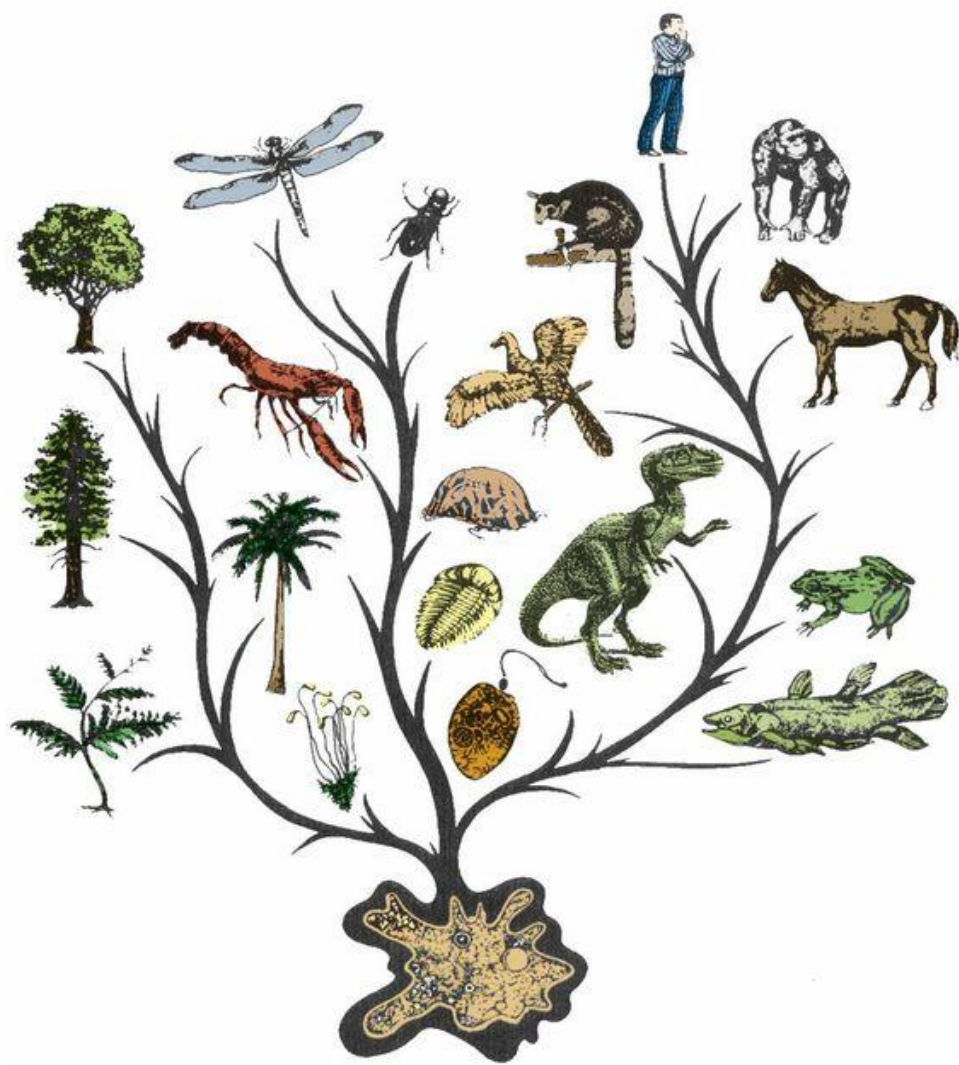
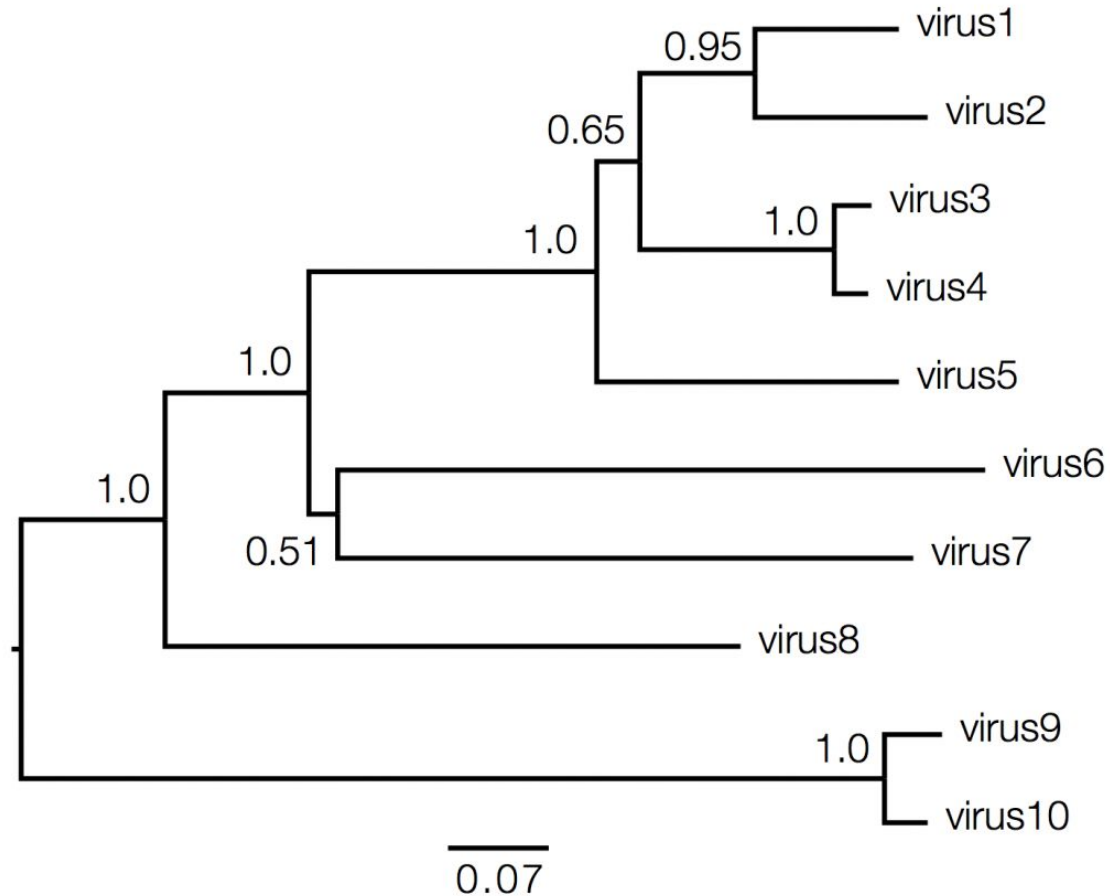


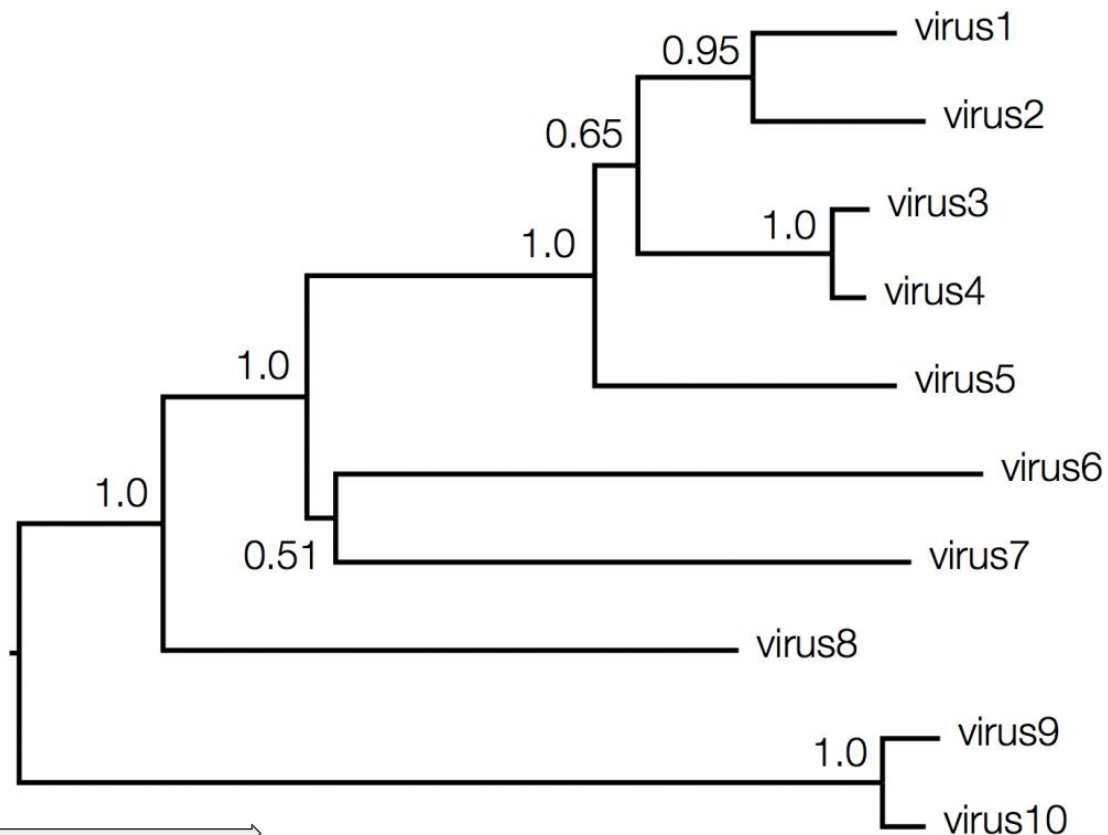
Филогенетические деревья



Пример



Пример

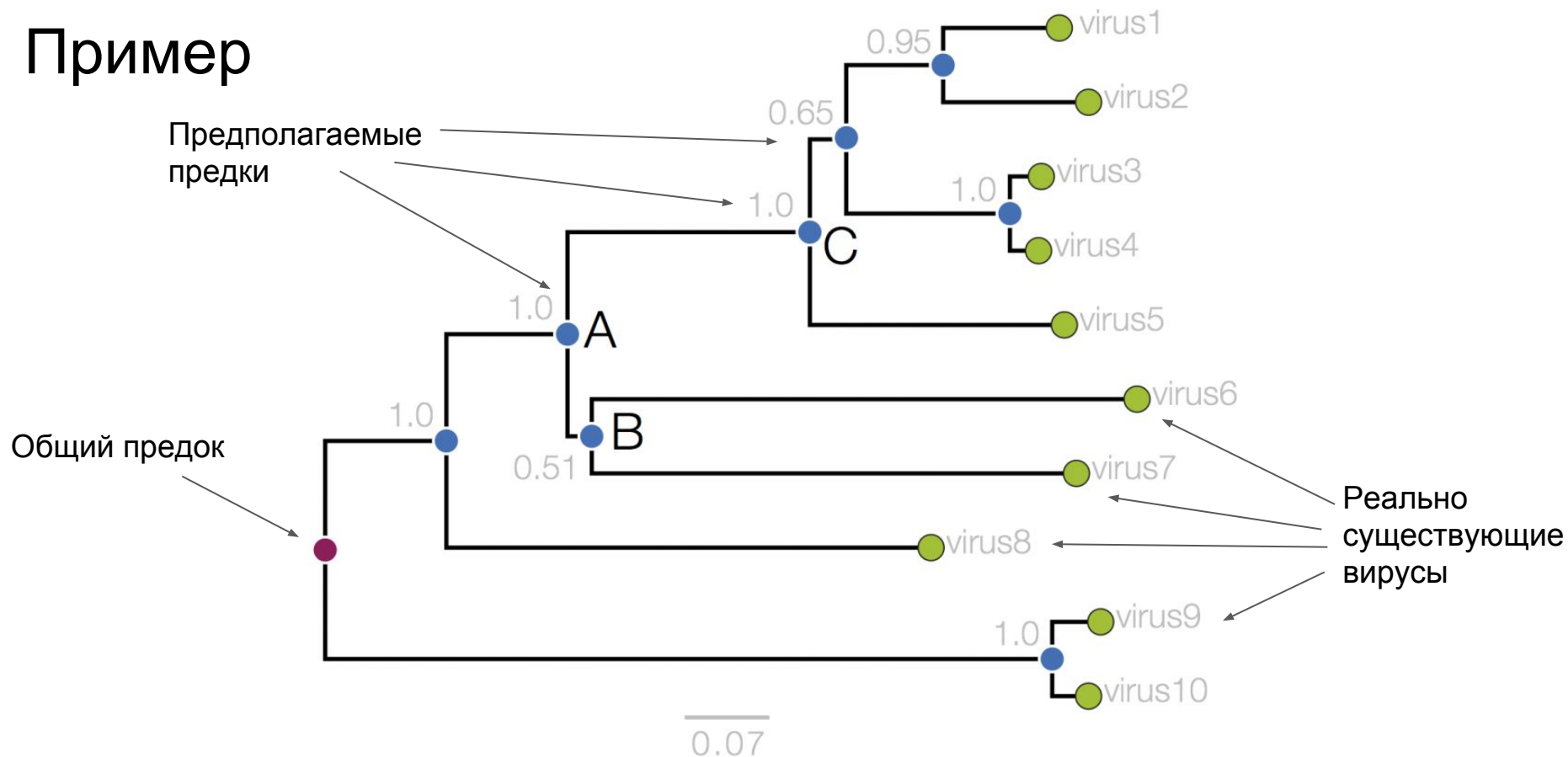


эволюционное расстояние

(расстояние между последовательностями)

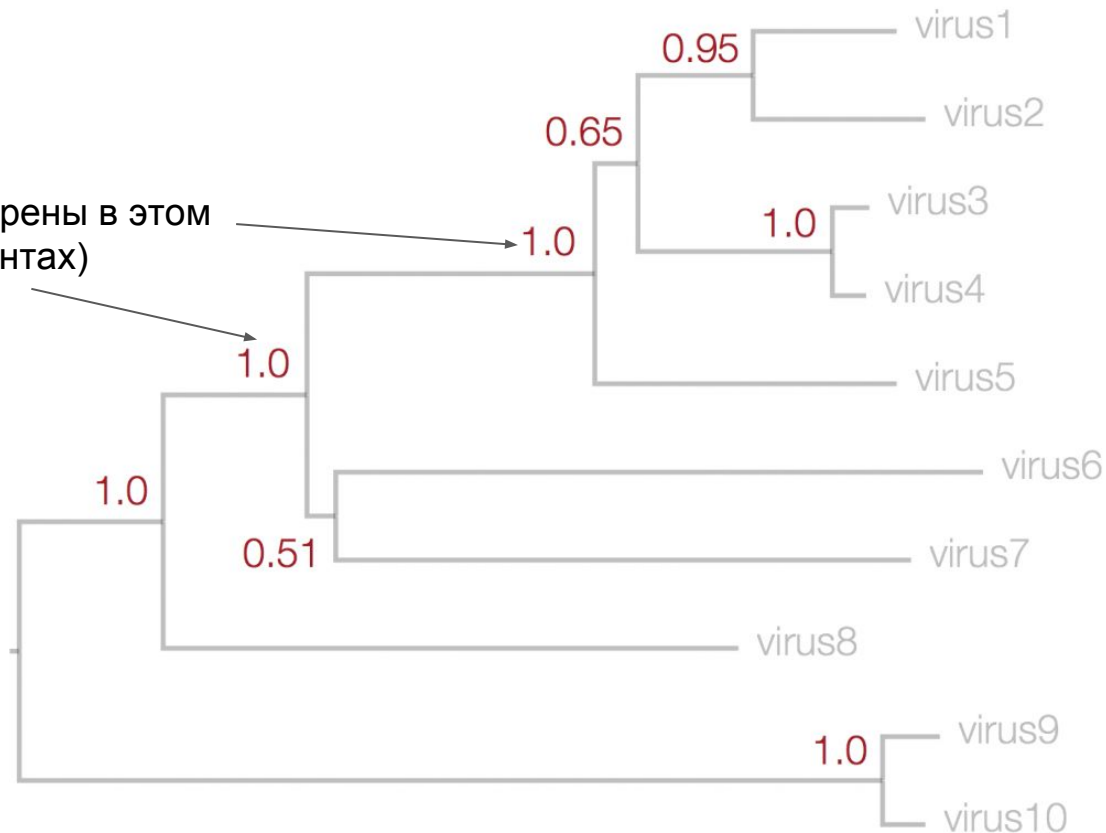
0.07

Пример



Пример

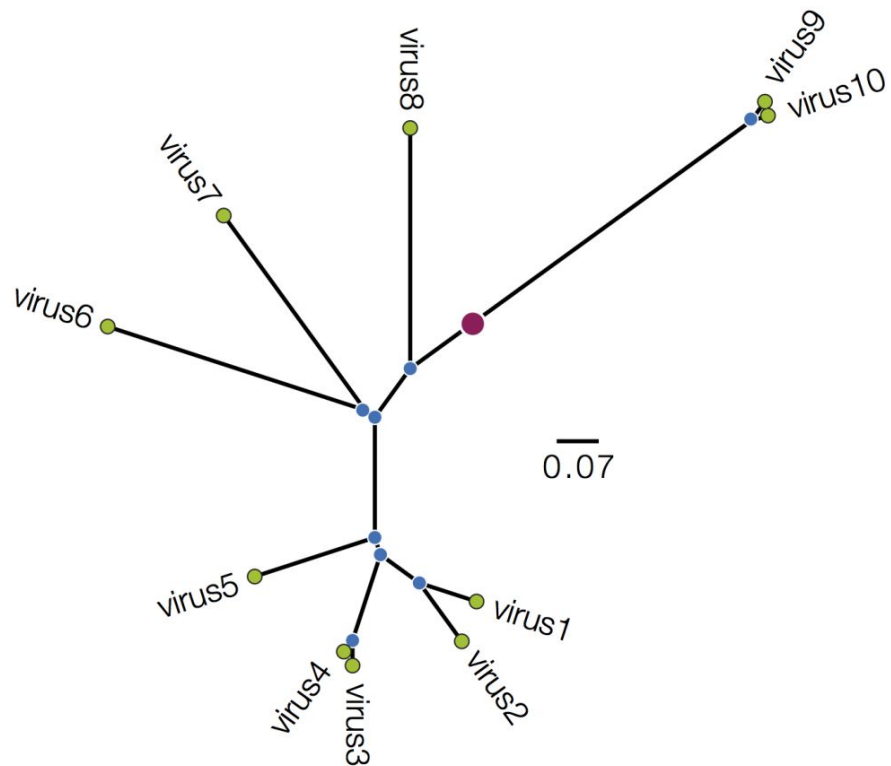
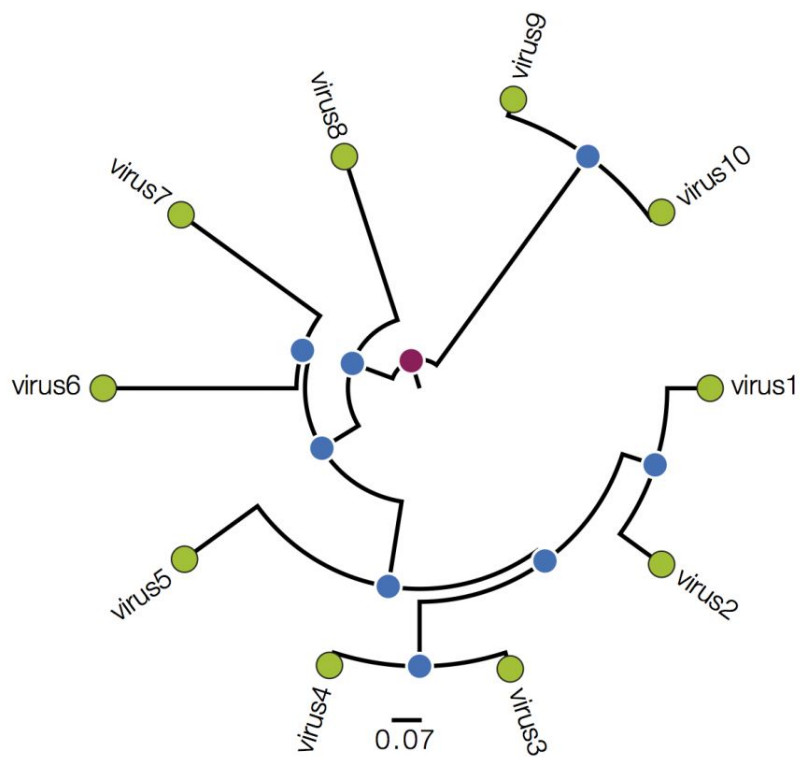
Насколько мы уверены в этом кластере (в процентах)



Масштаб по горизонтали

0.07

Пример

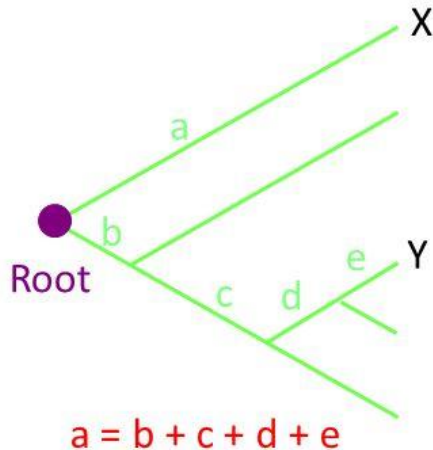


Гипотеза молекулярных часов

мутации во всех ветвях дерева накапливаются с одинаковой скоростью => общий предок будет равноудален от каждого из листьев

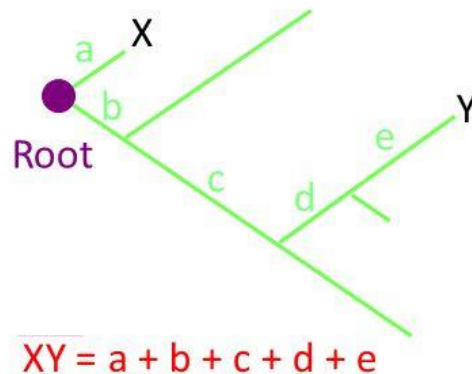
Ultrametricity

All tips are an equal distance from the root.



Additivity

Distance between any two tips equals the total branch length between them.



WPGMA

- weighted pair group method with arithmetic mean
- строит укорененное ультраметрическое дерево
- жадно выбирает две ближайшие вершины и объединяет их в новый кластер
- новое расстояние - среднее арифметическое двух старых

$$d_{(i \cup j), k} = \frac{d_{i, k} + d_{j, k}}{2}$$

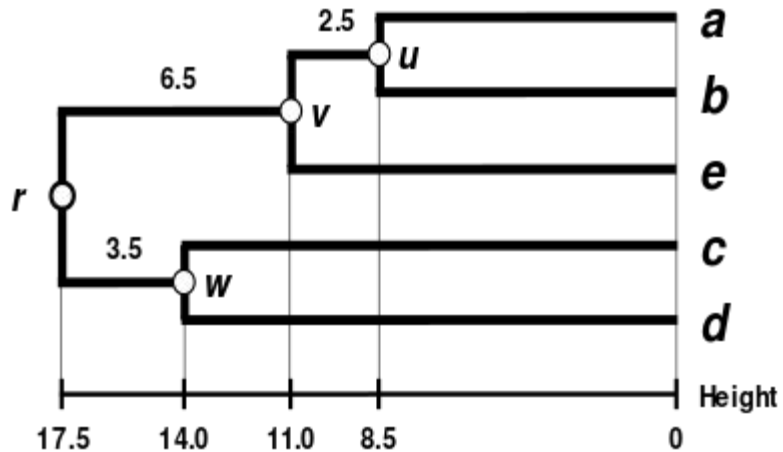
WPGMA

	a	b	c	d	e
a	0	17	21	31	23
b	17	0	30	34	21
c	21	30	0	28	39
d	31	34	28	0	43
e	23	21	39	43	0

	(a,b)	c	d	e
(a,b)	0	25.5	32.5	22
c	25.5	0	28	39
d	32.5	28	0	43
e	22	39	43	0

	$((a,b),e)$	c	d
$((a,b),e)$	0	32.25	37.75
c	32.25	0	28
d	37.75	28	0

	$((a,b),e)$	(c,d)
$((a,b),e)$	0	35
(c,d)	35	0



UPGMA

- unweighted pair group method with arithmetic mean
- строит укорененное ультраметрическое дерево
- жадно выбирает две ближайшие вершины и объединяет их в новый кластер
- новое расстояние учитывает размер объединяемых кластеров

$$d_{(A \cup B),k} = \frac{|A| \cdot d_{A,k} + |B| \cdot d_{B,k}}{|A| + |B|}$$

UPGMA

	a	b	c	d	e
a	0	17	21	31	23
b	17	0	30	34	21
c	21	30	0	28	39
d	31	34	28	0	43
e	23	21	39	43	0



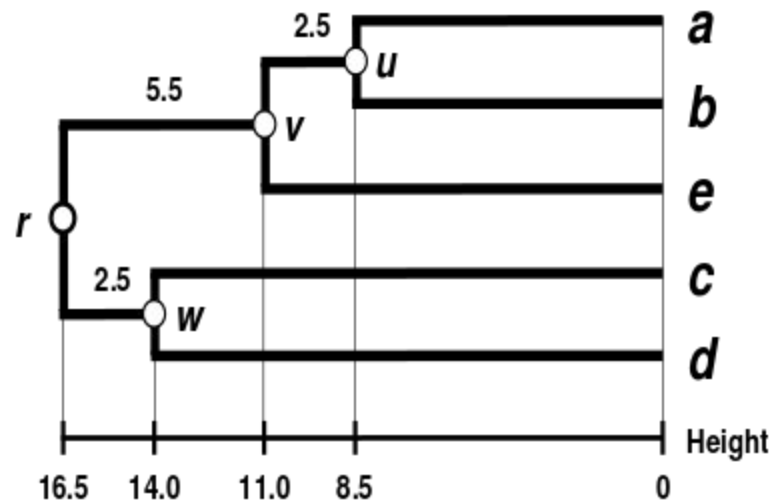
	(a,b)	c	d	e
(a,b)	0	25.5	32.5	22
c	25.5	0	28	39
d	32.5	28	0	43
e	22	39	43	0



	((a,b),e)	c	d
((a,b),e)	0	30	36
c	30	0	28
d	36	28	0

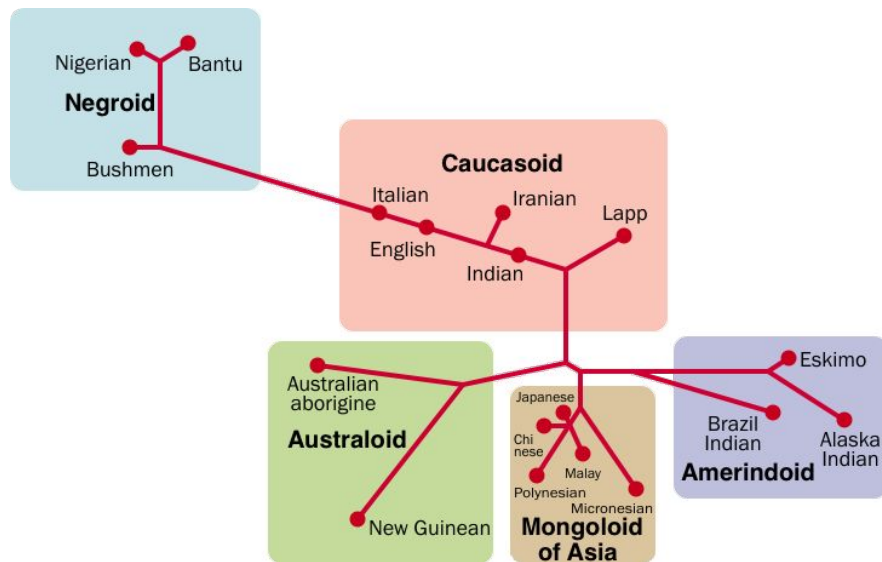


	((a,b),e)	(c,d)
((a,b),e)	0	33
(c,d)	33	0



Neighbor joining

- метод ближайших соседей
- требует выполнения аддитивности
- если гипотеза не работает
- выбираем пару, которая еще и далека от остальных
- результат – неукоренённое дерево



Neighbor joining

1. Рассмотрим все пары и выберем ту, для которой

$$\min(D(A, B) - M(A) - M(B))$$

↑
Расстояние
из матрицы

↑ ↑
Суммарное расстояние от
вершины до всех остальных

2. Объединяем $(A, B) = U$. Пересчитываем расстояния по формулам:

$$D(U, A) = 0.5 \cdot (D(A, B) + M(A) - M(B))$$

$$D(U, B) = 0.5 \cdot (D(A, B) + M(B) - M(A))$$

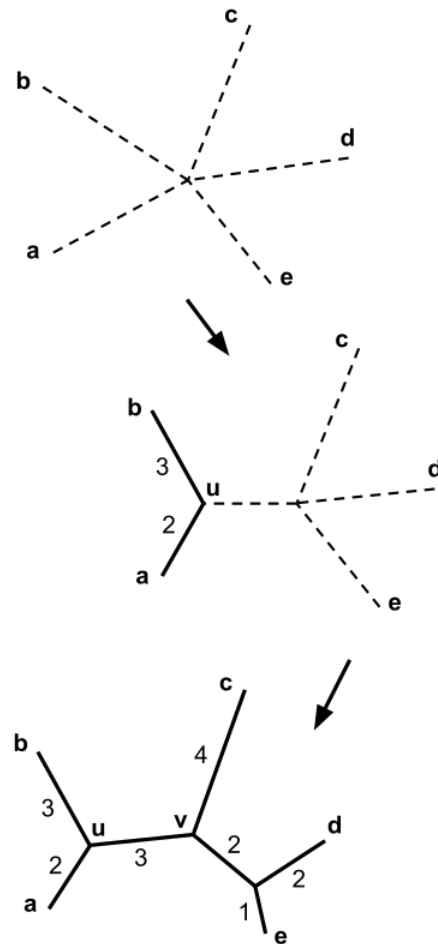
$$D(U, X) = 0.5 \cdot (D(A, X) + D(B, X) - D(A, B))$$

Neighbor joining

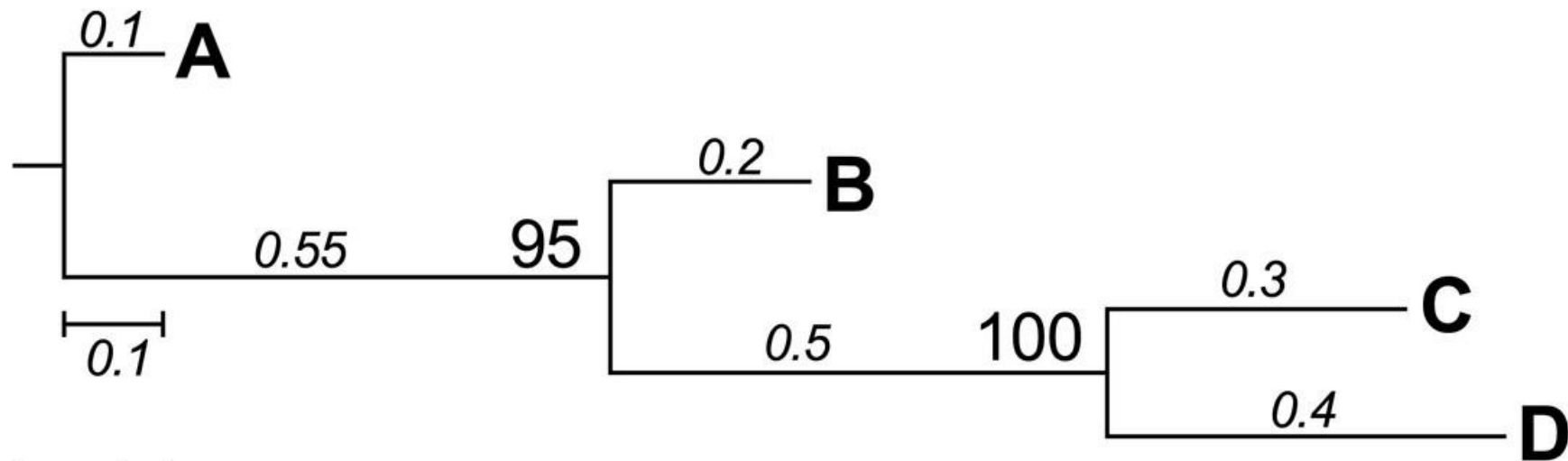
	a	b	c	d	e
a	0	5	9	9	8
b	5	0	10	10	9
c	9	10	0	8	7
d	9	10	8	0	3
e	8	9	7	3	0

	u	c	d	e
u	0	7	7	6
c	7	0	8	7
d	7	8	0	3
e	6	7	3	0

	v	d	e
v	0	4	3
d	4	0	3
e	3	3	0



Newick format



Newick:

```
(A:0.1, (B:0.2, (C:0.3,D:0.4) 100:0.5) 95:0.55) ;
```