Филогенетические деревья и таксономия организмов

Сравнение деревьев

Реконструкция филогении (общая схема)

Расстояния между последовательностями

Филогенетические деревья и таксономия организмов

http://www.ncbi.nlm.nih.gov/taxonomy

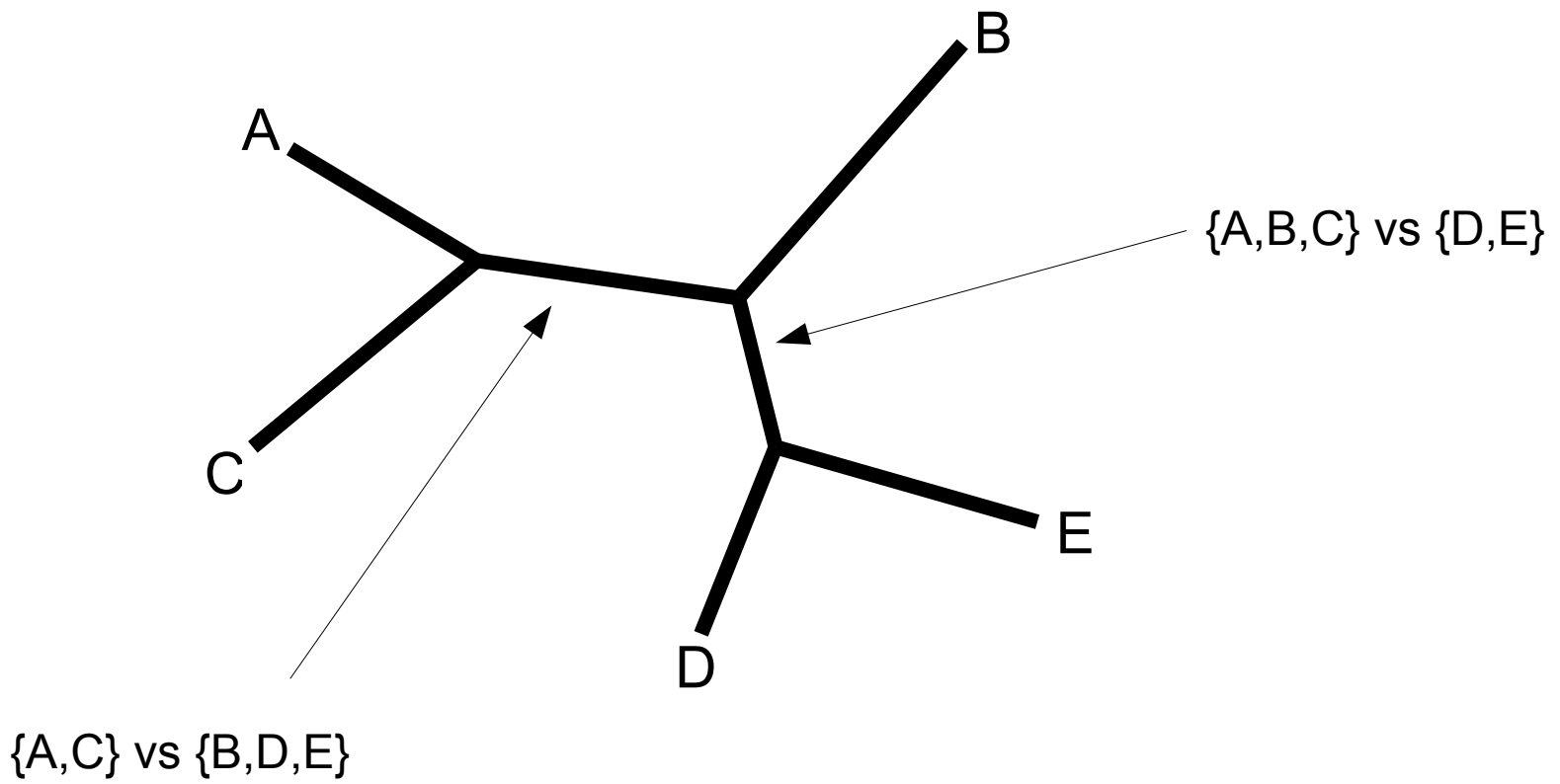
Сравнение деревьев

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

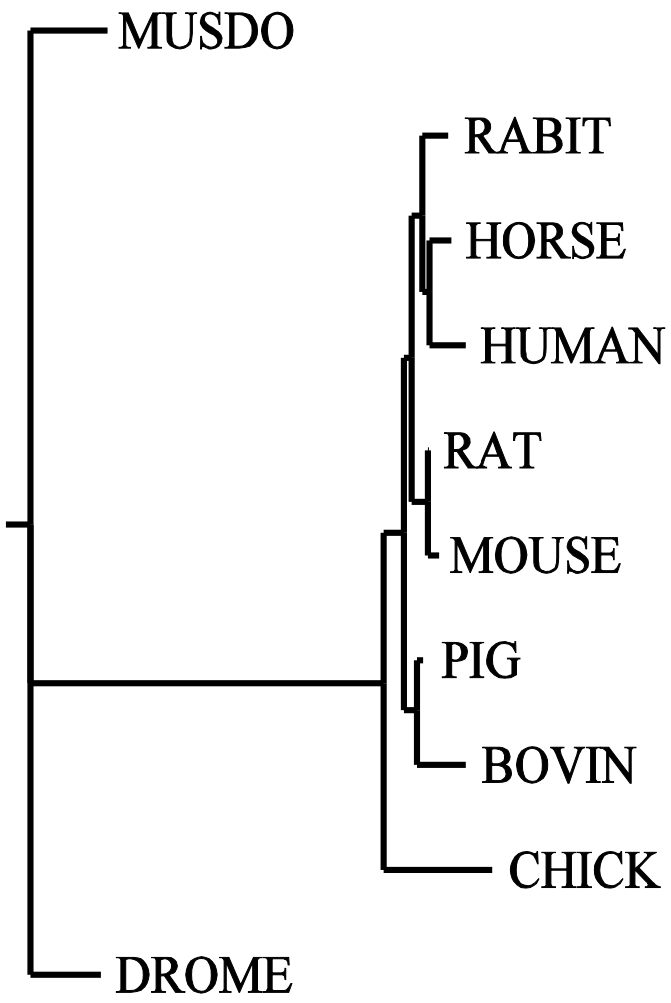
Поэтому (в частности) возможны различные варианты реконструкции по одним и тем же данным.

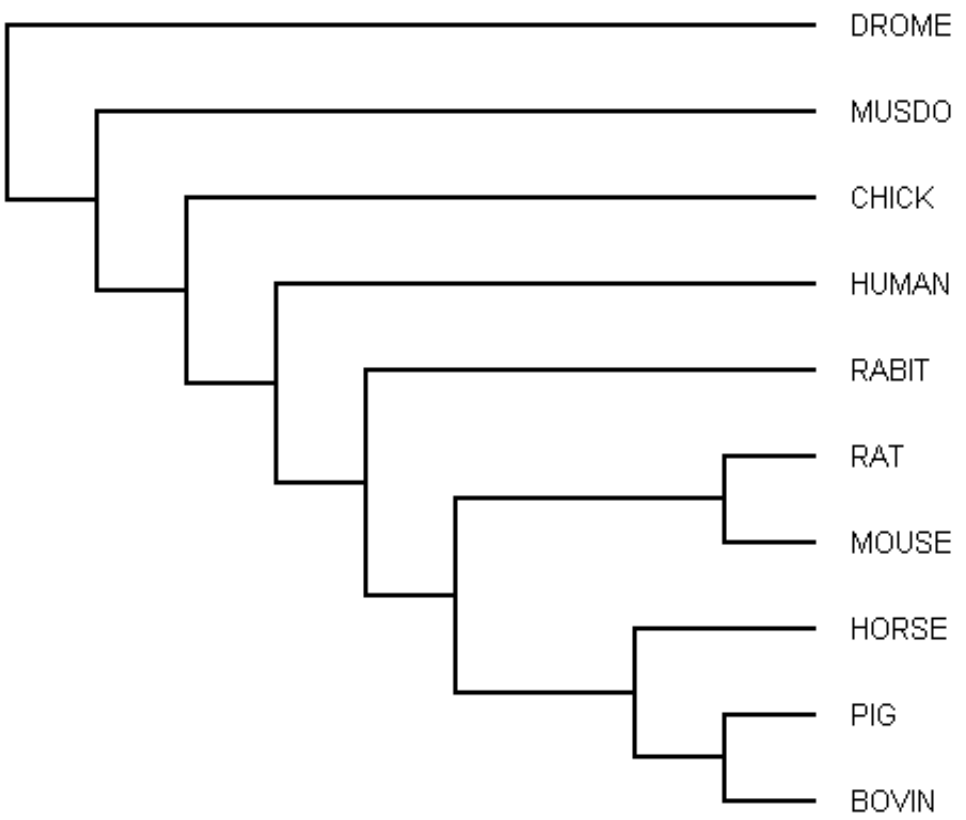
Встаёт задача сравнения различных деревьев с одним и тем же множеством листьев.

Напоминание: ветвь дерева как разбиение множества листьев

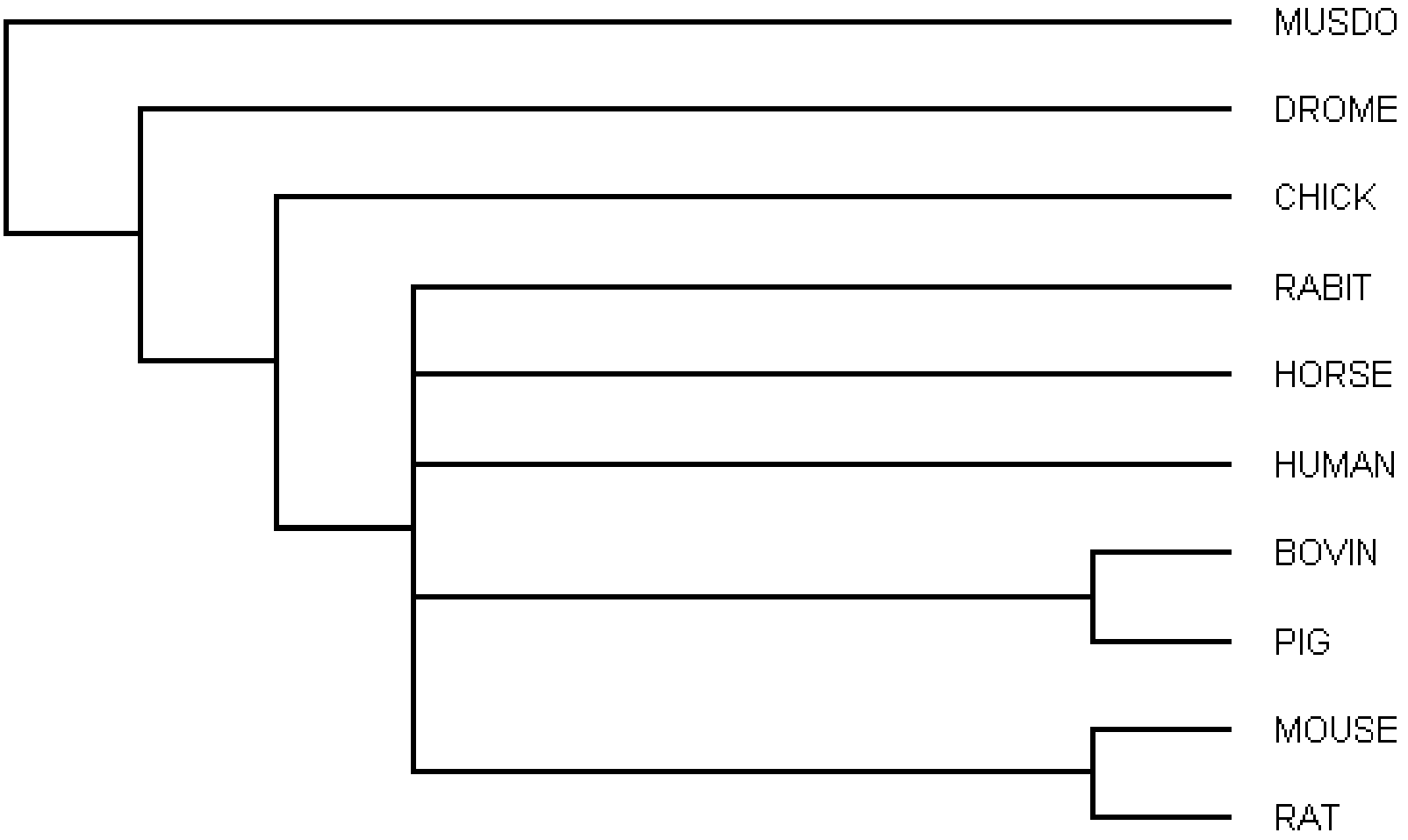


Что общего у этих двух деревьев?





Консенсусное дерево

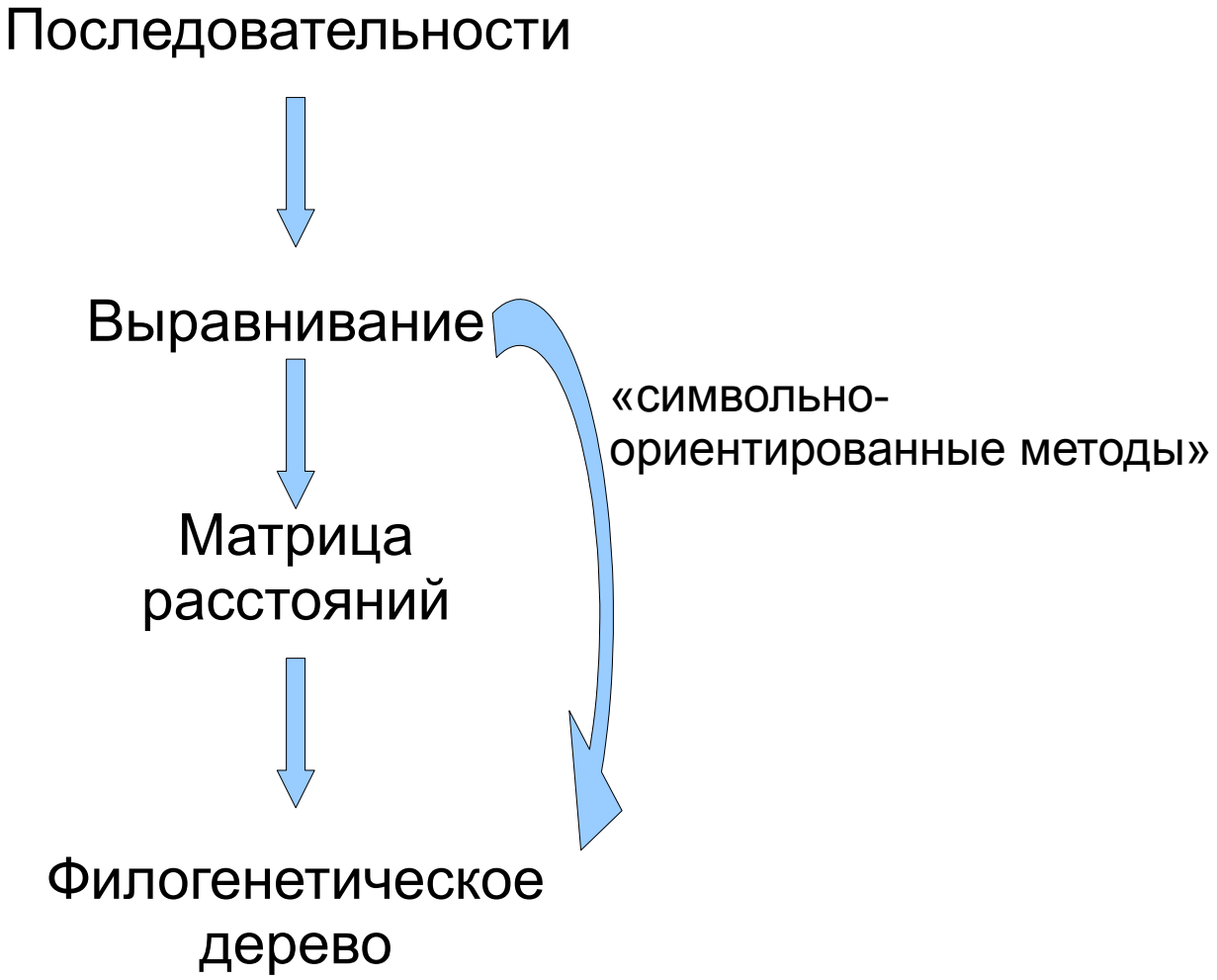


Дерево большинства (Majority-rule tree)

Строится по большому набору деревьев с одинаковым множеством листьев (например, деревья одного и того же набора бактерий, реконструированные по разным ортологическим рядам белков)

Включает только те ветви, которые встретились в большинстве деревьев исходного набора.

Схема реконструкции филогении по последовательностям



Матрица расстояний

	A	B	C	D
A	0	0.2	0.7	0.6
B	0.2	0	0.7	0.6
C	0.7	0.7	0	0.3
D	0.6	0.6	0.3	0

Множество объектов (последовательностей) превращается в **метрическое пространство**

Аксиомы метрического пространства:

1) d (A,A) = 0

2) d (A,B) > 0, если A $\neq$ B

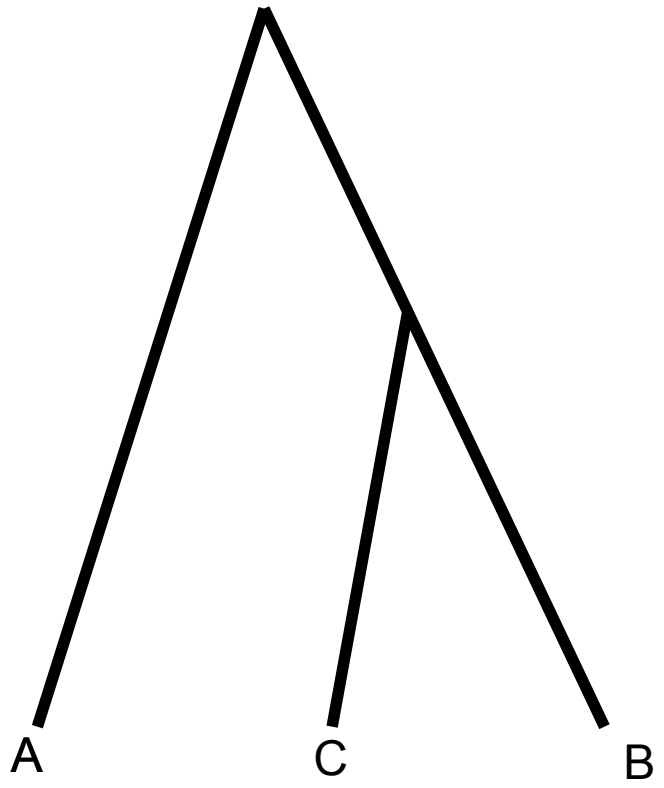
3) d (A,B) = d (B,A)

4) d (A,B) $\leq$ d (A,C) + d (B,C)

Если расстояние пропорционально эволюционному времени, то эти аксиомы выполняются.
Но верно и нечто большее:

4') d (A,B) $\leq$ max(d (A,C), d (B,C))
(«ультраметрическое пространство»)

Ультраметрическое расстояние



Если $d(A,B) > d(B,C)$, то $d(A,C) = d(A,B)$

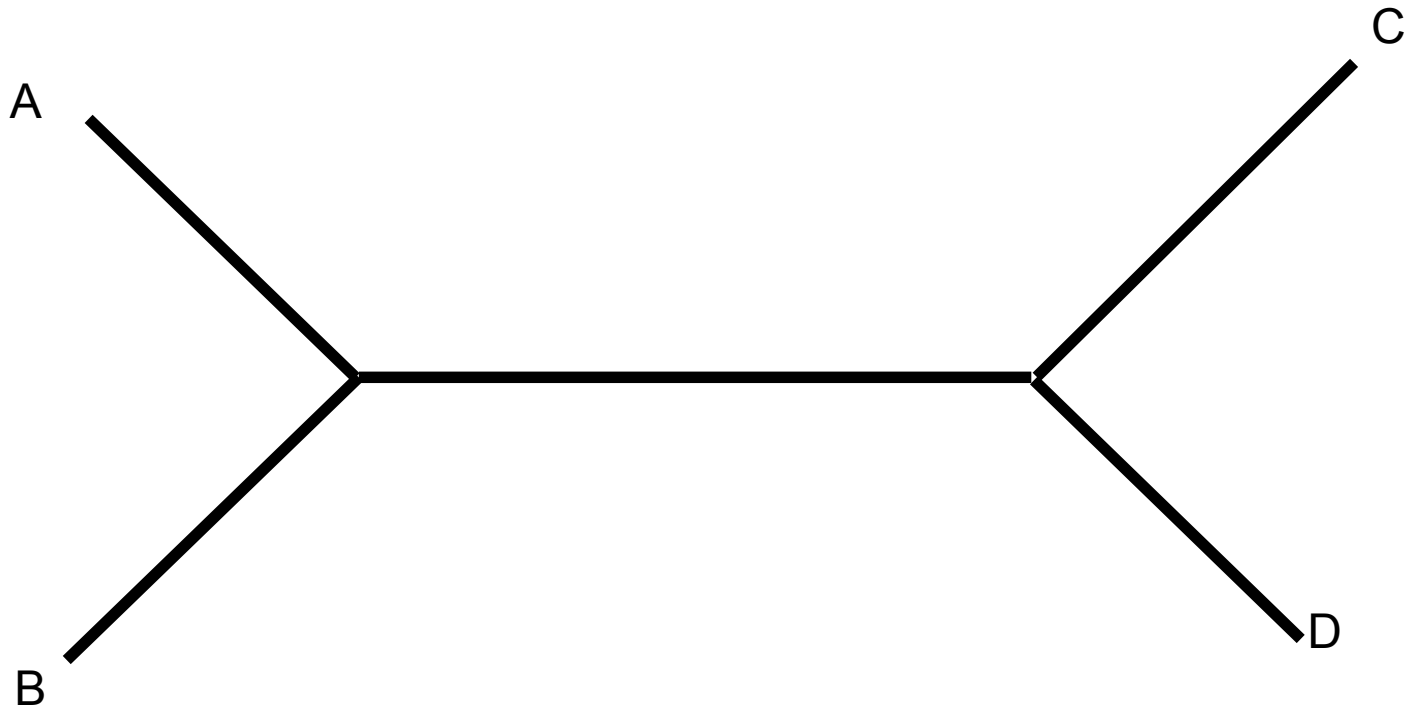
Или: из трёх расстояний между тремя объектами два всегда равны между собой и не меньше третьего
(это равносильно аксиоме ультраметричности)

Расстояние как число мутаций

Расстояние между последовательностями ультраметрично, если его понимать как эволюционное время...

Но если неверно предположение о «молекулярных часах», то больше информации несёт понимание расстояния как числа произошедших мутаций. **Такое расстояние не обязательно ультраметрично.**

Аддитивность: если есть четыре последовательности A,B,C,D, то из трёх сумм 1) $d(A,B) + d(C,D)$ 2) $d(A,C) + d(B,D)$ 3) $d(A,D) + d(B,C)$
две равны между собой и больше третьей.



Как оценить расстояние между последовательностями

По аддитивному набору расстояний дерево (с длинами веток) восстанавливается однозначно!

Но в реальности нам даны последовательности и требуется оценить число произошедших мутаций. Это не так просто, поскольку мутации могут происходить в одной и той же позиции.

Всё же простейшая оценка расстояния есть число различий, делённое на длину последовательности.
Более изощрённые методы учитывают тот факт, что чем больше наблюдаемое различие между последовательностями, тем больше можно ожидать повторных и возвратных мутаций в одинаковых позициях.

Программа protdist (fprotdist): оценка расстояния по методу наибольшего правдоподобия.

То, что получается, как правило не обладает (в точности) свойством аддиттивности!