

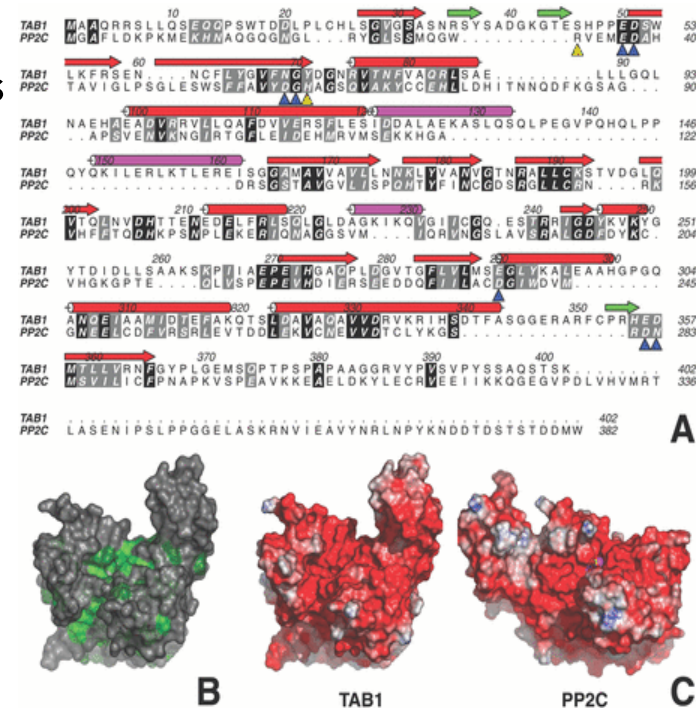
BIOINFORMATIKA – CV. 6

Lokální a globální zarovnávání dvou sekvencí, zarovnávání vícero sekvencí – rozšíření a doplnění přednášky

Příklad na globální zarovnání

- Máme sekvence: PRETERIT a VEITGEIST
- Nalezněte globální zarovnání těchto sekvencí s následujícími parametry:
 - „match“ = 1, „mismatch“ = 0 a „gap“ = -1 (tj. $d = 1$)

		P	R	E	T	E	R	I	T
	0	-1	-2	-3	-4	-5	-6	-7	-8
V	-1								
E	-2								
I	-3								
T	-4								
G	-5								
E	-6								
I	-7								
S	-8								
T	-9								



$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \end{cases}$$

Řešení (zobrazeny jen některé šipky)

		P	R	E	T	E	R	I	T
	0	-1	-2	-3	-4	-5	-6	-7	-8
V	-1	0	-1	-2	-3	-4	-5	-6	-7
E	-2	-1	0	0	-1	-2	-3	-4	-5
I	-3	-2	-1	0	0	-1	-2	-2	-3
T	-4	-3	-2	-1	1	0	-1	-2	-1
G	-5	-4	-3	-2	0	1	0	-1	-2
E	-6	-5	-4	-2	-2	1	1	0	-1
I	-7	-6	-5	-3	-2	0	1	2	1
S	-8	-7	-6	-4	-3	-1	0	1	2
T	-9	-8	-7	-5	-3	-2	-1	0	2

V E I T G E I S T
P R E T - E R I T

V E I T G E I S T
P R E T E R I - T

Poznámka

- Výpočet se zkomplikuje, pokud místo „linear gap penalty“ (kdy nerozlišujeme mezery následující za sebou a izolované mezery) použijeme „affine gap penalty“, ale pořád jej lze provést v kvadratickém čase, což neplatí pro obecnou „gap penalty“, kdy už je potřeba obecně kubický čas.

linear

$$w(g) = -g \times d$$

affine

$$w(g) = \begin{cases} -d - (g-1)e, & g \geq 1 \\ 0, & g = 0 \end{cases}$$

Lokální zarovnávání

- Lokální zarovnání nachází nejlepší „match“ mezi **podsekvencemi** zadaných dvou sekvencí
- Užitečné pro porovnávání sekvencí proteinů obsahujících podobný „motiv“ (evolučně konzervovanou podsekvenci) nebo společnou doménu (nezávisle „skládaná“ jednotka), ale liší se ve zbytku svých sekvencí

Lokální zarovnávání – algoritmus (1)

- Modifikovaná rekurzivní funkce

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \\ 0 \end{cases}$$

- $F(i, j)$ = skóre nejlepšího zarovnání suffixu sekvence $x[1 \dots i]$ a suffixu sekvence $y[1 \dots j]$

Lokální zarovnávání – algoritmus (2)

- Inicializace: první řádek a první sloupec inicializujeme nulami
- Nalezení optimálního zarovnání:
 - ▣ Nalezneme maximum funkce $F(i,j)$ kdekoliv v matici
 - ▣ Postupujeme podle zaznamenaných šipek, dokud nenarazíme na políčko s nulovou hodnotou $F(i,j)$

Lokální zarovnání - příklad

- Máme sekvence: AAGA a TTAAG
- Nalezněte lokální zarovnání těchto sekvencí s následujícími parametry:
 - „match“ = 1, „mismatch“ = 0 a „gap“ = -1 (tj. $d = 1$)

		A	A	G	A
	0	0	0	0	0
T	0				
T	0				
A	0				
A	0				
G	0				

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \\ 0 \end{cases}$$

Lokální zarovnávání - řešení

		A	A	G	A
	0	0	0	0	0
T	0	0	0	0	0
T	0	0	0	0	0
A	0	1	1	0	1
A	0	1	2	1	1
G	0	0	1	3	2

A A G
A A G