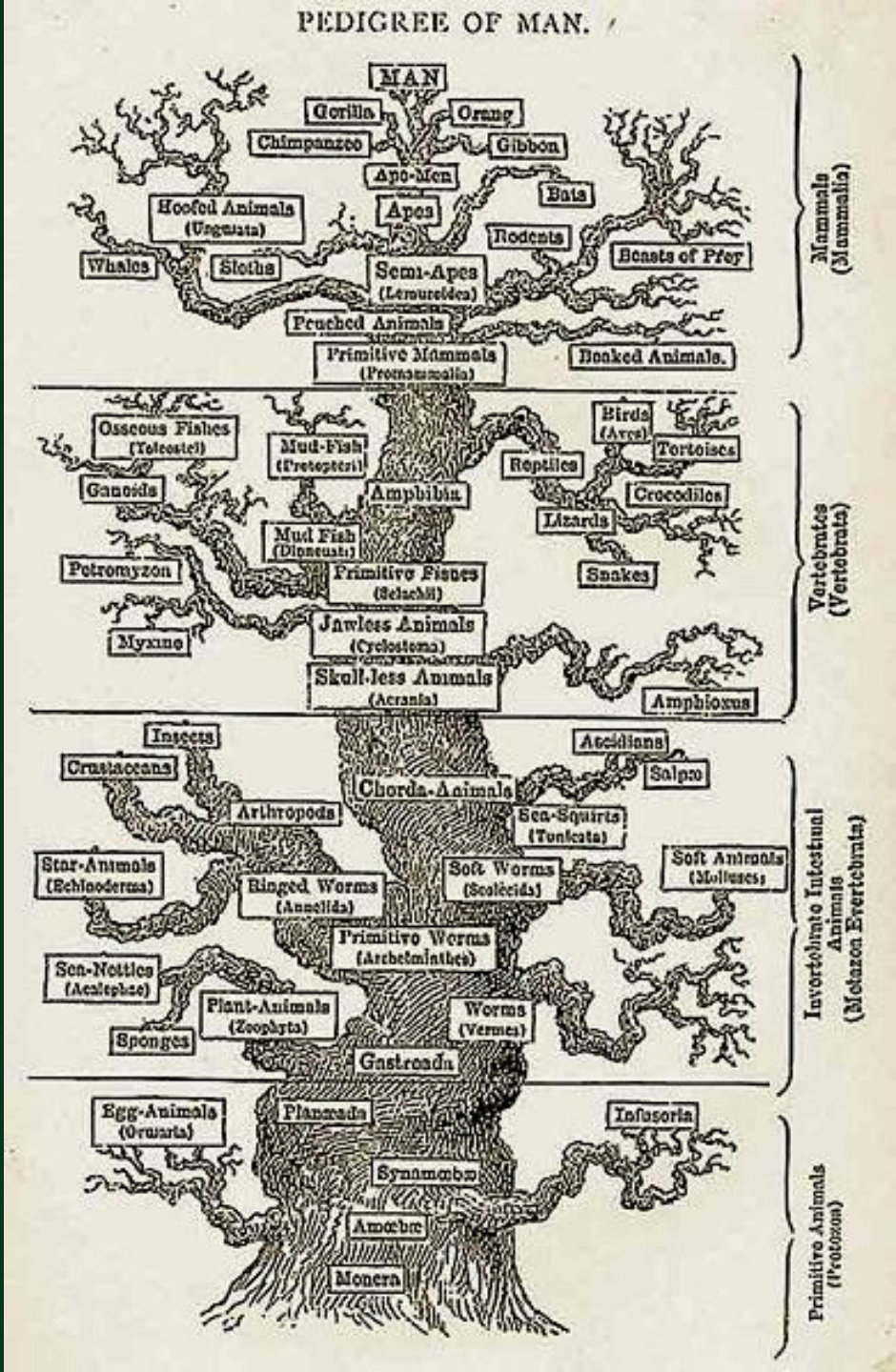
Филогенетические деревья

«…великое Дерево Жизни заполняет земную кору своими мертвыми и сломанными ветвями и покрывает поверхность вечно ветвящимися и прекрасными побегами»

Ч. Дарвин



ФББ, IV семестр, весна 2023

The time will come, I believe, though I shall not live to see it, when we shall have fairly true genealogical trees of each great kingdom of Nature.

Charles Darwin

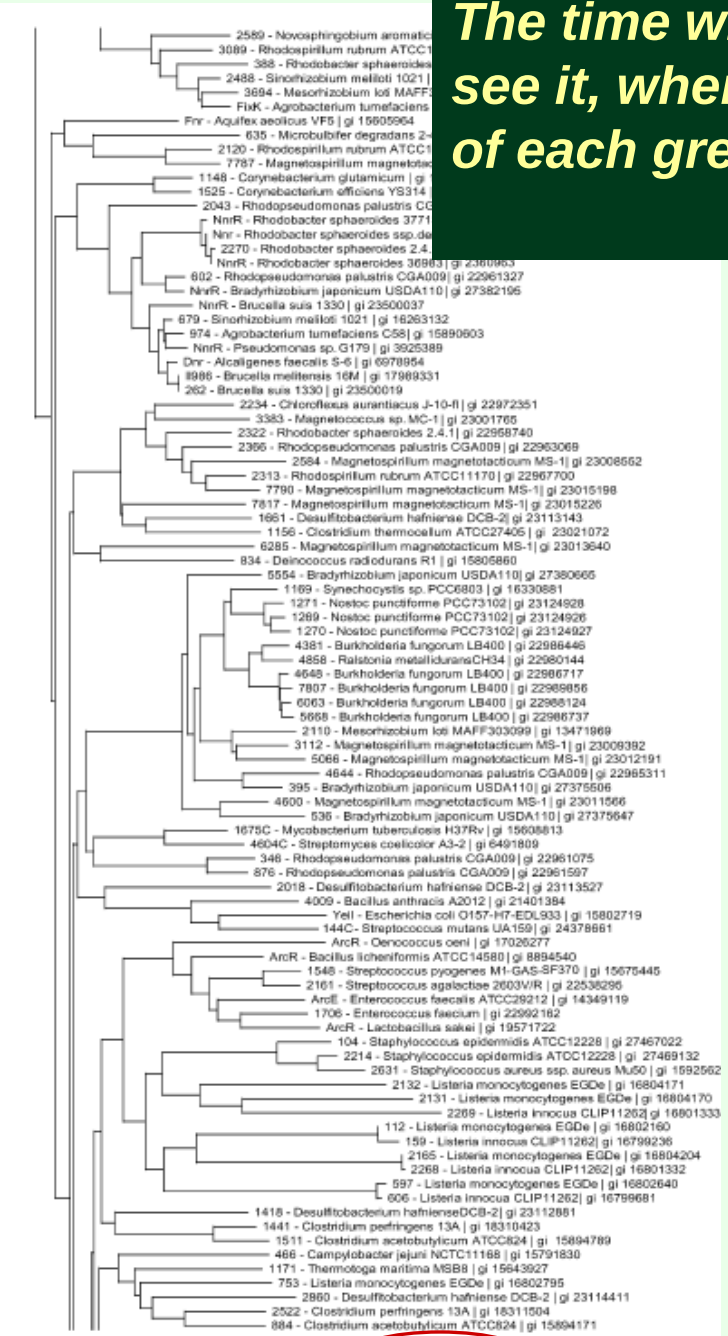
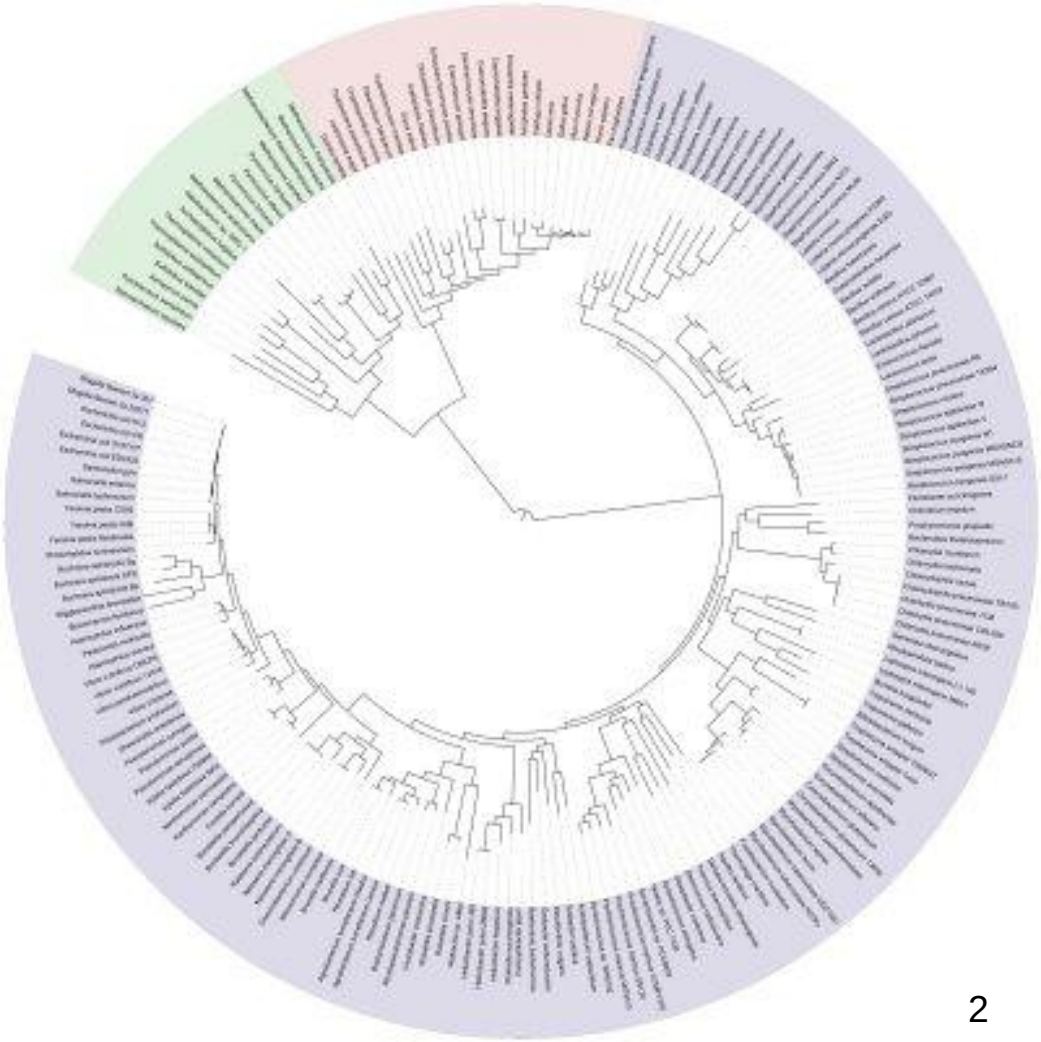


Fig. 1 (*Continued*).



http://itol.embl.de/

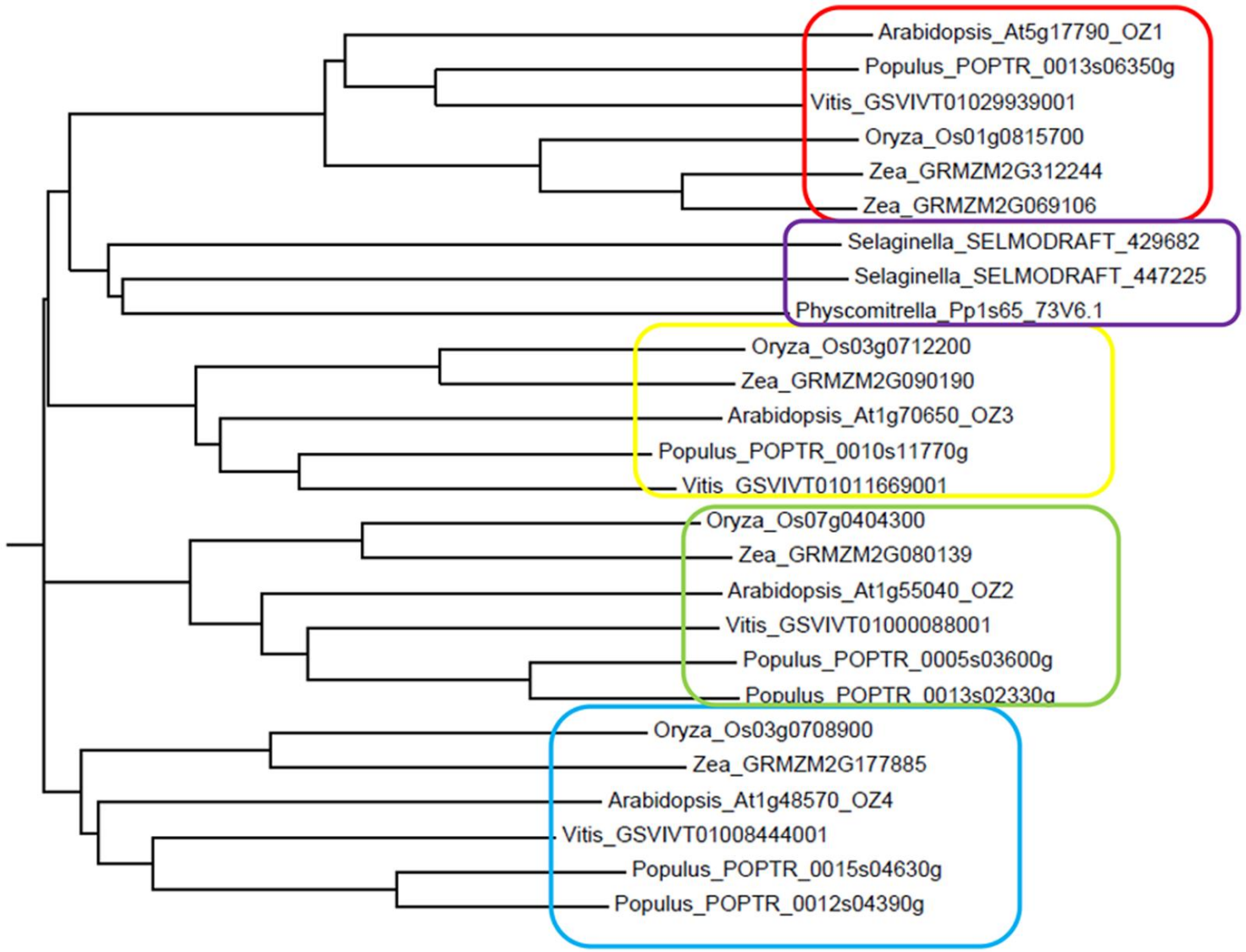
Эволюция геномов

Генотип определяет фенотип

Основой эволюции организмов и видов являются изменения геномов

- Наследственность
  - Геном потомка похож на геном предка
- Изменчивость
  - Геном потомка имеет отличия от генома предка:
    - ✓ комбинация генов родителей (у эукариот при половом размножении),
    - ✓ изменения генома клетки (у прокариот при горизонтальном переносе генов),
    - ✓ **мутации при репликации и репарации ДНК**
- Отбор
  - Разные генотипы (геномы) в популяции имеют разную распространенность среди потомков. Отбираются более приспособленные организмы.
  - Последовательности многих участков генома (в частности, почти все белок-кодирующие участки) испытывают действие отбора

Молекулярная филогения



Phylogenetic tree of the OZ family. Sun et al., 2014, PLOS Genet.

Кто строит деревья по последовательностям?

- Специалисты по систематике и филогенетике

находят ортологичные последовательности, строят деревья по ним, на основании этих деревьев судят о филогении организмов

- Исследователи, изучающие семейства белков со схожей функцией

дерево семейства в сочетании с деревом видов помогает судить о происхождении специализированных функций, предсказывать функцию и т. д.

- Эпидемиологи

исследуют происхождение штаммов вирусов и бактерий

- Исследователи злокачественных опухолей

важно понимать, как и когда возникают раковые клетки с опасными свойствами (например, способностью к метастазированию, устойчивостью к препаратам и т.п.)

- …

Востребованность

Упоминание некоторых филогенетических программ в аннотациях статей, вышедших после 1 января 2022 (по PubMed)

- RAxML 38 раз
- IQ-Tree 39 раз
- MrBayes 25 раз
- Neighbor-joining 204 раза
- UPGMA 108 раза
- Maximum parsimony 80 раз

Точечные замены в гене

… AATCCGTCAAGTCTA…

… Asn Pro Ser Ser Leu …

1) “молчащая”(синонимическая)мутация

… AATCCGTCGAGTCTA…

… Asn Pro Ser Ser Leu …

2) замена остатка на близкий по свойствам

… AATCCGACAAGTCTA…

… Asn Pro Thr Ser Leu …

3) замена остатка на остаток с иными свойствами

… AATCCGTCAAGACTA…

… Asn Pro Ser Arg Leu …

Мы видим лишь закрепившиеся мутации

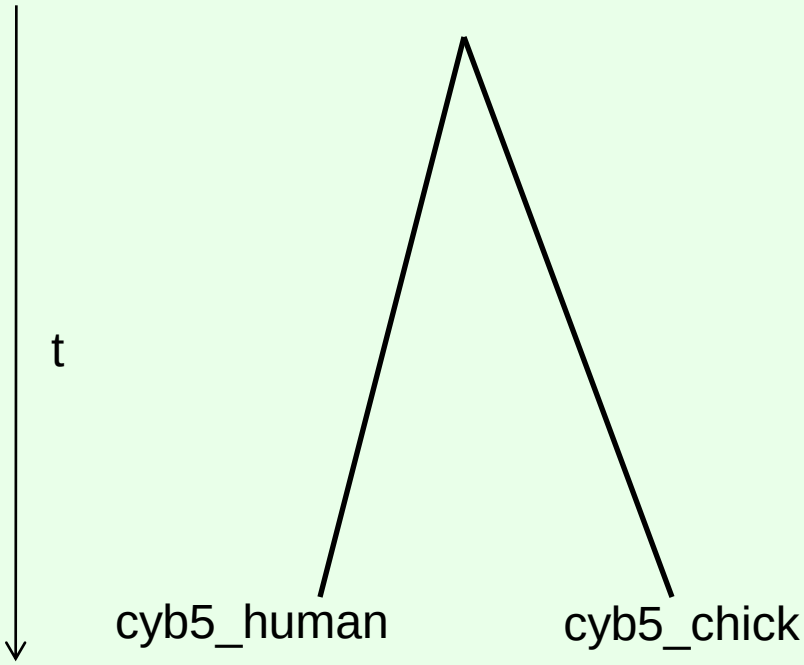
А шанс закрепиться есть лишь у безвредных мутаций…

```
CYB5_CHICK         1 MVGSSEAGGEAWRGRYYRLEEVQKHNNSQSTWIIVHHRIYDITKFLDEHP     50
                        .:|...||  .:||.|||:||||:|:|||:|:||::||:||||:|||
CYB5_HUMAN         1 ---MAEQSDEA--VKYYTLEEIQKHNHSKSTWLILHHKVYDLTKFLEEHP     45

CYB5_CHICK        51 GGEEVLREQAGGDATENFEDVGHSTDARALSETFIIGELHPDDRPKLQKP    100
                     ||||||||||||||||||||||||||.:|:|||||||||||||||.||
CYB5_HUMAN        46 GGEEVLREQAGGDATENFEDVGHSTDAREMSKTFIIGELHPDDRPKLNKP     95

CYB5_CHICK       101 AETLITTVQSNSSSWSNWVIPAIAAIIVALMYRSYMSE-    138
                     .||||||:.|:||.|:||||||:|:.||||||.||:|
CYB5_HUMAN        96 PETLITTIDSSSSWWTNWVIPAISAVAVALMYRLYMAED    134
```

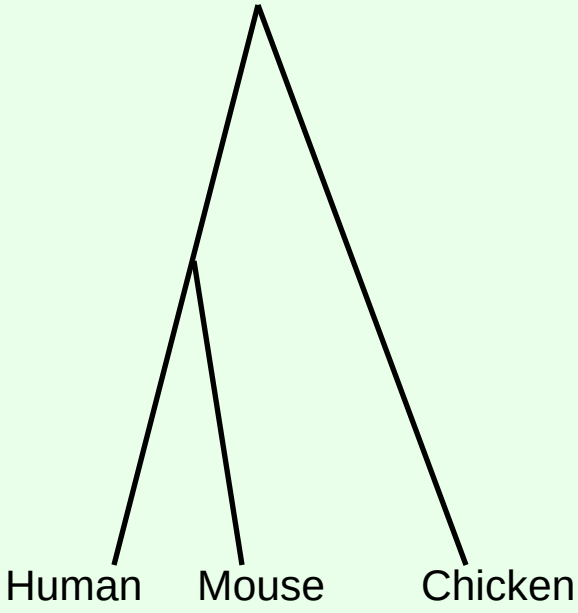
Путь эволюции

cyb5_human ——————————————————— cyb5_chick

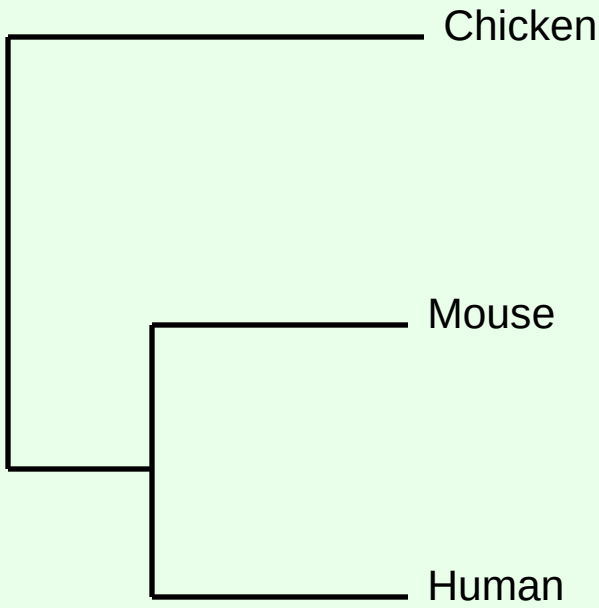
или



Филогенетическое дерево



Угловая форма



Прямоугольная форма

Эволюция видов и эволюция белков

Когда виды разделяются, то разделяются пути эволюции всех их белков…

В результате большинству белков одного вида соответствует **ортолог** в другом виде.

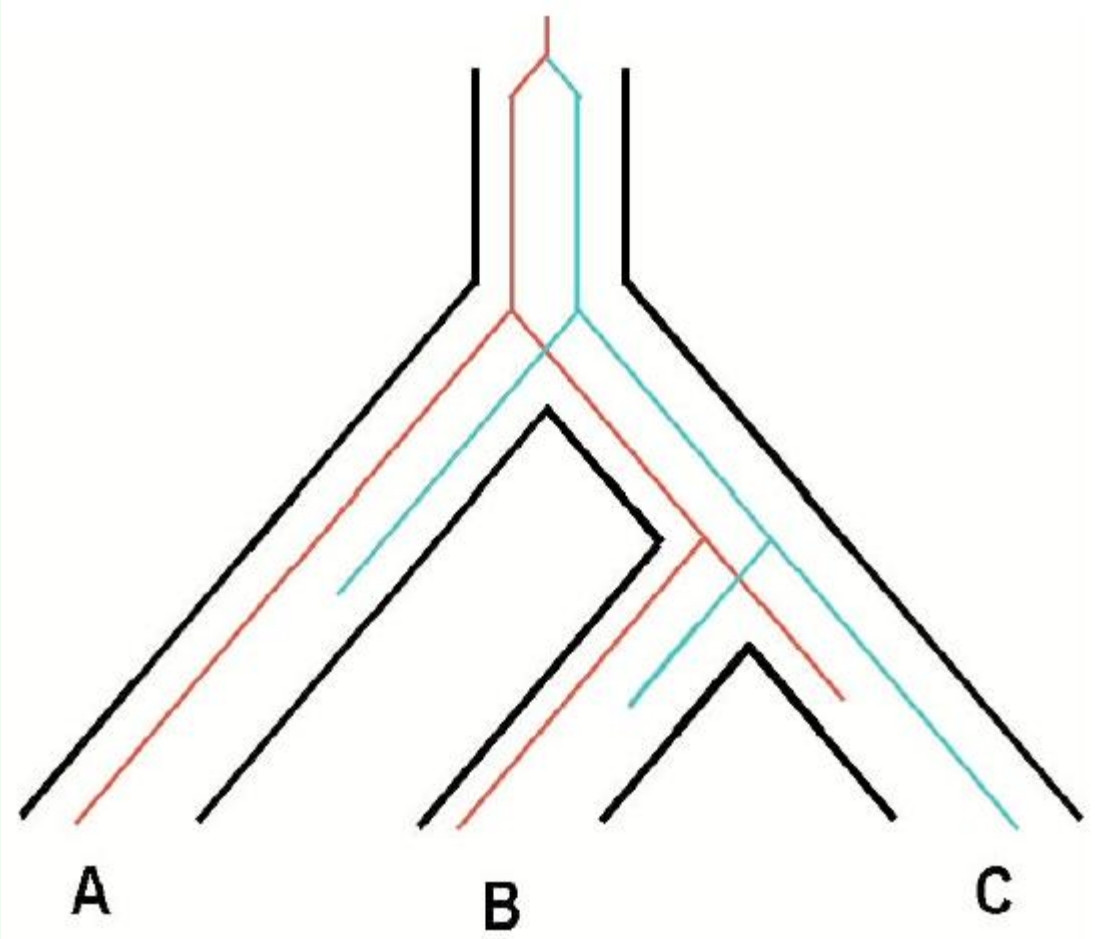
Но:

1) Бывают дупликации белков без разделения видов: два родственных белка существуют в одном геноме и эволюционируют (почти) независимо – такие белки называются **паралогами**.
2) Бывают потери генов.

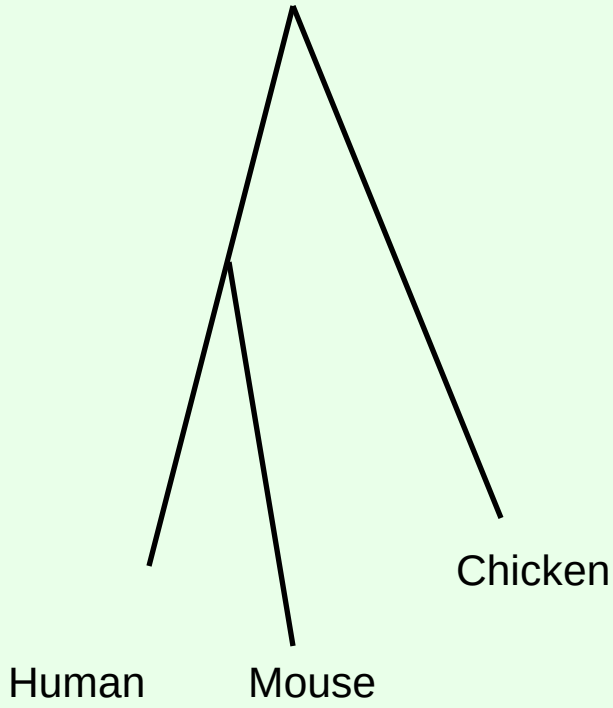
   Если в двух видах потерялись по одному белку из пары паралогов, то получается, что общий предок белков, которые выглядят как ортологи, «жил» существенно раньше, чем общий предок видов.
3) Бывает, что два белка объединяются в один многодоменный, и наоборот.

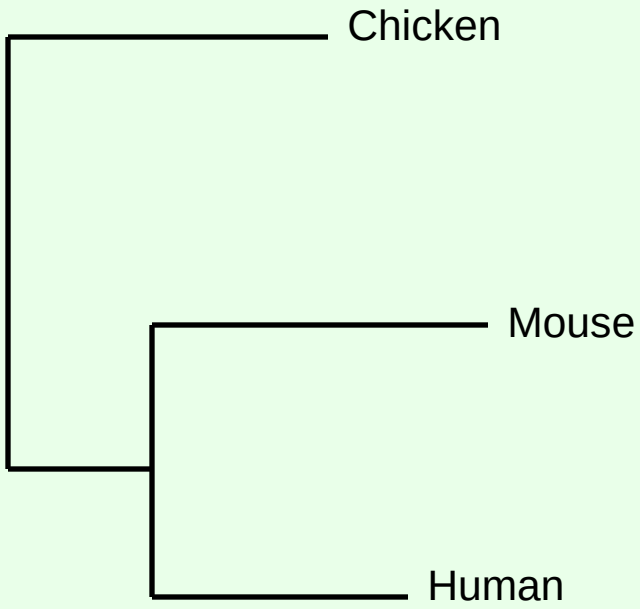
   Поэтому правильнее говорить об эволюции белковых доменов.

Дерево видов и дерево белков



«Молекулярные часы»: всегда идут, но иногда неточно

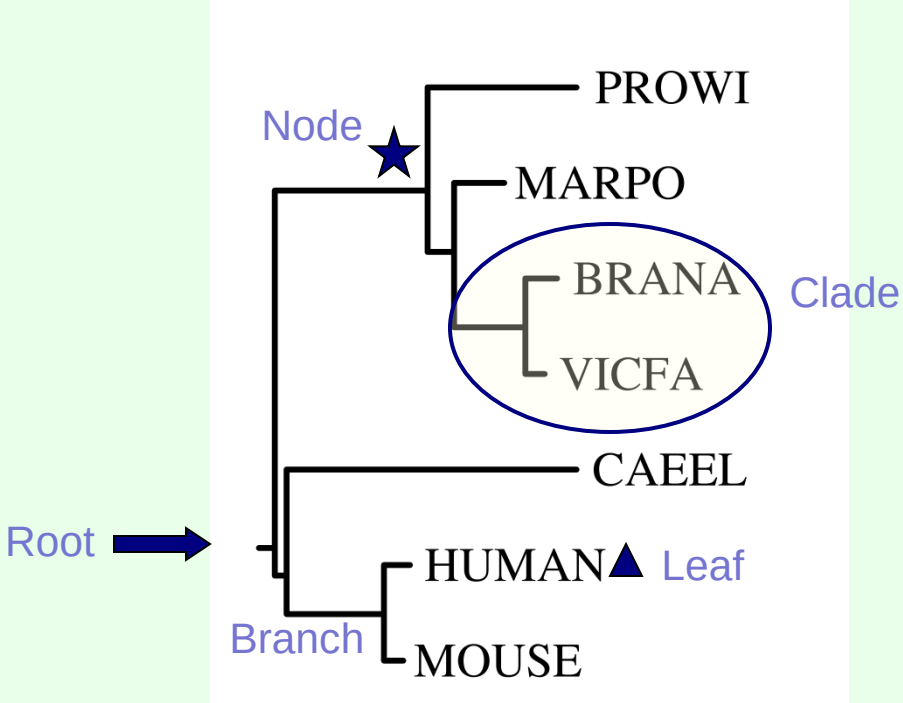




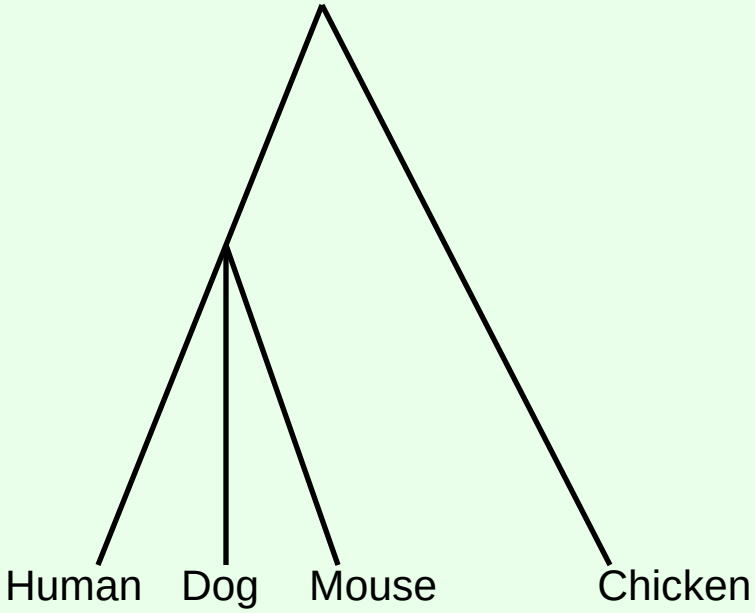
Когда хотят отразить разное число мутаций, произошедших на пути от общего предка, получается что-то вроде такого

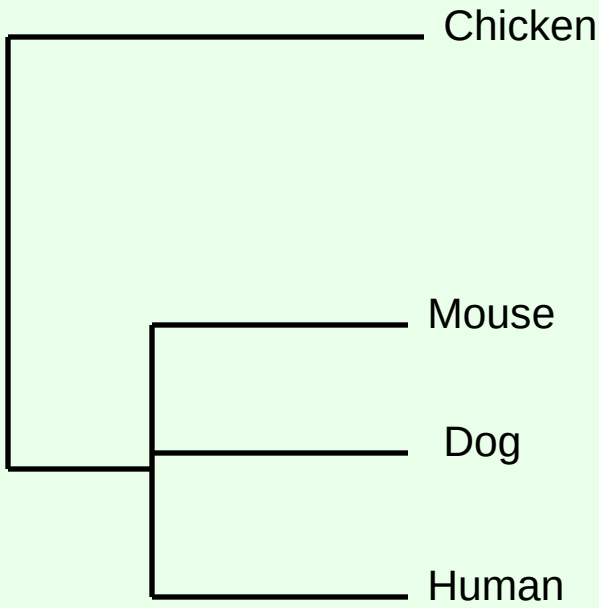
Описание структуры дерева (терминология)

- **Узел (node)** — точка разделения предковой последовательности (вида, популяции) на две независимо эволюционирующие. Соответствует внутренней вершине графа, изображающего эволюцию.
- **Лист (leaf)** — реальный (современный) объект; внешняя вершина графа.
- **Ветвь (branch)** — связь между узлами или между узлом и листом; ребро графа.
- **Корень (root)** — гипотетический общий предок.
- **Кла́да (clade)** — группа организмов, представляющая собой всех потомков некоторого общего предка.

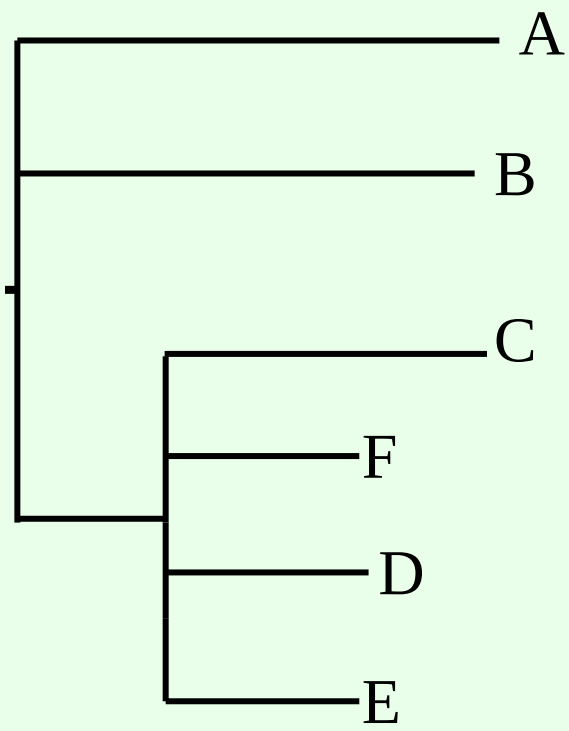


Небинарное дерево

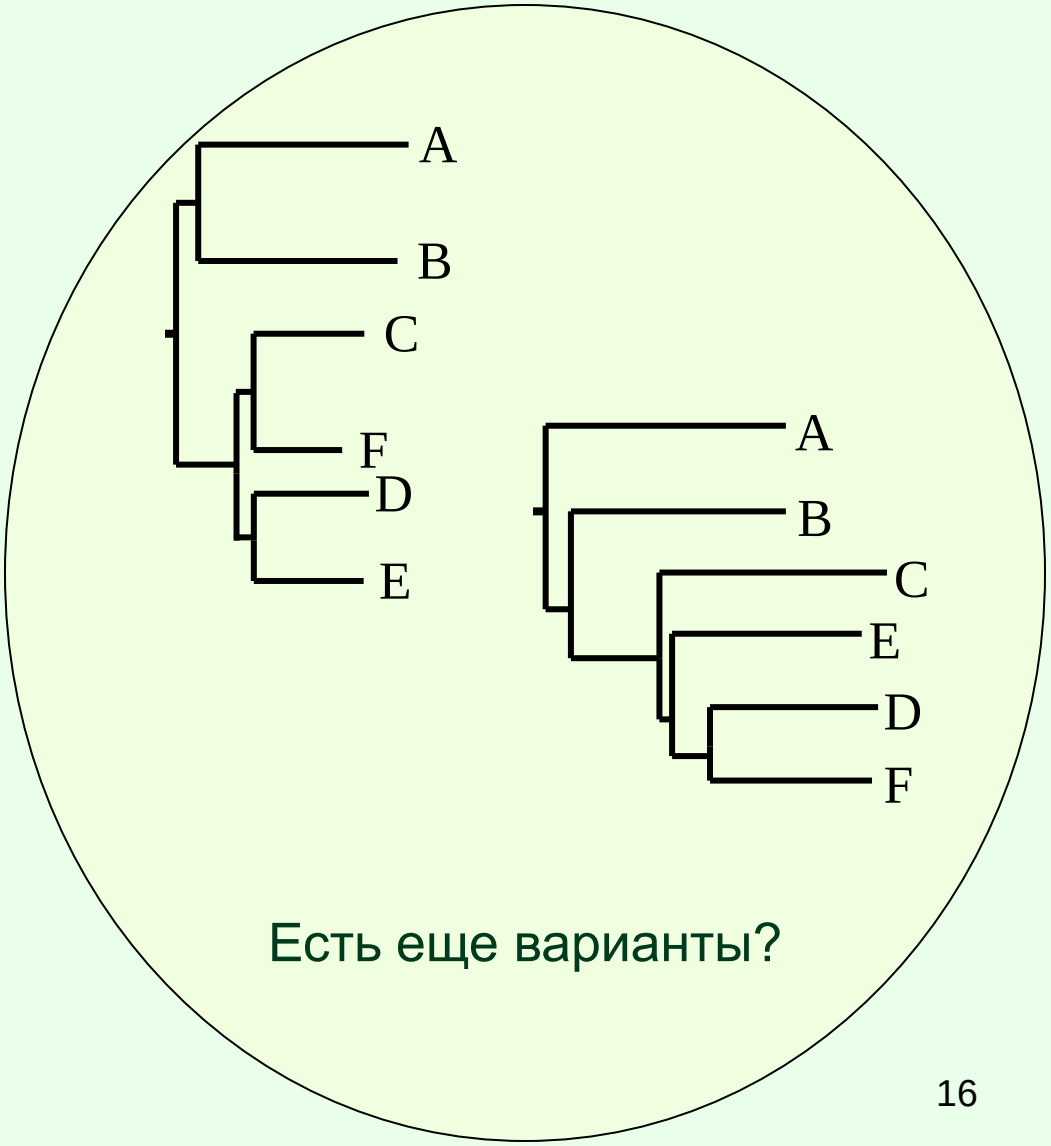




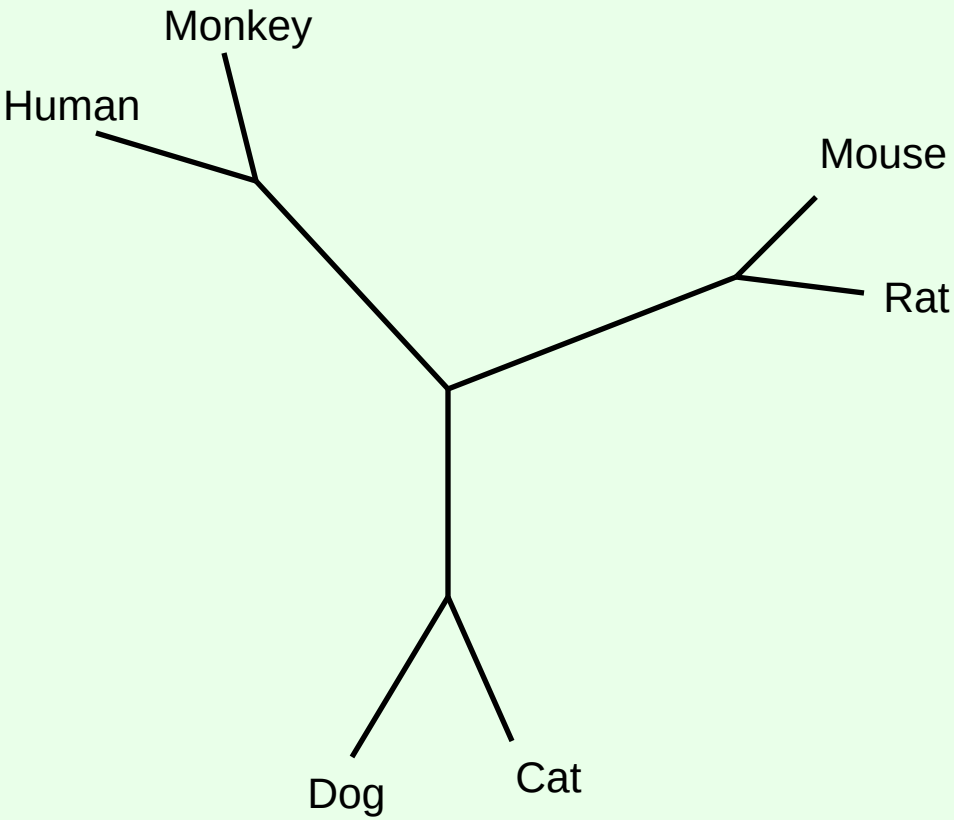
Небинарное дерево следует понимать как множество возможных «разрешений»



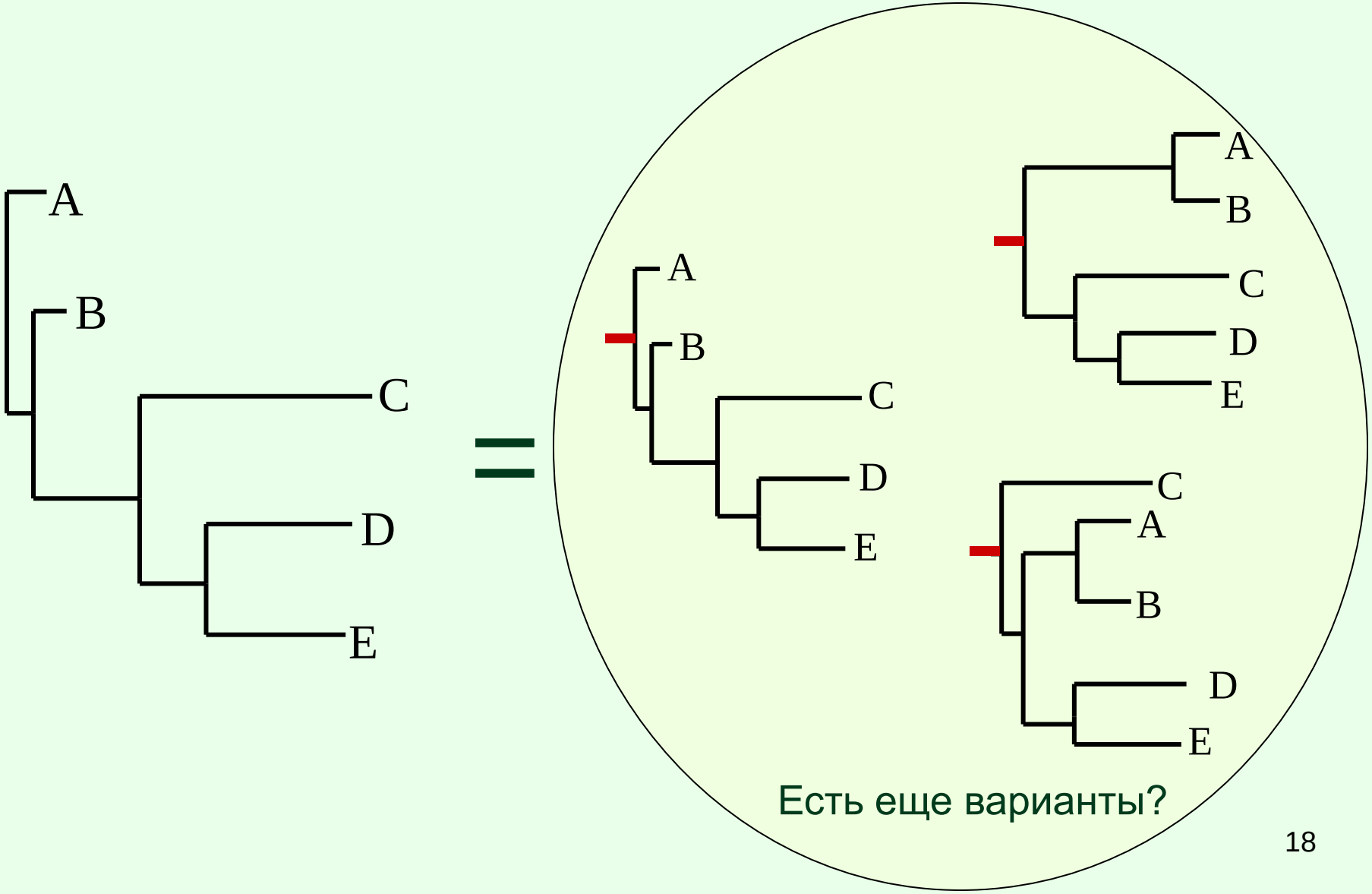
=



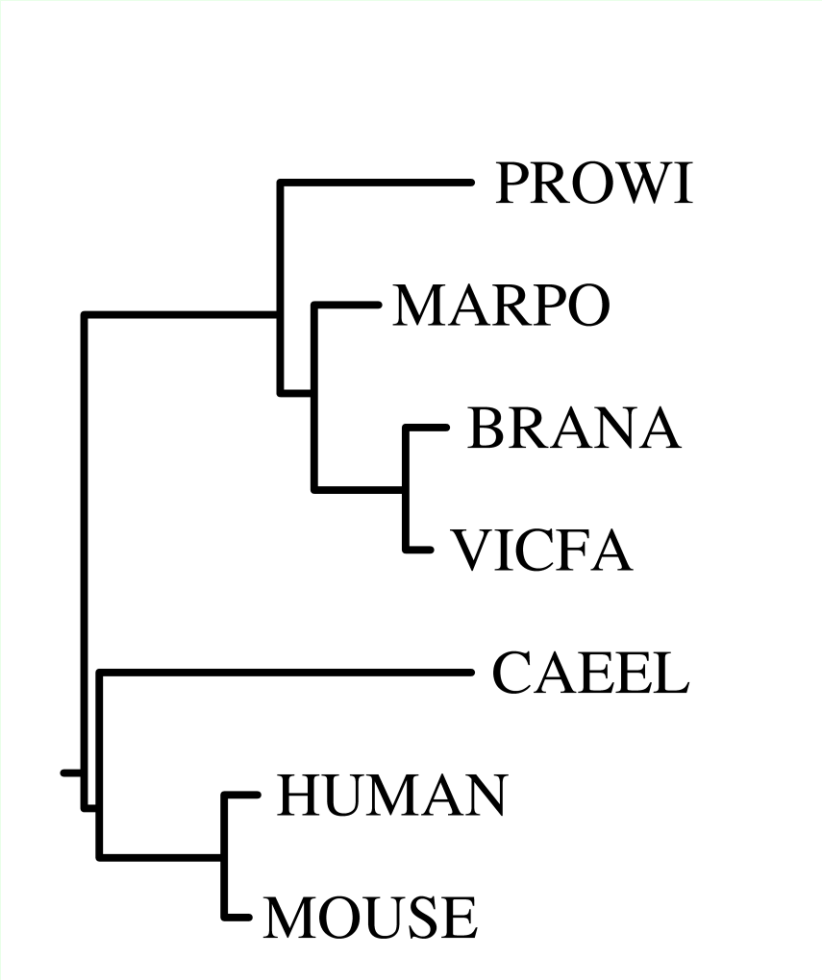
Неукоренённое дерево



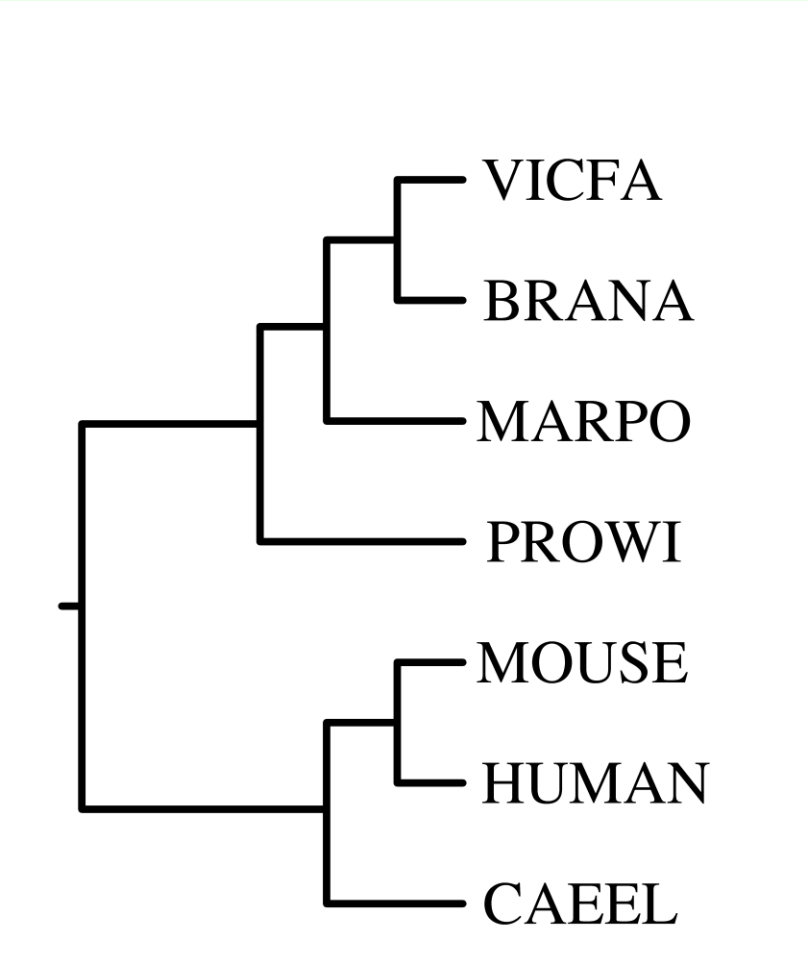
Неукоренённое дерево следует понимать как множество возможных укоренений



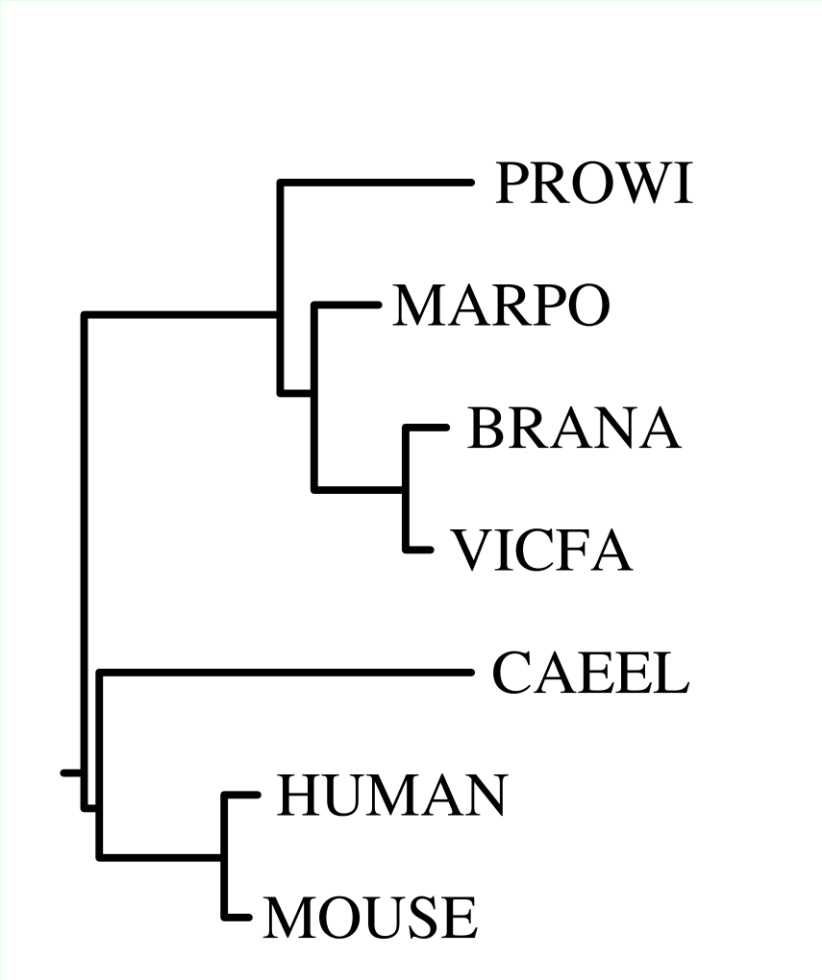
Топология дерева



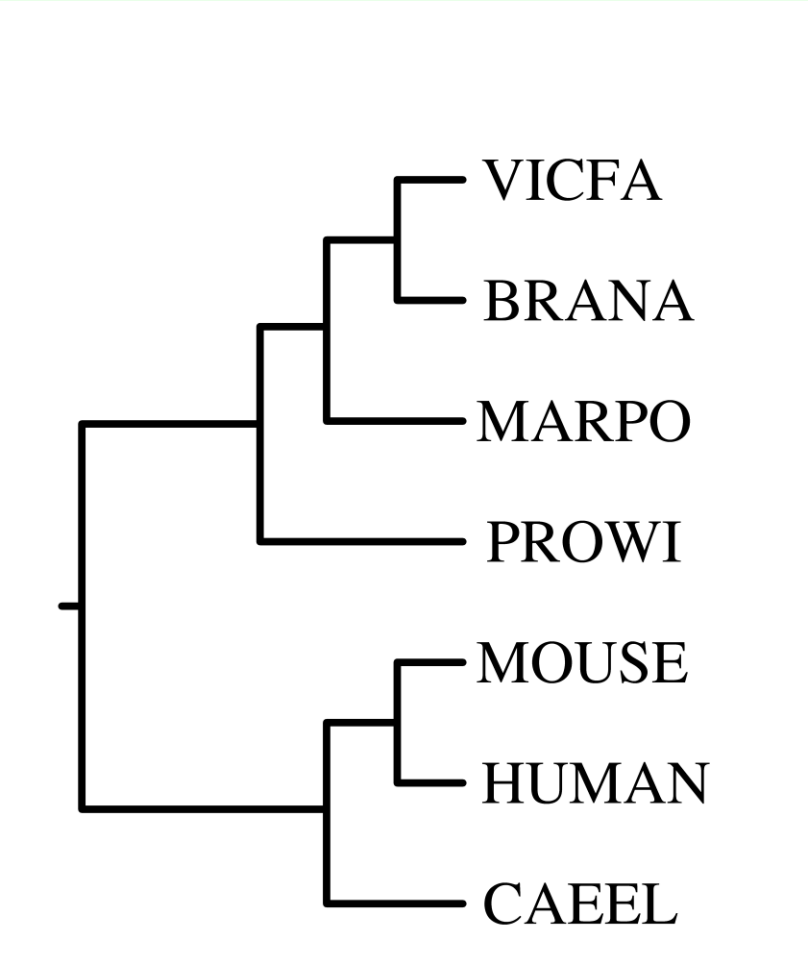
?



Топология дерева



=



Топология дерева

Каждая ветвь разбивает множество листьев на два.

В каждом дереве есть **тривиальные** ветви (отделяющие один лист от всех остальных), они не зависят от топологии.

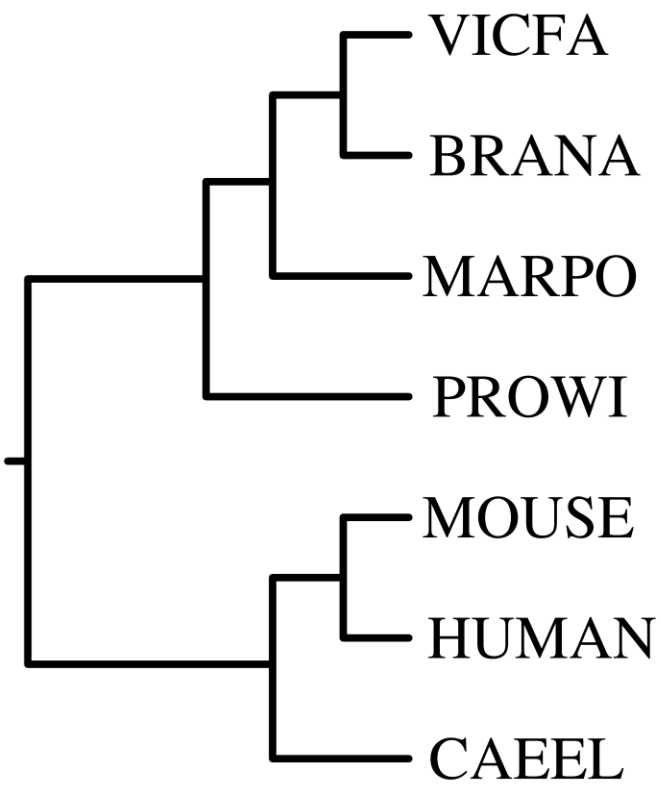
Топологию (неукоренённого) дерева можно однозначно записать набором нетривиальных разбиений. Например:

{HUMAN, MOUSE} vs {CAEEL,PROWI,MARPO,BRANA,VICFA}

{HUMAN, MOUSE,CAEEL} vs {PROWI,MARPO,BRANA,VICFA}

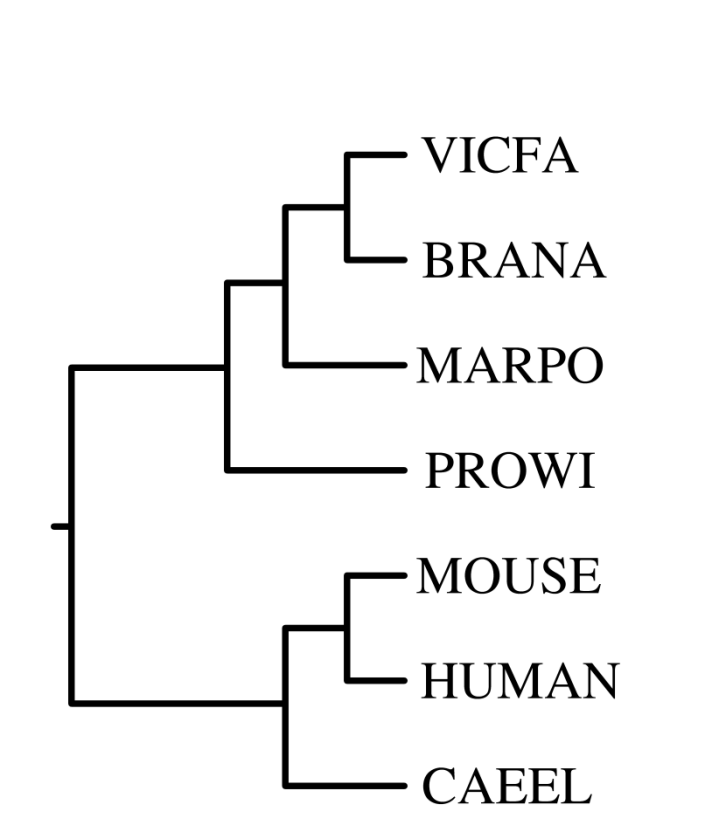
{HUMAN,MOUSE, CAEEL,PROWI} vs {MARPO,BRANA,VICFA}

{HUMAN,MOUSE,CAEEL,PROWI,MARPO} vs {BRANA,VICFA}



HUMAN	MOUSE	CAEEL	VICFA	BRANA	MARPO	PROWI
+	+	-	-	-	-	-
+	+	+	-	-	-	-
+	+	+	-	-	-	+
+	+	+	-	-	+	+

Скобочная формула



Каждая пара скобок, кроме внешней, отвечает узлу. Число после листа или закрывающей скобки означает длину ветви, идущей из данного листа или узла в вышестоящий узел или корень.

Чтобы записать скобочной формулой неукоренённое дерево, выбирают какое-нибудь его укоренение.

Newick Standard:

((((VICFA:3, BRANA:3):3, MARPO:6):2, PROWI:8):7, ((MOUSE:3, HUMAN:3):3, CAEEL:6):15);

«The reason for the name is that the second and final session of the committee met at Newick's restaurant in Dover, and we enjoyed the meal of lobsters.»

Joseph Felsenstein, http://evolution.genetics.washington.edu/phylip/newicktree.html

Программы визуализации деревьев

- MEGA https://www.megasoftware.net/
- FigTree http://tree.bio.ed.ac.uk/software/figtree/
- Archaeopteryx https://www.phylosoft.org/archaeopteryx/
- drawgram и drawtree из пакета PHYLIP https://evolution.genetics.washington.edu/phylip.html
- онлайн сервис iTOL https://itol.embl.de/upload.cgi
- см. также https://en.wikipedia.org/wiki/List_of_phylogenetic_tree_visualization_software