

MBI, 10.06.2026, czas: 60 min

Imię i nazwisko:

Proszę wpisywać odpowiedzi w ramkach

Zadanie 1 (3 pkt)

Badamy podobieństwo globalne (algorytm Nedlemana - Wunsch) dla sekwencji ACGAC i CAAA. Macierz podobieństwa pokazano obok, n jest liczbą liter Twojego nazwiska. Stosujemy liniową karę za przerwę $d = -2 * n$.

	A	C	G	T
A	$2 * n$	$-2 * n$	$-2 * n$	$-n$
C	$-2 * n$	$2 * n$	$-n$	$-2 * n$
G	$-2 * n$	$-n$	$2 * n$	$-2 * n$
T	$-n$	$-2 * n$	$-2 * n$	$2 * n$

Uzupełnij macierz podobieństwa, macierz pomocniczą. Podaj najlepsze dopasowanie, jeżeli jest kilka, podaj jedno z nich.

	A	C	G	T
A				
C				
G				
T				

$d =$

		A	C	G	A	C
C	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
A	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
A	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
A	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Rozwiązanie:

Zadanie 2 (3 pkt)

Sekwencje AAAAAA, BCBCBC, AAABCA oraz BCABCB użyto do budowy grafu de Bruijna, bez wag w krawędziach.

Narysuj graf 4-rzędu, w wierzchołkach są sekwencje o długości 3 symboli.

Czy istnieje rozwiązanie?

Sekwencja wyjściowa (jeżeli istnieje) to:

Zadanie 3 (3 pkt)

Zbuduj drzewo filogenetyczne na podstawie macierzy odległości pokazanej obok.

	B	C	D	E
A	6	4	2	6
B		6	6	2
C			4	6
D				6

Zadanie 4 (3 pkt)

Posługujemy się monetami A i B, obserwując sekwencje rzutów *RRRR* (orły i reszki). Podaj najbardziej prawdopodobną sekwencję użytych monet, zakładając, że to doświadczenie może być opisane ukrytym modelem Markowa pokazanym niżej.

$$Q = \{A, B\},$$

$$V = \{O, R\},$$

$$P_A = 0, P_B = 1.$$

	A	B
A	$\frac{4}{5}$	$\frac{1}{5}$
B	$\frac{1}{5}$	$\frac{4}{5}$

	O	R
A	0	1
B	$\frac{1}{2}$	$\frac{1}{2}$

Uzupełnij macierz pomocniczą w algorytmie Viterbiego

	R	R	R	R
A	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
B	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Podaj najbardziej prawdopodobną sekwencję użytych monet:

Pytania (2 pkt)

Jakie zagadnienie / zagadnienia warto byłoby dodać do MBI

Ile godzin poświęciłeś na MBI (sumarycznie wykłady, ćwiczenia, praca własna)

Twoim zdaniem, jakie mogą być korzyści ze stosowania metod przedstawionych na MBI

Notatki lub uwagi do R. Nowaka:

MBI, 10.06.2026, czas: 60 min

Imię i nazwisko:

Zadanie 5 (3 pkt)

Poniżej podano fragment sekwencji referencyjnej.

```
>chr3:101-121  
ATGCTGTCCGAGTACTAACCC
```

Zakładamy, że ramka odczytu zaczyna się od pierwszego nukleotydu podanej sekwencji referencyjnej (pozycja 101). W badanej próbce zidentyfikowano następujące trzy warianty jednogenowe (SNV):

1. chr3:106_G>A (wariant homozygotyczny),
2. chr3:108_C>G (wariant heterozygotyczny),
3. chr3:115_C>G (wariant heterozygotyczny).

Polecenia:

- Skonstruuj wycinek pliku SAM (podając wyłącznie kolumny POS i SEQ), składający się z dokładnie 4 odczytów o długości 10 nukleotydów każdy. Wygenerowane przez Ciebie odczyty muszą spełniać łącznie następujące warunki:
 - zapewniać głębokość pokrycia (depth of coverage) równą dokładnie 4x dla każdej z pozycji, w której występuje mutacja (pozycje: 106, 108, 115);
 - odzwierciedlać podaną zygotyczność. Dla heterozygot załóż idealny stosunek alleli zmutowanego do referencyjnego, równy 1:1.

- Dla każdego z trzech zidentyfikowanych wariantów:
 - Podaj kodon referencyjny oraz kodon zmieniony.

- Określ efekt wariantu na białko (np. zmiana synonimiczna, missense, nonsense/stop-gain, frameshift).

Przykładowy fragment pliku SAM (jeden odczyt):

101 ACTGACTGAC

Tablica kodonów:

		Second base					
		U	C	A	G		
First base	U	UUU } Phenylalanine F UUC } UUA } Leucine L UUG }	UCU } Serine S UCC } UCA } UCG }	UAU } Tyrosine Y UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine C UGC } UGA } Stop codon UGG } Tryptophan W	U C A G	
	C	CUU } Leucine L CUC } CUA } CUG }	CCU } Proline P CCC } CCA } CCG }	CAU } Histidine H CAC } CAA } Glutamine Q CAG }	CGU } Arginine R CGC } CGA } CGG }	U C A G	
	A	AUU } Isoleucine I AUC } AUA } AUG } Methionine start codon M	ACU } Threonine T ACC } ACA } ACG }	AAU } Asparagine N AAC } AAA } Lysine K AAG }	AGU } Serine S AGC } AGA } Arginine R AGG }	U C A G	
	G	GUU } Valine V GUC } GUA } GUG }	GCU } Alanine A GCC } GCA } GCG }	GAU } Aspartic acid D GAC } GAA } Glutamic acid E GAG }	GGU } Glycine G GGC } GGA } GGG }	U C A G	

Notatki lub uwagi do Tomasza Gambina:

