

LA BIOINFORMATICA

Tutto cominciò attorno al 1990 quando si è assistito a una crescita parallela della ricerca genetica e delle potenzialità dei computer. Con il contributo degli elaboratori elettronici, infatti, si è ottenuto in tempi brevi il completamento del *Human Genome Project*, conclusosi nel 2000 con il sequenziamento dell'intero genoma umano. Sorprendentemente, si scoprì che gli esseri umani hanno un numero di geni di gran lunga inferiore rispetto a quanto era stato previsto. In effetti, per quanto il numero preciso di geni non sia ancora noto, il genoma umano comprende forse solo 20 000-25 000 geni e non 50 000-100 000 come si credeva in passato. Ciò significa che i nostri geni sono solo circa due volte più numerosi di quelli di un moscerino della frutta (*Drosophila melanogaster*, 13 600 geni).

Produrre la sequenza di un intero genoma è stato come assemblare un enorme puzzle, milioni di piccoli pezzi dovevano essere composti in una sequenza ordinata. Il genoma umano è costituito infatti da più di 3 miliardi di nucleotidi (le unità base che costituiscono il DNA) e per essere sequenziato completamente, deve essere analizzato prima in piccoli pezzi e poi ricomposto, utilizzando le informazioni presenti alle estremità dei differenti frammenti di DNA; a rendere la cosa ancora più complessa, i «pezzi» di cui è composto il puzzle sono stati analizzati da diversi laboratori sparsi in tutto mondo.

Una volta disponibile la sequenza del genoma, resta a tutt'oggi ancora da completare il compito più difficile, cioè la sua interpretazione. Per fare ciò è necessario analizzare le sequenze disponibili utilizzando un approccio multidisciplinare che comprenda sia le scienze della vita (biologia molecolare, genetica molecolare, biochimica delle proteine, ecc.) sia le scienze matematico-statistiche e l'informatica. In sintesi è nata una nuova branca della scienza che viene chiamata «bioinformatica».

La bioinformatica è quindi una nuova disciplina che unisce competenze in campo biologico, biochimico e medico con competenze matematiche, statistiche e informatiche; i dati inseriti negli archivi elettronici (sequenze di DNA, di RNA o di aminoacidi) chiamati anche *database*, di facile reperibilità per la comunità scientifica, sono ottenuti dalla ricerca in laboratorio *in vivo* (cioè su cavie vive) o *in vitro* (cioè in provetta);

gli stessi dati vengono poi rielaborati al computer per ottenere previsioni (dati *in silico*, cioè nei chip di silicio) che devono essere comunque verificate in laboratorio.

Molteplici sono gli scopi della bioinformatica tra cui la possibilità di potere ricostruire le tappe evolutive delle varie specie, incluso l'uomo, e la possibilità di ricavare le caratteristiche delle biomolecole utili nella progettazione di nuovi farmaci. Grazie alla bioinformatica la ricerca può essere direzionata e mirata, con notevole riduzione nei costi attuativi.

Per diventare bioinformatici è necessario laurearsi in medicina o in biologia e seguire poi un master di altri 3 anni di specializzazione (per esempio, presso l'Università di Trento o l'Università Bicocca di Milano); è dunque un percorso di studi lungo e difficile, ma di grande soddisfazione e con evidenti ampi sbocchi professionali e possibilità di impiego ad alti livelli.

Le principali **banche dati biologiche** sono:

- EMBL datalibrary (Europa);
- GenBank (USA);
- DDBJ (Giappone).

Fra le tre banche dati è stato stipulato un accordo internazionale per cui il contenuto dei dati di sequenza presenti è quasi del tutto coincidente e le informazioni vengono scambiate giornalmente. Qualsiasi ricercatore può depositare, attraverso un apposito sistema *on line*, le proprie sequenze di DNA, RNA o aminoacidiche, che vengono controllate e inserite nella banca dati. Le banche dati che contengono sequenze nucleotidiche e aminoacidiche vengono dette **banche dati di primo livello**.

Esistono poi le **banche dati specializzate** che raccolgono informazioni specifiche (per esempio banche dati di enzimi di restrizione, banche dati di mutazioni, ecc.). Alcune di queste possono essere estremamente specializzate (per esempio banche dati che raccolgono le informazioni su uno specifico gene), altre invece contengono informazioni più ampie (per esempio banche dati di strutture tridimensionali di proteine, banche dati di malattie genetiche o di pubblicazioni scientifiche).

In questo elenco alcuni indirizzi tramite cui accedere a banche dati biologiche:

www.ncbi.nlm.nih.gov/ (primo livello)
<http://smart.embl-heidelberg.de/> (specializzata)
www.expasy.org/sprot/sprot-top.html (specializzata)
www.embl-heidelberg.de/ (primo livello)
<http://genome.ucsc.edu/> (primo livello)
www.ensembl.org/ (primo livello)

Come funzionano le banche dati

Ciascuna banca dati è organizzata attorno a un elemento centrale (nelle banche dati di sequenze nucleotidiche questo è per esempio la sequenza di DNA o di RNA) che di solito viene immesso dal ricercatore sulla base di esperimenti condotti in laboratorio. Attorno all'elemento centrale viene costruita la *entry*, ossia la richiesta alla banca dati, che può comprendere tutte le annotazioni utili a classificare (per esempio il nome della sequenza, la specie a cui appartiene) e a caratterizzare (per esempio la funzione della sequenza, le referenze bibliografiche) l'elemento stesso.

Utile e importante è considerare l'*allineamento* tra la «nuova» sequenza nucleotidica e quelle già presenti in banca dati.

L'*allineamento* tra sequenze nucleotidiche o aminoacidiche di una stessa specie o di specie diverse (anche molto distanti filogeneticamente) consente di mettere in luce l'esistenza di similarità di sequenza, che vengono «misurate» in base alla percentuale di identità fra le due (o più) sequenze allineate.

Gli scopi per cui è utile analizzare la similarità fra due sequenze nucleotidiche o aminoacidiche sono molteplici:

- identificazione di regioni conservate: domini funzionali nelle proteine, sequenze di regolazione dell'espressione genica;
- attribuzione di una funzione e/o identificazione di un nuovo gene/proteina;
- *homology modelling*, ovvero elaborazione di una struttura tridimensionale per una proteina in base alla omologia con proteine di cui la struttura 3D sia già stata determinata;
- classificazione dei geni in famiglie;
- studi di filogenesi molecolare, per ricostruire le relazioni evolutive fra le specie;
- studi di genetica di popolazione (migrazioni delle popolazioni umane, relazioni fra i vari gruppi umani);

- identificazione di mutazioni mediante confronto fra la sequenza mutata e la sequenza di riferimento o *wild type* (ovvero la sequenza normalmente presente in natura).

Navighiamo tra i genomi

In queste attività impareremo a esplorare il genoma umano utilizzando le informazioni contenute in alcune delle banche dati biomediche disponibili *on line*. Scopriremo le corrispondenze esistenti fra i cromosomi umani e quelli di altre specie animali e impareremo a trovare un gene all'interno del genoma e a scoprirne la struttura e la funzione nonché le malattie associate al suo malfunzionamento. È da segnalare che le banche dati bioinformatiche sono tutte in inglese. Il sito da cui partiremo per la nostra navigazione tra i genomi di specie diverse è www.ensembl.org.

Ensembl (un gioco di parole fra *ensemble* cioè «insieme» e EMBL la sigla del *European Molecular Biology Laboratory*) è un progetto sviluppato in collaborazione tra *Sanger Center* (uno dei più importanti centri di ricerca sul genoma a Cambridge) e EMBL-EBI (*European Bioinformatics Institute*) per sviluppare un sistema *software* di *annotazione* automatica dei genomi animali. Con il termine *annotazione* si intende l'inserimento di tutte le informazioni riguardanti la funzione di una determinata sequenza.

OBIETTIVI DELLE ESERCITAZIONI

- Saper consultare banche dati di interesse biomedico.
- Apprendere l'utilizzo di alcuni strumenti bioinformatici.
- Sapere confrontare genomi di organismi differenti.
- Saper correlare fra loro geni e malattie.

LEZIONI
Esercitazioni di bioinformatica

PER INIZIARE

Effettuate un collegamento a internet e aprite un *browser* di navigazione (per esempio *Internet Explorer* o *Firefox*). Visitate il sito all'indirizzo www.ensembl.org/

Nella Home Page si possono osservare *link* ai genomi di diverse specie animali. Attualmente sono note le sequenze di molte più specie rispetto a quelle qui elencate, sia di organismi procarioti che eucarioti; *Ensembl* tuttavia prende in considerazione soprattutto le sequenze degli organismi del regno animale. I genomi degli animali sono particolarmente importanti perché sono molto simili al nostro, perciò possono essere utilizzati come modelli più semplici per eseguire degli esperimenti che ci aiutino a capire più approfonditamente la biologia umana e le malattie. È importante prestare attenzione alla versione del *database*, che viene aggiornata ogni due mesi; in alto sulla sinistra si può trovare la data della versione più recente che si apre alla pagina iniziale, mentre in basso nella schermata, sulla sinistra, rimangono archiviate tutte le versioni passate, che si possono comunque usare.

Ensembl Archive

View previous release of page in Archive!

- e! 41: Oct 2006
- e! 40: Aug 2006
- e! 39: Jun 2006
- e! 38: Apr 2006
- e! 37: Feb 2006
- e! 36: Dec 2005
- e! 35: Nov 2005
- e! 34: Oct 2005

ESERCITAZIONE 1

Esploriamo il genoma umano

Per iniziare l'esplorazione del genoma umano, fate ora clic sulla casella *Homo sapiens* e si aprirà una pagina al link www.ensembl.org/Homo_sapiens/.

Osservando la pagina che si apre possiamo raccogliere informazioni, aggiornate ad agosto 2006, sul cariotipo dell'uomo: numero dei cromosomi autosomici, tipo di cromosomi sessuali, presenza di DNA mitocondriale (MT), posizione del centromero e classificazione dei cromosomi in base alla loro forma, presenza di bandeggi, numero di geni conosciuti, numero di paia di basi che li costituiscono, ecc.

Karyotype

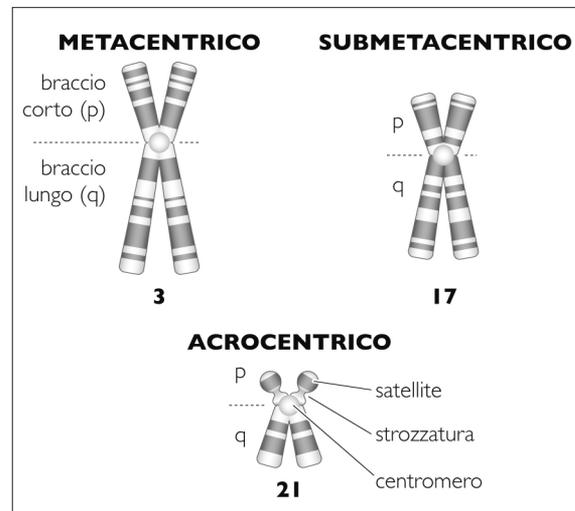
Click on a chromosome for a closer view



Statistics

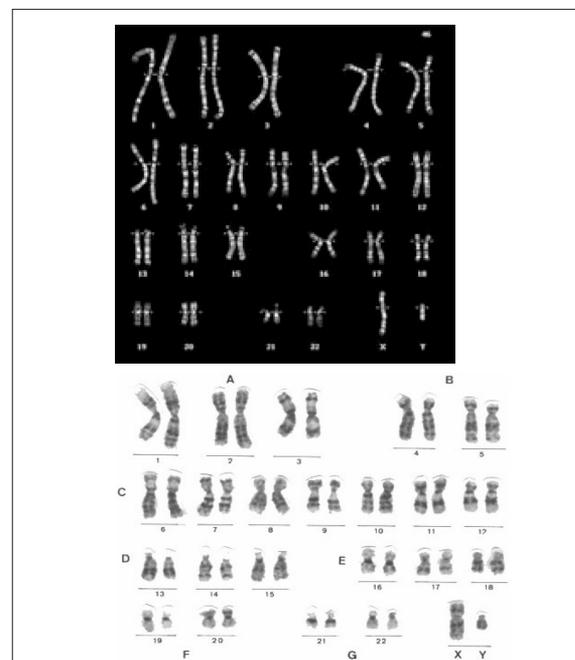
| | |
|---------------------------|-------------------|
| Assembly: | NCBI 36, Oct 2005 |
| Genebuild: | Ensembl, Aug 2006 |
| Database version: | 42.36d |
| Known genes: | 21,774 |
| Novel genes: | 1,036 |
| Pseudogenes: | 1,069 |
| RNA genes: | 3,994 |
| Genscan gene predictions: | 69,185 |
| Gene exons: | 270,661 |
| Gene transcripts: | 44,676 |
| Variations: | 11,287,592 |
| Base Pairs*: | 3,253,037,807 |
| Golden Path Length**: | 3,093,120,360 |

Osservate uno schema utile per classificare un cromosoma in base alla posizione del centromero, sotto il disegno è riportato il numero del cromosoma.



Cromosomi distinti in base alla posizione del centromero.

In queste due immagini al microscopio di cariotipi si possono osservare i cromosomi trattati con colorazioni specifiche per ottenere i bandeggi. Nei siti di bioinformatica di solito si ritrovano i cromosomi con le bande G, ottenute con la tecnica di colorazione con la sostanza Giemsa che si lega in modo specifico alle zone del cromosoma che contengono una percentuale maggiore di Citosina e di Guanina che diventano chiare.



Osservate anche la seguente tabella che riporta i numeri di cromosomi tipici di alcune specie che spesso si trovano nelle banche dati bioinformatiche.

| Nome della specie | N. cromosomi autosomici | Cromosomi sessuali |
|---------------------------------|-------------------------|--------------------|
| <i>Homo sapiens</i> | 44 | XY |
| <i>Mus musculus</i> | 38 | XY |
| <i>Rattus norvegicus</i> | 40 | XY |
| <i>Sus scrofa</i> | 36 | XY |
| <i>Drosophila melanogaster</i> | 6 | XY |
| <i>Saccharomyces cerevisiae</i> | 32 | |
| <i>Caenorhabditis elegans</i> | 10 | XX |
| <i>Canis familiaris</i> | 76 | XY |
| <i>Gallus gallus</i> | 76 | WZ |
| <i>Bos taurus</i> | 58 | XY |
| <i>Felis catus</i> | 36 | XY |
| <i>Macaca mulatta</i> | 40 | XY |
| <i>Ovis aries</i> | 52 | XY |
| <i>Equus caballus</i> | 62 | XY |
| <i>Pan troglodytes</i> | 46 | XY |
| <i>Zea mays</i> | 20 | |
| <i>Oryza sativa</i> | 24 | |
| <i>Arabidopsis thaliana</i> | 10 | |
| <i>Avena sativa</i> | 42 | |

Tornando alla pagina iniziale (cliccando sulla «e» di Ensembl), per imparare a navigare e a confrontare i genomi dei diversi animali, proviamo a rispondere alle seguenti domande relative alla specie *Homo sapiens*.

1. Quanti cromosomi ha l'uomo?
2. Quali cromosomi determinano il sesso?
3. Il centromero ha sempre la stessa posizione lungo il cromosoma?
4. Cosa significa la sigla MT posta sotto l'ultimo ideogramma di cromosoma?
5. Cosa significano i bandeggi neri, grigi e bianchi lungo i cromosomi?

Provate a rispondere alle stesse domande e a completare la tabella che segue per altre specie animali quali: scimpanzè (*Pan troglodytes*), topo (*Mus musculus*), cane (*Canis familiaris*), gallo (*Gallus gallus*), lievito (*Saccharomyces cere-*

visiae), moscerino della frutta (*Drosophila melanogaster*). Per navigare cliccate ogni volta sull'immagine dell'organismo o sul suo nome e osservate il suo cariotipo. Ricordate che tutti gli organismi in elenco sono diploidi e non sempre sono elencati tutti i loro cromosomi, possono esserci dei salti nella numerazione poiché sono raffigurati solo quelli su cui si è scoperto qualcosa.

- Infine alcune domande per riflettere su quanto appreso.
6. È importante il numero dei cromosomi nell'evoluzione?
 7. Lo scimpanzè ha lo stesso numero di cromosomi dell'uomo?
 8. Il sesso è sempre definito da X e Y?
 9. Il cromosoma mitocondriale è sempre presente?
 10. Il bandeggio è presente in tutte le specie e quindi è sempre significativo e utile?

| | UOMO | SCIMPANZÈ | TOPO | GALLO | CANE | LIEVITO | MOSCIERINO |
|-------------------------------|-------|-----------|-------|-------|-------|---------|------------|
| N° CROMOSOMI TOTALI | | | | | | | |
| N° CROMOSOMI SESSUALI | | | | | | | |
| TIPO CROMOSOMI SESSUALI | | | | | | | |
| PRESENZA CENTROMERO | | | | | | | |
| PRESENZA POSIZIONE CENTROMERO | | | | | | | |
| PRESENZA CROMOSOMA MT | | | | | | | |
| PRESENZA BANDEGGIO | | | | | | | |

ESERCITAZIONE 2

Osserviamo i cromosomi da vicino

Sempre sul sito www.ensembl.org/ proviamo a conoscere in modo più approfondito i vari cromosomi dell'uomo addentrando nel genoma di *Homo sapiens*.

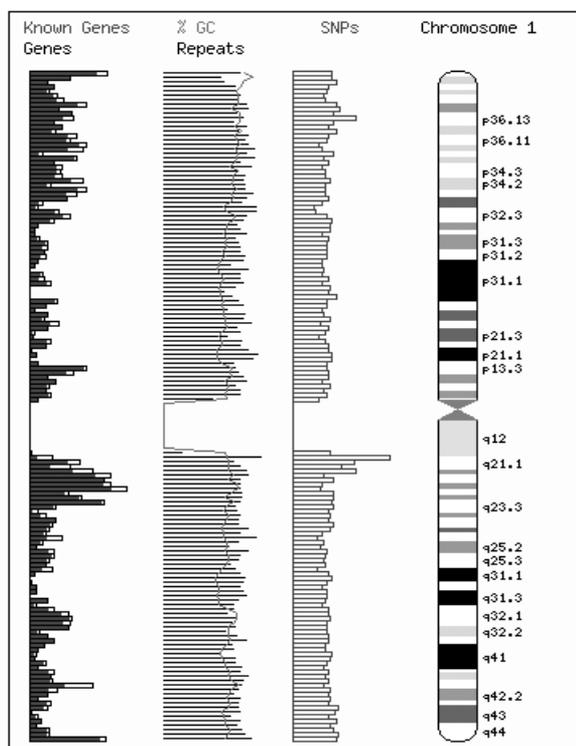
Alla pagina www.ensembl.org/Homo_sapiens/ cliccate sull'ideogramma del cromosoma 1 e osservate la figura. Sulla destra c'è l'immagine del cromosoma nella sua totalità con i bandeggi e le indicazioni delle varie zone sui due bracci in cui risulta diviso dal centromero. Si possono vedere i nomi che sono stati assegnati ad alcune delle bande, contrassegnati dalla lettera **p** (dal francese *petit* = piccolo) per il braccio corto del cromosoma e **q** (il nome deriva dall'inglese *queue* = coda, o dalla semplice conseguenza alfabetica) per l'altro braccio e con numeri crescenti dal centro verso le estremità. Le bande (che si ottengono con particolari colorazioni dei cromosomi) servono come punti di riferimento sul cromosoma per ritrovare una particolare regione e quindi un gene, e corrispondono alle coordinate di una carta geografica.

La colonna più a sinistra mostra la densità genica, ovvero quanti geni sono contenuti in una regione specifica. La parte in rosso degli istogrammi indica la frazione di geni noti sul totale dei geni previsti.

Notate che ci sono regioni con un numero elevato di geni e regioni che non ne contengono alcuno (per esempio nella zona del centromero). La colonna adiacente mostra la percentuale di sequenze di basi ripetute e (la linea in rosso) la sola percentuale di Guanina e Citosina; esiste una relazione tra il contenuto di GC ripetuto e le regioni del DNA ricche in geni. Inoltre esiste una corrispondenza tra il contenuto GC e la colorazione Geimsa utilizzata per fare risaltare il bandeggio. Infatti le regioni a basso contenuto di GC sono definite da bande scure mentre le regioni ad alto contenuto di GC sono definite da bande chiare; la colonna in blu indica la distribuzione degli **SNP (Single Nucleotide Polymorphism)**, ovvero dei polimorfismi di singolo nucleotide.

Un polimorfismo, cioè una variazione a livello di una sequenza di acidi nucleici, di un singolo nucleotide è un polimorfismo che si presenta tra individui della stessa specie, caratterizzata da una differenza a carico di un unico nucleotide.

Per esempio, se le sequenze individuate in due uomini sono AAGCCTA e AAGCTTA, è presente uno SNP che differenzia i due alleli C e T.



Chromosome 1

Length: 247,249,719 bps

For more information on gene statistics, see the [MapView help](#) page

Change Chromosome

Chromosome

Chromosome 1

Length: 247,249,719 bps

Known Protein-coding Genes: 2,173

Novel Protein-coding Genes: 2,365

Pseudogene Genes: 51

miRNA Genes: 42

rRNA Genes: 42

snRNA Genes: 178

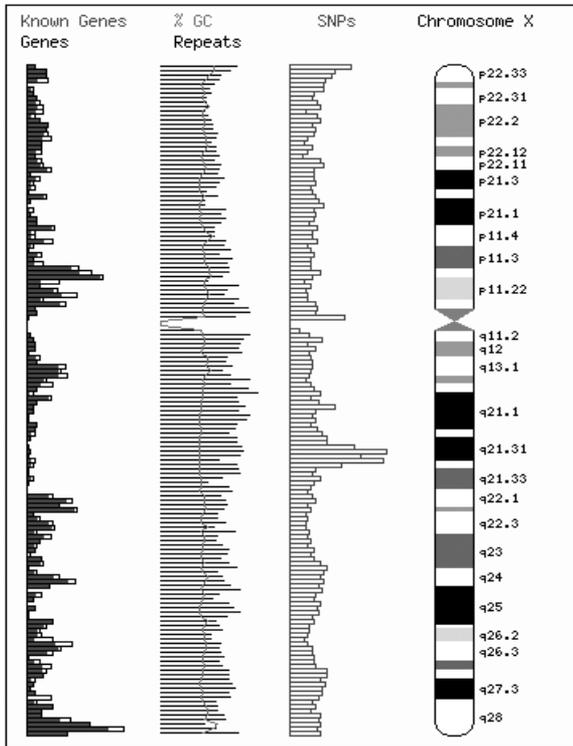
snoRNA Genes: 60

Misc RNA Genes: 93

SNPs: 936,633

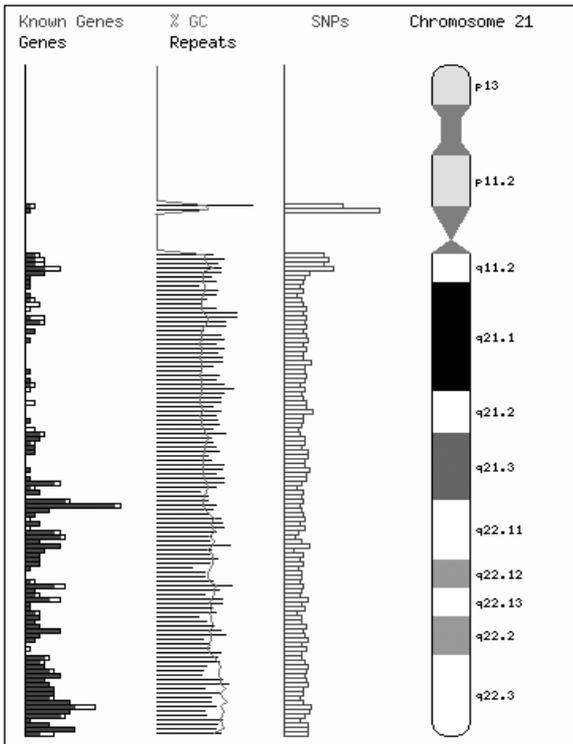
Nella parte destra della pagina di trovano le caratteristiche del cromosoma I e la possibilità di scegliere

un nuovo cromosoma da osservare ancor più nel dettaglio.



Chromosome X

| | |
|------------------------------------|-----------------|
| Length: | 154,913,754 bps |
| Known Protein-coding Genes: | 880 |
| Novel Protein-coding Genes: | 35 |
| Pseudogene Genes: | 80 |
| miRNA Genes: | 58 |
| rRNA Genes: | 19 |
| snRNA Genes: | 64 |
| snoRNA Genes: | 25 |
| Misc RNA Genes: | 48 |
| SNPs: | 413,743 |



Chromosome 21

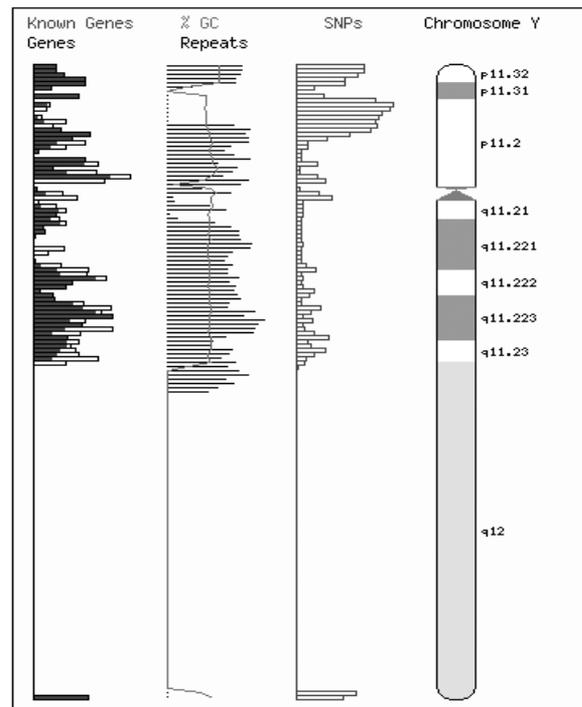
| | |
|------------------------------------|----------------|
| Length: | 46,944,323 bps |
| Known Protein-coding Genes: | 270 |
| Novel Protein-coding Genes: | 297 |
| Pseudogene Genes: | 9 |
| miRNA Genes: | 6 |
| rRNA Genes: | 3 |
| snRNA Genes: | 10 |
| snoRNA Genes: | 5 |
| Misc RNA Genes: | 6 |
| SNPs: | 163,398 |

LEZIONI
Esercitazioni di bioinformatica

| Chromosome Y | |
|-----------------------------|----------------|
| Length: | 57,772,954 bps |
| Known Protein-coding Genes: | 90 |
| Novel Protein-coding Genes: | 23 |
| Pseudogene Genes: | 2 |
| miRNA Genes: | 2 |
| rRNA Genes: | 6 |
| snRNA Genes: | 14 |
| snoRNA Genes: | 3 |
| Misc RNA Genes: | 2 |
| SNPs: | 50.447 |

La lunghezza del cromosoma viene espressa in bps cioè in paia di basi (in inglese *base pairs*).

- **Novel protein-coding genes** sono predizioni di geni al computer, che potrebbero codificare per proteine ma manca ancora la prova di laboratorio.
- **Pseudogene** sono geni che hanno subito nel corso dell'evoluzione moltissime mutazioni e hanno perso la loro capacità di codificare per proteine.
- **miRNA gene** rappresentano i geni che codificano per microRNA, molecole di RNA capaci di interferire con l'mRNA.
- **rRNA gene** rappresentano i geni che codificano per l'RNA ribosomiale.
- **snRNA gene** rappresentano i geni che codificano per RNA piccoli che stanno nel nucleo (*small nuclear RNA*).
- **snoRNA gene** rappresentano i geni che codificano per RNA piccoli che stanno nel nucleolo (*small nucleolar RNA*).



- **Misc RNA gene** rappresentano i geni che codificano per vari tipi di RNA (miscellanea di RNA).
- **SNPs** rappresentano i singoli nucleotidi che fanno differire due individui appartenenti alla stessa specie; nell'uomo si incontra un nucleotide di differenza ogni 1000.

Osservate ora il cromosoma 21, il cromosoma X e il cromosoma Y e notate la distribuzione dei bandeggi, la posizione dei centromeri e la densità genica; completate poi la tabella sottostante.

| Uomo (<i>Homo sapiens</i>) | Cromosoma 1 | Cromosoma 21 | Cromosoma X | Cromosoma Y |
|---------------------------------------|-------------|--------------|-------------|-------------|
| Lunghezza in paia di basi | | | | |
| Numero geni codificanti noti | | | | |
| Numero geni codificanti in previsione | | | | |
| Numero di SNPs | | | | |

Tornando alla pagina iniziale cliccando sulla «e» di Ensembl, fate la stessa cosa con i cromosomi dello scimpanzè e del topo e confrontate i dati ottenuti con quelli dell'uomo.

| Scimpanzè (<i>Pan troglodytes</i>) | Cromosoma 1 | Cromosoma X | Cromosoma Y |
|---------------------------------------|-------------|-------------|-------------|
| Lunghezza in paia di basi | | | |
| Numero geni codificanti noti | | | |
| Numero geni codificanti in previsione | | | |
| Numero di SNPs | | | |

È importante paragonare la struttura dei cromosomi delle varie specie poiché i genomi degli animali sono molto simili al nostro e possono essere utiliz-

zati come modelli più semplici per eseguire esperimenti per approfondire la biologia umana, lo studio delle malattie e l'uso di farmaci specifici.

| Topo (<i>Mus musculus</i>) | Cromosoma 1 | Cromosoma X | Cromosoma Y |
|---------------------------------------|-------------|-------------|-------------|
| Lunghezza in paia di basi | | | |
| Numero geni codificanti noti | | | |
| Numero geni codificanti in previsione | | | |
| Numero di SNPs | | | |

ESERCITAZIONE 3

I cromosomi conservano la loro struttura nelle diverse specie?

All'indirizzo www.ensembl.org/ possiamo scoprire se i cromosomi si mantengono inalterati nella loro struttura e se mantengono associati i geni in gruppi fissi, confrontando gli organismi tra loro e salendo nella scala evolutiva.

Alla pagina www.ensembl.org/Homo_sapiens/ cliccate sull'ideogramma del cromosoma 1 e nella parte in alto a sinistra della schermata cercate il link **View chr 1 sinteny** (ovvero osserva la sintenia del cromosoma 1). Con il termine *sintenia* si indica l'associazione di più geni su un tratto di cromosoma; in specie diverse, tratti di cromosoma contenenti gruppi di geni associati (gruppi di sintenia) risultano conservati. Per provare a confrontare i gruppi di sintenia presenti sul cromosoma 1 umano rispetto ai cromosomi di topo cliccate su *View Chr1 Synteny* e nel menu a tendina scegliete l'opzione *vs Mouse*.

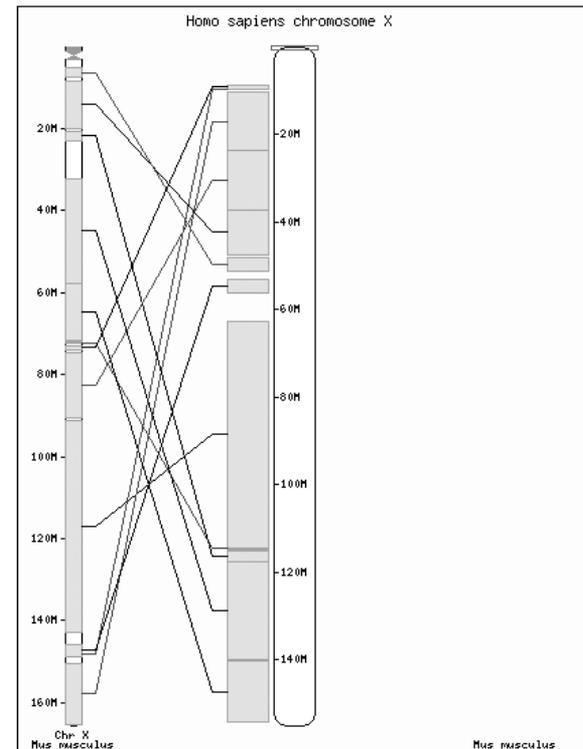
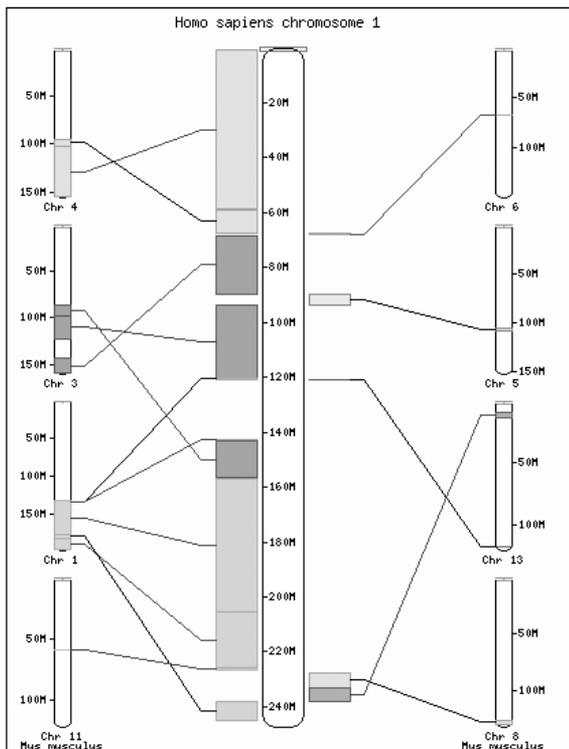
Scoprirete che i geni presenti sul cromosoma 1 umano sono distribuiti, nel topo, sui cromosomi 1, 3,

4, 5, 6, 8, 11 e 13. In questa visualizzazione il rettangolo rosso (che corrisponde alla zona che stiamo osservando) sull'ideogramma del cromosoma 1 umano è posizionato all'estremità del braccio corto. Tuttavia, se provate a spostarvi lungo il cromosoma 1, utilizzando il tasto **Downstream** in fondo sul lato destro della pagina, verrà visualizzato un elenco di geni omologhi fra uomo e topo.

Provate a cambiare cromosoma, per esempio scegliete il cromosoma X.

Notate che il cromosoma X umano, a differenza degli autosomi, presenta sintenia solo col cromosoma X di topo e lo stesso accade se provate a confrontare il cromosoma X dell'uomo con quello di altre specie. Questo è dovuto al fatto che i geni presenti sul cromosoma X sono presenti in singola copia nei maschi (che hanno un solo cromosoma X); nelle femmine, che hanno due cromosomi X, avviene l'inattivazione di uno dei due, così che solo una copia di ogni gene presente sul cromosoma X è funzionante.

Per questo motivo i geni posti sul cromosoma X tendono a rimanere localizzati insieme nell'evoluzione, perché se si spostassero su un autosoma, sarebbe complicato garantirne l'espressione a partire da uno solo dei due alleli.



Provate a cercare la distribuzione dei vari geni confrontando la sintenia tra cromosomi umani con cro-

mosomi di altre specie e completate le seguenti tabelle:

| Uomo | Topo |
|-------------|-------|
| Cromosoma 2 | |
| | |
| | |
| | |
| | |
| | |

| Uomo | Topo |
|-------------|-------|
| Cromosoma 4 | |
| | |
| | |
| | |
| | |
| | |

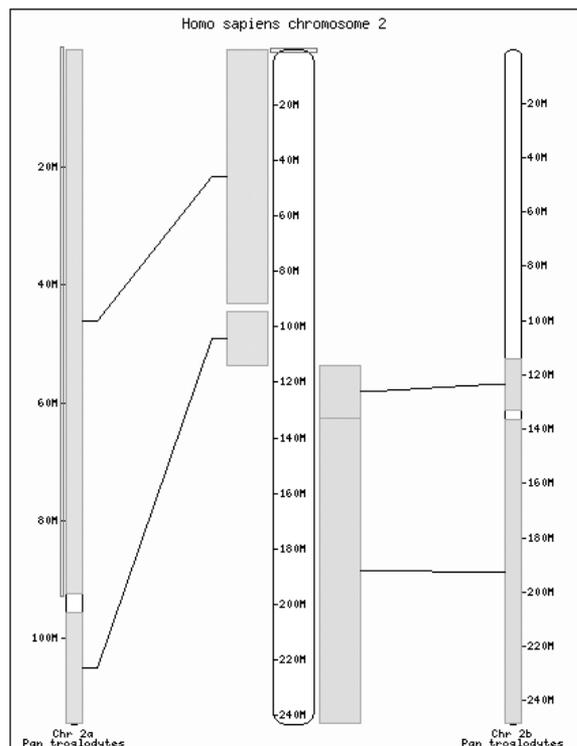
| Uomo | Topo |
|--------------|-------|
| Cromosoma 13 | |
| | |
| | |
| | |

| Uomo | Topo |
|-------------|-------|
| Cromosoma Y | |
| | |
| | |
| | |

Come avrete già scoperto, lo scimpanzé possiede una coppia di cromosomi in più rispetto alla specie umana. Utilizzando la funzione di ricerca della sintenia, cercate di spiegare l'origine del cromosoma in più presente nello scimpanzé.

Per farlo cliccate sulla versione di Ensembl di febbraio 2006 e cliccate sul cromosoma 2 umano; cliccando poi su **View chromosome 2 synteny** e confrontandolo con lo scimpanzé vi accorgete che il cromosoma 2 umano si è formato dall'unione del cromosoma 2A e 2B dello scimpanzé.

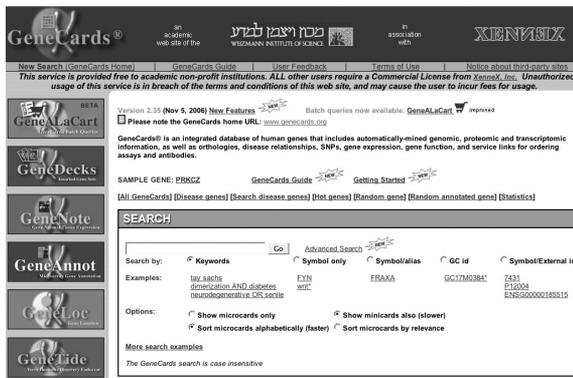
Osservate anche la sintenia tra i vari cromosomi umani e quelli dello scimpanzé e noterete che i gruppi di sintenia si mantengono nella maggior parte dei casi associati sugli stessi cromosomi; si può concludere che più le specie sono vicine dal punto di vista evolutivo, più i geni rimangono raggruppati e localizzati in modo simile sullo stesso cromosoma.



ESERCITAZIONE 4

Alla ricerca di un gene

Per trovare informazioni sui geni dell'uomo e sulle malattie a essi collegate per mutazioni della struttura del geni stessi, visitiamo il sito www.genecards.org. *Gene Card* è un database bioinformatico specializzato, che presenta in maniera organica le informazioni relative a tutti i geni noti.



Di solito le mutazioni in un gene causano la perdita totale o parziale della funzione del gene stesso (in inglese *loss of function*), per cui la proteina che deriva dalla trascrizione e traduzione della sequenza nucleotidica alterata o non si forma per niente, o non è completa, oppure non si trova nella corretta struttura tridimensionale. Spesso i geni prendono il nome dalla malattia che causano quando non funzionano più.

Raramente la mutazione nel gene causa un aumento della sua funzione (in inglese *gain of function*), e comunque anche questa situazione è spesso causa di malattie.

Cliccando sul link *Hot genes* che compare nella pagina iniziale al di sopra del campo *search*, si accederà ad un elenco dei geni più richiesti, ovvero a un elenco di geni che sono stati cercati e studiati da un gran numero di persone.

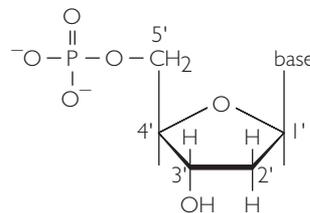
Come potete osservare il nome dei geni è preceduto da una sigla ed è rigorosamente in inglese.

GeneCards® Hot Genes

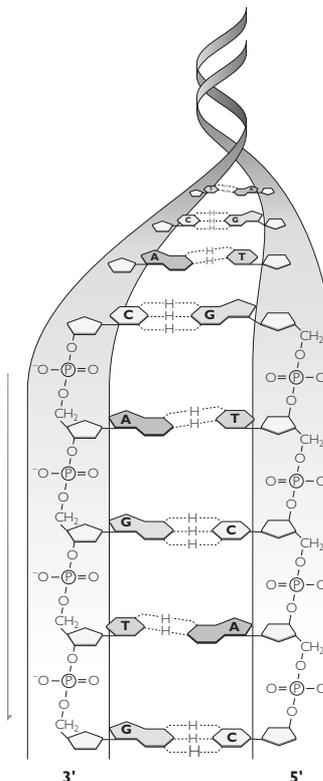
The following is a list of "hot" genes: those that have been most frequently requested at the GeneCards home site since Jan 1, 2006 sorted by frequency of access.

1. BRCA1 breast cancer 1, early onset
2. A2M alpha-2-macroglobulin
3. TP53 tumor protein p53 (Li-Fraumeni syndrome)
4. BTK Bruton agammaglobulinemia tyrosine kinase
5. CFTR cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)
6. ABL1 c-abl proto-oncogene tyrosine kinase
7. APP amyloid beta (A4) precursor protein (peptidase nexin-II, Alzheimer disease)
8. ALDH2 aldehyde dehydrogenase 2 family (mitochondrial)
9. AR androgen receptor (dihydrotestosterone receptor; testicular feminization; spinal and bulbar muscular atrophy; Kennedy disease)
10. FYN FYN oncogene related to SRC, FGR, YES
11. BRCA2 breast cancer 2, early onset
12. EGFR epidermal growth factor receptor (epithelial growth factor receptor; erb-B; erb-B oncogene homolog, avian)
13. ATRX ataxia telangiectasia mutated (includes complementation groups A, C and D)
14. GDF1 insulin-like growth factor 1 (somatomedin C)
15. HEXA hexosaminidase A (alpha polypeptide)
16. JAK2 Janus kinase 2 (a protein tyrosine kinase)
17. CASP8 caspase 8, apoptosis-related cysteine peptidase
18. TOP2A topoisomerase (DNA) II alpha 170kDa
19. F508 prostatic acid phosphatase 2 and variant metachromatic leukodystrophy
20. HBB hemoglobin, beta
21. BCL2 B-cell CLL/lymphoma 2
22. ABL1 v-abl Abelson murine leukemia viral oncogene homolog 1
23. COL2A1 collagen, type II, alpha 1 (primary osteoarthritis, spondyloepiphyseal dysplasia, congenital)
24. BAX BCL-2-associated X protein
25. AKT1 v-akt murine thymoma viral oncogene homolog 1
26. APOE apolipoprotein E
27. APC adenomatous polyposis coli
28. INSR insulin receptor

Cliccando poi su di un gene specifico, per esempio CFTR che è il gene responsabile della fibrosi cistica, possiamo osservare la localizzazione del gene sul cromosoma (il numero 7), in che braccio si trova (il braccio q), quanto è lungo in paia di basi (pb 188 702), se si trova sul filamento di DNA 5'-3' (detto *forward* o *plus* o *+*) oppure sul filamento di DNA 3'-5' (detto *reverse* o *minus* o *-*). Questo nome convenzionale deriva dal fatto che in una molecola di ribosio ci sono 5 atomi di carbonio e ciascuno di loro è numerato in base alla sua posizione. Ogni zucchero è attaccato al nucleotide precedente in posizione 3' e al nucleotide seguente in posizione 5'. Di solito la sequenza di nucleotidi viene letta in direzione 5' - 3'; l'altra sequenza di nucleotidi si dice antiparallela e viene letta in direzione 3' - 5'. La molecola di DNA è formata da due catene polinucleotidiche avvolte l'una intorno all'altra con andamento destrorso. Le due catene sono antiparallele, ossia i due singoli filamenti sono orientati in direzioni opposte.



Un nucleotide è formato da un gruppo fosfato, da uno zucchero a 5 atomi di carbonio e da una base azotata.



Schema di una molecola di DNA: le due catene hanno direzione 5' - 3' opposte, l'ordine delle basi azotate lungo una catena determina l'ordine delle basi lungo la catena complementare.

Chromosome: 7
 Entrez Gene cytogenetic band: **7q31.2** Ensembl cytogenetic band: **7q31.2**

Gene in genomic location: bands according to Ensembl, locations according to **GeneLoc** (and/or Entrez Gene and/or Ensembl if different)

GeneLoc location for GC07P116907: (about GC identifiers)
 Start: **116,907,253** bp from pter
 End: **117,095,955** bp from pter
 Size: **188,702** bases
 Orientation: **plus** strand

Cliccando ora su *ensembl cytogenetic band*, esattamente su *7q31.2*, si passa alla banca dati Ensembl che mostra il gene nel dettaglio con la struttura di esoni e introni (gli esoni sono rappresentati da trattini rossi verticali, collegati da linee spezzate orizzontali, ovvero gli introni).

ENST0000003084 ENSP0000003084 CFTR_HUMAN [Transcript info] [Exon info] [Peptide info]

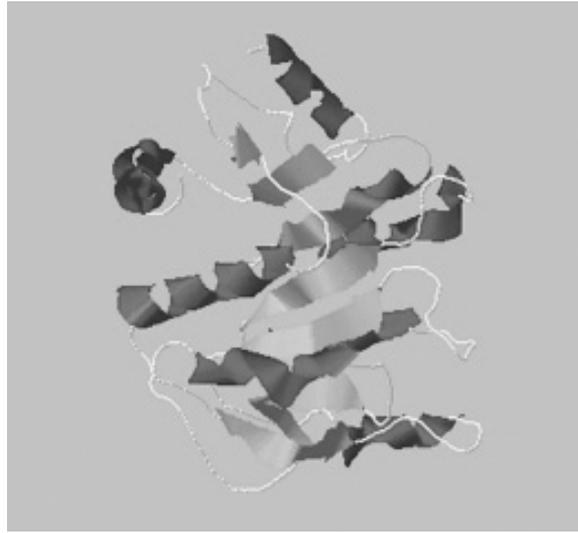
Features ▼

Provate ora a completare la seguente tabella con un elenco di geni le cui mutazioni causano gravi malattie.

| Sigla del gene | Nome completo | Malattia | Numero cromosoma | Lunghezza in bp (paia di basi) | Filamento di stampo |
|----------------|---------------|----------|------------------|--------------------------------|---------------------|
| BRCA2 | | | | | |
| F8 | | | | | |
| DMD | | | | | |
| FMRI | | | | | |
| G6PD | | | | | |

Sempre sul sito www.genecards.org per ogni gene elencato nella tabella sovrastante potete osservare la ricostruzione in 3 dimensioni della proteina che codifica cliccando su *3Dstructures* di uno qualsiasi di quelli che sono elencati. Si aprirà una nuova pagina che caricherà un programma di rielaborazione grafica in 3D, è gratuito e necessita dell'installazione di un *applet* spesso già presente nei sistemi dei computer (se non è presente installatelo seguendo le indicazioni che lo stesso sito vi darà).

L'immagine in 3D che vi si presenterà dopo alcuni secondi è la ricostruzione virtuale della proteina codificata dal gene in esame, dove sono indicati in giallo i foglietti beta e in rosa le strutture ad alfa elica. Passando con il mouse sulla figura potete ruotarla, tenendo cliccato il tasto **ctrl** potete zoomarla, ecc.



ESERCITAZIONE 5

Trova il gene

Il «dogma centrale della biologia» è basato sul principio che l'informazione genetica fluisca dal DNA verso l'RNA e che dall'RNA sia poi tradotta in proteine. Il DNA serve quindi come stampo per la propria duplicazione e per la sintesi di una molecola intermedia, detta RNA messaggero (mRNA), che contiene le informazioni necessarie per la sintesi delle proteine.

Questo processo è articolato in due tempi: nel primo, detto trascrizione, il DNA serve come stampo per la sintesi di un filamento di RNA complementare; nel secondo, detto traduzione, la sequenza di basi dell'mRNA è convertita in una sequenza di aminoacidi (ogni tre basi del filamento di mRNA corrispondono a un aminoacido) e dà luogo alla sintesi della proteina. *N.B. Nelle banche dati si usa il codice genetico con le triplette di DNA associate all'aminoacido corrispondente e quindi con la Timina al posto dell'Uracile (vedi schema seguente).*

| | T | C | A | G |
|---|--|--|--|--|
| T | TTT Phe (F) TTC Phe (F) TTA Leu (L) TTG Leu (L) | TCT Ser (S) TCC Ser (S) TCA Ser (S) TCG Ser (S) | TAT Tyr (Y) TAC Tyr (Y) TAA STOP TAG STOP | TGT Cys (C) TGC Cys (C) TGA STOP TGG Trp (W) |
| C | CTT Leu (L) CTC Leu (L) CTA Leu (L) CTG Leu (L) | CCT Pro (P) CCC Pro (P) CCA Pro (P) CCG Pro (P) | CAT His (H) CAC His (H) CAA Gin (Q) CAG Gin (Q) | CGT Arg (R) CGC Arg (R) CGA Arg (R) CGG Arg (R) |
| A | ATT Ile (I) ATC Ile (I) ATA Ile (I) ATG Met (M) | ACT Thr (T) ACC Thr (T) ACA Thr (T) ACG Thr (T) | AAT Asn (N) AAC Asn (N) AAA Lys (K) AAG Lys (K) | AGT Ser (S) AGC Ser (S) AGA Arg (R) AGG Arg (R) |
| G | GTT Val (V) GTC Val (V) GTA Val (V) GTG Val (V) | GCT Ala (A) GCC Ala (A) GCA Ala (A) GCG Ala (A) | GAT Asp (D) GAC Asp (D) GAA Glu (E) GAG Glu (E) | GGT Gly (G) GGC Gly (G) GGA Gly (G) GGG Gly (G) |

Gli aminoacidi possono essere rappresentati con una nomenclatura a una lettera (usata molto spesso nelle banche dati bioinformatiche) o con una nomenclatura a tre lettere che corrispondono alle prime tre lettere del loro nome esteso (vedi tabella).

| | | |
|-----|---|------------------|
| Ala | A | Alanina |
| Arg | R | Arginina |
| Asn | N | Asparagina |
| Asp | D | Acido aspartico |
| Cys | C | Cisteina |
| Gly | G | Glicina |
| Glu | E | Acido glutammico |
| Gln | Q | Glutamina |
| His | H | Istidina |
| Ile | I | Isoleucina |
| Leu | L | Leucina |
| Lys | K | Lisina |
| Met | M | Metionina |
| Phe | F | Fenilalanina |
| Pro | P | Prolina |
| Ser | S | Serina |
| Thr | T | Treonina |
| Trp | W | Triptofano |
| Tyr | Y | Tirosina |
| Val | V | Valina |

Trovate ora le tre catene polipeptidiche elencate, utilizzando la tabella e lo schema, nella sequenza di DNA riportata in fondo alla pagina.

Catene polipeptidiche da trovare:

1. Met Ser Ile Leu Leu Tyr Stop
2. Met Ser Ile Leu Val Glu Arg Stop
3. Met Cys Ala Thr Stop

| | | | | | |
|---------|------------|------------|------------|------------|------------|
| 1-50 | TGGTCCTGCA | GTCCTCTCCT | GGCGCCCCGG | GGGCGAGCGG | ATGTCGATT |
| 51-100 | TCGTGGAAAG | ATAGTCCC | TGCCTGCGGG | CGGAGGGACC | GTGCTGACCA |
| 101-150 | AGATGTACCC | GCGCGGCAAC | CACTGGGCGG | TGGGGCACTT | AATGGGGAAA |
| 151-200 | AAGAGCACAG | GGGAGTCTTC | TTCTGTATGT | TCTGAGAGAG | GGAGCCTGAA |
| 201-250 | GCAGCAGCTG | AGAGATGTGT | GCAACGTGAA | GTACATCAGG | TGGGAAGAAG |
| 251-300 | CTGCAAGGAA | TTTGCTGGGT | CTCATAGAAG | CTAAAGGAGA | ACAGAAACCA |
| 301-350 | CCAGCCACCT | CAACCAAGA | TGTCGATTCT | ACTTTATTAA | GCCCTGGGCA |
| 351-400 | ATCAGCAGCC | TTCGTGGGAT | ATGTCAGAGG | ATAGCAGCCA | ACTGAATAGC |

ESERCITAZIONE 6

Giochiamo con il codice genetico

Partiamo dai tre polipeptidi sottoelencati

- Met Ser Ile Leu Leu Tyr Stop;
- Met Ser Ile Leu Val Glu Arg Stop;
- Met Cys Ala Thr Stop.

Gli aminoacidi sono identificati dalla nomenclatura a tre lettere; sostituisci il nome degli aminoacidi con la nomenclatura a una lettera (vedi attività precedente); che parole nascondono? Ricordate che il primo aminoacido, la metionina, è sempre presente e non caratterizza la catena specifica della proteina così come il codone di stop, per cui nella parola che dovrete trovare omettete la metionina e lo stop.

| | | |
|---------|------------|-------------|
| 1-50 | TGGTCCTGCA | GTCCTCTCCT |
| 51-100 | TCGTGGAAAG | ATAGTCCC GC |
| 101-150 | AGATGTACCC | GCGCGGCAAC |
| 151-200 | AAGAGCACAG | GGGAGTCTTC |
| 201-250 | GCAGCAGCTG | AGAGATGTGT |
| 251-300 | CTGCAAGGAA | TTTGCTGGGT |
| 301-350 | CCAGCCACCT | CAACCCAAGA |
| 351-400 | ATCAGCAGCC | TTCGTGGGAT |

Cercate le tre parole in cui possono essere lette le tre sequenze aminoacidiche e individuate la frase di lettura esatta.

Compariranno sei diversi modi in cui il DNA può essere tradotto in proteine delle quali uno solo è quello giusto. Infatti ci sono sei modi in cui il DNA può essere letto, detti anche *frasi di lettura* (in inglese *ORF, open reading frame*), perché quando il DNA di un gene inizia la trascrizione apre la doppia elica e quindi entrambe le catene di nucleotidi possono essere lette. Ogni elica può essere letta a partire dalla prima base azotata, dalla seconda o dalla terza, facendo slittare la lettura di una tripletta ogni volta, per cui ci sono tre possibili traduzioni per ogni elica. Di solito una sola delle sei traduzioni possibili è quella che porta alla formazione di una proteina funzionante.

Usiamo ora il sito per la traduzione della sequenza nucleotidica in sequenza aminoacidica all'indirizzo: www.expasy.org/tools/dna.html e incolliamo nel campo bianco la sequenza di DNA che segue; clicchiamo poi su *Translate sequence* e il software ri elaborerà la sequenza proponendoci la sua traduzione in sequenze di aminoacidi scritti con la nomenclatura a una lettera.



Translate tool

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

Please enter a DNA or RNA sequence in the box below (numbers and blanks are ignored).

Output format:

or

| | | |
|-------------|------------|------------|
| GGCGCCCCGG | GGGCGAGCGG | ATGTCGATTC |
| TGCCCTGCGGG | CGGAGGGACC | GTGCTGACCA |
| CACTGGGCGG | TGGGGCACTT | AATGGGGAAA |
| TTCTGTATGT | TCTGAGAGAG | GGAGCCTGAA |
| GCAACGTGAA | GTACATCAGG | TGGGAAGAAG |
| CTCATAGAAG | CTAAAGGAGA | ACAGAAACCA |
| TGTCGATTCT | ACTTTATTAA | GCCCTGGGCA |
| ATGTCAGAGG | ATAGCAGCCA | ACTGAATAGC |

Rispondete poi alle seguenti domande.

1. Quale aminoacido viene letto da AGG?
2. Quali codoni codificano per la serina?
3. In quanti modi si può scrivere la sequenza dei codoni per: Cys Arg Thr?

ESERCITAZIONE 7

Le mutazioni puntiformi e le loro conseguenze

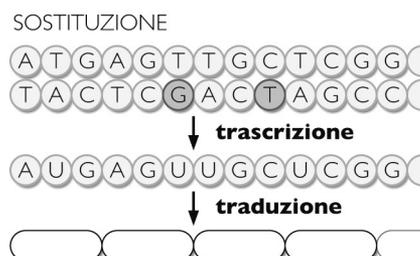
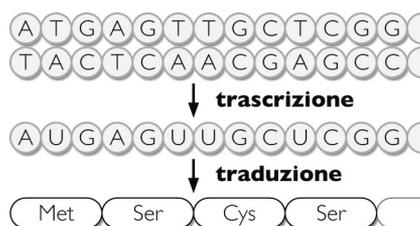
Una sequenza nucleotidica viene trascritta in una molecola di mRNA che va incontro al processo di traduzione in una catena di aminoacidi. Le mutazioni geniche o puntiformi consistono in cambiamenti nella sequenza delle basi del DNA. Le mutazioni puntiformi comprendono:

1. sostituzioni,
2. delezioni,
3. inserzioni.

Se una singola base viene sostituita, anche nell' mRNA corrispondente ci saranno delle sostituzioni che porteranno a modificazioni nella sequenza di aminoacidi finale.

La sostituzione di una base azotata può provocare:

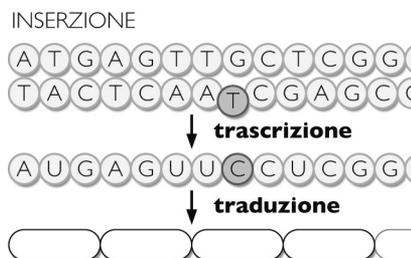
- a. una mutazione *sinonima* se la tripletta codifica per lo stesso aminoacido;
- b. una mutazione *missense* se l'effetto è quello di formare una tripletta che codifica per un aminoacido diverso da quello iniziale; la proteina che incorporerà quell'aminoacido può perdere o modificare la sua attività;
- c. una mutazione *non sense* se si forma una tripletta che non codifica per alcun aminoacido (tripletta di stop o di arresto), questo comporta l'arresto della sintesi proteica a livello ribosomiale perchè il tRNA non sarà in grado di opporre l'anticodone specifico per il codone sostituito nell'mRNA;



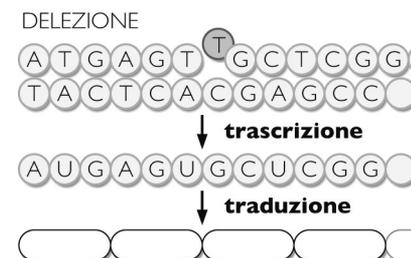
- d. una mutazione di *allungamento* se una tripletta di stop viene sostituita da una tripletta codificante.

Risolvi l'esercizio che precede inserendo il nome degli aminoacidi che vengono codificati dalla sequenza che presenta due sostituzioni; che polipeptide si forma? Che tipo di sostituzioni sono avvenute?

Se nella sequenza nucleotidica si inserisce una nuova base azotata, dal punto di inserzione in poi avviene uno slittamento nella traduzione in sequenza aminoacidica; risolvi l'esercizio che segue confrontando la sequenza ottenuta con quella di partenza.



Se la sequenza nucleotidica perde una base azotata, dal punto in cui avviene la delezione della base in poi c'è uno slittamento nella traduzione in sequenza aminoacidica; risolvi l'esercizio che segue confrontando la sequenza ottenuta con quella di partenza.



ESERCITAZIONE 8

Geni e malattie

Le mutazioni sono alterazioni del corredo genetico di un individuo; spesso derivano da errori nella replicazione del genoma e possono essere *spontanee* o *indotte*. Le mutazioni possono essere di tipo *somatico* e *germinale*. Quelle somatiche provocano danni all'individuo che le porta, invece quelle germinali, che interessano le cellule riproduttive, possono ripercuotersi sulla discendenza futura.

In generale le mutazioni possono essere di tre tipi:

- genomiche, alterazioni nel numero dei cromosomi tipico della specie;
- cromosomiche, alterazioni nella struttura dei cromosomi;
- geniche o puntiformi, alterazioni nella sequenza nucleotidica del gene.

Le mutazioni geniche o puntiformi consistono in cambiamenti nella sequenza delle basi del DNA. Le mutazioni puntiformi comprendono: le sostituzioni, le delezioni e le inserzioni.

Nello schema seguente sono mostrati diversi tipi di mutazioni puntiformi:

GAC-AAA-GGA-TGA-CTG SEQUENZA ORIGINALE
 GAC-AAA-CGA-TGA-CTG SOSTITUZIONE DI G CON C
 GAC-AAA-TGG-ATG-ACT-G INSERZIONE DI T
 GAC-AA-G-GAT-GAC-TG DELEZIONE DI A

Studiamo ora il caso dell'anemia falciforme: l'anemia falciforme, chiamata anche *Sickle Cell Disease* (SCD), è una malattia ereditaria autosomica recessiva caratterizzata dalla classica forma a falce che hanno i globuli rossi degli individui malati. Questa malattia è causata da una mutazione nel gene che codifica la catena beta dell'emoglobina (Hb). Nelle catene beta dell'emoglobina degli individui malati, l'acido glutammico nella posizione 6 è sostituito dall'aminoacido valina determinando così la formazione di una emoglobina patologica, chiamata *emoglobina S* (HbS); la mutazione HbS è una sostituzione di una base nucleotidica nel gene SCD del cromosoma 11, esattamente di A con T nella sequenza CCT GAG GAG, che fa cambiare il significato del 6° codone del gene della catena beta dell'Hb (Glu viene sostituito da Val).

Per trovare informazioni sulla malattia in banche dati bioinformatiche clicchiamo su <http://www.ncbi.nlm.nih.gov/> il sito del *National Center of Biotechnology Information* e clicchiamo su *OMIM*, la banca dati dedicata alle malattie umane che si trasmettono seguendo le leggi di Mendel; nel campo *search* scriviamo «Sickle cell anemia» e si aprirà una pagina con l'elenco delle varie forme di anemia, con tutte le caratteristiche cliniche e le pubblicazioni mediche più aggiornate sull'argomento. Cliccando poi sul numero vicino a *HbS* potremo trovare le notizie riguardanti la mutazione tipica dell'anemia falciforme.

Sempre dalla pagina internet all'indirizzo <http://www.ncbi.nlm.nih.gov/> cliccate sul link nella colonna di destra dal titolo *Genes & Disease*; comparirà una pagina ricca di informazioni sulle malattie genetiche divisa per tipologia; cliccando su ciascuna voce si otterranno notizie sul tipo di malattia in generale e alcune curiosità sull'argomento.

Per ogni argomento troverete una spiegazione dettagliata sull'apparato o sistema colpito da mutazioni e quindi da malattie (che sono elencate sulla destra) e notizie o link di interesse collegati all'argomento (sulla sinistra).

Cliccando poi su una singola malattia (per esempio la fibrosi cistica) la schermata che apparirà vi offrirà notizie ancora più dettagliate e sulla sinistra potrete trovare i link per la localizzazione genica, articoli e pubblicazioni ecc.

Andate alla ricerca di tre malattie metaboliche, tre tumori e tre malattie del sistema nervoso trovando informazioni su: tipo di malattia, cromosoma che porta la mutazione, effetti clinici, cure e terapie.

Clickando invece sull'icona del cromosoma completate la seguente tabella.

| | Malattia | Lunghezza in bp | N° geni | Effetti | Terapie |
|--------------|----------|-----------------|---------|---------|---------|
| Cromosoma 1 | | | | | |
| Cromosoma 4 | | | | | |
| Cromosoma 7 | | | | | |
| Cromosoma 12 | | | | | |
| Cromosoma 13 | | | | | |
| Cromosoma 21 | | | | | |

ESERCITAZIONE 9

Disegniamo l'albero filogenetico di alcuni Primati

Proviamo a disegnare un albero filogenetico partendo da cinque sequenze nucleotidiche omologhe,

lunghe 46 basi azotate; le sequenze appartengono a cinque primati.

In biologia, il termine *omologia* ha il significato particolare di indicare che due strutture, per esempio due organi, hanno una origine evolutiva comune. L'omologia presuppone quindi l'esistenza di un antenato comune da cui le strutture omologhe si sono evolute.

Il concetto di omologia si applica anche ai geni che derivano da uno stesso gene ancestrale.

| | | |
|---|---------------------|---|
| N | Neanderthal | TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC |
| H | <i>Homo sapiens</i> | TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC |
| S | scimpanzé | TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGC GGTTGTCC |
| G | gorilla | TGGACCTGCAGTCATCTTCTGCCGCCCGAGCGCTTGCCGATGTCC |
| O | orango | ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC |

Contiamo ora le differenze tra le basi azotate confrontando a due a due tutti gli organismi.

| | | |
|---|---------------------|--|
| N | Neanderthal | TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC |
| H | <i>Homo sapiens</i> | TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC |

Differenze 3

| | | |
|---|-------------|---|
| N | Neanderthal | TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC |
| S | scimpanzé | TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGC GGTTGTCC |

Differenze 4

| | | |
|---|-------------|--|
| N | Neanderthal | TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC |
| G | gorilla | TGGACCTGCAGTCATCTTCTGCCGCCCGAGCGCTTGCCGATGTCC |

Differenze ...

| | | |
|---|-------------|--|
| N | Neanderthal | TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC |
| O | orango | ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC |

Differenze ...

| | | |
|---|---------------------|---|
| H | <i>Homo sapiens</i> | TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC |
| S | scimpanzé | TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGC GGTTGTCC |

Differenze ...

| | | |
|---|---------------------|--|
| H | <i>Homo sapiens</i> | TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC |
| G | gorilla | TGGACCTGCAGTCATCTTCTGCCGCCCGAGCGCTTGCCGATGTCC |

Differenze ...

| | | |
|---|---------------------|--|
| H | <i>Homo sapiens</i> | TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC |
| O | orango | ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC |

Differenze ...

| | | |
|---|-----------|---|
| S | scimpanzé | TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGC GGTTGTCC |
| G | gorilla | TGGACCTGCAGTCATCTTCTGCCGCCCGAGCGCTTGCCGATGTCC |

Differenze ...

S scimpanzé TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGC GGTTGTCC
 O orango ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

Differenze ...

G gorilla TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC
 O orango ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

Differenze ...

Completate la seguente tabella scrivendo il numero di differenze trovate.

| | Neanderthal | <i>Homo sapiens</i> | scimpanzé | gorilla | orango |
|---------------------|-------------|---------------------|-----------|---------|--------|
| Neanderthal | 0 | 3 | 4 | | |
| <i>Homo sapiens</i> | 3 | 0 | | | |
| scimpanzé | 4 | | 0 | | |
| gorilla | | | | 0 | |
| orango | | | | | 0 |

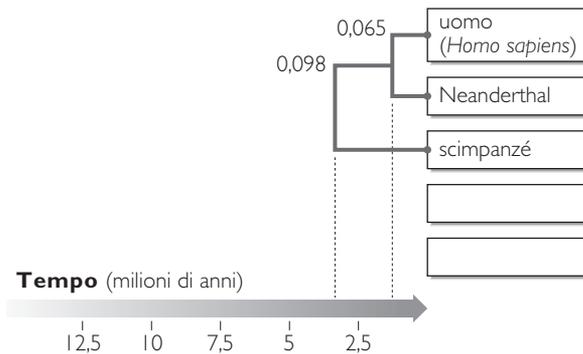
I due organismi più simili sono l'*Homo sapiens* e l'uomo di Neanderthal, la cui differenza è di 3 basi su 46 cioè il 6,5%.

Cominciamo a disegnare allora l'albero evolutivo; uniamo *Homo sapiens* e Neanderthal con una linea verticale; la lunghezza del ramo dovrebbe rappresentare il tempo che queste due specie hanno impiegato a divergere da un antenato comune; se consideriamo che per cambiare una sequenza di DNA del 100% ci vuole un arco di tempo di 20 milioni di anni, noi ci troviamo di fronte a un cambiamento di 3 basi su 46 quindi del 6,5%.

Con una semplice proporzione troviamo che:
 $100 : 20\ 000\ 000 = 6,5 : X$
 $X = 1\ 300\ 000$ anni

Sull'asse x rappresentiamo il tempo in milioni di anni e il progenitore comune a *Homo sapiens* e Neanderthal lo poniamo a 1,3 cm.

Riuniamo ora in un solo gruppo i due organismi più simili, cioè il Neanderthal e *Homo sapiens* e confrontiamoli (sommando le differenze che entrambi hanno con l'organismo che prendiamo in considerazione e facendone una media dividendo per 2) con gli altri organismi.



| | Neanderthal/ <i>Homo s.</i> | scimpanzé | gorilla | orango |
|-----------------------------|-----------------------------|------------------------|------------------|------------------|
| Neanderthal/ <i>Homo s.</i> | 0 | $(4+5)/2=4,5/46=0,098$ | $(11+12)/2=11,5$ | $(16+17)/2=16,5$ |
| scimpanzé | $(4+5)/2=4,5/46=0,098$ | 0 | 11 | 14 |
| gorilla | $(11+12)/2=11,5$ | 11 | 0 | 14 |
| orango | $(16+17)/2=16,5$ | 14 | 14 | 0 |

Il gruppo Neanderthal/*Homo sapiens* ha 4,5 basi di differenza con lo scimpanzé cioè $4,5/46 = 0,098$ e quindi il 9,8%.

Con una semplice proporzione possiamo trovare che:
 $100 : 20\ 000\ 000 = 9,8 : X$

Per calcolare la distanza tra lo scimpanzé e il gruppo Neanderthal/*Homo sapiens* si deve calcolare $9,8 \times 20\ 000\ 000 / 100 = 1\ 960\ 000$ anni (1,96 cm).

Se vogliamo invece calcolare la distanza di *Homo sapiens* con lo scimpanzé dobbiamo sommare $1\ 300\ 000 + 1\ 960\ 000 = 3\ 260\ 000$ anni.

Poniamo ora in un solo gruppo Neanderthal, *Homo sapiens* e scimpanzé e confrontiamo il gruppo con gli altri organismi.

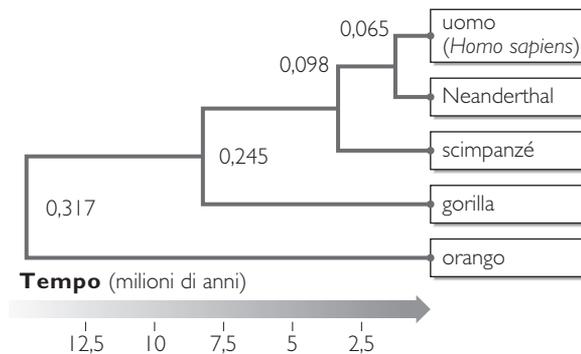
| | Neanderthal/ <i>Homo s./scimpanzé</i> | gorilla | orango |
|---------------------------------------|---------------------------------------|---------------------|---------------------|
| Neanderthal/ <i>Homo s./scimpanzé</i> | 0 | $(11,5+11)/2=11,25$ | $(16,5+14)/2=15,25$ |
| gorilla | $(11,5+11)/2=11,25$ | 0 | 14 |
| orango | $(16,5+14)/2=15,25$ | 14 | 0 |

Poniamo infine in un solo gruppo Neanderthal, *Homo sapiens*, scimpanzé e gorilla e confrontiamo il gruppo con l'ultimo organismo, l'orango.

| | Neanderthal/ <i>Homo s./scimpanzé/gorilla</i> | orango |
|---|---|-----------------------|
| Neanderthal/ <i>Homo s./scimpanzé/gorilla</i> | 0 | $(15,25+14)/2=14,625$ |
| orango | $(15,25+14)/2=14,625$ | 0 |

Compilando la tabella riassuntiva tenendo conto dei gruppi che insieme vengono confrontati con gli organismi rimasti da esaminare, saremo in grado di completare l'albero graficamente.

| | differenze | % di differenza | anni | cm sull'asse X |
|--|------------|-------------------------|--|----------------|
| Neanderthal/ <i>Homo s.</i> | 3 | $3 / 46 = 6,5\%$ | 1 300 000 | 1,3 |
| Neanderthal/ <i>Homo s./scimpanzé/</i> | 4,5 | $4,5 / 46 = 9,8\%$ | $(6,5 + 9,8) \times 20\ 000\ 000 / 100 = 3\ 260\ 000$ | 3,26 |
| Neanderthal/ <i>Homo s./scimpanzé/gorilla</i> | 11,25 | $11,25 / 46 = 24,45\%$ | $(6,5 + 9,8 + 24,45) \times 20\ 000\ 000 / 100 = 8\ 150\ 000$ | 8,15 |
| Neanderthal/ <i>Homo s./scimpanzé/gorilla/orango</i> | 14,625 | $14,625 / 46 = 31,79\%$ | $(6,5 + 9,8 + 24,45 + 31,79) \times 20\ 000\ 000 / 100 = 14\ 500\ 000$ | 14,5 |



ESERCITAZIONE 10

Recensioni

<http://50annidna.scienze.unipd.it/DFTB/index.htm>

Il sito dal titolo *DNA from the beginning*, creato in collaborazione con il *Dolan DNA learning center* dell'Università della Carolina, è organizzato in tre grandi sezioni che sono:

- la genetica classica
- le molecole della genetica
- controllo e organizzazione dei geni

che sono ipertesti completi, corretti scientificamente (il sito è coordinato dall'Università di Padova) e arricchiti da animazioni, gallerie di immagini, problemi, bibliografie e link ad altri siti collegati.

Le video interviste sono per ora in inglese ma il sito è in via di traduzione.

In ogni sezione è presente sulla sinistra un indice degli argomenti contenuti con la possibilità di ritornare su temi già trattati oppure di controllare sempre a che punto del percorso ci si trova durante la navigazione.

<http://www.molecularlab.it/>

Il sito *Molecularlab* è organizzato in modo da offrire notizie aggiornate, contenuti e attività per esperimenti di laboratorio, video, ipertesti e animazioni, bibliografie sia per studenti che per ricercatori e scienziati. La navigazione è molto semplice e immediata, con la possibilità di cliccare sia sulla barra in alto nella pagina iniziale, sia lateralmente sulla sinistra con l'elenco delle sezioni, sia centralmente sulle icone delle quattro macrosezioni in cui dividono il sito.

Particolarmente interessante la sezione multimedia dove si possono trovare alcune animazioni interattive da svolgere solo *on line* e altre che si possono scaricare gratuitamente, sia per Pc che per Mac.

http://www.gene-abc.ch/index_i.html

Il sito *L'ABC dei geni* in italiano, francese e tedesco, nato dalla collaborazione con il Fondo nazionale svizzero per la ricerca scientifica, presenta il mondo dei geni, del DNA ricombinante, dei cromosomi, delle malattie e dei vaccini in modo divertente ma scientificamente corretto e molto approfondito (non fatevi ingannare dall'uso delle animazioni e dei pupazzi; il contenuto scientifico è di alto livello, quasi universitario!)

Cliccando sull'icona del mondo, in alto a sinistra, ci si addentra nel sito con la possibilità di scegliere tra i diversi argomenti; un bambino o una bambina a fondo pagina danno sempre consigli per la navigazione.

Per la navigazione si possono utilizzare anche i link in alto nella pagina che rimandano alle stesse pagine delle icone:

- **il mondo dei geni**, che contiene tutti i percorsi riguardanti i geni, le biotecnologie, i cromosomi, i vaccini, la storia delle scoperte in campo biologico e genetico, le tecnologie usate in laboratorio biologico; tutti gli argomenti sono presentati come un ipertesto, con pagine che si possono visitare una dopo l'altra e che contengono a loro volta link
 - al glossario;
 - all'indice dell'argomento;
 - a ingrandimenti di immagini o ad animazioni;
 - ad approfondimenti.

Per tornare al percorso e continuarlo ricordarsi sempre di cliccare sull'indice e riprendere dall'argomento dove si era arrivati. Possono comparire anche altri pulsanti che permettono di spostare o ruotare gli oggetti.

Molto interessante la sezione dedicata ai quiz di valutazione con risposta immediata e soluzione delle domande sbagliate.

Nella sezione attività è possibile fare domande agli scienziati, avere la risposta alla propria domanda pubblicata, scaricare materiale utile anche per esperienze di laboratorio e una lista di link utili all'approfondimento

www.dnai.org

Il sito *DNAinteractive* è ricco di animazioni, filmati audio, esercizi e link di approfondimento, è dedicato ai docenti di scienze per la preparazione di argomenti di biologia molecolare e di genetica, nel modo più ricco e aggiornato possibile.

Le varie sezioni in cui risulta diviso il sito riguardano:

- la linea del tempo con la storia delle scoperte in campo biologico molecolare e genetico arricchita di immagini e video;
- il codice genetico con simulazioni e giochi;
- la manipolazione genetica;
- il genoma;
- le applicazioni (dal test del DNA all'albero genealogico per ritrovare mutazioni e malattie ereditarie);
- le implicazioni bioetiche e gli sviluppi futuri.