

ESEMPI DI DOMANDE PRESI DA APPELLI 2018

Per le domande a risposta multiple, ogni domanda ha un'unica risposta corretta possibile. Non verranno applicate penalità per risposte errate.

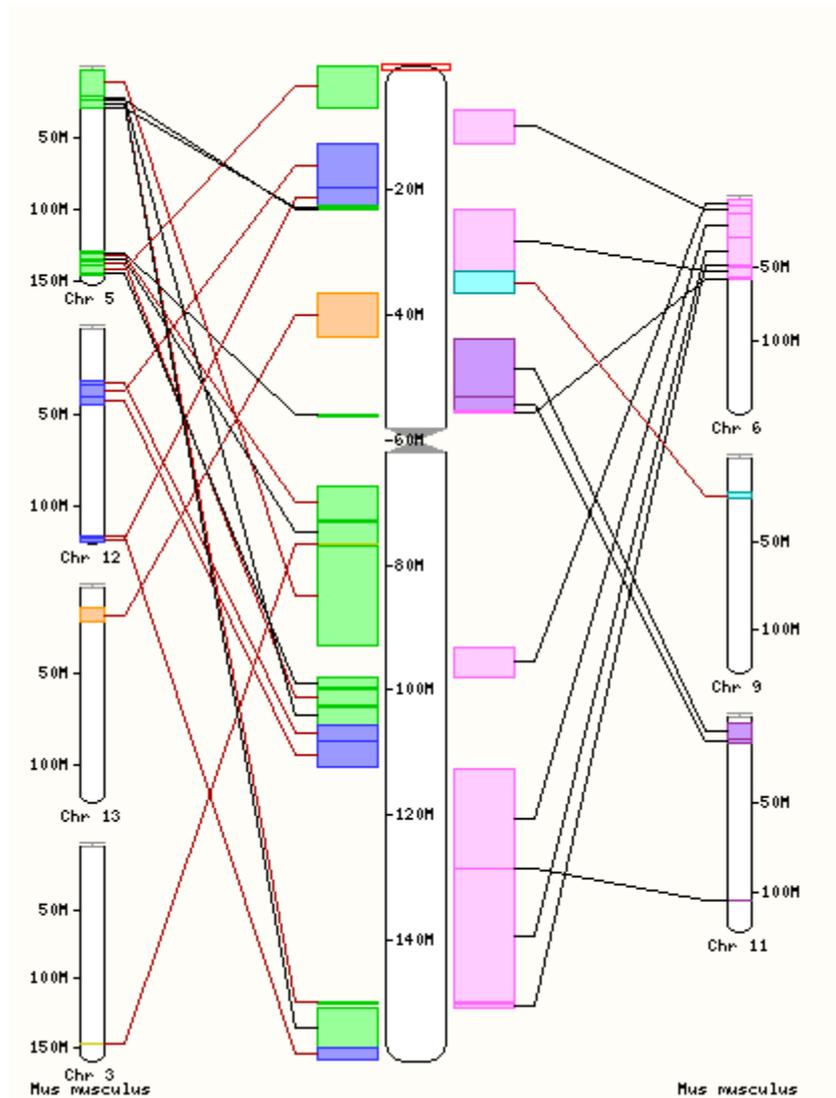
1. Sequenze geniche definibili come paraloghe:
 - a) Sono il prodotto di un evento di speciazione
 - b) Devono essere necessariamente due, non di più
 - c) Sono il prodotto di un evento di duplicazione genica
 - d) Devono necessariamente trovarsi sul medesimo cromosoma

2. Indicate quale tra i seguenti potrebbe essere un corretto header di un file FASTA:
 - a) >ciao, sono la risposta corretta, barrami
 - b) [AM501548.1 Hypsibius klebelsbergi partial mRNA for actin la D mente]
 - c) <AM501548.1/sono/la/risposta/sbagliata>
 - d) AM501548.1|partial_mRNA_for_actin[non_ti_fidare_della_risposta_A]

3. Gli output grezzi di sequenziamento di una analisi genomica di next generation sequencing devono essere depositati in un database chiamato:
 - a) GenBank
 - b) NCBI genomes
 - c) Ensembl
 - d) SRA

4. Devo fare un BLASTp per una ricerca di similarità tra una sequenza query ed un database proteico di una specie molto lontana filogeneticamente, per cui mi aspetto similarità piuttosto remote. Che matrici dovrei usare?
 - a) PAM o BLOSUM entrambe con indice alto
 - b) PAM con indice alto oppure BLOSUM con indice basso
 - c) PAM o BLOSUM entrambe con indice basso
 - d) PAM con indice basso oppure BLOSUM con indice alto

5. Il parametro ktup nell'algoritmo FASTA è fondamentale per settare l'accuratezza, ma anche per definire la velocità di calcolo. In una ricerca di similarità con FASTA, il numero di k-tuples indicizzati sarà uguale a:
 - a) 20^{ktup}
 - b) $20 * ktup$
 - c) $ktup^{20}$
 - d) $(L * ktup) / 20$, dove L = lunghezza della sequenza query



6. Date uno sguardo all'immagine riportata sopra, che mostra il risultato di un'analisi di sintenia tra un cromosoma di una specie ignota ed il genoma di topo. Indicate quale tra le seguenti affermazioni è errata:
- Si tratta di un'analisi che può essere effettuata tramite il portale Ensembl
 - E' un'analisi basata sulla posizione di geni ortologhi in specie diverse
 - In questo caso la specie ignota è probabilmente filogeneticamente molto affine al topo (ad esempio ratto, criceto oppure un altro roditore).
 - I 7 cromosomi mostrati alla destra ed alla sinistra del cromosoma centrale appartengono alla stessa specie
7. Se effettuassi un BLASTx, utilizzando come query la sequenza proteica della acido nitrico sintasi umana contro il genoma del topo, otterrei come migliori risultato:
- La sequenza ortologa di topo
 - La sequenza parologa di topo
 - La sequenza analoga di topo
 - Otterrei un messaggio di errore

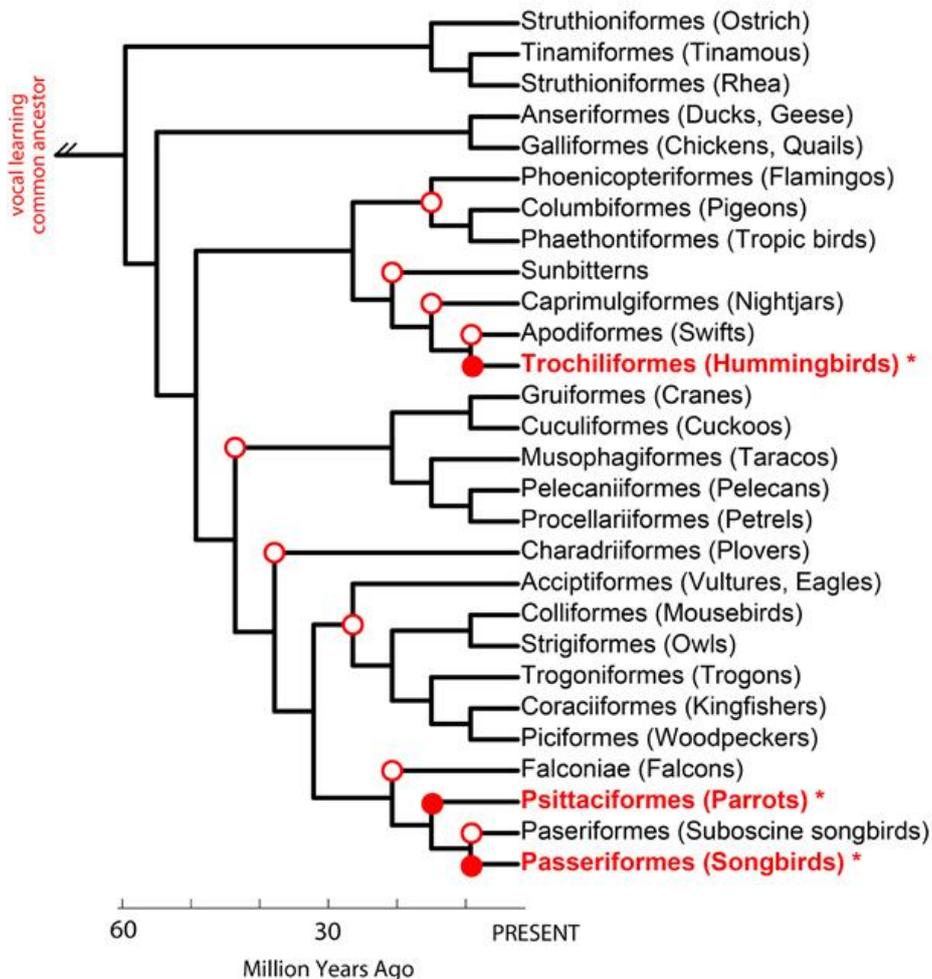
8. Gli approcci euristici sono spesso e volentieri utilizzati in ambito bioinformatico. Ma che cosa si intende per approccio euristico?
- E' un sinonimo di approccio esaustivo: esplora tutte le possibili soluzioni individuando la migliore
 - Un approccio basato sul principio del rasoio di Occam: la soluzione più semplice è individuata come quella maggiormente probabile
 - Un approccio basato sull'interrogazione di database e può quindi anche essere definito un approccio knowledge-base
 - Un approccio non rigoroso alla soluzione dei problemi, che li risolve in modo sub-ottimale semplificando la complessità di calcolo richiesta
9. ViennaRNA è:
- Una importante conferenza tematica sullo splicing alternativo che si tiene annualmente a Vienna
 - Un pacchetto di tools per la predizione della struttura secondaria dell'RNA
 - Uno strumento che permette di identificare 5' e 3' UTR e coding region in un RNA messaggero
 - Un algoritmo per allineare tra loro diverse varianti di splicing alternativo originate dallo stesso gene
10. Come funziona il contest CASP? Valuta quanto i modelli strutturali generali di un algoritmo si avvicinano a:
- Strutture 3D di proteine bersaglio predette da I-TASSER
 - Strutture 3D di proteine bersaglio scelte casualmente tra quelle già disponibili e depositate nel database PDB
 - Strutture 3D di proteine bersaglio la cui struttura sia nota ma non ancora resa pubblica
 - Strutture secondarie di proteine bersaglio scelte casualmente tra quelle già disponibili e depositate nel database PDB
11. Che edit distance c'è tra le parole TRANELLO ed ANELLO?
- 0
 - 2
 - Approssimativamente 2500
 - Non ho gli elementi necessari per calcolarla
12. La statistica Q3 identifica:
- L'accuratezza di un allineamento multiplo di sequenza
 - L'accuratezza della predizione della struttura secondaria di una proteina
 - L'accuratezza della predizione della struttura terziaria di una proteina
 - L'accuratezza della predizione della struttura terziaria di una molecola di RNA
13. Un sito sotto selezione positiva sarà caratterizzato da un parametro omega (dN/dS)
- >1
 - <1
 - <0
 - >0
14. Data la lunghezza di un allineamento (L) ed un numero di sequenze (N), la complessità del calcolo di un allineamento multiplo (O) sarà definita dalla formula:

- a) $O = L * N$
- b) $O = L^N$
- c) $O = N^L$
- d) $O = (L/6) * N$

15. Immaginiamo di voler effettuare una analisi filogenetica “esplorativa” con un numero molto elevato di sequenze, per cui non siamo interessati almeno inizialmente ad una grande accuratezza, ma in cui ci interessa fondamentale ottenere un risultato indicativo nel minor tempo possibile. Quale tra le seguenti combinazioni tra algoritmo di allineamento multiplo ed analisi filogenetica utilizzereste?

- a) Espresso + Maximum Likelihood
- b) Probcons + BEAST
- c) MUSCLE + NJ
- d) MUSCLE + MrBayes

AVIAN FAMILY TREE (Vocal learners)



16. L’albero riportato sopra ci mostra la filogenesi degli uccelli, evidenziando in rosso i “vocal learners”, cioè quegli uccelli in grado di imparare a riprodurre suoni uditi nell’ambiente. L’albero mostrato è definibile come:

- a) Cladogramma
- b) Filogramma
- c) Cronogramma
- d) Aviogramma

17. Sempre con riferimento all'albero sopra, gli uccelli vocal learner sono un gruppo:

- a) Monofiletico
- b) Polifiletico
- c) Parafiletico
- d) Pseudofiletico

18. Sempre con riferimento all'albero sopra, quale vi sembra l'outgroup più appropriato da inserire per la radicazione (in questo momento l'albero è radicato sul gruppo più basale, gli struzzi)?

- a) Coccodrillo
- b) Uomo
- c) *Escherichia coli*
- d) *Drosophila melanogaster*

Saranno poi presenti semplici domande VERO/FALSO. In questo caso risposte errate verranno penalizzate (se non le sapete meglio non rispondere). +1 risposta corretta, 0 nessuna risposta, -0,5 risposta errata

Altra tipologia di domanda possibile sarà quella in cui vi verrà chiesto di cancellare la risposta sbagliata tra due alternative (come nel vero/falso gli errori verranno penalizzati)

ESEMPI

-In un BLASTp, la modifica del parametro *word size* da 6 a 2 renderà la ricerca piu stringente/meno stringente

-L'approccio utilizzato da I-TASSER usa un approccio può essere definito come un metodo di predizione ab initio/di fold recognition

-Un sito sotto selezione purificante mostrerà una preponderanza di sostituzioni sinonime/non sinonime

Altre domande potrebbero essere a semplice completamento (nessuna penalità per risposte errate in questo caso)

ESEMPI:

-Data una matrice di sostituzione (avrete la tabella in una figura), calcolate lo score dell'allineamento senza gap tra queste due sequenze: ADFTRDR e AKFLMDR: _____

-Il principale ente che si occupa della gestione dei dati bioinformatici in Giappone è chiamato _____

-L'utilizzo di un metodo di predizione strutturale basato su omologia strutturale richiede l'identificazione di strutture che fungano come template all'interno del database _____

-L'omoplasia è un fenomeno che è ritenuto di scarso rilievo ed è quindi trascurato dai metodi filogenetici
