Per aiutarvi ho elaborato (frettolosamente) questi quesiti che dovrebbero aiutarvi ad individuare gli argomenti importanti del corso ed a darvi un'idea delle domande che potrebbero esservi poste all'esame.
Ripeto: ho scritto le domande frettolosamente solo per fornire una idea generale….. potrebbero essere presente alcune imprecisioni

Provate rispondere alle domande, se ci riuscirete, sarete pronti a superare l'esame per quanto riguarda la parte di bioinformatica.

Se ci sono problemi, se non riuscite a risponde ad alcune domande, fatemelo sapere.

I quesiti scritti in rosso, non sono stati trattati in modo adeguate durante le lezioni, quindi non saranno domande d'esame (sarebbe comunque opportuno che voi cercaste di rispondere)

BUON LAVORO

-

DATABASE: Possibile tipo di domanda:
Viene data una figura rappresentante alcuni record in formato flat-file (non viene specificato il tipo di DB (tabella o testuale), bisogna riconoscerlo)
Domande possibili:
- Quanti record sono rappresentati ?
- Quanti campi?
- Riconoscere i separatori di record, di campo?
- è possibile riconoscere se il relativo database è indicizzato o ordinato?
- contiene cross-reference?

Ecc. ecc.

<u>Possibili argomenti per le domande sui database:</u>
- Differenze/analogie tra schede e attributi di un archivio rispetto ai record e campi di un database
- Cosa sono i database ordinati
- Cosa sono gli indici dei database
- Come agisce una ricerca dicotomica (binaria)
- cosa è l'identificatore di un record (o Accession Number)
- differenze tra flat-file e database relazionale

Negli acidi nucleici:
Nel DNA le basi azotate sono: Adenina, Citosina, Uracile, Guanina?
Nel DNA le basi azotate sono: Adenina, Citosina, Timina, Guanina?

Il DNA si distingue dall'RNA:
- Perché il DNA è più stabile dell'RNA?
- perché contiene la base azotata guanina al posto dell'adenina?
- perché contiene la base azotata timina al posto dell'adenina?
- perché contiene la base azotata timina al posto della citosina?

---

la sintesi proteica inizia dall'estremità:
-Aminica?
- Carbossilica?
- la sua traduzione inizia dal 5' della sequenza dell'mRNA?
- la sua traduzione inizia dal 3' della sequenza dell'mRNA?

La trascrizione è il processo per cui:
- da una molecola di RNA si ricava la relativa proteina?
- Da una molecola proteica si ricava il relativo RNA?
- Da una molecola di DNA si ricava il relativo RNA?

Data la stringa:
MEQTEGNSSEDGTTVSPTAGNLETPGSQGIAEEVAEGTVGTSDKEGPSDWAEHLCKAASKSGESGGSPGEASILDELKTD
LQGEARGKD
E' una sequenza proteica o una sequenza nucleotidica?
Se rappresenta una proteina completa (nella traduzione inizia dal codone di start e finisce con uno stop, il primo aminoacido è quello atteso?

Data la stringa:
CCNNTCTTGAAGGTNNGGGGTGGGGTGGAAGCAAGNGAGGTGGGACTTTGAGCTGAA
GGCTCCAAGCCACCAGAGCTTAAAGAACGATGAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAA

Rappresenta una sequenza proteica, di DNA, di RNA o di cDNA?
Come giudicate la qualità della sequenza (scarsa o buona)?

Data la sequenza:
5' AGAGCTCCGAGCTGCAG 3'
scrivere la sua inversa complementare (3' ← 5')
scrivere la sua complementare (5' → 3')

o individuatele tra queste due sequenze:
TCTCGAGGCTCGACGTC
CTGCAGCTCGGAGCTCT

Data la sequenze TGAACTGAGCACA,

Questa sequenza ACTTGACTCGTGT
Rappresenta la sua: complementare? Inversa? Inversa complementare?

Mentre quest'altra sequenza TGTGCTCAGTTCA
Rappresenta la sua: complementare? Inversa? Inversa complementare?

Mentre quest'altra sequenza ACACGAGTCAAGT
Rappresenta la sua: complementare? Inversa? Inversa complementare?

Dato il record (un po' manomesso):

```
LOCUS       NM_001103X               4528 bp    mRNA    linear   PRI 11-FEB-2008
DEFINITION  Homo sapiens actinin, alpha 2 (ACTN2), mRNA.
ACCESSION   NM_001103
```

…… omissis ……..

```
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from DC378038.1, BC051770.1,
            CB153006.1, AL359185.25 and M86406.1.
```

…… omissis ……..

```
FEATURES             Location/Qualifiers
     source          1..4528
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q42-q43"
     gene            1..4528
                     /gene="ACTN2"
                     /note="actinin, alpha 2"
                     /db_xref="GeneID:88"
```

…… omissis ……..

```
     CDS             205..2889
                     /gene="ACTN2"
```

…… omissis ……..

```
     exon            331..445
                     /gene="ACTN2"
                     /inference="alignment:Splign"
                     /number=2
     exon            446..565
                     /gene="ACTN2"
                     /inference="alignment:Splign"
                     /number=3
```

…… omissis ……..

```
     exon            2572..2730
                     /gene="ACTN2"
                     /inference="alignment:Splign"
                     /number=20
     exon            2731..4528
                     /gene="ACTN2"
                     /inference="alignment:Splign"
                     /number=21
```

```
ORIGIN
        1 attaagccgc gcggcagctg ctcgcagccg gagctggtgc ttcgcccgag acccagcgcc
       61 caggcgtgtc gccccgagag gagccgcgcg aaggtcaccc cgcgcccgcc gcccgccgcc
      121 cgccgcctcc gtgggtccgt ttgccagtca gcccgtgcgt ccgagcccct cgcgccccgc
      181 cgcagccccg gccaaccgag cgcc
      205 atgaac cagatagagc ccggcgtgca gtacaactac
      241 gtgtacgacg aggatgagta catgatccag gaggaggagt gggaccgcga cctgctcctg
      301 gacccagcct gggagaagca gcagaggaag accttcactg cctggtgtaa ctcccaccta
  …… omissis ……..
     4201 gcaggaattc tttttatatc tgagtcctta atgatcagga aatggattac aaacaaacaa
     4261 aaacagttgc aaacagcaat taaagcatta ctagagaagc agtaacatgt ctgttaccta
     4321 cagtctgttt gctcagattt ttatagattt ccaccaacag ctggaaatgt aatatttcca
     4381 ttccaagctg tctgacagct tgactcttct tcctacactg ggccatttag aaccttcaac
     4441 ctttatgaat ttcacttaat ttgcttcatt attaagtaat tggtagtttc cagattcctg
     4501 aaatatttga atccttgagt ttaatgcc
//
```

Possibili domande:

- E' un record ricavato dal database nucleotidico dell'EMBL o dell'NCBI?
- La sequenza è stata ricavata da una unica reazione di sequenziamento?
- Sono indicati gli esoni?
- Si possono individuare gli introni?
- E' una sequenza genomica ?
- E' una sequenza di cDNA (quindi rappresenta un mRNA) ?
- E' rappresentata anche la regione 3'UTR?
- E' rappresentata anche la regione 5'UTR?
- La relativa proteina è lunga più di 2500 aa?
- Il primo codone della regione codificante rappresenta una metionina?

Questa figura:

```
5′ TGAACTGAGCACAGTATAGGGGTAGCAGTTGGGGTTCA 3′
   ||||||||||||||||||||||||||||||||||||
3′ ACTTGACTCGTGTCATATCCCCATCGTCAACCCCAAGT 5′
```

- Rappresenta una regione di DNA a doppia elica?
- Sono evidenti dei *mismatch* (disappaiamenti)?
- La seconda sequenza è l'inversa complementare della prima?
- La prima sequenza è l'inversa complementare della seconda?

```
Date le stringhe a e b
a) GAGCACAGTATAGGGG
b) 3′ GAGCACAGTATAGGGG 5′
```

- rappresentano la stessa sequenza nucleotidica?

```
date le stringhe:
```

TTVSPTAGN
NGATPSVTT
Rappresentano delle proteine?
Rappresentano dei peptidi?
Osservando che una è scritta inversamente all'altra, rappresentano comunque la stessa sequenza?
La seconda è l'inversa complementare della prima?

Data una sequenza di DNA dalla quale vogliamo ricavare la potenziale regione codificante.
In quale filamento del DNA cercheresti la sequenza codificante:

- solo sul filamento '+' ?
- solo sul filamento '-' ?
- su ambedue i filamenti?
- In questo ultimo caso quanti frame di lettura dovresti considerare?

Data la sequenza:
AGAACGAUGAGAGUGAGGCAGCUCGGCUAGAACGAGGCAAGAGCUCCGAGCUGCAGGCAGAACCCC
ACAGACAUGGCCCUCAUCCACAAGUUUUUUCCAGAAGCCUUUGACUAUGCAGAGCUGGACCCAGCA
AAGCGCCGGCACAACUUCACCCUAGCUUUCUCGACUGCAGAGAAACUAGCCGACUGUGCCCAGCUG
CUGGAGGUGGAUGACAUGGUGCGGCUGGCAGUGCCCGAUUCCAAGUGUGUCUACACCUACAUCCAG
GAGCUAUACCGUAGCCUUGUGCAGAAAGGGCUGGUAAAGACGAAAAAGAAAUGAAUGGCAAAGCCC
UCCGGUGAACCAAAAAAAAAAAAAAAAAAAAAAAAA

Si ricavano i seguenti frame di traduzione

5'3' Frame 1
R T **Met** R V R Q L G **Stop** N E A R A P S C R Q N P T D **Met** A L I H K F F P E A F D Y A E L
D P A K R R H N F T L A F S T A E K L A D C A Q L L E V D D **Met** V R L A V P D S K C V
Y T Y I Q E L Y R S L V Q K G L V K T K K K **Stop** **Met** A K P S G E P K K K K K K K K

5'3' Frame 2
E R **Stop** E **Stop** G S S A R T R Q E L R A A G R T P Q T W P S S T S F F Q K P L T **Met** Q S
W T Q Q S A G T T S P **Stop** L S R L Q R N **Stop** P T V P S C W R W **Met** T W C G W Q C P I
P S V S T P T S R S Y T V A L C R K G W **Stop** R R K R N E W Q S P P V N Q K K K K K K K
K
5'3' Frame 3
N D E S E A A R L E R G K S S E L Q A E P H R H G P H P Q V F S R S L **Stop** L C R A G P S
K A P A Q L H P S F L D C R E T S R L C P A A G G G **Stop** H G A A G S A R F Q V C L H L
H P G A I P **Stop** P C A E R A G K D E K E **Met** N G K A L R **Stop** T K K K K K K K K K
3'5' Frame 1
F F F F F F F F G S P E G F A I H F F F V F T S P F C T R L R Y S S W **Met** **Stop** V **Stop** T H
L E S G T A S R T **Met** S S T S S S W A Q S A S F S A V E K A R V K L C R R F A G S S S A
Stop S K A S G K N L W **Met** R A **Met** S V G F C L Q L G A L A S F **Stop** P S C L T L I V
3'5' Frame 2
F F F F F F F F L V H R R A L P F I S F S S L P A L S A Q G Y G I A P G C R C R H T W N R A
L P A A P C H P P P A A G H S R L V S L Q S R K L G **Stop** S C A G A L L G P A L H S Q R L
L E K T C G **Stop** G P C L W G S A C S S E L L P R S S R A A S L S S F

3'5' Frame 3
F F F F F F F F W F T G G L C H S F L F R L Y Q P F L H K A T V **Stop** L L D V G V D T L G I
G H C Q P H H V I H L Q Q L G T V G **Stop** F L C S R E S **Stop** G E V V P A L C W V Q L C I
V K G F W K K L V D E G H V C G V L P A A R S S C L V L A E L P H S H R S

- Quale di questi frame sembra la più probabile traduzione?
- Date le tue conoscenze, era necessario analizzare tutte e 6 i frame di lettura o era sufficiente analizzarne
 - solo 1 ?
 - solo 2 ?
 - solo 3 ?

(per rispondere a questa domanda bisogna cercare di capire cosa rappresenta la sequenza nucleotidica data: rappresenta DNA a doppia elica oppure rappresenta il filamento codificante di cDNA o rappresenta una molecola di mRNA?)

Cosa sono le EST ?
-Peptidi
-Corte sequenze proteiche 50 – 100 aa
-Corte sequenze genomiche (100 – 500 bp)
-Corte sequenze genomiche (200 – 5000 bp)
-Corte sequenze ricavate da mRNA (100 – 500 bp)
-Corte sequenze ricavate da mRNA (200 – 5000 bp)

- Esiste un database pubblico in cui sono contenute solo EST?
- In GeneBank sono state depositate milioni di sequenze EST?

Le sequenze EST:

- Vengono fatte per completare la sequenza genomica di un organismo?
- per ricavare la sequenza completa di un gene?
- per capire se un determinato gene è espresso in un particolare tessuto?
- per determinare l'espressione genica (quali e quanto i geni sono espressi) in un determinato tessuto o in un determinato momento cellulare ?

Il termine trascrittoma si riferisce:
-Del DNA nel suo complesso (tutti geni presenti in un organismo, le sequenze intergeniche, le sequenze regolatrici, ecc).?
-all'insieme degli RNA messaggeri nel loro complesso ?
-allo studio di un particolare trascritto?

---

per espressione genica, si intende:

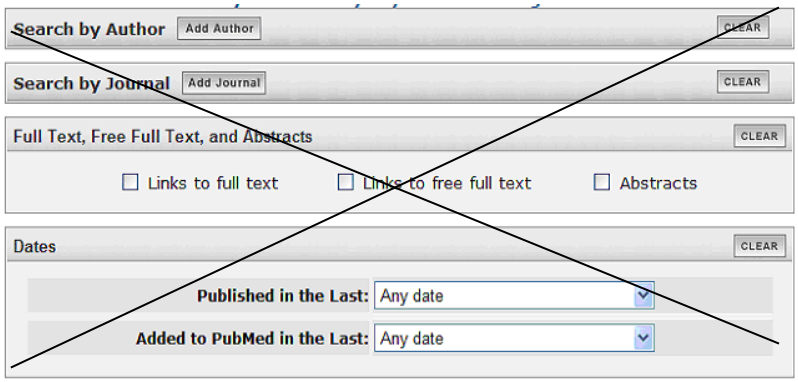
- tutto il set di geni presenti in un organismo?
- l'insieme di geni espressi in un determinato tipo cellulare e/o in un determinato momento ?
- la traduzione dei geni in proteine?

Domande sui trasctti:
- Nel trascritto primario possono essere presenti oltre che gli esoni anche gli introni?
- Gli introni vengono ‘saltati’ durante la prima fase della trascrizione?
- Gli introni vengono escissi (splicing) solo nella maturazione dell’mRNA?
- Durante la fase di maturazione possono essere escissi alcuni esoni in modo differente, producendo così delle isoforme alternative dello stesso gene?
- nel trascritto maturo sono presenti anche sequenze non codificanti?

Che cosa si ottiene dalla seguente ricerca nel database NCBI Nucleotide?: (Homo sapiens [organism] OR Mus musculus [organism]) AND telethonin [Gene Name]

N.	Testo risposta
1	la sequenza nucleotidica del gene telethonin in uomo e in topo
2	niente, la query non è corretta
3	tutte le sequenze nucleotidiche in uomo o in topo
4	la sequenza proteica del gene telethonin in uomo e in topo

QUESTA FIGURA NON E' PIU' VALIDA (è relativa a Pubmed che è cambiato) potrei mettere un'altra figura di ricerche effettuate in laboratorio da individuare
Selezionare la/le affermazioni corrette relative alla seguente figura che rappresenta una form di interrogazione ai database



-appare quando si sceglie 'Preview' in Entrez –PubMed
-appare quando si sceglie 'History' in Entrez –PubMed
-appare quando si sceglie 'Limits' in Entrez –PubMed
-appare quando si sceglie 'Limits' in Entrez-Nucleotide
-L'uso di questa form permette di limitare la ricerca al contenuto di specifici campi

---

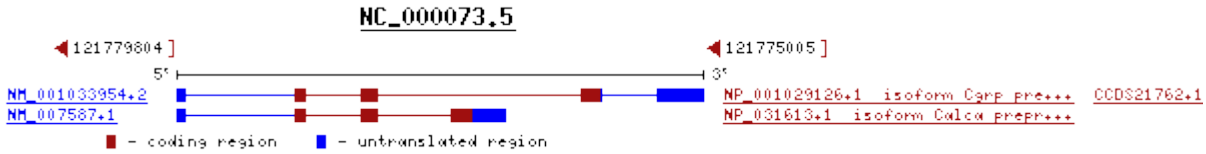
Seleziona la/le affermazioni corrette relative a PubMed

1	E'un database di articoli scientifici
2	E' un database di termini medici
3	E' un database di sequenze nucleotidiche
4	E' un database di sequenze proteiche
5	E' un database di riviste scientifiche
6	Consente di effettuare ricerche per similarità
7	E' un database di termini scientifici

(fare lo stesso per MeSh, Medline)

Nella figura sottostante sono rappresentate due isoforme dello stesso gene?
Nella prima isoforma manca l'esone 4?
L'esone 4 contiene una sequenza che può determinare la fine della trascrizione?
Le isoforme sono dovute a splicing alternativo?
Il filamento codificante è il forward ('+') oppure il reverse ('-')

Calcitonin/calcitonin-related polypeptide, alpha



Relative sequenze in formato FASTA:

MM_001033954:Mus musculus calcitonin/calcitonin-related polypeptide, alpha (Calca), transcript variant 2

TCAAGAGTCACCGCTTCGCAAGCACTGCCTGGCTCCATCAGGATCCCCGCAGGCTCAGCTCCAAGGCACCGCTCACCAGGAAGGCATCATGGGCTTCCTG
AAGTTCTCCCCTTTCCTGGTTGTCAGCATCTTGCTCCTGTACCAGGCATGCAGCCTCCAGGCAGTGCCTTTGAGGTCAATCTTGGAAAGCAGCCCAGGCA
TGGCCACTCTCAGTGAAGAAGAAGTTCGCCTGCTGGCTGCACTGGTGCAGGACTATATGCAGATGAAAGCCAGGGAGCTGGAGCAGGAGGAAGAGCAGGA
GGCTGAGGGCTCTAGTGTCACTGCTCAGAAGAGATCCTGCAACACTGCCACCTGTGTGACCCATCGGCTGGCAGGTCTGCTGAGCAGATCAGGAGGTGTG
GTGAAGGACAACTTTGTTCCCACCAATGTGGGCTCTGAAGCCTTCGGCCGCCGCCGCAGGGACCTTCAGGCCTGAACAGATAACAGCCCCAGAATGAAGG
TTACACAATAAAGATAAACTCTAATTCTATTCATGTATAATTAAAGTTATGTATAAGAAAGGCTGATGAAAGACACATATATTTGCATCCTTCTTAGTAT
TGAAAAACCCTTCTCCCTTTGACAGGAGCTAAAGCTAAGTGCAGAATAAGTTTGCCTATTGTGCATCGTGTTGTATGTGACTCTGTATCCAATAAACATG
ACAGCATGGTTCTGGCTTATCTGGTAGCAAATATGGTCCCCATAAACCATCCTGTTGATGTTGATGACTCTGCTAAACCTCAAGGGGATATGAAACACTG
CCTCTTGCTCTTCTGGGGACACATGGTAATGTTGTGACTCAATGGAACCATATGCTTAAAGAACTCTTAATATTGTCACTTGTGAATGTAATCAAAATTA
AAAACAAATGTATTTCATAAAAAAAAAAAAAAAAAA

NM_007587: Mus musculus calcitonin/calcitonin-related polypeptide, alpha (Calca), transcript variant 1

AAGAGTCACCGCTTCGCAAGCACTGCCTGGCTCCATCAGGATCCCCGCAGGCTCAGCTCCAAGGCACCGCTCACCAGGAAGGCATCATGGGCTTCCTGAA
GTTCTCCCCTTTCCTGGTTGTCAGCATCTTGCTCCTGTACCAGGCATGCAGCCTCCAGGCAGTGCCTTTGAGGTCAATCTTGGAAAGCAGCCCAGGCATG
GCCACTCTCAGTGAAGAAGAAGTTCGCCTGCTGGCTGCACTGGTGCAGGACTATATGCAGATGAAAGCCAGGGAGCTGGAGCAGGAGGAAGAGCAGGAGG
CTGAGGGCTCTAGCTTGGACAGCCCCAGATCTAAGCGGTGTGGGAATCTGAGTACCTGCATGCTGGGCACGTACACACAAGACCTCAACAAGTTTCACAC
CTTCCCCAAAACTTCAATTGGGGTTGAAGCACCTGGCAAGAAAAGGGATGTGGCCAAGGACTTGGAGACAAACCACCAATCCCATTTTGGCAACTAAGCT
CCTTCTCTCCTTTCTAGTTTCCTTCTTGCTTTCTTCCTATAACTTGATGCATGTAGTTCCTCTCTGGTTGCTCTCCAGGCTATTACTGGTTGCTTTCCTG
AGGCAAAGAATGGTATCTGAAATCCCCAGTGGGTGAGGAGAAAGTCCCACAGGCTAAAAGAGAATCACCCAGGAAGATGGCAGAGAGCAAGGGCACACTC
AGGAAGATGGCAGAGAGCAAGGGCAGTCATCTGGCTTCCTAGTAGAGCTTCTAGTCTTGCTTCTGGAAGTGTTGGTTGTTTGGGAAATAAAACTATTTTT
TAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Swiss-Prot :
È un database primario di sequenze di acidi nucleici accuratamente controllate?
È un database primario di sequenze proteiche accuratamente controllate?
È un database primario di sequenze proteiche ottenute automaticamente dalle sequenze di mRNA?
È un database derivato di particolari sequenze proteiche?
È ordinato secondo le funzioni proteiche?

Trembl:
È un database primario di sequenze di acidi nucleici accuratamente controllate?
È un database primario di sequenze proteiche accuratamente controllate?
È un database primario di sequenze proteiche ottenute automaticamente dalle sequenze di mRNA?
È un database derivato di particolari sequenze proteiche?
È ordinato secondo le funzioni proteiche?

Entrez ed SRS:
Sono dei database primari di molecole biologici?
Entrez è gestito dall'NCBI?
SRS è gestito dall'EMBL (EBI)?
Sono due sistemi integrati che permettono di ricavare particolari dati biologici (sequenze proteiche, nucleotidiche domini conservati ecc)?

Il database 'nucleotide' dell'NCBI :
-È in formato flat-file testuale?
-È in formato flat-file tipo tabella?
-È un database relazionale?
- i campi sono contraddistinti dagli 'headers' ?
- possono contenere crosslink con altri database?
- la sigla 'CDS' indica i due nucleotidi da cui inizia e finisce la sequenza codificante?
- la sigla 'CDS' indica l'inizio e la fine di un introne?
- la diciture ' mol_type="mRNA" ' indica che la molecola a cui si riferisce la sequenza è stata ricavata da una sequenza di mRNA?

La sequenza qui rappresentata è in formato FASTA?

```
ORIGIN
        1 tcaagagtca ccgcttcgca agcactgcct ggctccatca ggatccccgc aggctcagct
       61 ccaaggcacc gctcaccagg aaggcatcat gggcttcctg aagttctccc ctttcctggt
      121 tgtcagcatc ttgctcctgt accaggcatg cagcctccag gcagtgcctt tgaggtcaat
      181 cttggaaagc agcccaggca tggccactct cagtgaagaa gaagttcgcc tgctggctgc
      241 actggtgcag gactatatgc agatgaaagc cagggagctg gagcaggagg aagagcagga
      301 ggctgagggc tctagtgtca ctgctcagaa gagatcctgc aacactgcca cctgtgtgac
      361 ccatcggctg gcaggtctgc tgagcagatc aggaggtgtg gtgaaggaca actttgttcc
      421 caccaatgtg ggctctgaag ccttcggccg ccgccgcagg gaccttcagg cctgaacaga
      481 taacagcccc agaatgaagg ttacacaata aagataaact ctaattctat tcatgtataa
      541 ttaaagttat gtataagaaa ggctgatgaa agacacatat atttgcatcc ttcttagtat
      601 tgaaaaaccc ttctcccttt gacaggagct aaagctaagt gcagaataag tttgcctatt
      661 gtgcatcgtg ttgtatgtga ctctgtatcc aataaacatg acagcatggt tctggcttat
      721 ctggtagcaa atatggtccc cataaaccat cctgttgatg ttgatgactc tgctaaacct
      781 caaggggata tgaaacactg cctcttgctc ttctggggac acatggtaat gttgtgactc
      841 aatggaacca tatgcttaaa gaactcttaa tattgtcact tgtgaatgta atcaaaatta
```

```
901 aaaacaaatg tatttcataa aaaaaaaaaa aaaaaa
```

RefSeq è:

- un database di sequenza nucleotidiche primario
- un database con tutte le sequenze di trascritti conosciute
- un database con tutte le sequenze rappresentate una sola volta (genomiche, RNA)
- un database con tutte le sequenze dei trascritti rappresentate una sola
- sono rappresentati anche gli introni
- sono rappresentate anche le proteine

Unigene è:

- un database di sequenze nucleotidiche primario
- un database con tutte le sequenze di trascritti conosciute
- un database con tutte le sequenze rappresentate una sola volta (genomiche, RNA)
- un database con tutte le sequenze dei trascritti rappresentate una sola
- sono rappresentati anche gli introni

Altri database da ricordare:

- Protein
- Gene
- Pfam

Gene Ontology:
- è un database di malattie genetiche
- è un database di articoli scientifici
- è un database di sequenze proteiche non ridondanti
- è un database di termini biologici per uniformare il modo di descrivere i processi biologici, le funzioni molecolari e le localizzazioni cellulare?

Il database OMIM:
Contiene i nomi ufficiali dei geni?
Sono archiviate e descritte le malattie genetiche conosciute?
Sono archiviate le simbologie dei geni e le loro funzioni?

Dominio proteico:
- Un dominio proteico è una parte di una proteina che rappresenta una regione funzionale con una propria struttura tridimensionale?
- Un dominio è un insieme di proteine cha hanno la medesima funzione?
- SwissProt è un database di domini proteici?
- Pfam è un database di domini proteici?
- Uno stesso dominio proteico può essere presente in differenti proteine
- Gli stessi Domini proteici presenti in differenti proteine hanno necessariamente la stessa sequenza nucleotica o la stessa sequenza proteica?

I database di proteine :
- SwissProt contiene solo sequenze ottenute con il sequenziamento diretto
- Trembl contiene solo sequenze ottenute con il sequenziamento diretto
- Trembl contiene solo sequenze ottenute con la traduzione automatica delle sequenze degli mRNA?

- SwissProt contiene solo sequenze ottenute con la traduzione automatica delle sequenze degli mRNA?
- SwissProt contiene sequenze proteiche che sono state attentamente vagliate?

Una sostituzione di un nucleotide in una sequenza codificante:
- fa cambiare sempre il relativo aminoacido?
- Sono necessari almeno due mutazioni affiancate per cambiare aminoacido?
- Sono necessari almeno tre mutazioni affiancate per cambiare aminoacido?

Esistono aminoacidi codificati da 4 differenti codoni?
Esistono aminoacidi codificati da 2 soli differenti codoni?
Esistono aminoacidi codificati da un solo codone?
Tutti i codoni codificano per gli aminoacidi?
Quanti codoni segnalano la fine della traduzione?
Quanti codoni esistono per la Metionina

Una mutazione nucleotidica può far variare la sequenza proteica se avviene:
- nella regione 3'UTR?

- nella regione CDS?
- nella regione 5'UTR?

In una sequenza codificante un gene, quali di questi tipi di mutazione (nella regione codificante) hanno maggior probabilità di generare, nella traduzione, una proteina non 'funzionante'
- una mutazione puntiforme (sNPS)
- una inserzione di un solo nucleotide
- una inserzione di due nucleotidi
- una inserzione di tre nucleotidi
- una delezione di un solo nucleotide
- una delezione di due nucleotidi
- una delezione di tre nucleotidi

---

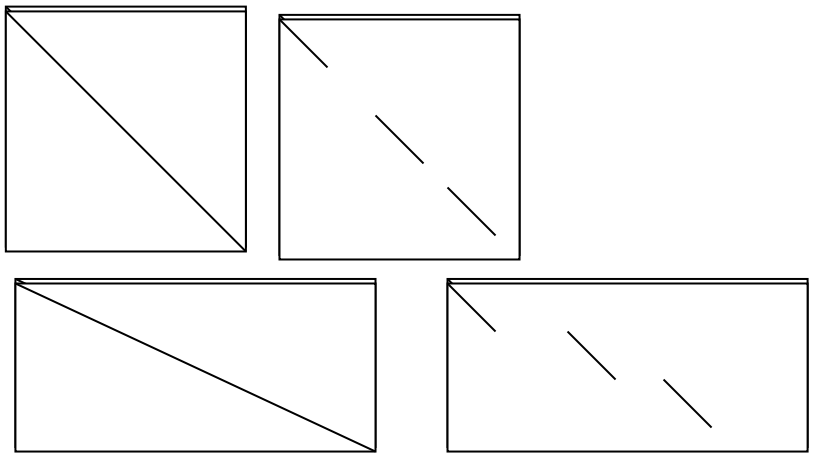
omologia: quali frasi sono giuste:

- considerando che l'uomo condivide il 98% del genoma con lo scimpanzé,
  - è giusto dire che lo scimpanzé è omologo all'uomo al 98%
  - nei due organismi, sicuramente, quasi tutti i geni sono omologhi

- due geni omologhi hanno probabilmente la sequenza simile
- due geni omologhi hanno sicuramente la stessa sequenza
- due organi con funzioni simili, sono sicuramente omologhi

Similarità:

Allineando la sequenza di un gene (comprese le sequenze introniche) con la relativa sequenza dell' mRNA maturo Quale matrice di allineamento si ottiene:



Similarità:

```
ACTTCACCCTAGCTTTCTCGACTGCAGAGAAACTAGCCGA
CAGCACGTGGCTTACTCACTACCAGTTCTCACAGAATGCA
Le due sequenze hanno ambedue 11 A, 9 T, 13 C, 7 G;
```
possono essere considerate simili?

Un allineamento di sequenze nucleotidiche, per essere considerato casuale deve aver all'incirca:

- 10% di identità
- 25% identità
- 50% di identità

Due sequenze proteiche 'casuali' allineate hanno mediamente una identità del:

- 1 %
- 5 %
- 10 %
- 25 %

Per valutare, con l'allineamento, se due sequenze proteiche possono avere funzioni simili;

- è sufficiente determinare il grado di identità ?
- Nell'allineamento devo utilizzare delle appositi matrici per valutare l'effetto di una eventuale sostituzione aminoacidica?

Ricerche testuali e per similarità:

Nelle ricerche per similarità eseguite con l'allineamento si usano gli operatori booleani (AND OR NOT)?
Le ricerche testuali sono in grado di selezionare anche termini simili ?
Le ricerche testuali evidenziano solo i record che contengono esattamente i termini cercati?
Per ricercare similarità tra sequenze è necessario utilizzare appositi programmi?

C'è la necessità di ricercare omologie tra sequenze. Che tipo di ricerca effettuerebbe?

- tipo testuale
- di similarità

I gap negli allineamenti di sequenza sono dovuti a mutazioni tipo:

- sostituzioni?
- delezioni?
- Inversioni?
- Inserzioni?

Le sequenze di geni omologhi, nei differenti organismi, sono maggiormente conservate:
-Negli esoni?
-Negli introni?
-Nelle regioni intergeniche?
-Nelle sequenze relative a domini funzionali?

Blast: Affermazioni (vere e false)
-Blast effettua lo scorrimento della sequenza query con tutte le sequenza presenti nel database
-Blast applica il metodo dot-matrix
-Blast è basato sulla indicizzazione di corte parole
-Blast valuta l'allineamento in modo globale oppure,
-Blast valuta l'allineamento in modo locale
-L'allineamento di due sequenze mediante scorrimento usa un algoritmo più complesso rispetto a quello adottato da Blast (ricordo che la complessità è in relazione al numero di operazioni che l'algoritmo deve applicare per arrivare alla soluzione)

---

differenti tipologie di figure di blast che possono essere mostrate(una sola figura)
figura A)

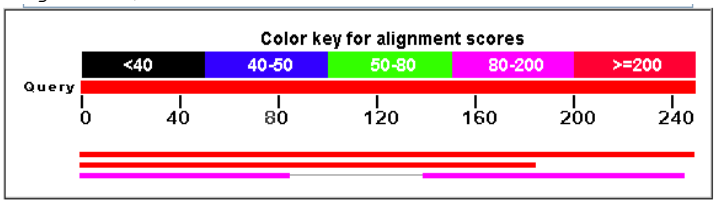
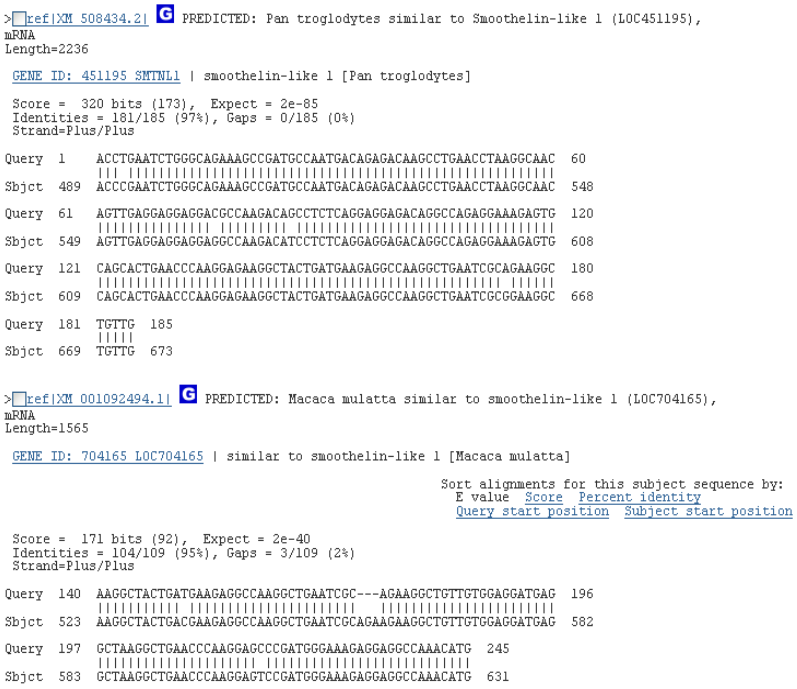


figura B)

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
NM_001105565.1	Homo sapiens smoothelin-like 1 (SMTNL1), mRNA	460	460	100%	9e-128	100%	U G
XM_508434.2	PREDICTED: Pan troglodytes similar to Smoothelin-like 1 (LOC451195), mRNA	320	320	74%	2e-85	97%	G
XM_001092494.1	PREDICTED: Macaca mulatta similar to smoothelin-like 1 (LOC704165), mRNA	171	312	76%	2e-40	96%	G

figura C)



Domande possibili:
- è evidente l'allineamento grafico fornito da blast
- è evidente il listato delle sequenza 'simili
- sono evidenti i dettagli degli allineamenti
- gli allineamenti con 'expect' (corrisponde anche ad 'e-value') maggiore sono più simili
- gli allineamenti con 'score' maggiore sono più simili

Figura con un allineamento di blast tra due sequenze (esempi) (viene data una sola delle due figure)

Figura 1:

```
 Score =  372 bits (498), Expect = e-108
 Identities = 249/249 (100%)
 Strand = Plus / Plus


Query: 1   acctgaatctgggcagaaagccgatgccaatgacagagacaagcctgaacctaaggcaac 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1   acctgaatctgggcagaaagccgatgccaatgacagagacaagcctgaacctaaggcaac 60


Query: 61  agttgaggaggaggacgccaagacagcctctcaggaggagacaggccagaggaaagagtg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 61  agttgaggaggaggacgccaagacagcctctcaggaggagacaggccagaggaaagagtg 120


Query: 121 cagcactgaacccaaggagaaggctactgatgaagaggccaaggctgaatcgcagaaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 121 cagcactgaacccaaggagaaggctactgatgaagaggccaaggctgaatcgcagaaggc 180


Query: 181 tgttgtggaggatgaggctaaggctgaacccaaggagcccgatgggaaagaggaggccaa 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 181 tgttgtggaggatgaggctaaggctgaacccaaggagcccgatgggaaagaggaggccaa 240


Query: 241 acatggtgc 249
           |||||||||
Sbjct: 241 acatggtgc 249
```

Figura 2)

```
 Score =  353 bits (472), Expect = e-102
 Identities = 244/249 (97%), Gaps = 2/249 (0%)
 Strand = Plus / Plus


Query: 1   acctgaatctgggcagaaagccgatgccaatgacagagacaagcctgaacctaaggcaac 60
           |||||||||||||||||| ||||||  ||||||||||||||||  ||||||||||||
Sbjct: 1   acctgaatctgggcagaaacccgatgaaaatgacagagacaagcct--acctaaggcaac 58


Query: 61  agttgaggaggaggacgccaagacagcctctcaggaggagacaggccagaggaaagagtg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 59  agttgaggaggaggacgccaagacagcctctcaggaggagacaggccagaggaaagagtg 118


Query: 121 cagcactgaacccaaggagaaggctactgatgaagaggccaaggctgaatcgcagaaggc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 119 cagcactgaacccaaggagaaggctactgatgaagaggccaaggctgaatcgcagaaggc 178


Query: 181 tgttgtggaggatgaggctaaggctgaacccaaggagcccgatgggaaagaggaggccaa 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 179 tgttgtggaggatgaggctaaggctgaacccaaggagcccgatgggaaagaggaggccaa 238


Query: 241 acatggtgc 249
           |||||||||
Sbjct: 239 acatggtgc 247
```

domande possibili:

- l'identità è del 100%
- non ci sono mismatch tra le due sequenze
- si evidenziano 'x' gap
- si evidenziano 'x' mismatch (disappaiamenti)

Allineamenti con blast
Volendo allineare con blast una sequenza nucleotidica contro un database di sequenze proteiche

- è necessario tradurre nel corretto frame di lettura la sequenza nucleotidica e poi lanciare blast
- è necessario tradurre in tutti i sei possibili frame di lettura e poi lanciare blast con le differenti sequenze proteiche così ottenute

- lanciando semplicemente blastp (che allinea sequenze proteiche) si può ottenere l'allineamento voluto
- lanciando semplicemente blastn: il programma automaticamente allinea anche le sequenze proteiche presenti nel database
- è necessario lanciare blastx che traduce preventivamente la sequenza nucleotidica in sequenza proteica

---

BLAST:
-esiste un programma di blast che allinea una sequenza proteica contro un database nucleotidico, trasformando preventivamente la sequenza proteica in sequenza nucleotidica.
-Esiste un programma di blast che allinea una sequenza nucleotidica contro un database nucleotidico traducendo preventivamente le sequenza nucleotidiche in sequenze proteiche in tutti i frame possibili

---

Domanda 'subdola':
Quante possibili sequenze nucleotidiche codificanti una determinata proteina si possono ottenere partendo dalla stessa sequenza proteica?
-una
-tre
-sei
-più di sei

Sostituzioni aminoacidiche
-Nell'evoluzione tutte le sostituzione aminoacidica hanno la stessa importanza
-Esistono aminoacidi con caratteristiche simili che possono essere facilmente scambiati durante l'evoluzione
-Dalle matrici di sostituzione (PAM o BLOSUM) si può ricavare la facilità con cui possono essere scambiati, nell'evoluzione, gli aminoacidi all'interno di una proteina

---

figura di UCSC Genome Browser (vedere la figura dell'esercitazione)
possibili domande:
-Quale o quali geni sono evidenti nell'allineamento (suggerimento fare riferimento alla riga 'known genes based on Swiss-Prot, Trembl, Refseq')
- quanti esoni relativi al gene 'xxxx' sono evidenti?
- quanti introni relativi al gene 'xxxx' sono evidenti?
- sono evidentemente rappresentate delle sequenze est?
- sono rappresentate anche le sequenze di RNA messaggeri
- quale è la direzione della sequenza codificante ('+','forward',senso oppure '-',reward,antisenso) ?
- la riga 'Conservation' mostra le regioni codificanti?
- la riga 'Conservation' mostra le sequenze conservate tra i geni omologhi nei differenti organismi?
- la riga 'Conservation' mostra le sequenze conservate tra i geni omologhi all'interno dello stesso organismo?
- i picchi neri più alti, della riga 'Conservation', dimostrano le regioni che risultano maggiormente conservate dei geni omologhi nei differenti organismi?

Affermazioni (vere e false) relative ad UCSC Genome-Browse
- Per posizionarsi sul genoma si deve conoscere la localizzazione esatta (cromosoma e posizione)?
- Si può posizionarsi e vedere l'allineamento delle sequenze conoscendo solo il nome del gene.
- Si può posizionarsi sul genoma e vedere l'allineamento delle sequenze conoscendo solo 'accession number' di una sequenza nucleotidica
- sono presenti link al database di malattie genetiche
- sono presenti link al database di articoli scientifici
- è possibile recuperare l'intera sequenza genomica del gene voluto
- fornisce solo la sequenza codificante del gene voluto
- si può esplorare solo il genoma umano
- si può scegliere di 'navigare' in differenti genomi

Questa NON è una domanda d'esame, abbiamo trovato un esempio simile durante una esercitazione, una riflessione sull'argomento…. non fa mai male:
- Abbiamo un allineamento fatto con blast di una sequenza di mRNA (eucariote) contro un database di sequenze genomiche.
Il possibile all'allineamento tra la sequenza query e la sequenza genomica si presenta come:

- una unico allineamento continuo che copre tutta la sequenza dell'mRNA
- più allineamenti discontinui che comunque coprono tutta la sequenze query

suggerimento: