

Bioinformatik I

Multiples Alignment, UPGMA

Roland Eils, Rainer König, Benedikt Brors
IPMB
Universität Heidelberg



Algorithmen zur Baumkonstruktion: *unweighted pair group method with arithmetic mean*

- Das Problem der Konstruktion eines optimalen Baums aus einer Distanzmatrix ist NP-hart. Verschiedene Heuristiken sind in Gebrauch.■
- Beispiel: UPGMA-Methode (*unweighted pair group method with arithmetic mean*).■
- Gegeben sei eine Distanzmatrix zwischen allen Objekten. Man findet zunächst die beiden ähnlichsten Objekte (mit der geringsten Distanz) und vereinigt sie zu einem neuen Objekt.



UPGMA (Forts.)

- Der Abstand des neuen Objekts zu allen übrigen Objekten muß neu berechnet werden. Dazu ersetzt man die Abstände durch den Mittelwert der Abstände zu den in dem neuen Objekt enthaltenen Objekten.■
- Man fährt solange fort, bis keine Objekte mehr übrig sind.



Beispiel

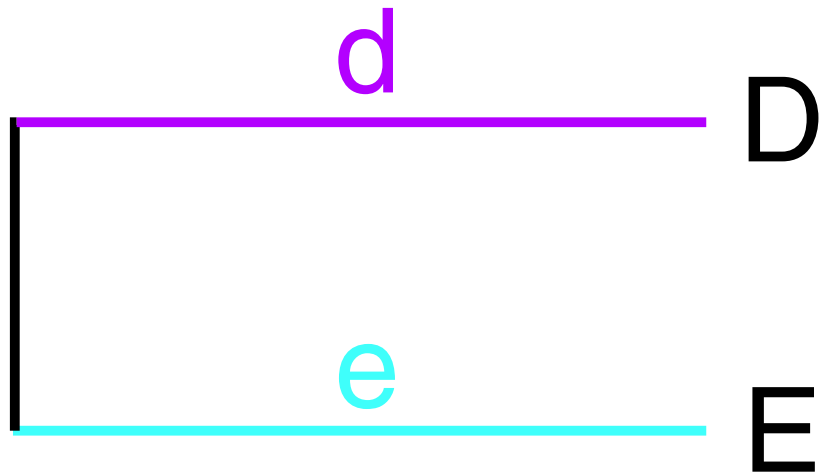
Beispiel: 5 Sequenzen A, B, C, D und E.

Distanzmatrix:

	A	B	C	D	E
A	0	22	39	39	41
B		0	41	41	43
C			0	18	20
D				0	10
E					0

Ähnlichste Sequenzen: D, E





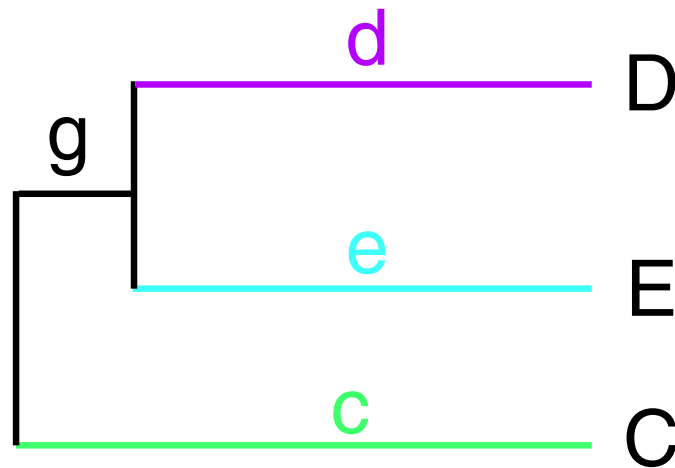
$$d = e = 10/2 = 5$$

Neue Distanzmatrix:

	A	B	C	(DE)
A	0	22	39	40
B		0	41	42
C			0	19
(DE)				0



Nächst ähnliches Paar: $C \leftrightarrow (DE)$.



$$c = 19/2 = 9.5$$

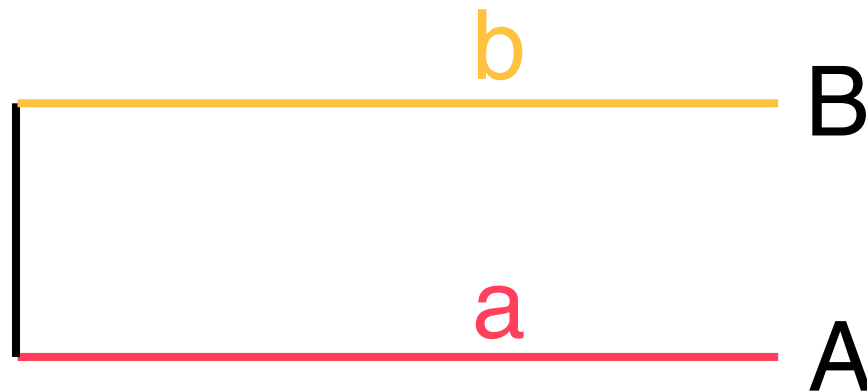
$$g = 19/2 - d = 9.5 - 5 = 4.5$$

Neue Distanzmatrix:

	A	B	(CDE)
A	0	22	39.7
B		0	41.7
(CDE)			0



Nächst ähnliches Paar: A — B.



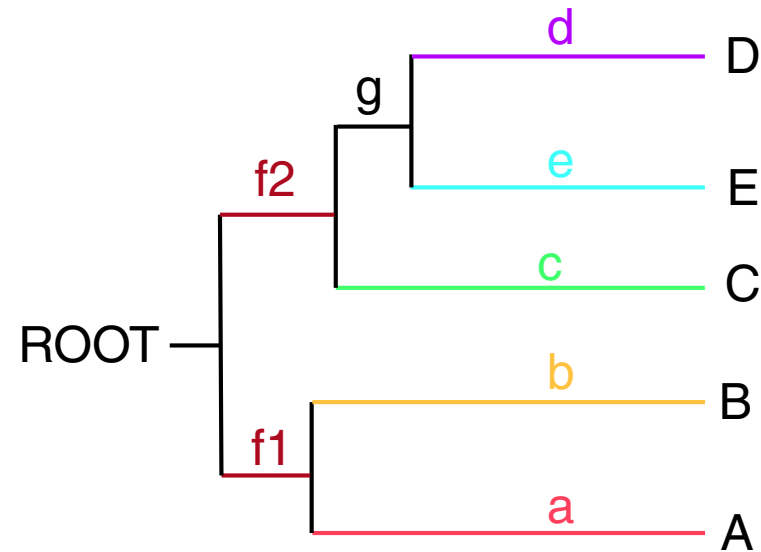
$$a = b = 22/2 = 11$$

Neue Distanzmatrix:

	(AB)	(CDE)
(AB)	0	40.7
(CDE)		0



Fertiger Baum:



$$f_2 + g + d = 40.7/2$$

$$f_1 + a = 40.7/2$$

$$\Rightarrow f_1 = 9.35, \quad f_2 = 10.85$$

