

B. Bioinformatica si Biostatistica.

Nume:.....

Sectia:.....

Punctaj: 0.5 pct/ intrebare **Obs:** întrebările pot avea 1,2 sau mai multe răspunsuri corecte

1. Secvențele ADN conțin simboluri din alfabetul: (a) {A,C,G,U}; (b) {A,N, D}; (c) {A,C,G,T}; (d) {A,C,G,T,U}
2. Numărul de aminoacizi (triplete de baze) din codul genetic este: (a) 4; (b) 20; (c) 64; (d) 128.
3. Pentru a stabili daca o secvență de nucleotide este generată aleator sau nu, se poate folosi: (a) testul chi-patrat; (b) testul Student; (c) testul F
4. Se consideră două secvențe ADN aliniate având fiecare câte 200 de nucleotide. In ipoteza simplificatoare că nucleotidele din fiecare secvența sunt independente scrieți formula probabilității ca cele două secvențe să coincidă în 5 din cele 200 de poziții:

R:

5. Pentru a verifica ipoteza privind independența nucleotidelor succesive într-o secvența ADN se poate folosi testul chi-pătrat cu: (a) 9 grade de libertate; (b) 16 grade de libertate; (c) 4 grade de libertate.
6. Se considera secvențele ADN de mai jos. Construiți matricea profil, sablonul consensual și scorul corespunzător.

A G T C A T T	<i>Matrice:</i>
G G A C T C G	
A C A T C T A	
A T C A T A G	
T G A C G G C	

Sablon: *Scor:*.....

7. Distanța de editare dintre secvențele ATCT și TGA este:
8. Se considera alinierea de mai jos. Calculați scorul alinierii știind că scorul unei potriviri este 1, scorul unei nepotriviri este -3 iar scorul asociat inserției unui gap este -2.

ATC_GA
TGAG

R:

9. Se considera matricea de scor S construită, pe baza relației de recurență, de către algoritmul Smith-Waterman pentru două secvențe de lungimi M respectiv N. Scorul alinierii este: (a) S(1,1); (b) S(M,N); (c) S(1,N); (d) valoarea maximă din matricea M.
10. Se consideră matricea de scor S construită, pe baza relației de recurență, de către algoritmul Needleman-Wunsch pentru două secvențe de lungimi M respectiv N. Scorul alinierii este: (a) S(1,1); (b) S(M,N); (c) S(1,N); (d) valoarea maximă din matricea M.

11. Care dintre următoarele afirmații este(sunt) adevărate pentru algoritmul Smith-Waterman: (a) este algoritm de aliniere globală; (b) are ordinul de complexitate $O(m+n)$; (c) are ordinul de complexitate $O(mn)$; (d) este algoritm de aliniere locală.
12. Care dintre următoarele afirmații sunt adevărate pentru o matrice de puncte (dot matrices):
 - (a) se poate folosi doar pentru calculul scorului unei alinieri; (b) se folosește pentru analiza vizuală a similarității dintre secvențe; (c) pentru două secvențe de lungimi m respectiv n matricea de puncte are $(m+n)$ linii și $(m+n)$ coloane; (d) pentru două secvențe de lungimi m respectiv n matricea de puncte are m linii și n coloane.
13. Construiți matricea de puncte corespunzătoare secvențelor: ATCAG și CAT

14. În cazul algoritmilor de aliniere ce folosesc o schemă afină de penalizare a gap-urilor scorul unei succesiuni de 5 gap-uri este (pentru un scor de inițiere a succesiunii egal cu -4 și un scor de extindere a succesiunii egal cu -2) este:
15. Algoritmii BLAST și FASTA sunt: (a) algoritmi exacti (optimali) de aliniere a perechilor de secvențe; (b) algoritmi de construire a arborilor filogenetici; (c) algoritmi utilizați în căutarea în bazele de date biologice; (d) algoritmi de grupare.
16. În analiza valorilor statistice generate de algoritmi de tip BLAST, E-valoarea sugerează că există o potrivire semnificativă dacă: (a) are o valoare suficient de mare; (b) are o valoare suficient de mică.
17. Dendrograma este: (a) o ierarhie de clustere; (b) o structură de date construită prin algoritmi ierarhici de grupare; (c) o structură de date construită prin algoritmi partiționali de grupare; (d) o structura arborescentă corespunzătoare terminațiilor dendritice ale unui neuron.
18. În contextul algoritmilor de grupare un centroid al unui cluster este: (a) media aritmetică a elementelor din clusterul respectiv; (b) mijlocul segmentului de dreapta care unește cele mai îndepărtate două elemente ale clusterului; (c) „centrul de greutate” al clusterului.
19. În care dintre următoarele variante distanța dintre doi clusteri se definește ca fiind distanța dintre cele mai îndepărtate două elemente aparținând celor doi clusteri: (a) complete link; (b) average link; (c) single link.
20. Se consideră următorul set de date bidimensionale: $x_1=(1,1)$, $x_2=(2,0)$, $x_3=(4,2)$, $x_4=(3,1)$, $x_5=(2,4)$. (a) Să se construiască matricea de distanțe folosind distanța Manhattan. (b) Să se construiască dendrograma obținută folosind un algoritm aglomerativ de tip single link.

Matrice:

Dendrograma: