

FAC SIMILE ESAME DI BIOINFORMATICA

Tipologia di domande ed istruzioni generali

-Domande a risposta multipla: un'unica risposta possibile, nessuna penalizzazione in caso di errore (+1 per risposta corretta)

-Domande vero falso e domande con due possibili alternative: +1 in caso di risposta corretta, -0,5 in caso di risposta errata. Nel caso di due possibili alternative, CANCELLARE la risposta errata

-Domande a completamento: nessuna penalizzazione in caso di errore (+1 per risposta corretta)

- 1) I genome browser sono stati sviluppati per permettere l'esplorazione di interi genomi, che sarebbero altrimenti di difficile gestione. Nel caso dei vertebrati, di quale ordine di grandezza parliamo tipicamente?
 - a) Decine di Kb
 - b) Alcune Gb
 - c) Centinaia di Gb
 - d) Decine di Mb

- 2) La maggior parte delle sequenze depositate nel database proteico UniProtKB/TrEMBL derivano da predizioni in silico da genomi/sono state validate sperimentalmente

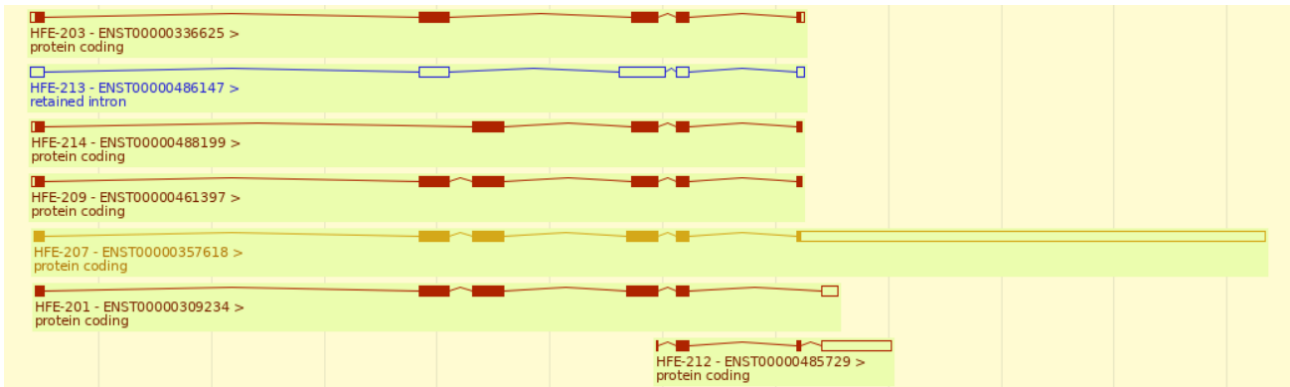
- 3) Una ricerca complessa all'interno di un database solitamente fa uso degli operatori booleani, ovvero _____, _____ e _____

- 4) Associate i seguenti termini ai database corrispondenti: dati di sequenziamento NGS / motivi riconosciuti da fattori di trascrizione / mutazioni somatiche / sequenze proteiche / giornali open access / domini proteici / dati di espressione genica / classificazione tassonomica di organismi marini / classificazione strutturale di proteine
 - a) DOAJ: _____
 - b) SRA: _____
 - c) SCOP: _____
 - d) JASPAR: _____
 - e) TrEMBL: _____
 - f) Prosite: _____
 - g) WoRMS: _____
 - h) COSMIC: _____
 - i) GTEEx: _____

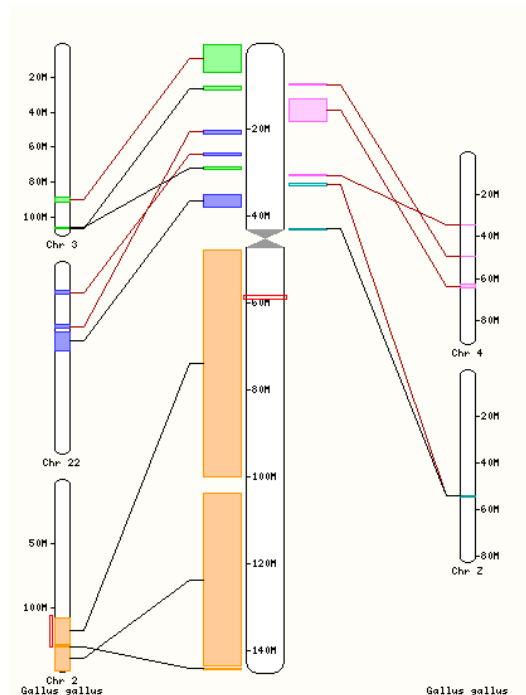
- 5) Vista la disparità nel numero di citazioni che possono essere ottenute da articoli pubblicati in ambiti molto inflazionati come quello biomedico, rispetto a quelle ottenibili da articoli che si occupano di argomenti di nicchia, a volte è utile osservare in quale _____ la rivista venga individuata dal Scientific journal Ranking, piuttosto che guardare soltanto l'impact factor

- 6) Tramite il genome browser di Ensembl, posso ricavare informazioni relative a:
 - Cromosoma sul quale è presente un gene di interesse VERO FALSO
 - Geni fiancheggiati il gene di interesse VERO FALSO
 - La frequenza di osservazione di una mutazione genetica nella popolazione VERO FALSO
 - La sequenza di una regione genomica di interesse in formato FASTA VERO FALSO

Prendiamo in esame questo esempio relativo al gene HFE umano preso da ENSEMBL



- 7) Nella figura sopra vengono riportate diversi geni/diversi trascritti. L'isoforma HFE-207 mostra un _____ più lungo rispetto alle altre, che include la totalità del 5'UTR/3'UTR.
- 8) Sei delle 7 isoforme in questione codificano una proteina: VERO FALSO
 L'isoforma HFE-203 include 4 introni VERO FALSO
 Le isoforme HFE-207 ed HFE-201 mostrano lo stesso numero di esoni VERO FALSO
 Le isoforme HFE-209 ed HFE-214 codificano proteine identiche VERO FALSO



- 9) Questa immagine mostra la _____ tra uomo e *Gallus gallus*, per quanto riguarda il cromosoma 8 umano. Sembra che un singolo gene umano trovi il proprio ortologo sul cromosoma Z di pollo. Se facessi un'analisi al contrario, verificando la presenza di blocchi sintenici del cromosoma Z di pollo in uomo, mi aspetterei di trovare maggiori corrispondenze:
- a) Nel cromosoma 8 umano
 b) Nel cromosoma 2 umano
 c) Nel cromosoma X umano
 d) Non ho elementi a sufficienza per dirlo

10) Tendenzialmente il livello di sintonia tra uomo e pollo dovrebbe essere maggiore di quello tra uomo e ratto VERO FALSO

E' corretto affermare che l'immagine riportata sopra è stata ottenuta grazie alle informazioni sulla posizione dei geni paraloghi nelle due specie VERO FALSO

Possiamo dire che le regioni del cromosoma 8 umano senza corrispondenze con quelli del pollo siano prive di geni VERO FALSO

11) Le matrici PAM/BLOSUM derivano dall'analisi dei dati contenuti nella banca dati BLOCKS

Prendiamo in esame questa matrice di sostituzione:

A	5																			
R	-1	6																		
N	-3	0	6																	
D	-3	-3	1	6																
C	-1	-5	-7	-8	9															
Q	-1	1	0	0	-8	5														
E	-2	0	-1	2	-7	2	4													
G	-1	-3	0	-2	-4	-3	-4	7												
H	-2	0	1	-2	-5	0	0	-4	8											
I	-2	-4	-5	-7	-2	-4	-5	-7	-5	4										
L	-1	-3	-4	-7	-2	-2	-5	-5	-3	3	4									
K	-1	2	0	-2	-7	1	1	-3	-1	-4	-3	5								
M	0	-1	-5	-5	-2	-1	-3	-4	-3	2	3	-2	5							
F	-2	-3	-4	-5	-6	-6	-6	-5	-2	0	0	-4	1	7						
P	-2	-3	-4	-3	-5	-4	-3	-2	-4	-6	-6	-3	-4	-4	7					
S	2	-1	2	0	-3	0	0	0	-1	-3	-3	0	-1	-3	-2	4				
T	0	-1	0	-2	-2	-1	-2	-3	-3	-1	-1	-1	-2	-3	-2	2	6			
W	-6	-6	-6	-10	-5	-5	-5	-4	-5	-5	-2	-7	-2	0	-11	-3	-4	10		
Y	-2	-3	-4	-6	-4	-2	-4	-5	1	-2	-2	-3	-1	4	-6	-3	-3	1	7	
V	0	-4	-5	-6	-3	-4	-4	-4	4	2	-3	1	-1	-4	-4	0	-6	-1	4	
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

12) Calcolate il punteggio del seguente allineamento tra due sequenze, sapendo che non sono attribuite penalità per la presenza di gaps: _____

D-AEN-ERYS

DEADSHE-IS

13) Nel caso invece in cui le penalità per inserzione ed estensione dei gaps siano -1 e -0,1 rispettivamente, il punteggio del seguente allineamento sarà:

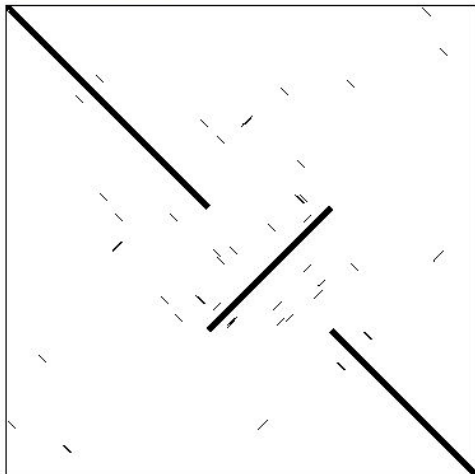
KASTA

K--TA

- a) 14,9
- b) 14,8
- c) 15
- d) 15,8

14) Indicate nella matrice sopra cerchiandoli i valori attribuiti al ritrovamento di amino acidi identici in un allineamento. Se la matrice fosse riportata come una matrice simmetrica, essa conterrebbe _____ valori

15) Questa dot matrix potrebbe essere ottenuta dal confronto:



- a) Esclusivamente tra sequenze proteiche
- b) Esclusivamente tra sequenze nucleotidiche
- c) Sia con sequenze proteiche che con sequenze nucleotidiche
- d) Solo con sequenze nucleotidiche, purché codificanti

Il tratto della retta a pendenza positiva potrebbe indicare _____

16) Indicate quale tra le seguenti affermazioni riguardo a tBLASTn è errata:

- a) Richiede una sequenza query amino acidica
- b) Traduce il database nucleotidico in 3 frame di lettura
- c) Può utilizzare un genoma come database per la ricerca
- d) Richiede una maggiore complessità di calcolo rispetto a BLASTp

17) Quale tra le seguenti affermazioni riguardo a BLASTp è errata?

- a) Il numero di "positives" è sempre minore o uguale al numero di "identities"
- b) L'e-value viene calcolato a partire da matrici PAM o BLOSUM
- c) E' possibile che il risultato con la percentuale di identità più elevato non sia quello con l'e-value migliore
- d) E' possibile ottenere dei risultati con e-value = 0

18) Ordinate i seguenti p-value dal migliore al peggiore: 0 / 0,1 / 100 / 0,00005

- (i) _____
- (ii) _____
- (iii) _____
- (iv) _____

19) Indicate quale tra le seguenti affermazioni relative a DIAMOND è errata:

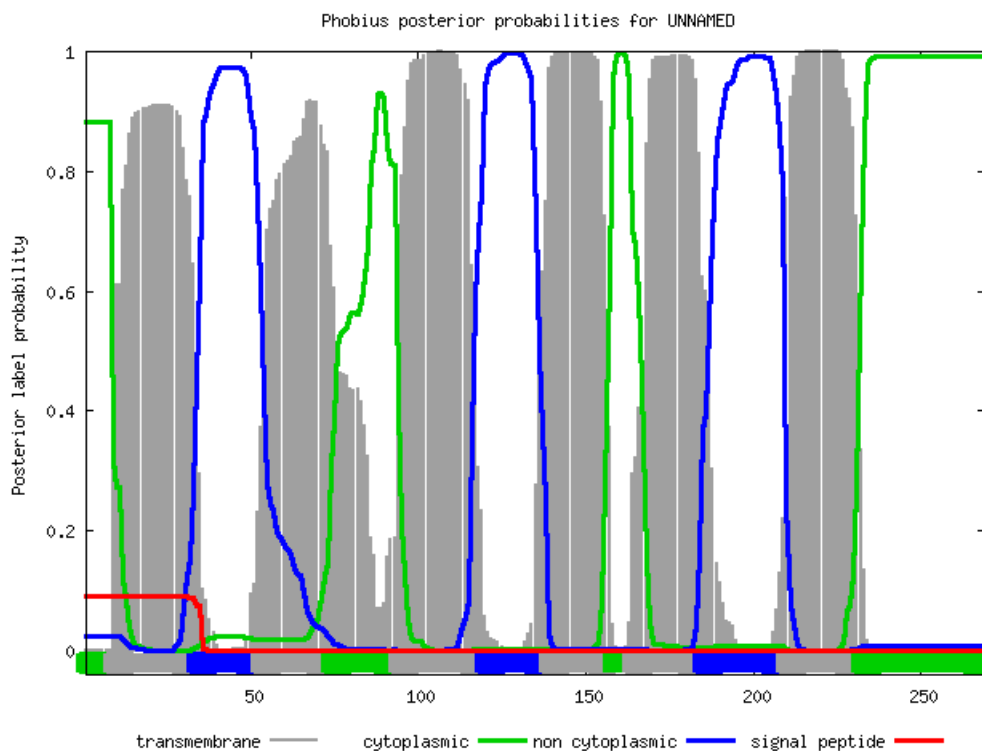
- a) E' fino a 20000 volte più rapido di BLASTx
- b) E' molto utile per l'annotazione di dataset di sequenze molto grandi, come i trascrittomi
- c) Ha soppiantato BLASTx in quanto maggiormente accurato
- d) E' stato sviluppato più recentemente rispetto a BLAST

20) Indicate quale tra le seguenti affermazioni relative a PSI-BLAST è errata:

- a) Nella prima fase di analisi utilizza un normale BLASTp
- b) Può essere reiterato N volte
- c) Cerca match solamente all'interno del database PDB
- d) Genera matrici PSSM

21) Un approccio di pattern discovery potrebbe avere senso se applicato ad un set di sequenze localizzate a monte del sito di inizio trascrizione di geni coregolati VERO FALSO

22) Questo output grafico, relativo all'aquaporina-1 umana, potrebbe essere il risultato di un'analisi effettuata con:



- a) SignalP
- b) Disembl
- c) Phobius
- d) Repeatscout

23) Ci indica che questa proteina possiede/non possiede un peptide segnale, che possiede sei _____ e che la porzione C-terminale è localizzata nella regione _____

24) La presenza di un segnale di secrezione non canonico potrebbe essere rilevato tramite

- a) TargetP
- b) SecretomeP
- c) ProP
- d) SignalP

25) I metodi di threading possono individuare nuove tipologie di folding, differenti da tutte quelle precedentemente depositate in database strutturali VERO FALSO

26) Due proteine possono mostrare una similarità strutturale significativa a causa di convergenza evolutiva VERO FALSO
Nel caso in cui si ritrovi una similarità a livello di sequenza che non è legata ad una comune origine evolutiva, ma piuttosto a convergenza evolutiva, questa situazione è definibile con il termine _____

27) L'assegnazione di termini Gene Ontology ad una sequenza solitamente avviene inferendoli dai risultati più significativi trovati con BLAST all'interno di un database di sequenze con funzione nota VERO FALSO

28) La differenza principale tra PSI-COFFEE ed EXPRESSO sta nel fatto che il primo/il secondo generi matrici PSSM, mentre l'altro basi l'allineamento su template strutturali derivati dal database _____

29) Quale tra i seguenti è un esempio di allineamento multiplo "block-based"?

- a) Probcons
- b) MAFFT
- c) DIALIGN2
- d) LALIGN

30) La selezione purificante

- a) Di solito si osserva in una regione ipervariabile di una proteina
- b) Può riguardare il sito attivo di un enzima
- c) E' stata studiata in modo particolare in relazione a virus come l'HIV
- d) E' detta anche selezione positiva

31) I nodi terminali di un albero filogenetico possono essere definiti _____

32) Fornite un esempio di un gruppo tassonomico che corrisponda alle seguenti definizioni:

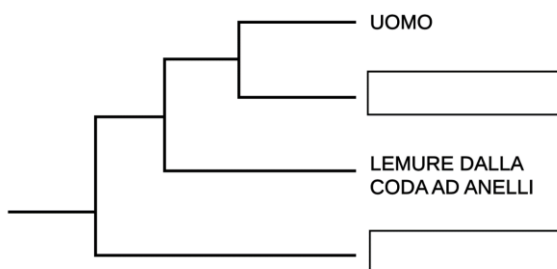
-MONOFILETICO: _____

-PARAFILETICO: _____

-POLIFILETICO: _____

33) Un albero filogenetico in cui siano presenti tre taxa può generare _____ alberi non radicati e _____ alberi radicati. Questo numero aumenta linearmente/esponenzialmente al crescere del numero di taxa.

34) Completate il seguente albero filogenetico con due taxa che ritenete possano essere appropriati



Il taxa del ramo basale di questo albero può essere definito _____