



Algoritmi v bioinformatiki - 2. DOMAČA NALOGA

Maksimalno število točk: 30. Rok za oddajo: 14.5.2025

1. Dano imamo naslednje zaporedje izidov metov kovanca

$$V = C C C G C G G C C G C,$$

pri čemer  $C$  označuje, da je bil izid meta cifra,  $G$  pa da je bil izid meta grb. Za mete imamo na voljo 3 kovanke  $A$ ,  $B$  in  $C$ , veljajo naslednje verjetnosti (v %):

Prehod:	%	A	B	C
	A	40	30	30
	B	30	40	30
	C	30	30	40

Izpis:	%	C	G
A	75	25	
B	80	20	
C	20	80	

Katera od možnosti je najbolj verjetna:

- (a) za vse mete smo uporabili kovanec  $A$
- (b) za vse mete smo uporabili kovanec  $B$
- (c) za vse mete smo uporabili kovanec  $C$
- (d)  $\Pi = AAACBCCBBCA$

Odgovor ustrezno utemeljite.

(10 točk)

2. Dani imamo zaporedji  $s = \text{GAAGTCA}$  in  $t = \text{TAGTCCA}$  ter vrednostno funkcijo s parametroma  $\mu = 4$ ,  $\sigma = 2$  in nagrado za ujemanje 2.

- (a) Z uporabo Needleman–Wunsch-evega algoritma za globalno poravnavo smo dobili naslednjo tabelo:

		-	G	A	G	T	A	C	A
		0	1	2	3	4	5	6	7
-	0	0	-2	-4	-6	-8	-10	-12	-14
T	1	-2	-4	-6	-8	-4	-6	-8	-10
G	2	-4	0	-2	-4	-6	-8	-10	-12
A	3	-6	-2	2	0	-2	-4	-6	-8
T	4	-8	-4	0	-2	2	0	-2	-4
T	5	-10	-6	-2	-4	0	-2	-4	-6
A	6	-12	-8	-4	-6	-2	2	0	-2
C	7								
A	8								

Dopolnite tabelo tako, da poračunate vrednosti (in ustrezne puščice) za zadnji dve vrstici. (3 točke)

(b) Koliko optimalnih globalnih poravnav dobite? Izpišite vse rešitve. (3 točke)

3. Dano imamo naslednjo matriko izražanja:

	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$
$g_1$	2	2	6	2	3	4
$g_2$	3	7	3	1	9	3
$g_3$	2	2	7	2	6	3
$g_4$	3	2	3	2	1	3
$g_5$	2	1	5	1	0	4
$g_6$	3	5	5	8	2	3
$g_7$	1	3	2	5	4	2
$g_8$	5	4	2	4	7	5

(a) Določite gruče z uporabo metode voditeljev, če je začetna množica voditeljev enaka  $X = \{g_1, g_5, g_6\}$ . (5 točk)

(b) Izračunajte drevo hierarhičnega gručenja z uporabo algoritma UPGMA. (9 točk)