Il genoma

Vietate copia riproduzione e modifica

Goals del progetto genoma umano:

-Determinare la sequenza dell'intero genoma umano

-Identificare tutti i geni

-Immagazzinare queste informazioni su database

-Sviluppare programmi per l'analisi di questi dati

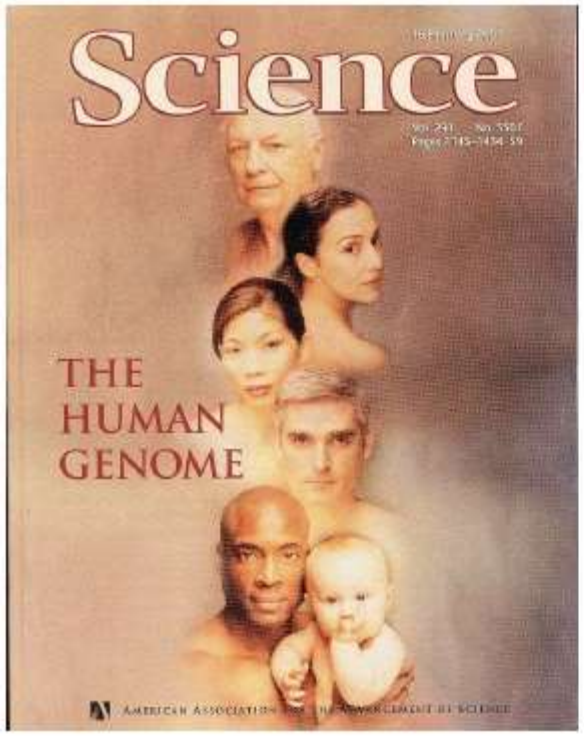
-Stabilire principi etici e legali per l'utilizzo di questi dati

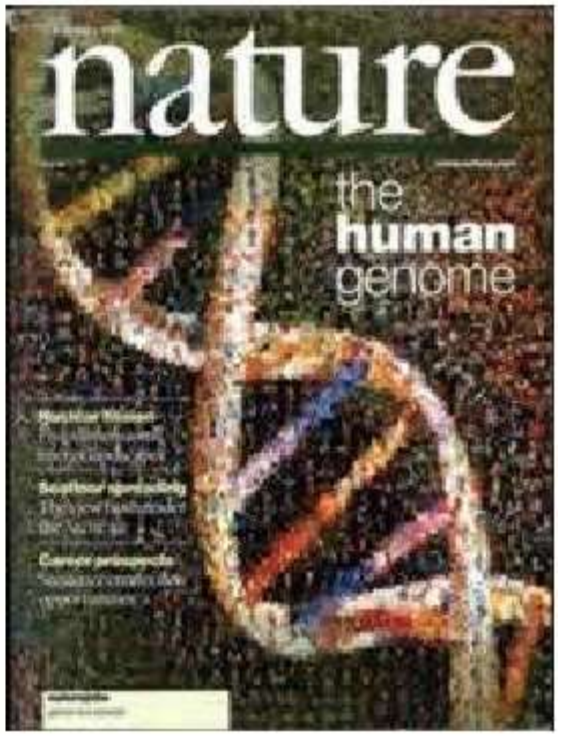
Vietate copia riproduzione e modifica

Consorzio Pubblico internazionale HGP

(USA, UK, Francia Germania, Cina e altri)

Celera Genomics di Craig Venter





Nel 2003, dopo 13 anni dalla istituzione del consorzio HGP, si dà l'annuncio ufficiale del sequenziamento del genoma umano

Vietate copia riproduzione e modifica

Bisogna PERO' tenere a mente che sebbene sia prassi comune parlare della sequenza del genoma umano ci sono in realtà molte sequenze perché

ogni individuo, eccetto i gemelli identici, ha la propria versione

Vietate copia riproduzione e modifica

- Tipi di variazione genetica

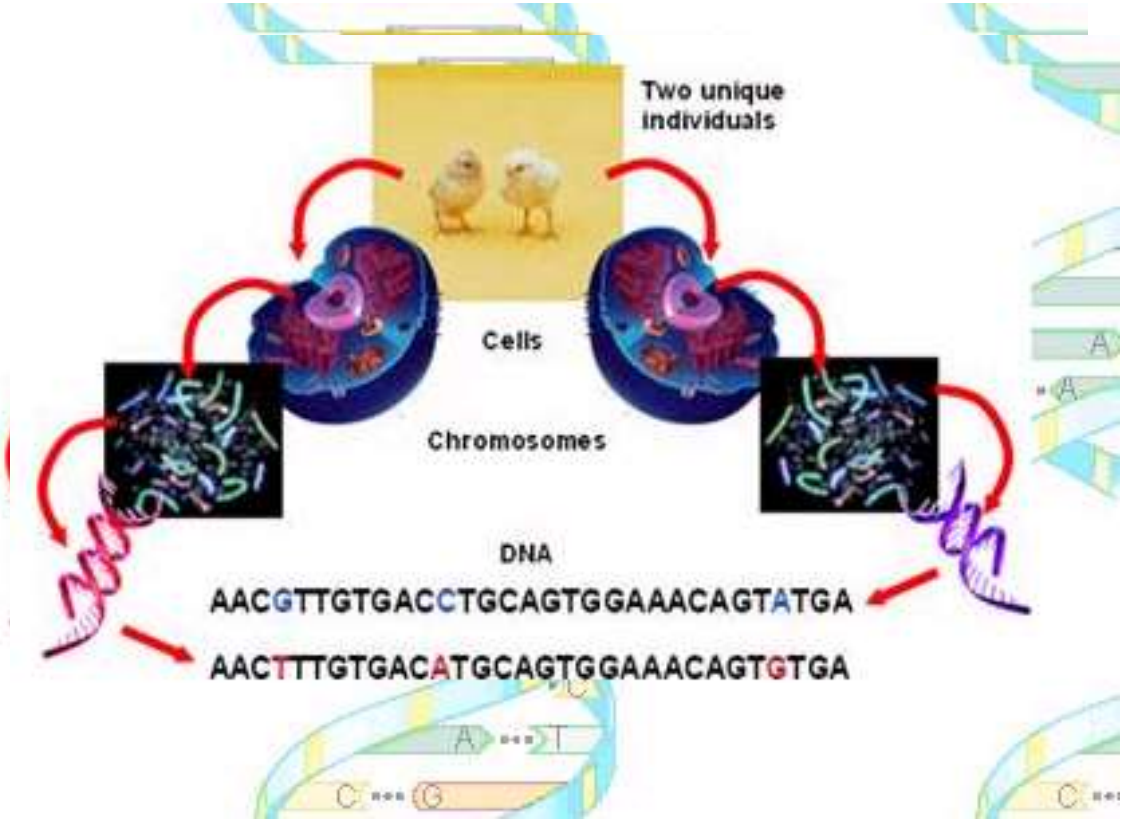
1) SNP

2) CNV = variazione nel numero di copie di un gene o di* sequenze ripetute (satelliti)

3) Indel = inserzioni/delezioni di nucleotidi, raramente a livello di geni, più spesso in zone intergeniche;

- Gene dell'amilasi salivare AMY1
- Gene per la chemochima CCL3 e suscettibilità all'AIDS
- La maggior parte delle CNVs non sembra essere adattativo né svantaggioso

Vietate copia riproduzione e modifica



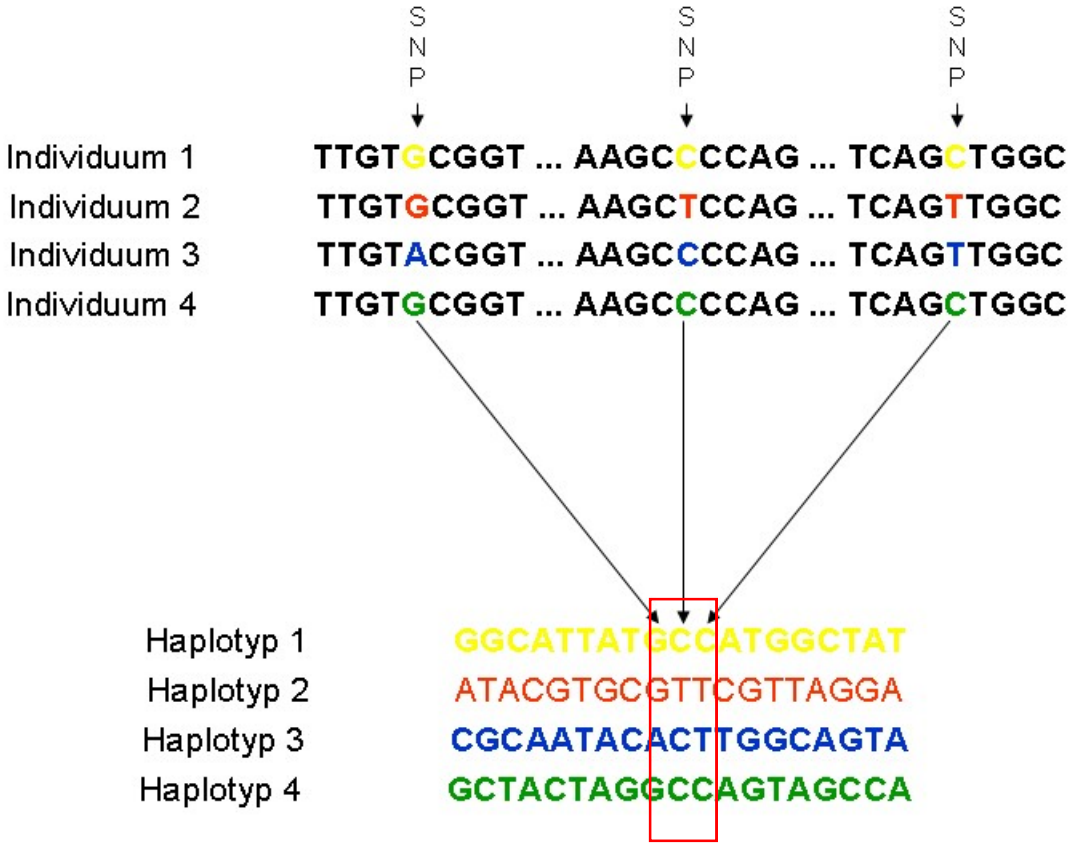
Sono stati identificati più di **3 milioni di SNPs**, una media di circa 1 ogni 1000 coppie di basi (alcuni autori valutano le differenze 1/300).

Essi costituiscono il 90% delle variazioni genetiche individuali.

Vietate copia riproduzione e modifica

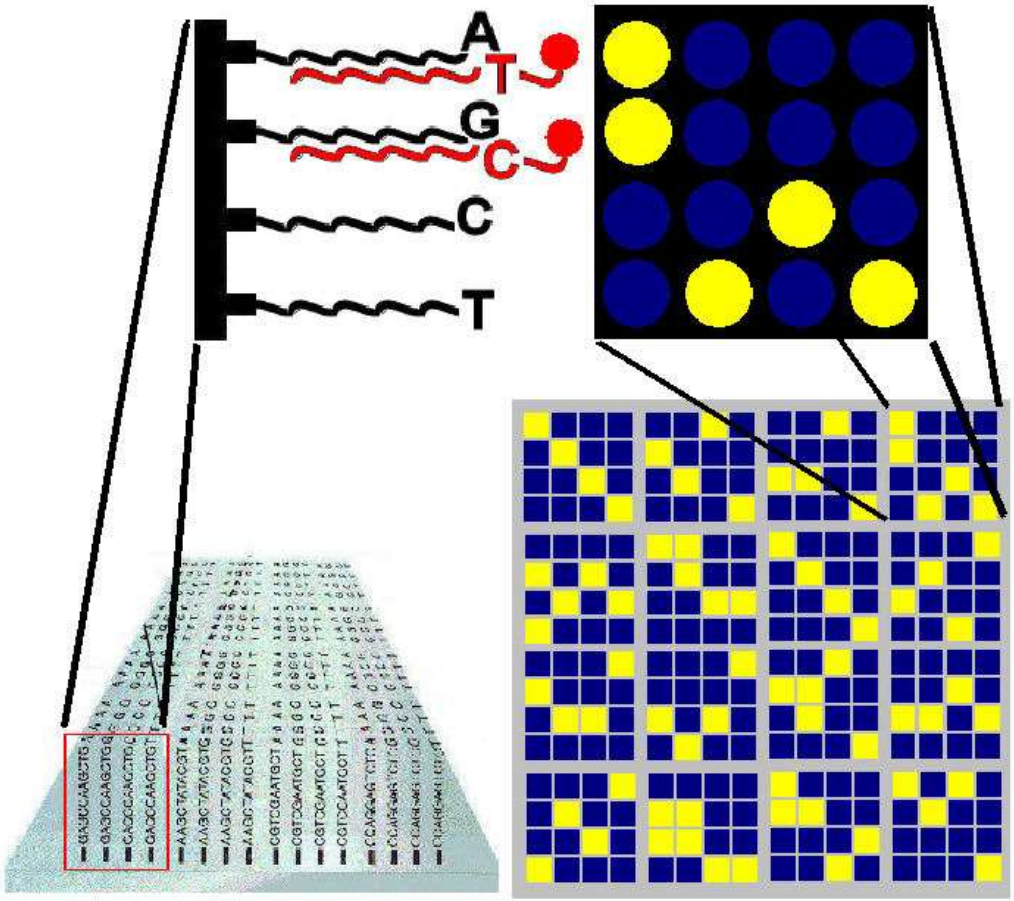
La combinazione di un certo numero di snp costituisce un **aplotipo**.

Con il termine **aplotipo** si definisce la combinazione di varianti alleliche lungo un cromosoma o segmento cromosomico contenente loci in linkage disequilibrium, cioè strettamente associati tra di loro e che, in genere, vengono ereditati insieme.



Vietate copia riproduzione e modifica

Più che di aplotipi semplici oggi possiamo parlare di veri e propri **profili di SNPs** grazie allo sviluppo di nuove tecnologie per la miniaturizzazione e l'automatizzazione di questo tipo di analisi basata sugli esperimenti con "microchips" di DNA che permettono l'analisi simultanea di milioni di SNPs.



Le sonde sul supporto sono allele-specifiche e le loro posizioni sono stabilite

Nella figura sono riportate le analisi su Chip relativi a 4 diversi snp (linee verticali)*

*ogni sonda allele-snp/specifica (disposta sulla linea verticale) termina con uno dei 4 nt anche se ogni snp ha solo due alleli

Vietate copia riproduzione e modifica

Molti SNPs non hanno effetto sulla funzionalità del genoma mentre altri sì.

Per esempio 60.000 SNPs si trovano all'interno di geni ed hanno un impatto sulla loro attività, determinando quelle differenze che rendono ognuno di noi un organismo unico.

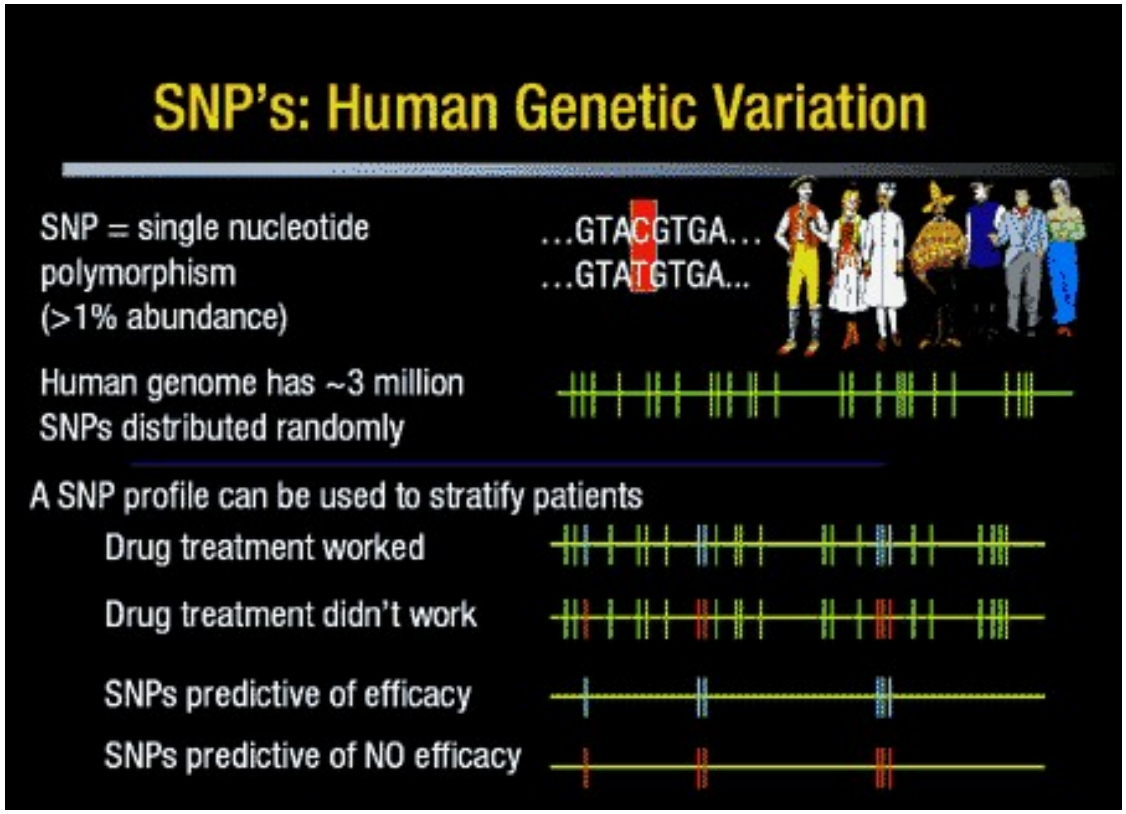
Gli SNPs che non si trovano in sequenza codificante possono avere consequenze sullo splicing o sul legame di fattori di trascrizione e quindi influenzare l'espressione genica

Vietate copia riproduzione e modifica

Lo studio degli SNP è molto utile poiché variazioni anche di singoli nucleotidi possono influenzare lo **sviluppo di patologie o la risposta a patogeni, ad agenti chimici, a farmaci**.

Gli SNPs possono avere una grande importanza nello **sviluppo di nuovi farmaci e nella pianificazione di protocolli terapeutici**: gli SNPs presenti nel gene responsabile della metabolizzazione di un farmaco consentono di predire l'effetto che esso potrà avere su quell'individuo.

Vietate copia riproduzione e modifica



Individuazione di specifici profili di SNPs in pazienti sensibili o resistenti ad una terapia **(Farmacogenetica)**

Vietate copia riproduzione e modifica

I Genomi degli Eucarioti
Compattezza

I genomi degli eucarioti hanno una densità genica molto ridotta.
Geni per proteine: 2-4% dell'intero genoma.
La struttura discontinua dei geni, con introni che nei mammiferi possono raggiungere dimensioni intorno a 20-30 kb (ed oltre) e la presenza di elementi ripetuti sono alla base della scarsa compattezza.

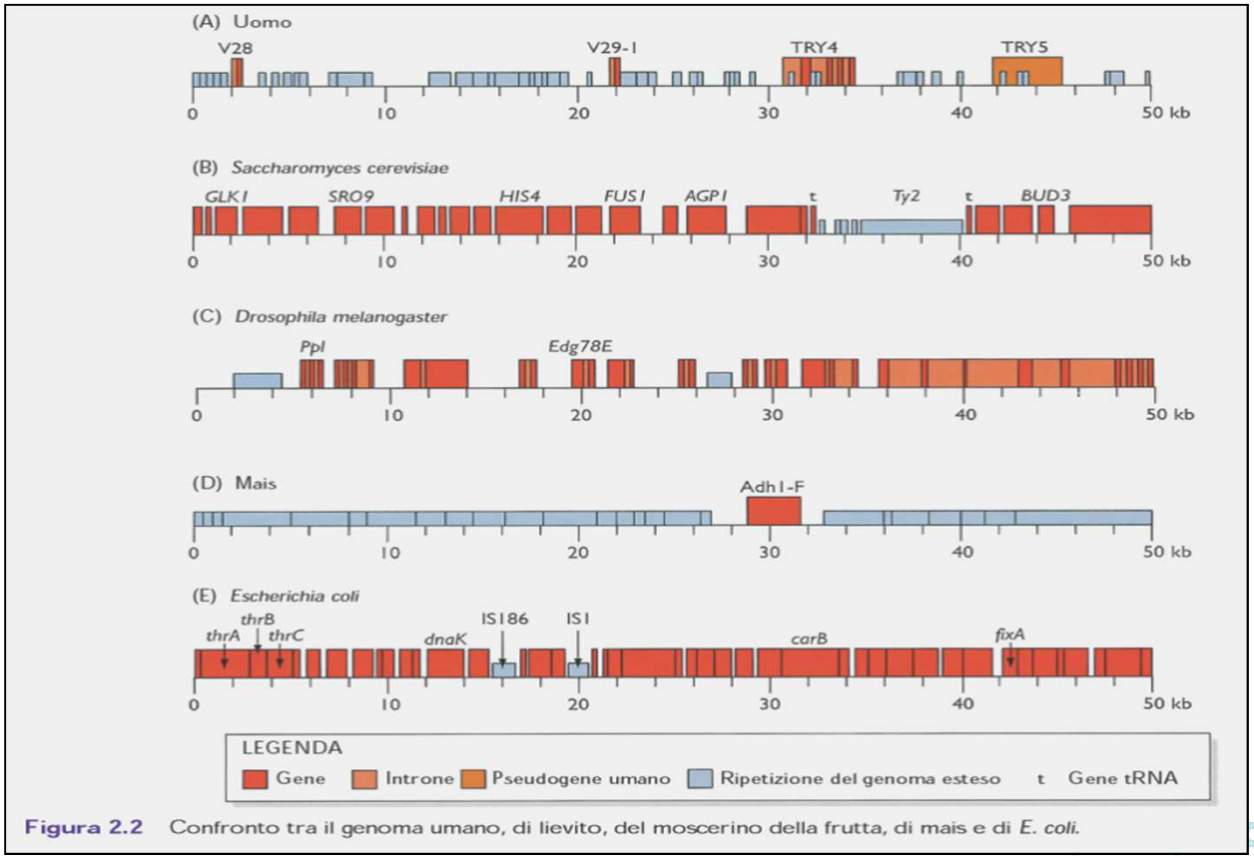


Figura 2.2 Confronto tra il genoma umano, di lievito, del moscerino della frutta, di mais e di *E. coli*.

Densità genica in un segmento di 50 kbp

duzione e modifica

Compattezza di alcuni genomi eucariotici

Proprietà del genoma	***S.cerevisiae***	***D.melanogaster***	***H. sapiens***
Densità genica (numero medio di geni per Mb)	479	79	11
Introni per gene (media)	0,04	3	9
% del genoma occupata dalle ripetizioni intersperse	3,4%	12%	44%

Vietate copia riproduzione e modifica

Come si riconoscono i geni?

Vietate copia riproduzione e modifica

Localizzazione dei geni in una sequenza di DNA

- Esame della sequenza con il computer nel tentativo di identificare caratteristiche spesso associate a geni

Bioinformatica

- Identificare un gene mediante **approcci sperimentali**

Vietate copia riproduzione e modifica

Dedurre dalla sequenza genomica i geni che codificano per le proteine

Individuazione dei moduli di lettura aperti (ORF) mediante analisi computazionale

Ogni ORF presenta un co
(ATG) ed un codone di terr
TAG, TGA)

Vietate copia riproduzione e modifica

Il punto chiave della corretta ricerca delle ORF sta nella frequenza con cui i codoni di terminazione appaiono in una sequenza di DNA

Vietate copia riproduzione e modifica

Se il DNA ha una sequenza casuale di nucleotidi ed un contenuto in CG del 50%, le sequenze corrispondenti ai 3 codoni di terminazione si presenteranno con una frequenza di 4^3 (64nt).

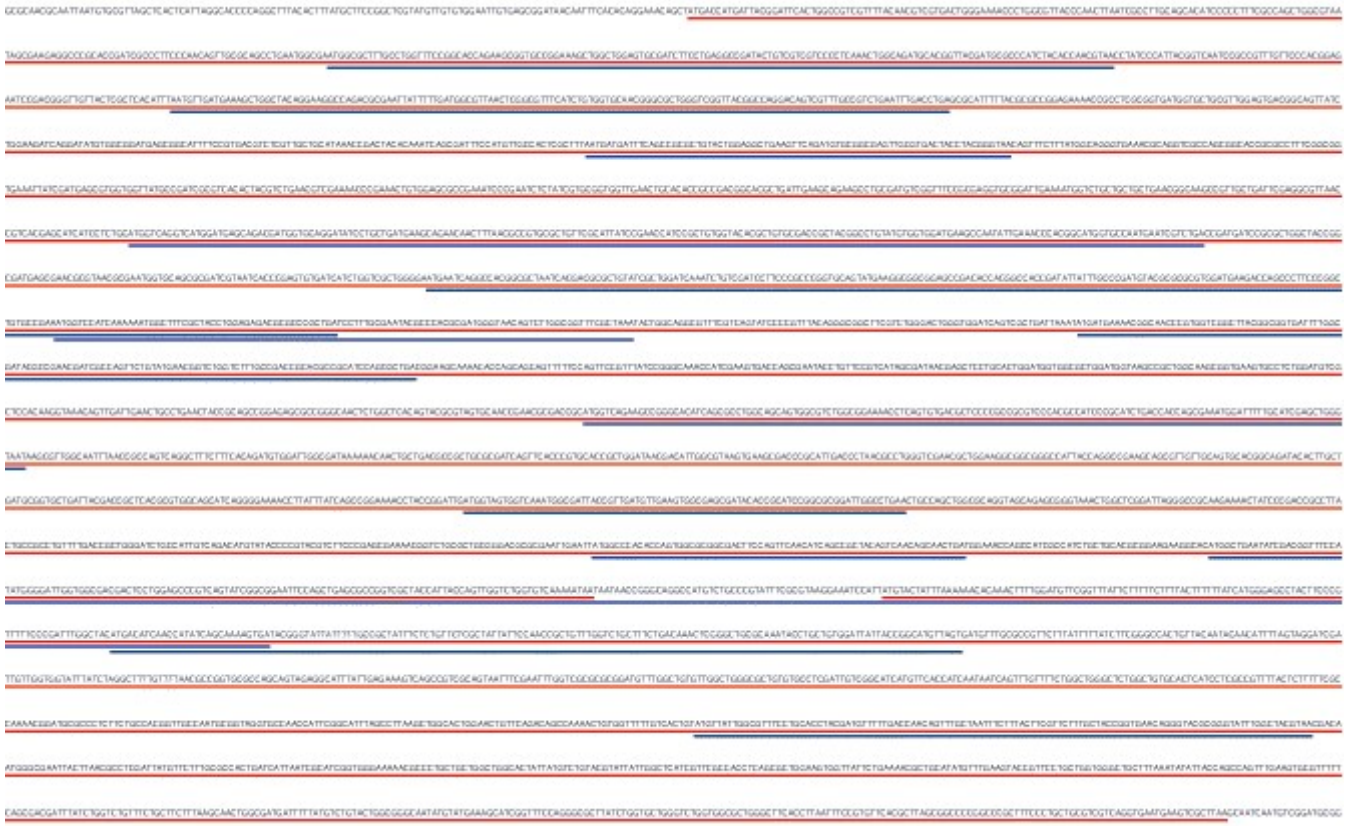
Se CG>50% i tre stop (ricchi in AT) avranno una frequenza minore (100-200nt ca.)

PERO'….

Vietate copia riproduzione e modifica

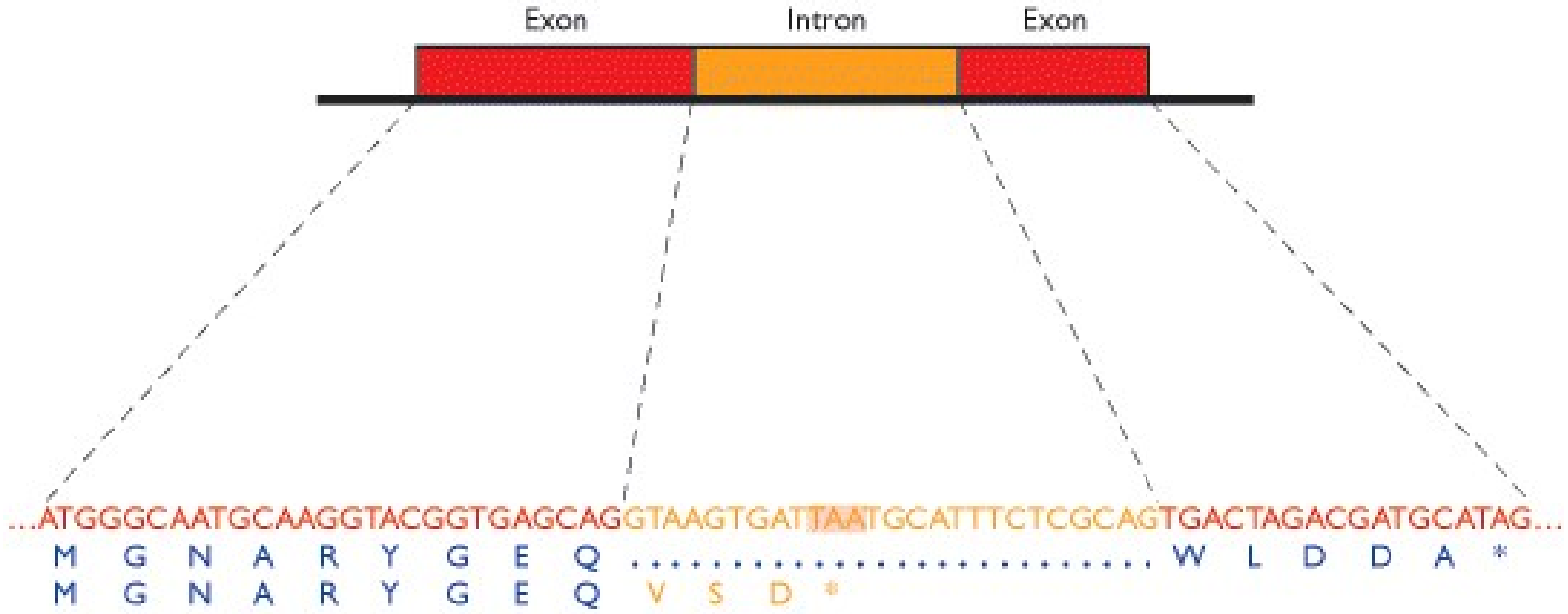
La scansione di una sequenza alla ricerca di una ORF permette di localizzare i geni nel genoma batterico.

Il diagramma mostra 4522 bp dell'operone del lattosio di E. Coli in cui risultano sottolineate tutte le ORF di lunghezza superiore ai 50 codoni. La sequenza contiene due geni reali - *lacZ* e *lacY* – sottolineati in rosso. Questi geni sono facilmente ed inequivocabilmente identificati in quanto molto più lunghi delle ORF spurie sottolineate in blu.



Vietate copia riproduzione e modifica

L'analisi delle ORF è più problematica con il DNA



Vietate copia riproduzione e modifica

Modifiche alla procedura di analisi bioinformatica delle ORF introdotte per minimizzare il disturbo dovuto alla presenza degli introni

1) I software per l'analisi delle ORF prendono in considerazione i **codoni preferenziali.**

Preferenzialità (specie-specifica) nell'uso dei codoni (*codon bias*)

Es. leucina TTA, TTG, CTT, CTC, CTA, CTG ma nell'uomo in genere solo CTG; valina GTG più frequentemente di GTA

SI INSERISCONO NEI SOFTWARE PER L'ANALISI DELLE ORF I CODONI PREFERENZIALI DEGLI ORGANISMI DA STUDIARE (I QUALI APPARIRANNO NELLE ORF Più FREQUENTEMENTE DI QUANTO LA CASUALITA' IMPONGA E DI QUANTO APPARIRANNO I SINONIMI)

Vietate copia riproduzione e modifica

2) ricerca delle **giunzioni di splicing**

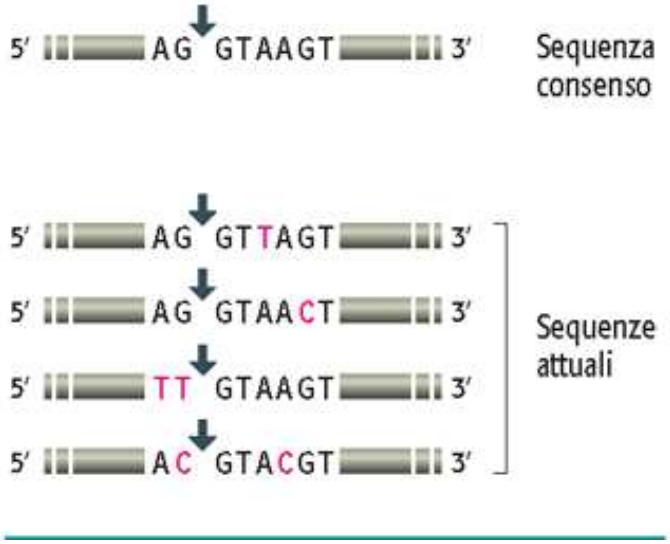
esone-introne

5′ AG|GTAAGT 3′
Consensus

introne-esone

5′ PyPyPyPyPyPyN CAG | 3
Consensus

rappresentano "consensus" con pochi nt veramente invariabili



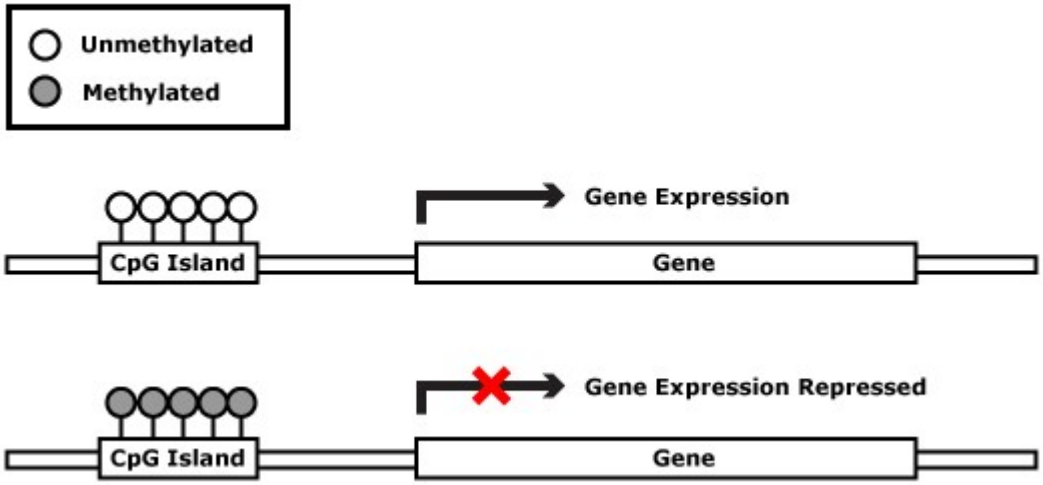
Nelle giunzioni esone-introne al 5′ soltanto la sequenza "GT" che segue immediatamente il sito di splicing è conservata!!!

Vietate copia riproduzione e modifica

3) ricerca di **regioni regolative al 5' dei geni**

-Siti di riconoscimento di fattori trascrizionali

-Isole CpG (nei vertebrati e nell'uomo per ca. il 50% dei geni)*



*In una sequenza di DNA di un vertebrato queste isole sono un forte indizio che un gene inizi nella regione subito a valle.

Vietate copia riproduzione e modifica

Isole CpG

Identificare un'isola CpG significa con ogni probabilità incontrare poco più a valle un gene.

Le isole CpG, infatti, sono elementi di controllo epigenetico dell'espressione genica

Le isole CpG controllano l'espressione genica dei geni a valle sulla base del loro stato di metilazione

Vietate copia riproduzione e modifica

Isole CpG

La metilazione di citosine è una delle più comuni modificazioni epigenetiche osservate nei genomi eucariotici. Nei vertebrati e nelle piante risulta metilato rispettivamente il 10% e il 30% delle citosine.

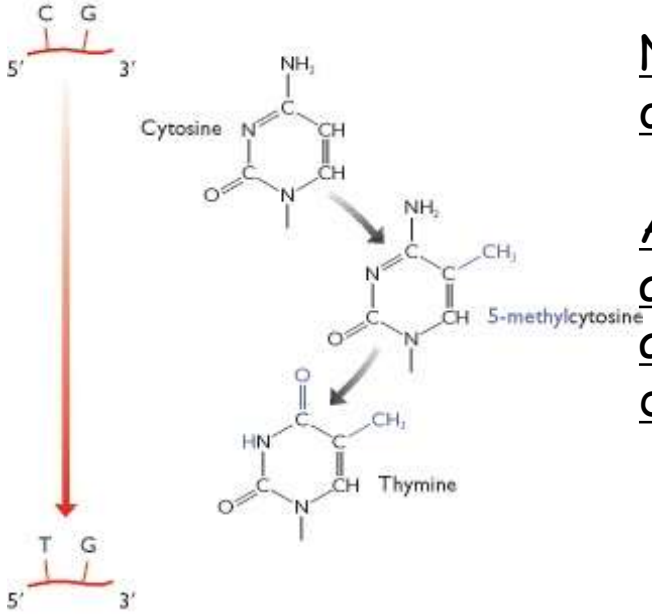
Le citosine metilate sono generalmente quelle presenti nel dinucleotide 5'-CpG-3'.

MA

La metilazione delle Citosine non è una cosa buona per il genoma

Vietate copia riproduzione e modifica

Le isole CpG



La 5-metil-citosina è soggetta a **deaminazione** formando timina. Il risultato di questo processo è che il dinucleotide CpG è generalmente evitato nel genoma dei vertebrati e delle piante.

Nel genoma umano infatti la frequenza osservata di CpG è circa 1/5 di quella attesa.

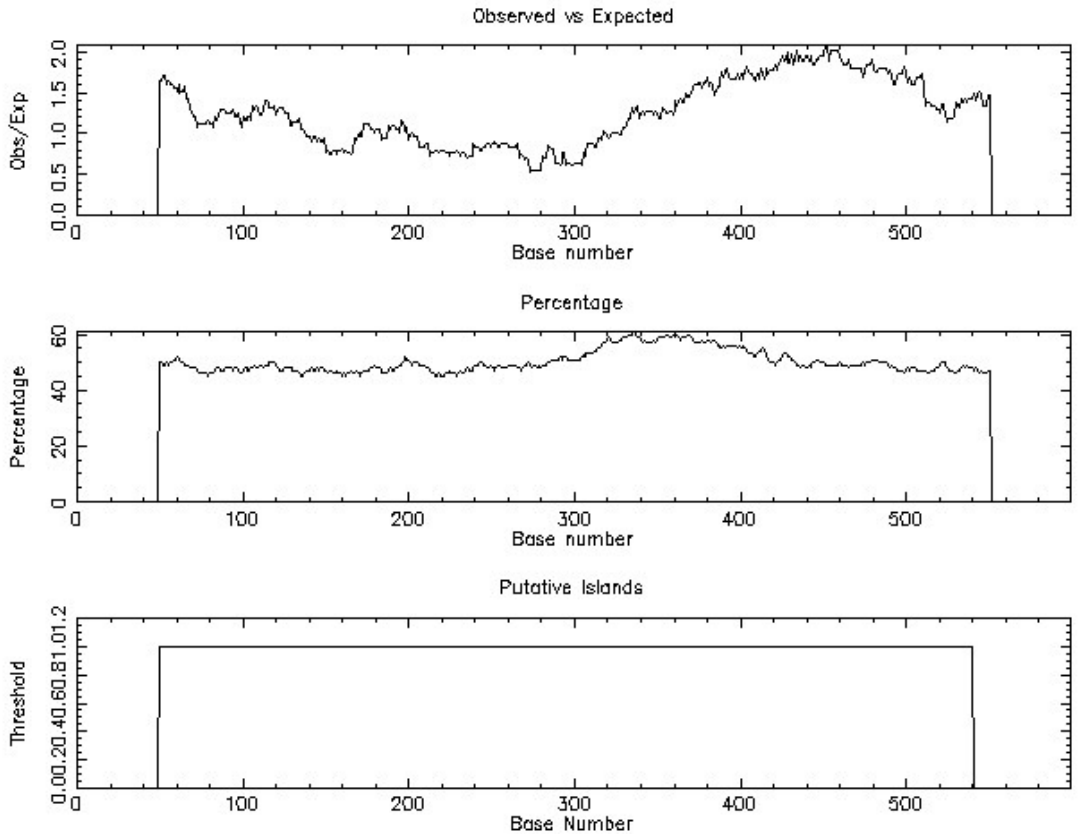
Anche nelle isole, dove la frequenza è più alta, essa è comunque inferiore a quella attesa sulla base di una distribuzione casuale dei nucleotidi (circa 1/3)

Vietate copia riproduzione e modifica

Isole CpG

Nota la sequenza genomica è possibile predire la localizzazione delle isole GpG con programmi bioinformatici. La definizione operativa che viene comunemente utilizzata per la definizione di un'isola CpG nei mammiferi è la seguente:

- L > 500 bp
- C+G% > 55%
- CpG Obs/Exp* > 0,65



Vietate copia riproduzione e modifica

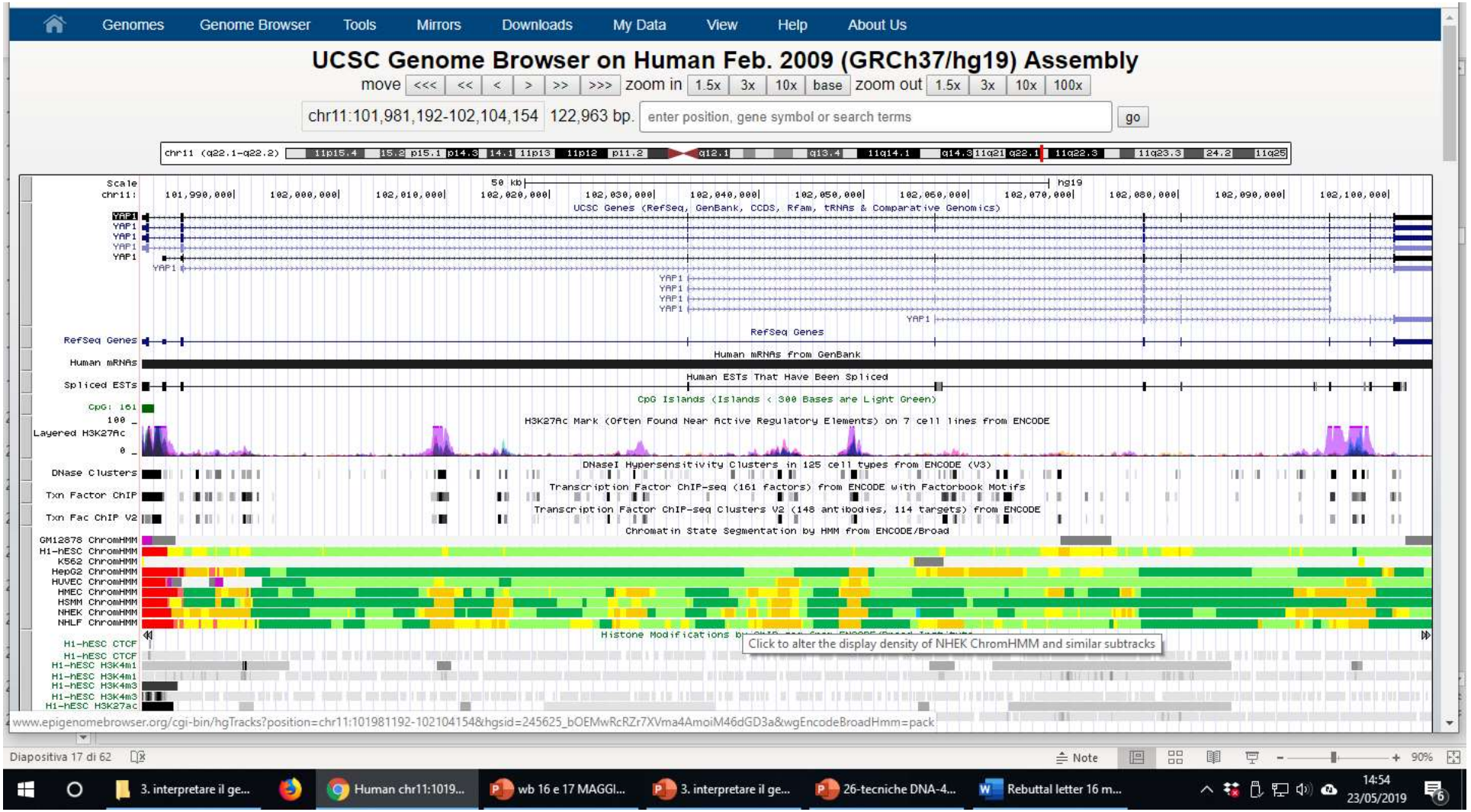
Le isole CpG

Le isole CpG sono localizzate nella regione del promotore di circa il 50% dei geni umani, la maggior parte dei quali di tipo costitutivo (housekeeping, espressi in molti tessuti diversi = isole CpG non metilate).

I geni con isole CpG metilate sono generalmente tessuto specifici (si esprimono dove le loro isole non sono metilate).

Vietate copia riproduzione e modifica

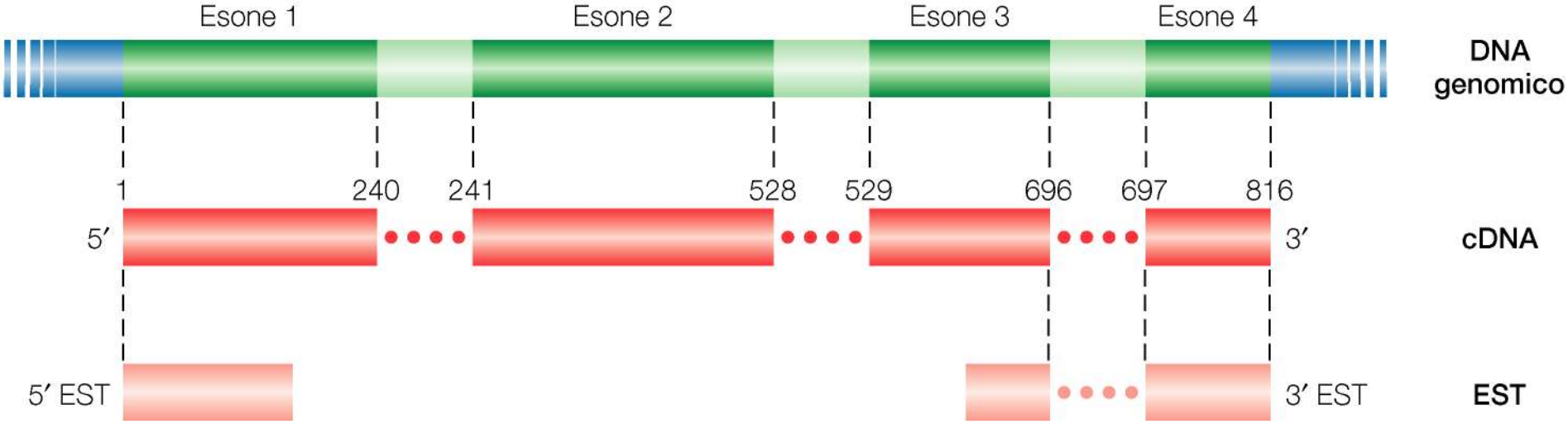
La ChIP seq analysis ha permesso di mappare il genoma con siti per fattori trascrizionali e con modifiche epigenetiche



Vietate copia riproduzione e modifica

Le ORF possono essere pescate da collezioni di cDNA o di EST (evidenza diretta della trascrizione di una sequenza)

EST: Espressed Sequence Tags, parte di geni espressi che possono essere utilizzati come sonde, ad es. nei microArray oppure nella scoperta di nuovi geni, nella determinazione della loro sequenza e della loro posizione nel genoma.



Allineamento con il DNA genomico di un cDNA completamente sequenziato e di sequenze EST

IDENTIFICARE LA SEQUENZA DELL'UNITA' DI TRASCRIZIONE

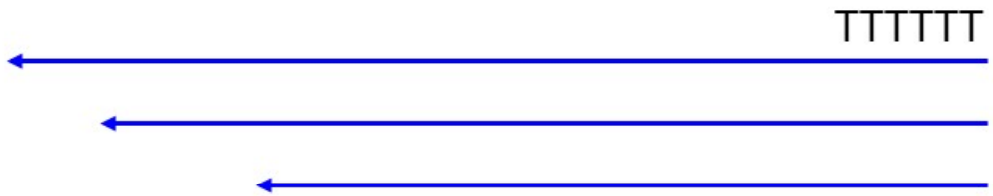
SPLICING

LOCALIZZARE L'ORF

Vietate copia riproduzione e modifica

La costruzione del cDNA

Le sequenze di cDNA ottenute dall'mRNA sono generalmente tronche



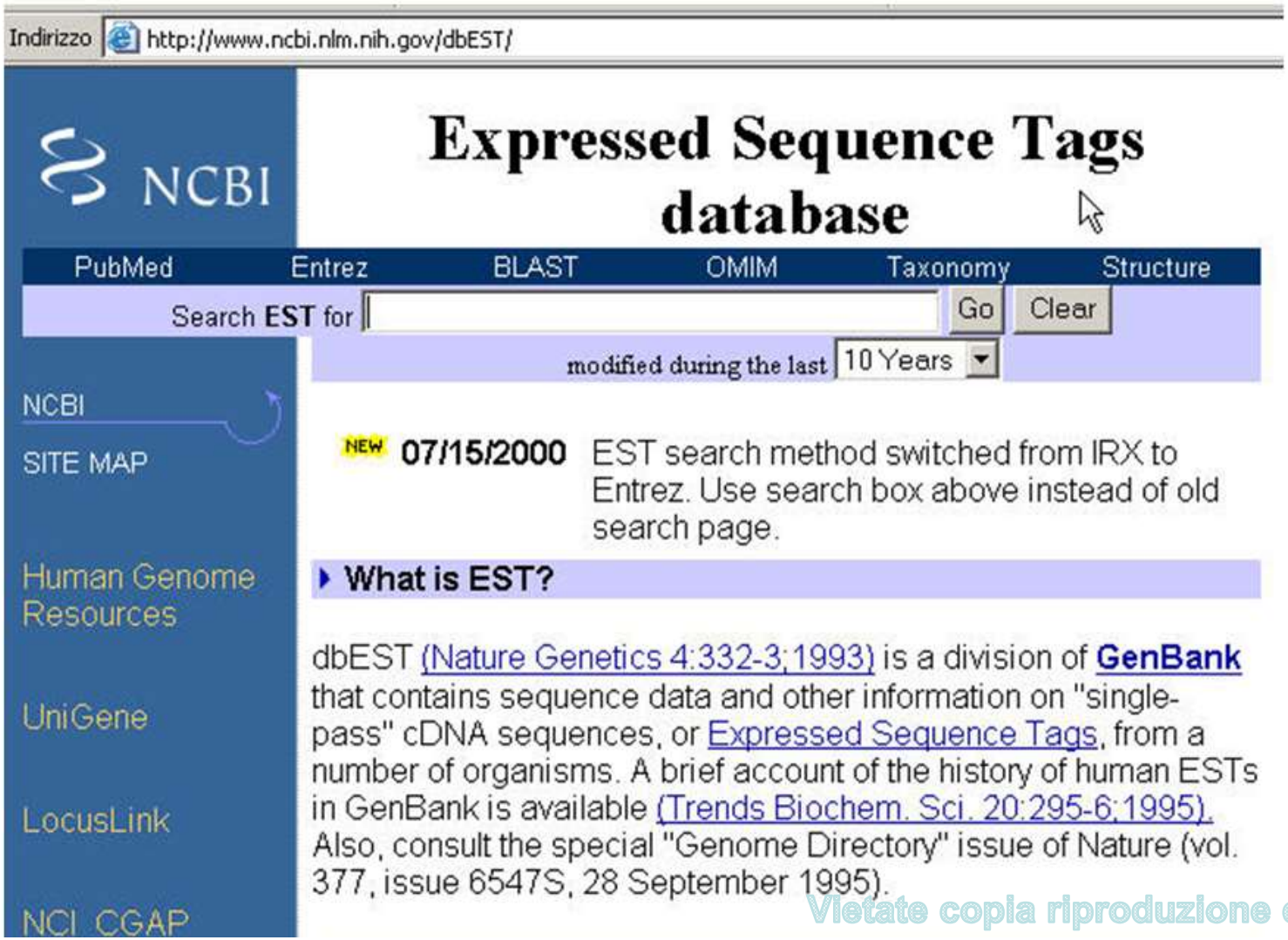
ona

Vietate copia riproduzione e modifica

cDNA, EST e banche dati

dbEST (pronuncia 'the best')

Divisione di GenBank che contiene tutte le sequenze EST, classificate per specie, tessuto, patologia…

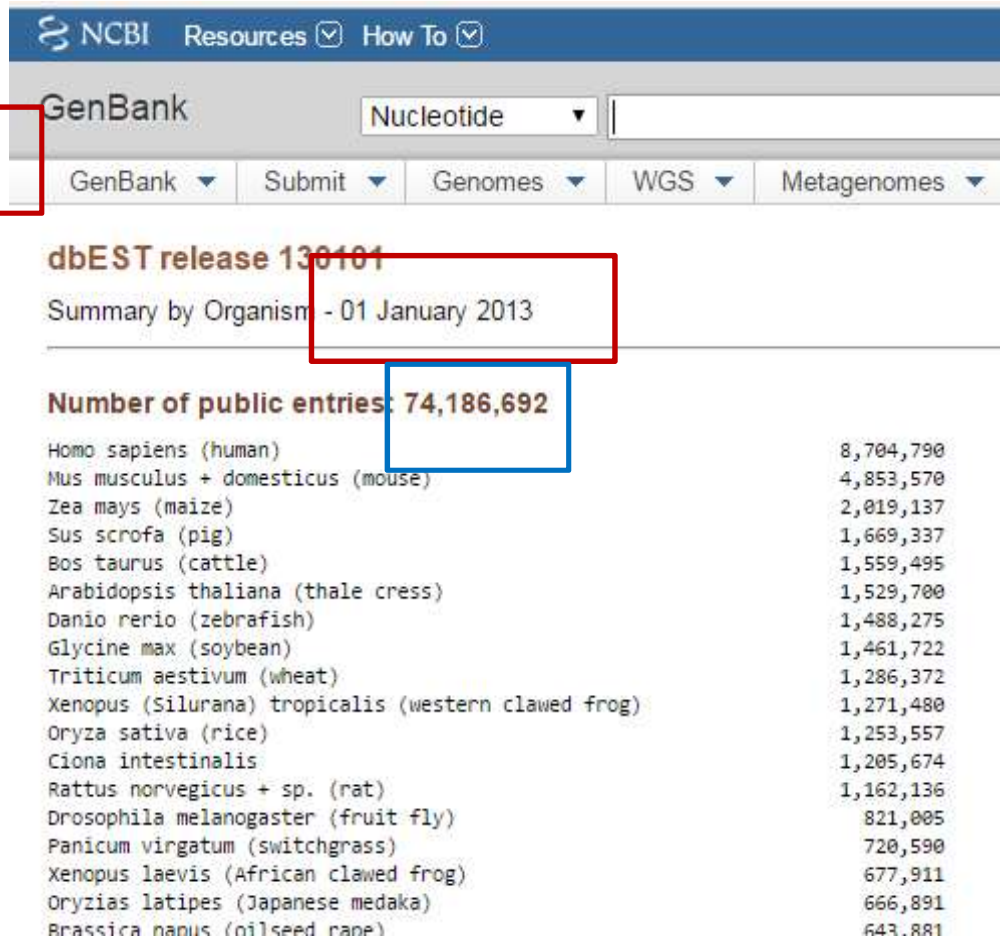


Vietate copia riproduzione e modifica

dbEST release 103103
Summary by Organism - October 31, 2003

Number of public entries: 18,971,362

Homo sapiens (human)	5,427,521
Mus musculus + domesticus (mouse)	3,915,334
Rattus sp. (rat)	538,251
Triticum aestivum (wheat)	500,902
Ciona intestinalis	492,488
Gallus gallus (chicken)	451,565
Zea mays (maize)	383,759
Danio rerio (zebrafish)	362,445
Hordeum vulgare + subsp. vulgare (barley)	348,233
Xenopus laevis (African clawed frog)	344,747
Glycine max (soybean)	341,578
Bos taurus (cattle)	329,387
Drosophila melanogaster (fruit fly)	261,414
Oryza sativa (rice)	260,890
Saccharum officinarum	246,301
Caenorhabditis elegans (nematode)	215,200
Silurana tropicalis	209,240
Arabidopsis thaliana (thale cress)	190,732
Medicago truncatula (barrel medic)	187,763
Sus scrofa (pig)	171,920



Vietate copia riproduzione e modifica

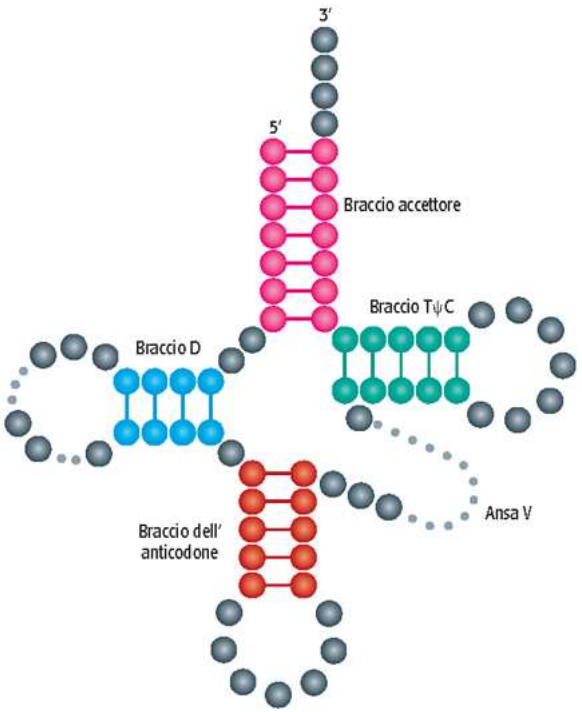
Individuazione di geni che codificano per RNA funzionali

Questi geni non contengono ORF.

Possibilità di APPAIAMENTO INTRAMOLECOLARE

Vietate copia riproduzione e modifica

(A) Struttura a quadrifoglio del tRNA



Tutti i tRNA si ripiegano in una caratteristica struttura a quadrifoglio che è stabilizzata da appaiamenti intramolecolari in quattro regioni diverse.

L'individuazione di queste regioni di complementarietà permette di individuare con una certa facilità i geni per i tRNA.

(B) Sequenza di uno dei geni di *Escherichia coli* $tRNA^{leu}$

5′ GCCGAAGTGGCGAAATCGGTAGTCGCAGTTGATTCAAAATCAACCGTAGAAATACGTGCCGGTTCGAGTCCGGCCTTCGGCACCA 3′

Vietate copia riproduzione e modifica

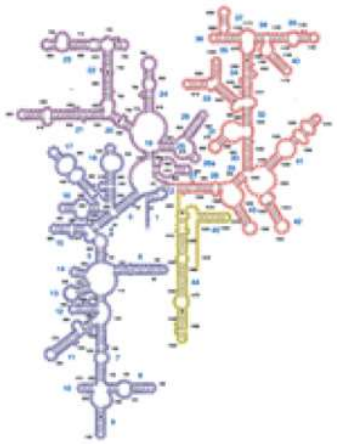
Anche gli rRNA e alcuni piccoli RNA funzionali adottano strutture secondarie con una complessità sufficiente da permettere di identificare i geni corrispondenti nel genoma.

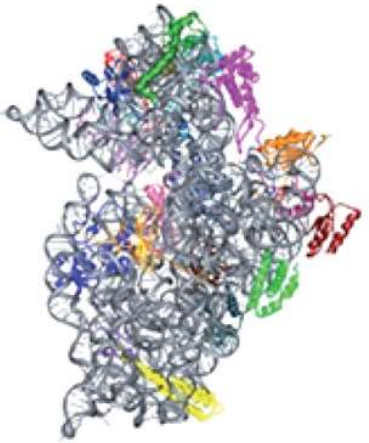
Vietate copia riproduzione e modifica

A) struttura secondaria dell'rRNA

B) struttura tridimensionale dell'rRNA

C) struttura della subunità ribosomale

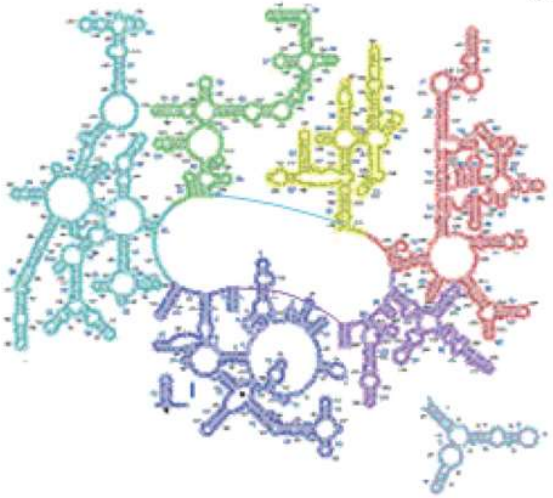


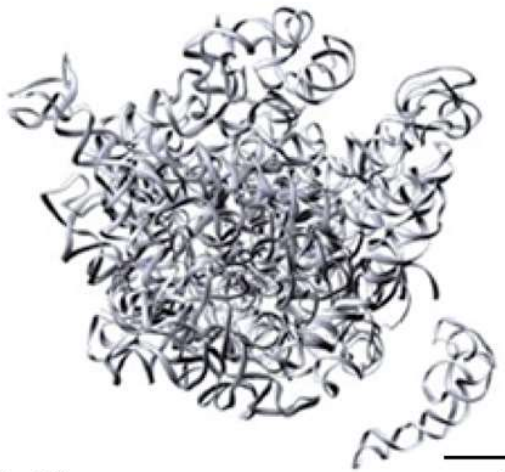


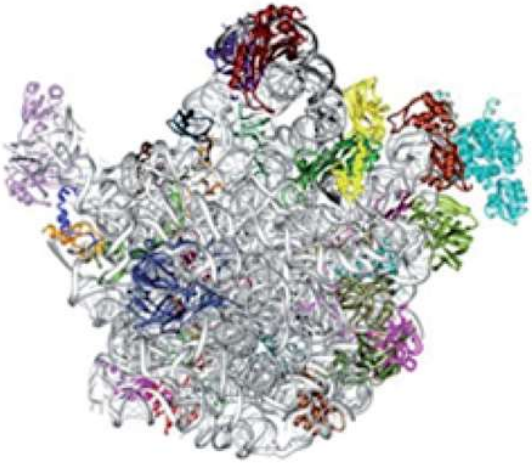
rRNA 16S

+21 r-proteine

subunità 30S





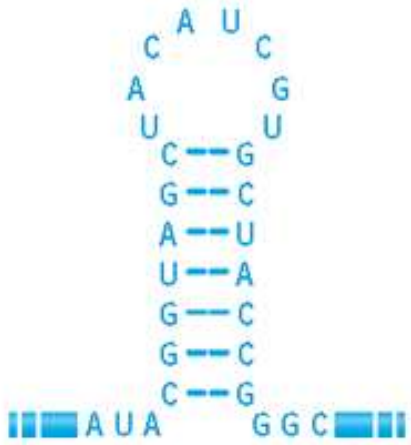


rRNA 23S+5S

+34 r-proteine

subunità 50S

Vietate copia riproduzione e modifica



Struttura a stem-loop di una molecola di RNA

Anche gli RNA funzionali che non assumono strutture secondarie complesse sono comunque caratterizzati da più o meno sequenze in grado di creare strutture a forcina.

Una sequenza compatibile con una forcina abbastanza stabile (un minimo di contenuto in CG) rappresenta un indicatore della presenza di un gene per RNA

Vietate copia riproduzione e modifica

Anche nel caso di geni per RNA funzionali si possono individuare **sequenze regolative**, che sono diverse da quelle dei geni codificanti proteine. Alcune possono trovarsi anche all'interno del gene.

Nei genomi compatti bisogna fare molta attenzione al DNA che rimane dopo una estensiva ricerca e individuazione di geni codificanti proteine: **gli spazi "vuoti"** sono spesso occupati da geni per RNA

Vietate copia riproduzione e modifica

Le analisi di omologie di sequenza sono un ulteriore strumento per l'identificazione dei geni

Confronto tra la sequenza in esame e tutte le sequenze presenti in banca dati, cercando similarità o identità con geni già sequenziati.

L'omologia può indicare geni correlati evolutivamente.

L'analisi, oltre a permettere la validazione di esoni della cui ORF non si è certi, serve anche ad assegnare funzioni ad un gene appena scoperto

Vietate copia riproduzione e modifica

L'individuazione dei geni deve considerare anche la

complessità dell'organizzazione di singoli loci genici

Vietate copia riproduzione e modifica

Un gene può avere più inizi, più terminazioni e più processamenti del suo RNA

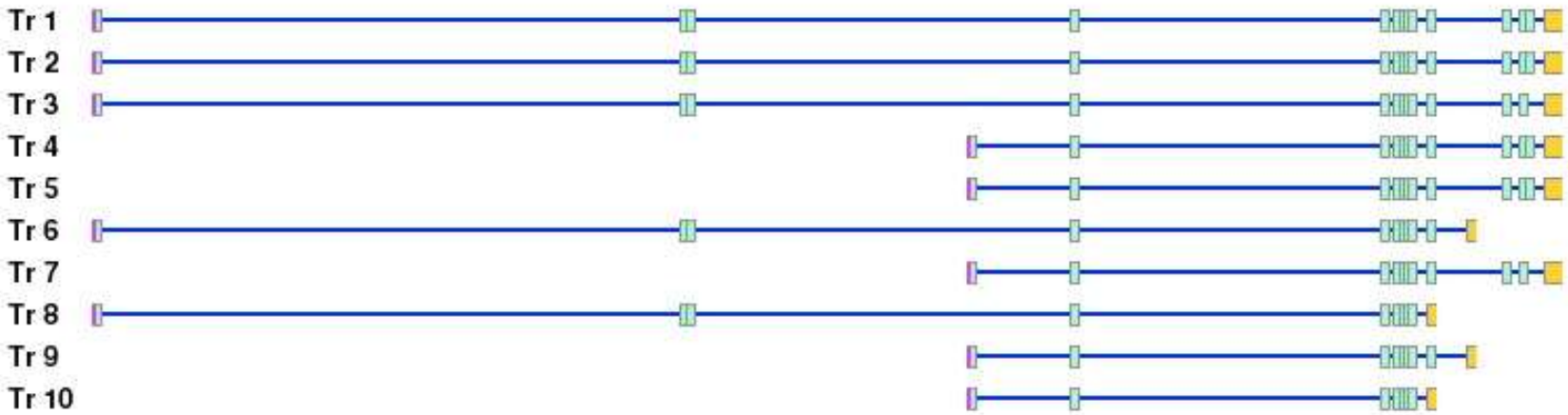
Per giungere ad una definizione il più possibile corretta di **GENE** è necessario conoscerne le caratteristiche principali.

- Un gene può utilizzare diversi promotori

- La trascrizione di un gene si può arrestare in corrispondenza di diversi terminatori

- I trascritti espressi da un gene possono subire splicing alternativo che generano trascritti che differiscono sia nelle regioni non tradotte (5' e 3'UTR) che nella regione codificante

Vietate copia riproduzione e modifica



Il gene per tp73L (fattore trascrizionale della famiglia p53)
codifica per **10 trascritti alternativi**
utilizza **2 promotori** e **3 diversi terminatori della trascrizione**
(predizione ottenuta dal programma ASPIC).

Vietate copia riproduzione e modifica

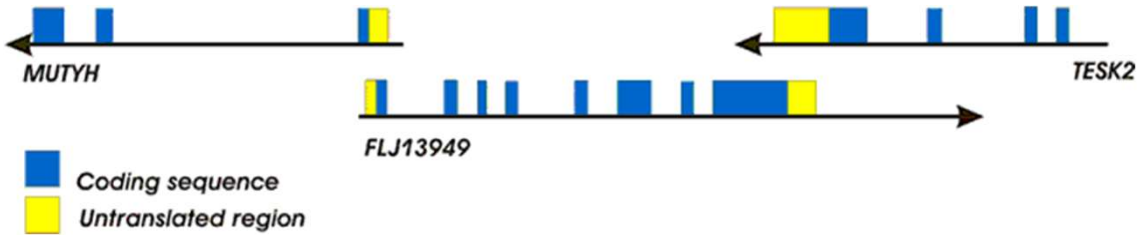
Quindi uno stesso gene con tanti promotori e tanti terminatori.

Ma un gene con lo stesso promotore e lo stesso terminatore può dare comunque trascritti diversi: lo splicing alternativo

Vietate copia riproduzione e modifica

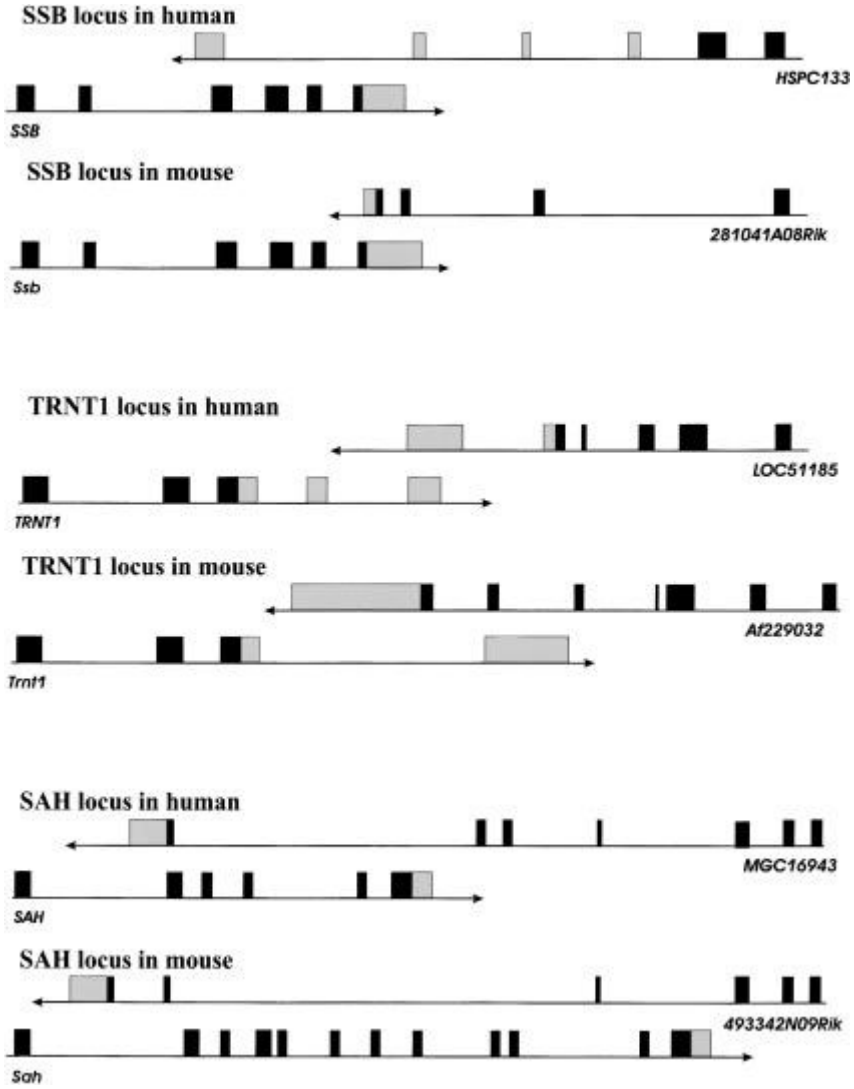
I geni possono essere sovrapposti

I geni possono essere sovrapposti tra loro, nello stesso orientamento o in orientamento opposto.



vedi: http://posnania.biotec.psu.edu/research/overlapping_genes.html

Vietate copia riproduzione e modifica

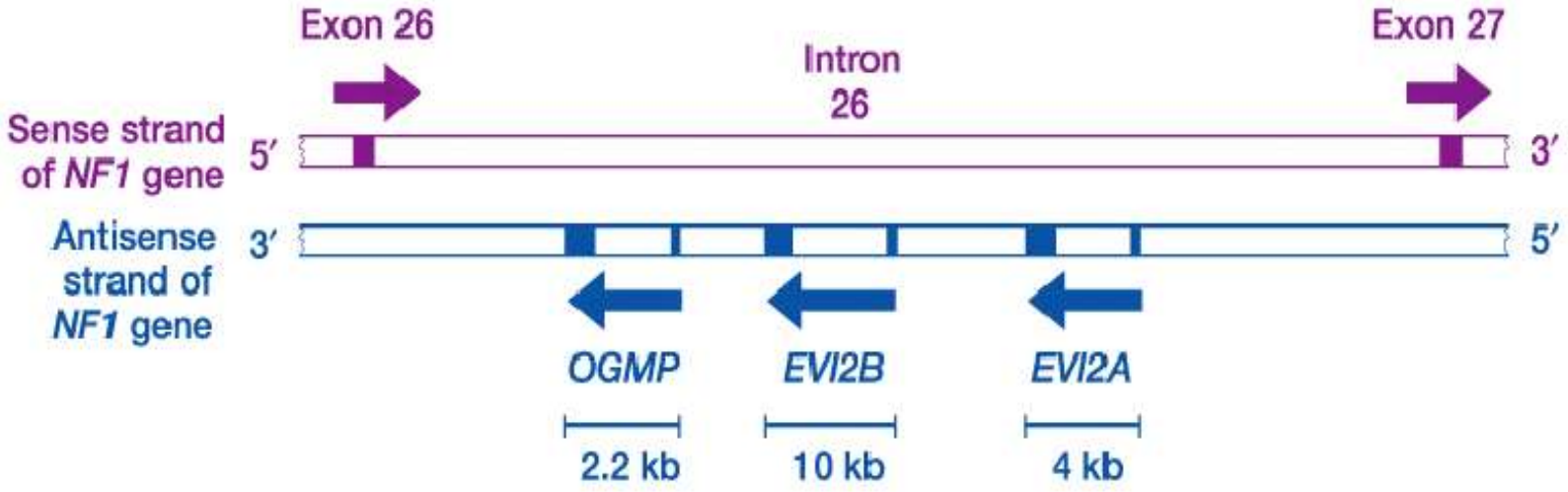


Esempi di geni con differenti pattern di sovrapposizione in uomo e topo (geni ortologhi).

Box neri: sequenze codificanti
Box grigi: sequenze non tradotte

Vietate copia riproduzione e modifica

L'introne 26 del gene *neurofibromatosis type I* (NF1) contiene 3 geni diversi nell'orientamento opposto (OMGP, EVI2A, EVI2B).

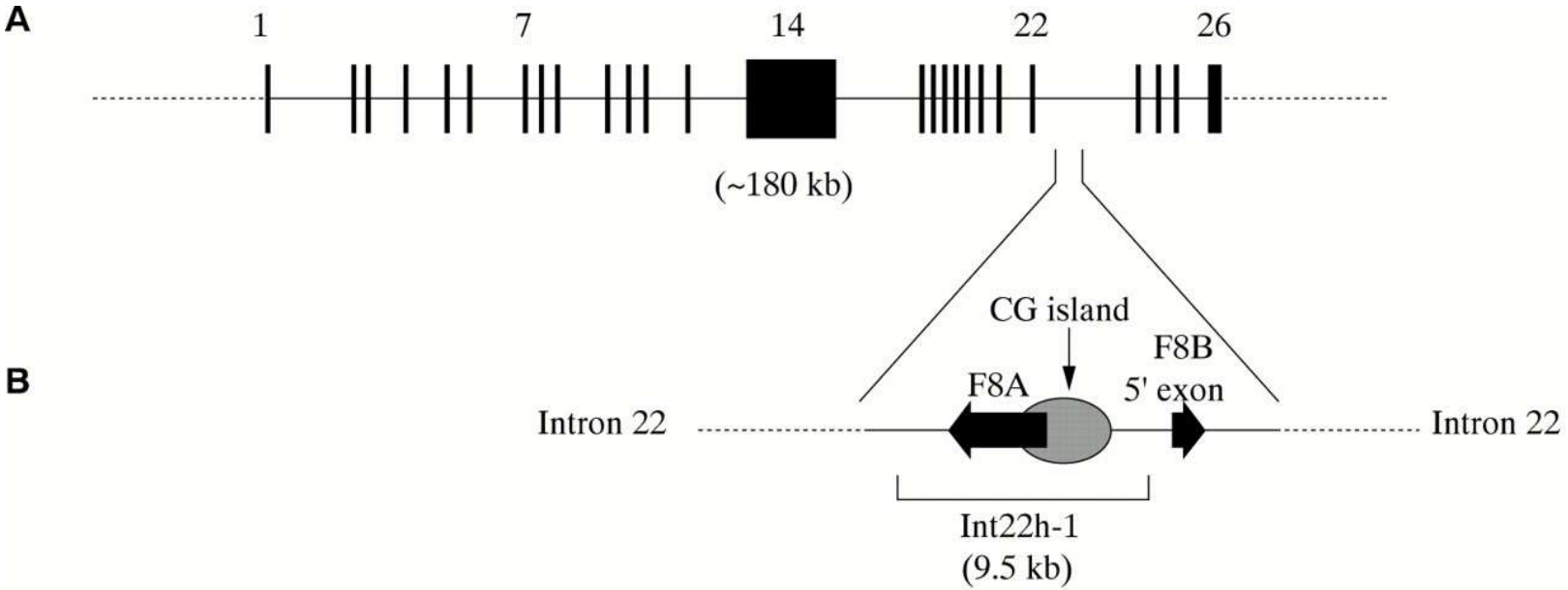


Mol Cell Biol. 1991 Feb;11(2):906-12.

The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene.

Viskochil D[1], Cawthon R, O'Connell P, Xu GF, Stevens J, Culver M, Carey J, White R.

Vietate copia riproduzione e modifica



Fattore VIII della coagulazione: l'introne 22 contiene due geni che utilizzano la stessa isola CpG nelle due direzioni. Il gene F8A rimane nell'introne 22 e viene abbondantemente trascritto in molti tipi cellulari ed utilizzando il filamento opposto a F8; è molto conservato (funzione).

F8B sintetizza un corto mRNA che ha un esone nuovo + gli esoni dal 23 al 26 di F8 (proteina più corta dell'F8 canonico)

Vietate copia riproduzione e modifica

I geni che codificano per i microRNAs sono spesso dei geni nei geni

La loro sequenze codificanti occupano spesso gli introni di geni codificanti per polipeptidi

Più del 50% dei geni per i microRNAs sono in cluster e possono essere trascritti in un unico RNA policistronico successivamente processato

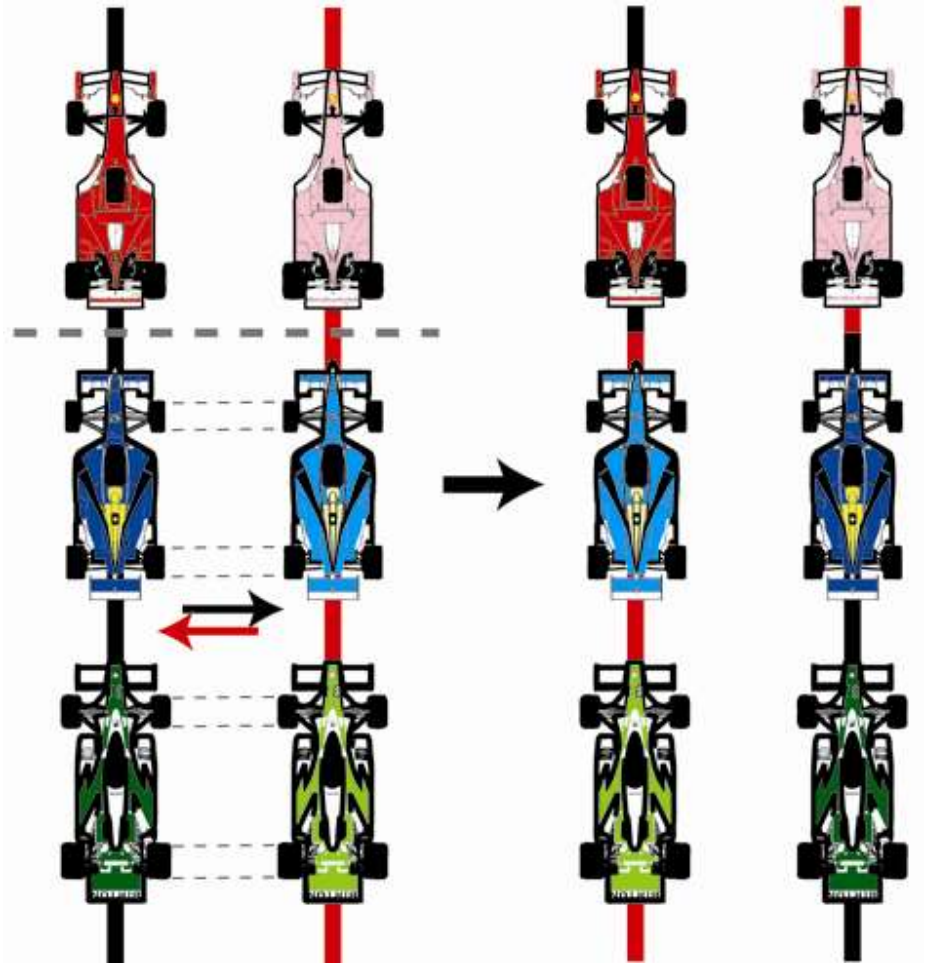
Vietate copia riproduzione e modifica

Nel genoma ci sono anche geni finti, gli Pseudogeni

(Pseudogeni duplicati o non processati
e Pseudogeni retrotrasposti o
processati)

Vietate copia riproduzione e modifica

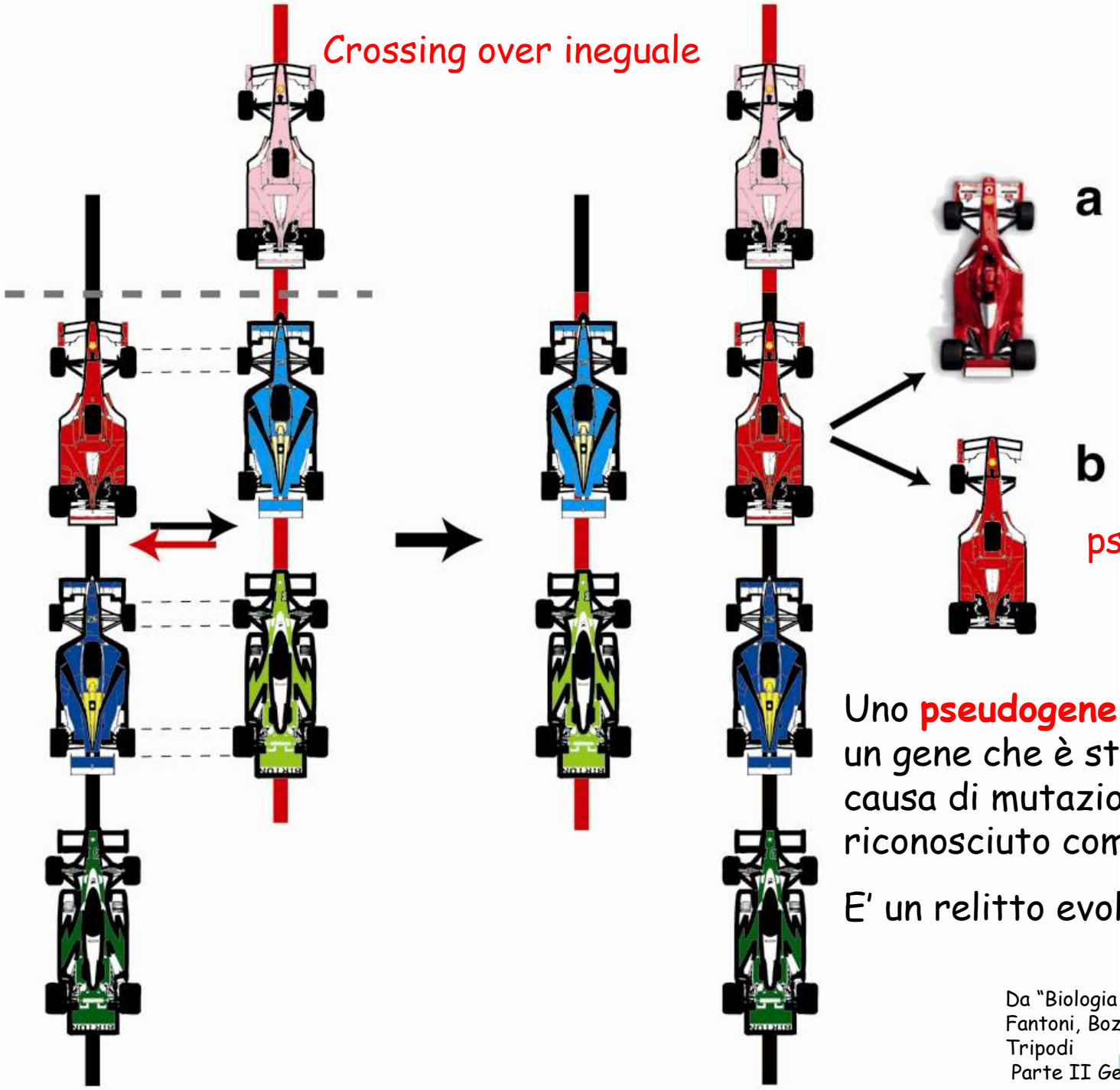
Gli **pseudogeni duplicati, o non processati,** (circa 11000 nel genoma umano) derivano da duplicazione in tandem o da crossing-over ineguale



Crossing over normale

Da "Biologia cellulare e genetica"
Fantoni, Bozzaro, Del Sal, Ferrari, Tripodi
Parte II Genetica, Piccin

Vietate copia, riproduzione e modifica



Uno **pseudogene convenzionale** è un gene che è stato inattivato a causa di mutazioni (non viene più riconosciuto come gene).

E' un relitto evolutivo.

Da "Biologia cellulare e genetica"
Fantoni, Bozzaro, Del Sal, Ferrari, Tripodi
Parte II Genetica, Piccin

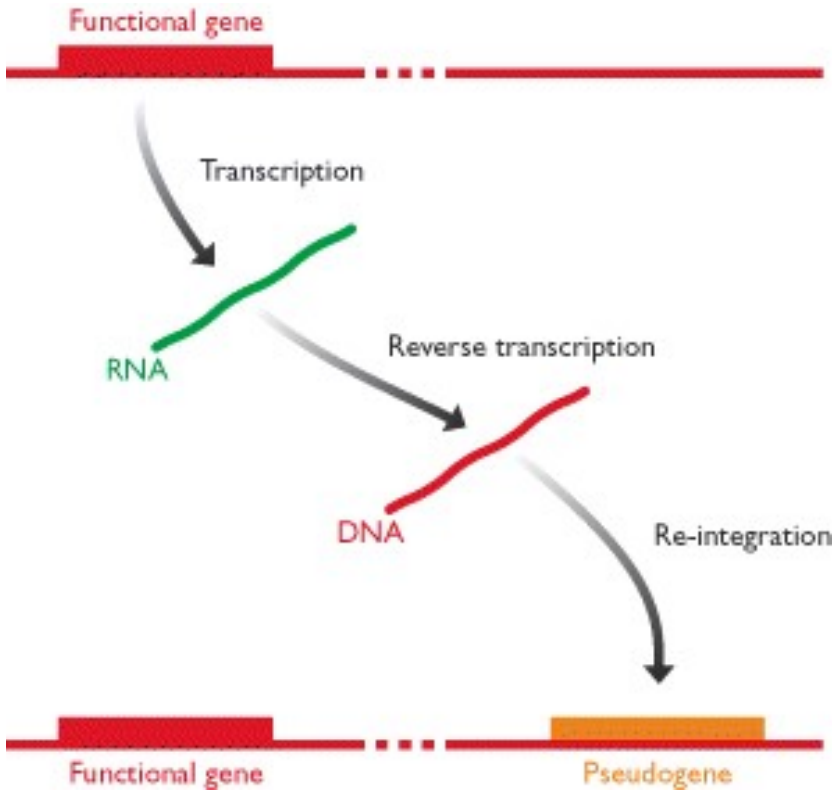
Lucidi a modifica

Uno **pseudogene retrotrasposto, o processato,**

deriva dall'mRNA di un gene su cui viene sintetizzata una copia di DNA che successivamente viene reinserita nel genoma.

Uno pseudogene processato non contiene le sequenze introniche e le sequenze 5-UTR che regolano l'espressione del gene.

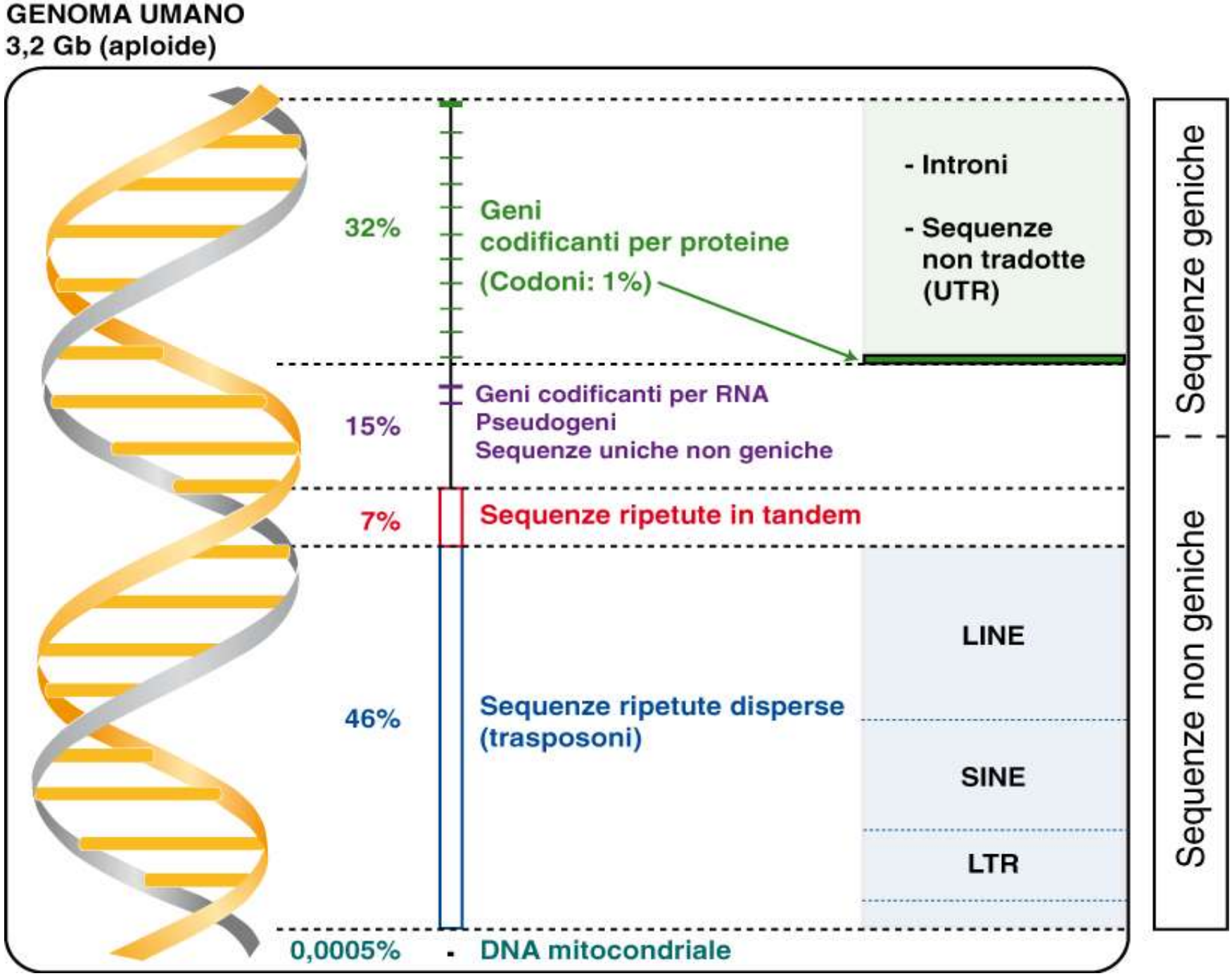
Uno pseudogene processato è inattivo.



Vietate copia riproduzione e modifica

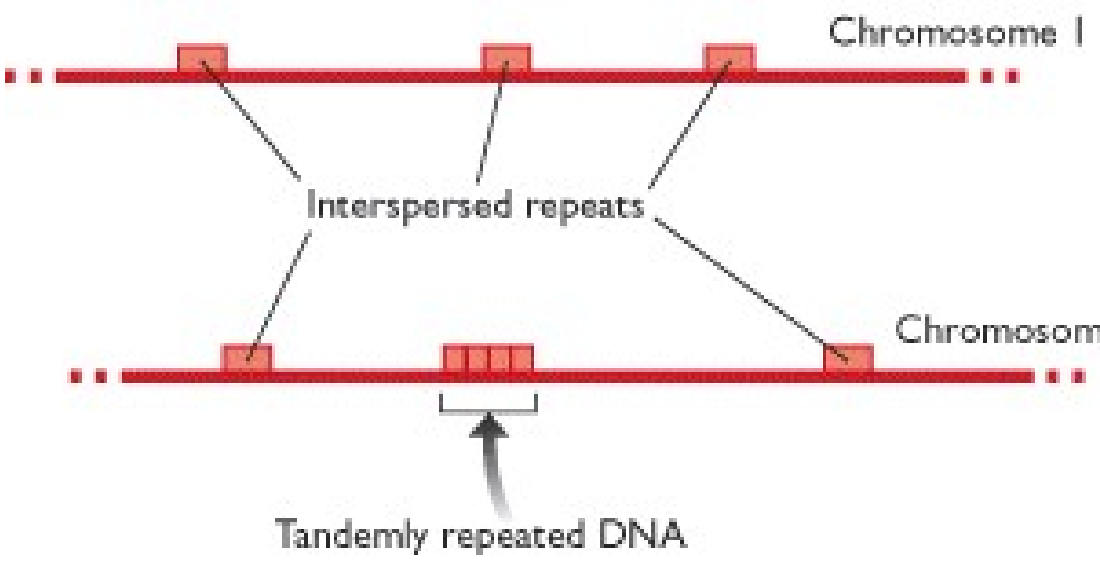
4) Sequenza ripetute

Vietate copia riproduzione e modifica



Da "Biologia cellulare e genetica" Fantoni, Bozzaro, Del Sal, Ferrari, Tripodi Parte II Genetica, Piccin

Vietata copia riproduzione e modifica



ere distinto in
:

cui unità sono
nodo casuale e
genoma

Vietate copia riproduzione e modifica

Sequenze ripetute a tandem

DNA satellite: costituisce la maggior parte delle regioni di eterocromatina, in particolare di quella pericentrica e centromerica. Unità ripetute di 5-171 nt.

DNA minisatellite: ipervariabile (in dimensioni), telomerico. L'unità ripetuta può essere lunga fino a 64 nt.

DNA microsatellite: STR= Short Tandem Repeats (1-6 bp che si susseguono nell'ambito di piccoli blocchi di lunghezza inferiore a 150 bp).
Uniformemente interdispersi in tutti i cromosomi.
Variabilità molto elevata nel numero di unità ripetute (VNTR=Variable Number of Tandem Repeats) riscontrabile tra diversi individui.

Vietate copia riproduzione e modifica

Sequenze ripetute a tandem: funzione?

- Si sa che derivano da errori nel processo di copiatura del genoma durante la divisione cellulare mitotica (scivolamento replicativo) o meiotica (crossing over ineguale) e potrebbero essere prodotti inevitabili della replicazione del genoma

- si ritiene che la loro funzione sia di natura strutturale e che quindi abbiano un qualche ruolo nell'organizzazione dei cromosomi.

Vietate copia riproduzione e modifica

Ripetizioni Intersperse

L'elemento di sequenza che ricorre più volte
è distribuito nel genoma in tante
localizzazioni diverse.

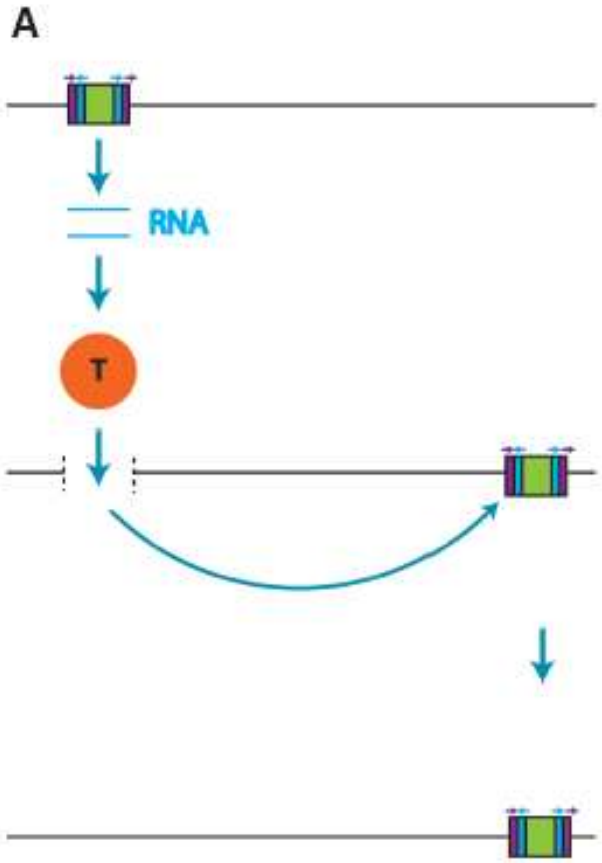
Vietate copia riproduzione e modifica

Trasposoni a DNA: tratti di DNA escissi e reinseriti in altri punti del genoma. Possono essi stessi codificare per enzimi che li rendono autonomi nel processo di trasposizione. Si tratta di una trasposizione non replicativa.

Retrotrasposoni: derivano da trascritti di RNA copiati dalla Trascrittasi Inversa e integrati nel genoma. Si tratta di una trasposizione replicativa.

Vietate copia riproduzione e modifica

Trasposoni a DNA:

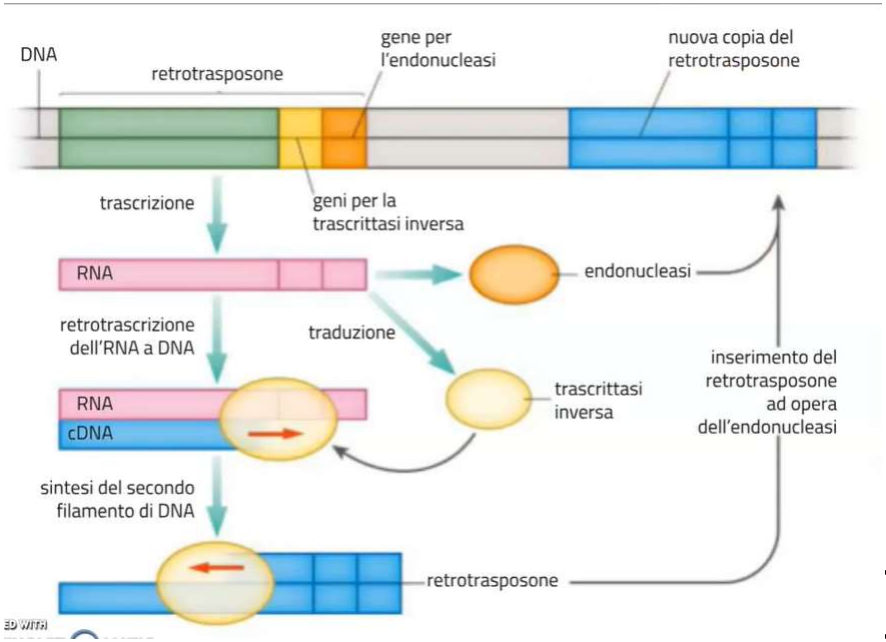


La sequenza del trasposone viene trascritta e tradotta in una proteina ad attività enzimatica (trasposasi). Questa rientra nel nucleo e rimuove la sequenza del trasposone dalla sua localizzazione originaria e la inserisce in una nuova. L'interruzione nel DNA viene riparata.

Sono elementi mobili del genoma che non creano sequenze ripetute e che quindi non sono causa di espansione del genoma.

Vietate copia riproduzione e modifica

Retrotrasposoni (derivano da retrovirs)
La sequenza del trasposone viene trascritta e tradotta nelle proteine orf1 e orf2 (endonucleasi e RT) .



L'RNA rientra nel nucleo associato alle proteine a cui ha dato origine. La proteina orf2 ha attività di trascrittasi inversa e di endonucleasi per cui converte l'RNA in cDNA e lo reinserisce in una nuova localizzazione.

Le attività enzimatiche possono essere utilizzate anche da trasposoni che non codificano per essi (elementi non autonomi)

Vietate copia riproduzione e modifica

Quasi metà del genoma umano deriva da elementi trasponibili

Nel genoma umano la quantità di DNA derivante da elementi trasponibili è **20 volte** la quantità di DNA che codifica per tutte le proteine umane

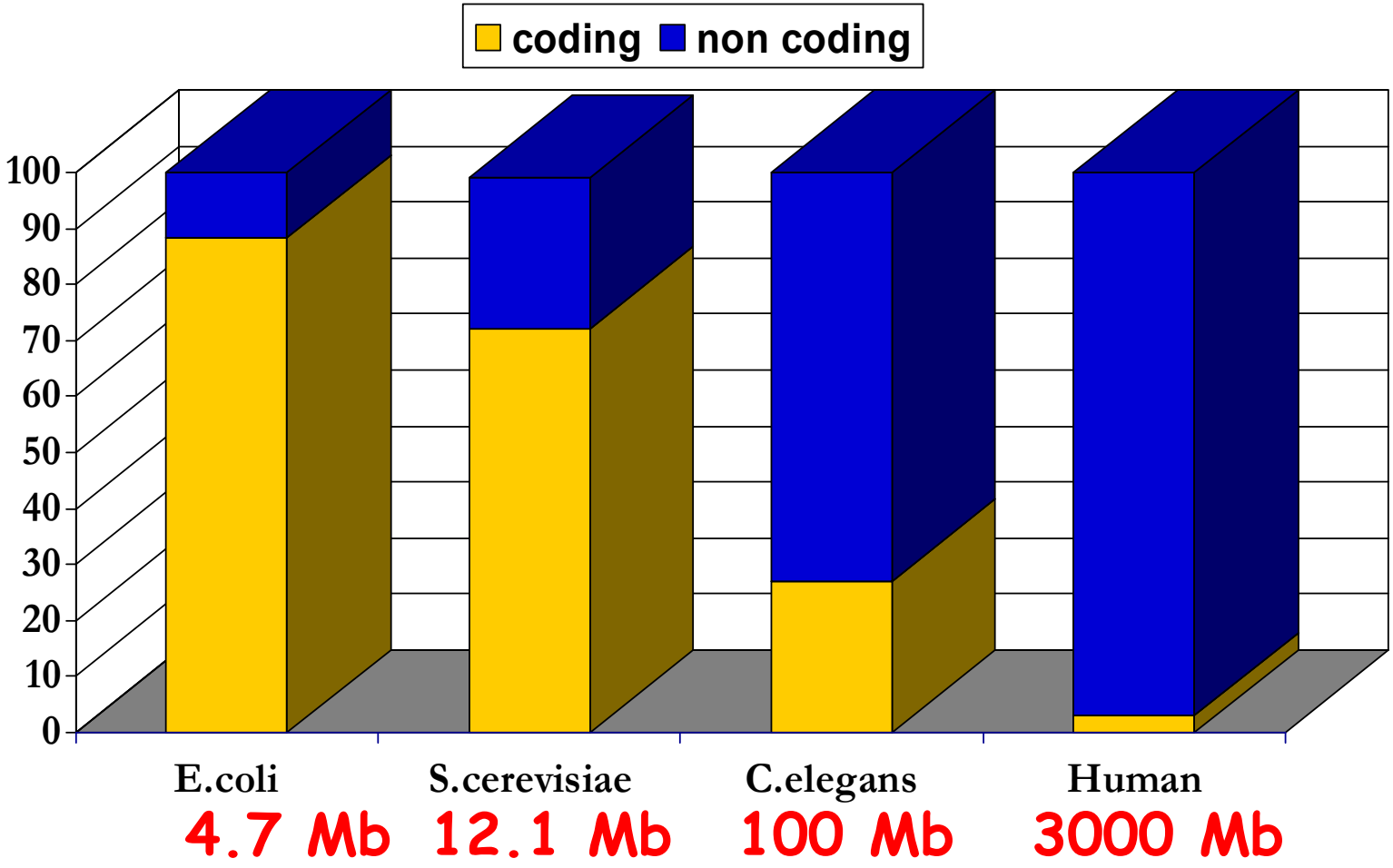
Vietate copia riproduzione e modifica

Ishihara's test per il daltonismo.
Siamo talmente impegnati nella ricerca di nuovi geni che non riusciamo a vedere ciò che gene non è!

Non possiamo ignorare le sequenze non codificanti alle quali dobbiamo attribuire importanti funzioni regolatorie.

Vietate copia riproduzione e modifica

La porzione non codificante dei genomi eucariotici



L'annotazione funzionale delle porzioni non-codificanti del genoma è una delle sfide principali dell'era post-genomica.

Vietate copia riproduzione e modifica

Ipotesi sull'origine e la funzione del DNA non codificante

Queste regioni potrebbero essere rimasugli di pseudogeni, che nel corso dell'evoluzione avrebbero perso la loro funzione, anche a causa di eventuali frammentazioni della sequenza codificante.
Il *DNA non codificante* potrebbe avere una funzione protettiva nei confronti delle regioni codificanti. Dal momento che il DNA è continuamente esposto a danni casuali da parte di agenti esterni, infatti, una tanto alta percentuale di DNA non codificante permette di pensare che le regioni ad essere statisticamente più danneggiate siano in realtà non codificanti.
Il DNA non codificante potrebbe anche essere una sorta di *riserva di sequenze* al momento non codificate, ma dalle quali potrebbe emergere un qualche gene in grado di conferire vantaggio all'organismo. Da questo punto di vista, dunque, tali regioni costituirebbero le vere basi genetiche dell'evoluzione.
Parte del *DNA non codificante* è ritenuto essere, più semplicemente, un elemento spaziatore tra geni. In questo modo gli enzimi che hanno rapporti con il materiale genetico avrebbero la possibilità di complessare più agevolmente il DNA. Il DNA non codificante così potrebbe avere una funzione fondamentale pur essendo composto di una sequenza assolutamente casuale.
Alcune regioni di DNA non codificante potrebbero avere una funzione regolatoria sconosciuta: potrebbero ad esempio controllare l'espressione di alcuni geni o lo sviluppo di un organismo dallo stato embrionale fino a quello adulto.
Nel *DNA non codificante* potrebbero essere contenute numerose sequenze trascritte in non coding RNA (si ritiene possano essere molti di più di quelli attualmente noti).

Alcune teorie puntano invece a confermare che tale DNA non abbia in effetti alcuna funzione. In un recente esperimento è stata rimossa una quantità di *DNA non codificante* dal genoma murino pari all'1%. I topi sottoposti al trattamento non hanno mostrato alcun fenotipo. Ciò può comunque essere interpretato in due modi: il DNA non codificante non ha effettivamente nessuna funzione, oppure i ricercatori non sono stati in grado di sviluppare un metodo di rilevazione tale da osservare cambiamenti fenotipici nei topi.

Vietate copia riproduzione e modifica

The ENCODE Project Consortium

La "Encyclopedia of DNA Elements (ENCODE)" è un progetto di ricerca pubblico promosso da US National Human Genome Research Institute (NHGRI) nel settembre 2003.

Il progetto ENCODE punta ad identificare tutti gli elementi funzionali del genoma umano, al di là dei geni codificanti per proteine.

Il progetto coinvolge un consorzio mondiale di gruppi di ricerca.

I dati ottenuti sono accessibili attraverso database pubblici.

Vietate copia riproduzione e modifica

Obiettivo del progetto è l'identificazione del cosiddetto

Reguloma, cioè di quella varietà di elementi del DNA (promotori, enhancer, silencer, regioni della cromatina suscettibili di intense modificazioni epigenetiche, geni per trascritti regolatori…) che possono regolare l'espressione dei geni codificanti proteine.

La disfunzione del reguloma può essere alla base di molte patologie alle quali ancora non è stata ancora attribuita una base genetica e molecolare.

Vietate copia riproduzione e modifica