну да ладно :)

Анализ дерева группы гомологичных белков помогает выявить ортологи и реконструировать эволюционные события: видообразования и геной дупликации.

Пример дерева с паралогами



игрунка

- OPSD – родопсин
- OPSV – фиолетовый опсин
- OPSB – синий опсин
- OPSG – зелёный опсин
- OPSR – красный опсин
- OPSL – длинноволновой опсин

5.

ЧТО НУЖНО ЗАПОМНИТЬ

Что нужно помнить при реконструкции филогении по последовательностям

1. Последовательности должны быть гомологичны по всей длине.
2. Если последовательности нуклеотидные, надо убедиться, что часть из них не представлена комплементарными вариантами.
3. Последовательности необходимо выровнять.
Кстати, по виду выравнивания можно оценить, действительно ли последовательности такие, как надо: должно быть много консервативных колонок и мало хаотично расположенных гэпов. Помните: программа выравнивания выдаст результат даже для совершенно неродственных последовательностей, но смысла в этом результате не будет!
4. Большинство программ реконструируют неукоренённое дерево (даже если оно выглядит как укоренённое). Определение положения корня – отдельная задача.
5. Результат реконструкции – не абсолютная истина. Достоверность той или иной ветви можно оценить путём сравнения результатов разных программ и/или бутстрепа. Как правило, чем короче выравнивание, тем хуже качество реконструкции.