

Anatomia del genoma

Organizzazione dell'informazione genetica

- Procarioti
- Eucarioti



Anatomia del genoma: Progetto Genoma Umano

- La conoscenza della sequenza dei geni ha portato all'identificazione e migliore comprensione della attività degli stessi. Dal 1970 i biologi molecolari sono riusciti ad ottenere le sequenze sempre più complete di DNA di diversi organismi.
- Il primo genoma ad essere sequenziato fu quello dell'eubatterio *Haemophilus influenzae* nel 1995, dopo E. Coli nel 1997, etc..
- Il progetto Genoma Umano fu proposto in 1984 e approvato nel 1989 dal Congresso Americano, iniziando nel 1990. Fu condotto da un consorzio internazionale costituito da ricercatori americano, inglesi, francesi, Tedeschi e cinesi.
- Uno studio parallelo fu sviluppato privatamente dalla società Celera Genomics of Maryland, USA - in 1998.
- I risultati di questi studi furono pubblicati in 2 giornali scientifici: **Nature** e **Science** nel **2001**. Lo studio non era ancora completo e includeva il 83-84% del genoma pero si pensava che rappresentava le regioni più importanti del genoma. Il 16-17% rimanente apparteneva ai telomeri e centromeri.



Anatomia del genoma:

- Fino a oggi si conoscono le sequenze di circa **2360 virus, 2950 batteri e 670 organismi eucarioti**
- Lo studio delle sequenze ha confermato la divisione degli organismi in :

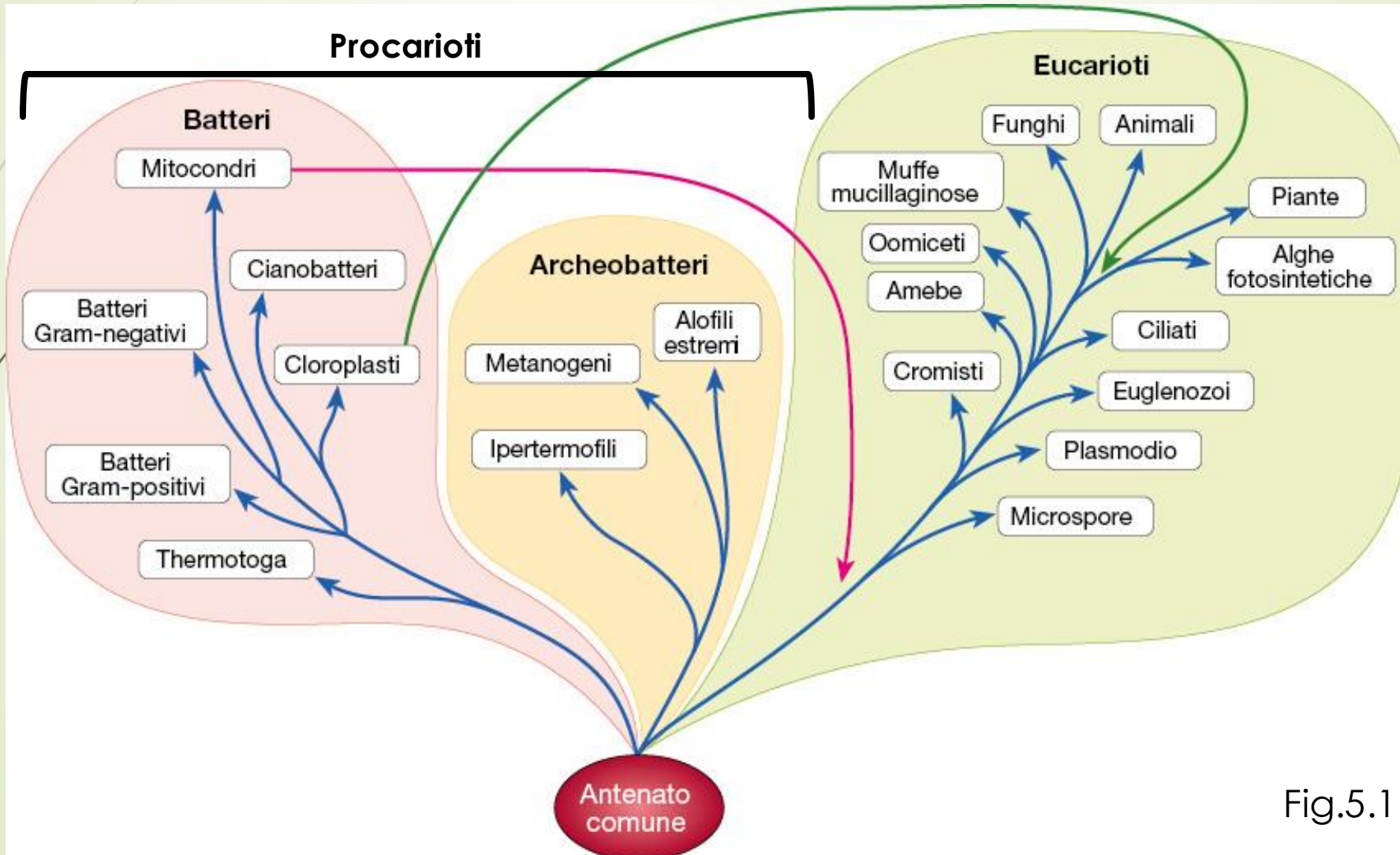


Fig.5.1



Genoma Procariotico

Genoma Procariotico

- Un **singolo cromosoma (circolare o lineare)**. Possono contenere anche dei piccoli DNA circolari extracromosomici (**plasmidi**).
- I plasmidi generalmente codificano per **proprietà vantaggiose** come può essere la resistenza ad antibiotici, anche se però sono dispensabili
- In generale i genomi procariotici sono **più piccoli** degli eucariotici ma molto variabili in dimensione. Nei batteri: 144kpb-13000kpb; più stabili negli Archeo.
- Il genoma procariotico è «**plastico**». All'interno di una stessa specie i genomi di diversi individui possono mostrare grande variabilità in dimensione. La parte comune di questo genoma all'interno della specie si chiama **ENDOGENOMA**
- La lunghezza del genoma dipende di quanto sia specializzato il batterio:
 - più specializzato → meno geni
 - Meno specializzato → più geni
- La parte variabile del genoma può essere conseguenza di un **trasferimento genico laterale**, da un individuo a un altro

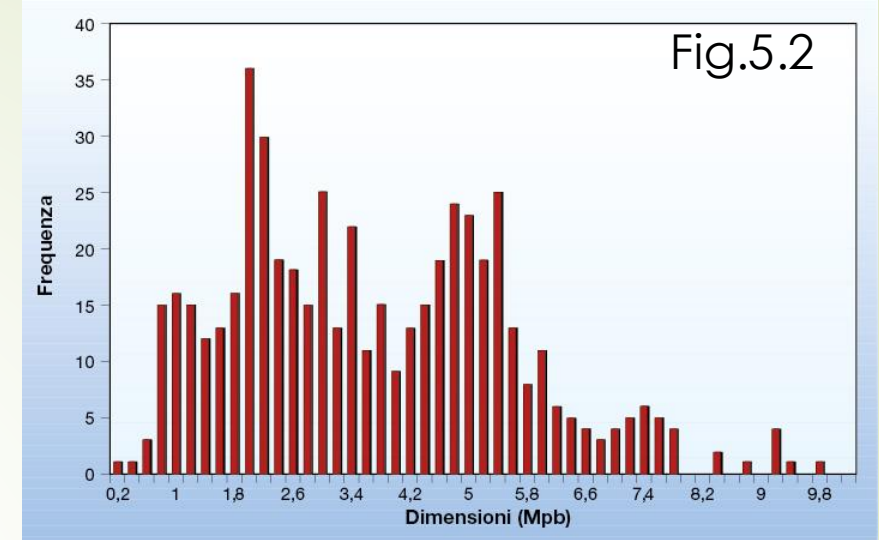


Tabella 5.1 Dimensioni e numero di geni di alcuni genomi procariotici. È anche riportato e l'identificativo (Accession) della banca dati NCBI/EMBL.

Specie	Cromosomi	Dimensione genoma (nt)
Candidatus Hodgkinia cicadicola	1	143 795
Mycoplasma genitalium	1	580 076
Rickettsia prowazekii	1	1 111 523
Borrelia burgdorferi	22	1 519 856
Helicobacter pylori	1	1 576 758
Haemophilus influenzae	1	1 813 033
Streptococcus pneumoniae	1	2 111 882
Neisseria meningitidis	1	2 153 416
Staphylococcus aureus	1	2 820 462
Escherichia coli str. K12 DH10B	1	4 686 137
Clostridium botulinum	1	3 863 450
Sorangium cellulosum	1	13 033 779

Trasferimento genico laterale (orizzontale):

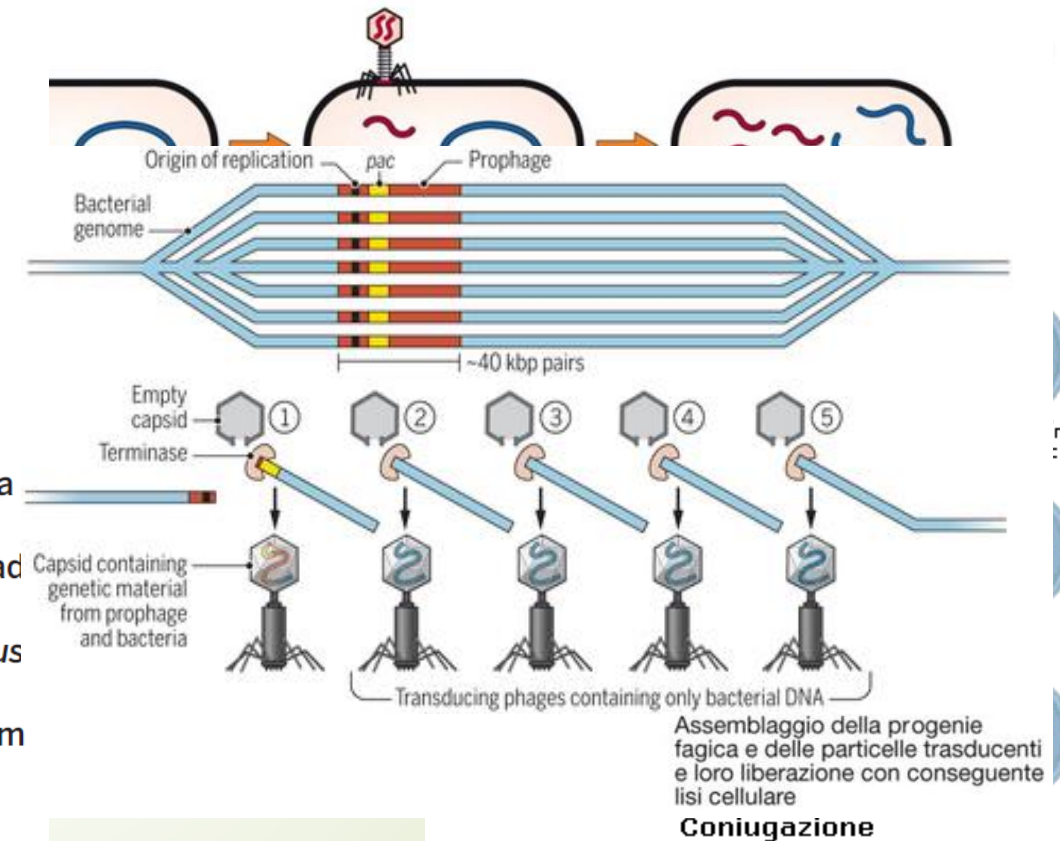
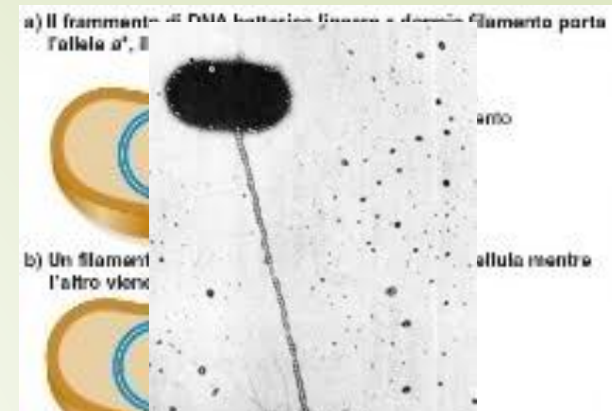
- **Trasformazione:** la cellula «competente» prende dal esterno un frammento del DNA rilasciato dalla

Genome hypermobility by lateral transduction

John Chen^{1*†}, Nuria Quiles-Puchalt^{2*}, Yin Ning Chiang^{1*}, Rodrigo Bacigalupe³, Alfred Fillol-Salom², Melissa Su Juan Chee¹, J. Ross Fitzgerald³, José R. Penadés^{2,4,5†}

Genetic transduction is a major evolutionary force that underlies bacterial adaptation. Here we report that the temperate bacteriophages of *Staphylococcus aureus* engage in a distinct form of transduction we term lateral transduction. Staphylococcal prophages do not follow the previously described excision-replication-packaging pathway but instead excise late in their lytic program. Here, DNA packaging initiates in situ from integrated prophages, and large metameric spans including several hundred kilobases of the *S. aureus* genome are packaged in phage heads at very high frequency. In situ replication before DNA packaging creates multiple prophage genomes so that lateral-transducing particles form during normal phage maturation, transforming parts of the *S. aureus* chromosome into hypermobile regions of gene transfer.

La sequenza genica trasferita si potrà inserire nel cromosoma della cellula mediante ricombinazione genica



Coniugazione



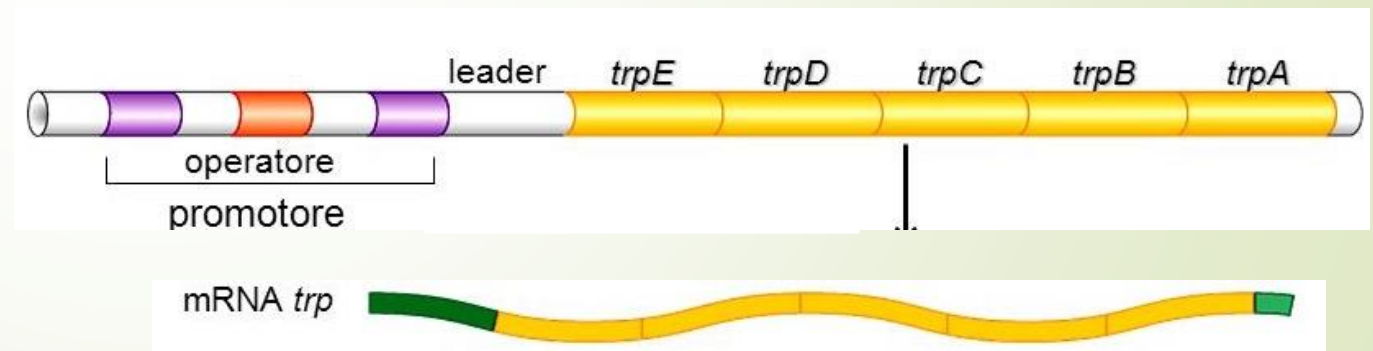
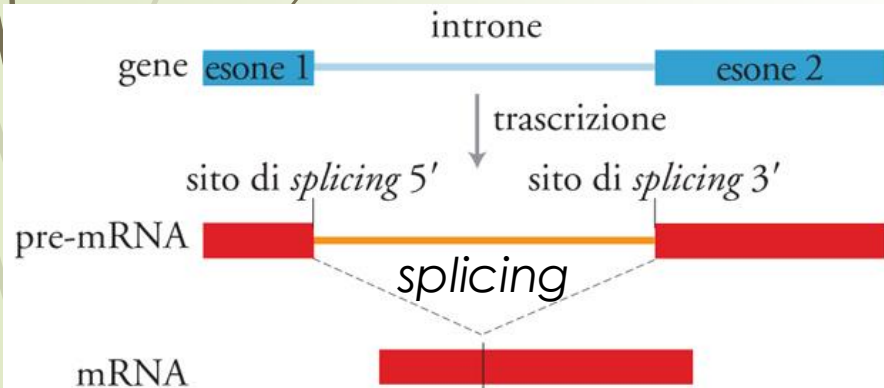
Genoma Procariotico

- I geni **non contengono introni** e sono organizzati in **Operoni**, la cui espressione da luogo a unità trascrizionali **policistroniche**.

Introne: regioni polinucleotidiche non codificanti all'interno del gene che vengono trascritte nel RNA ma eliminate durante la sua maturazione in mRNA

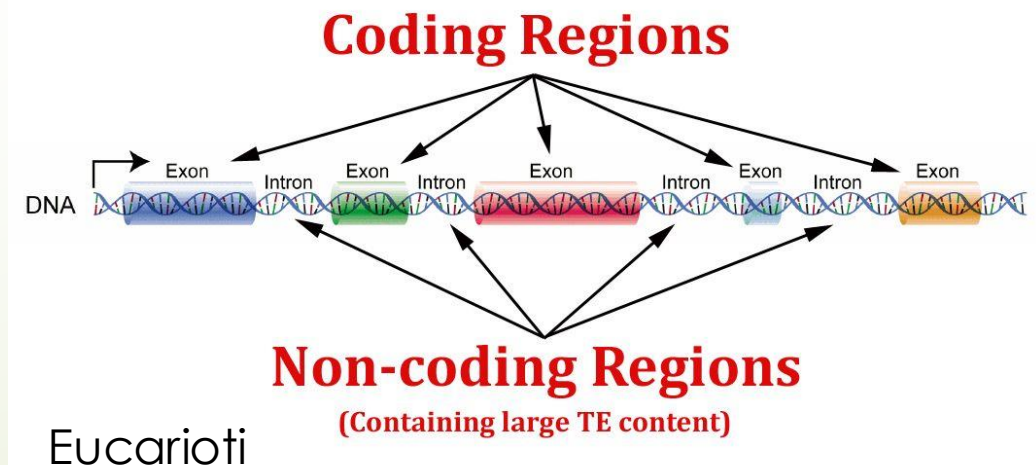
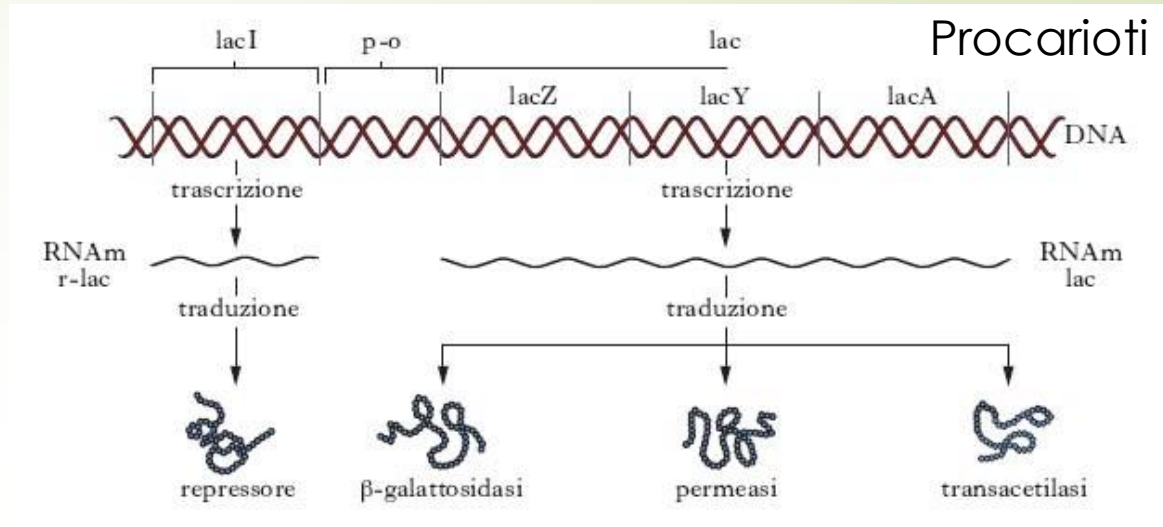
Operone: gruppo di geni adiacenti, che sono trascritti sotto il controllo dello stesso promotore. Danno luogo a una singola molecola di mRNA e sono pertanto soggetti a una regolazione coordinata.

Policistronico: diversi geni sotto il controllo di un solo promotore sono trascritti in una singola molecola di mRNA (policistronica!) capace di tradurre per più catene polipeptidiche, riducendo la presenza di sequenze non codificanti regolatrici.

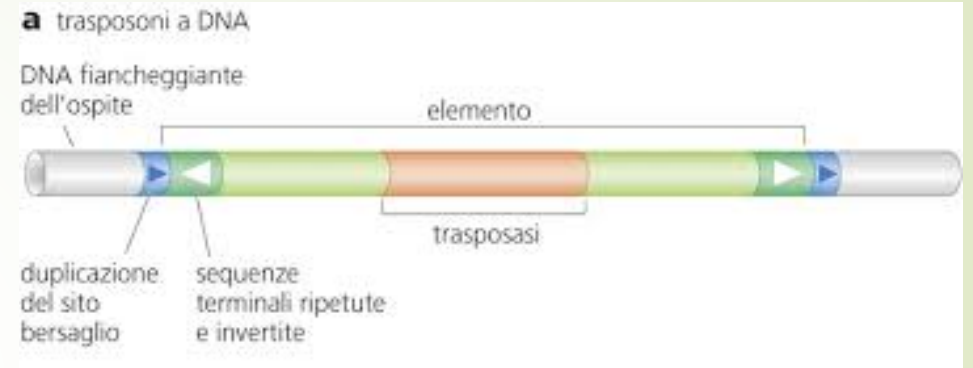


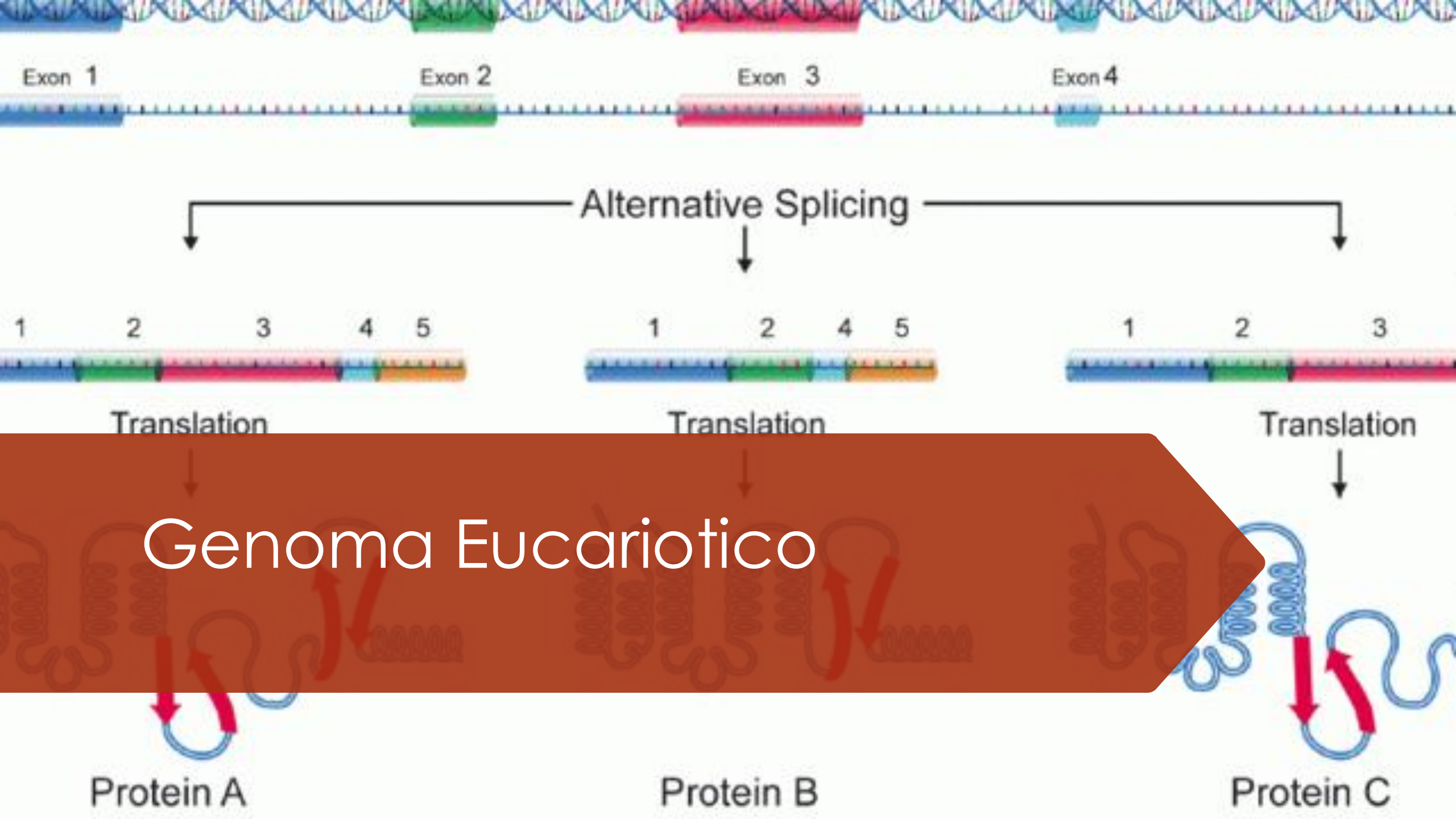
Genoma Procariotico

- I geni **non contengono introni** e sono organizzati in unità trascrizionali **policistroniche**, che deriva dalla organizzazione in **Operoni** dei geni.
- L'organizzazione in operoni e l'assenza di introni implica che la **lunghezza del genoma sia correlata alla loro complessità metabolica**.
- I procarioti hanno **geni più corti** che gli eucarioti, incluso dopo l'eliminazione degli introni negli eucarioti.
- Contengono **poche sequenze non codificanti** che occupano un 15% del genoma, principalmente **rRNA, tRNA** e altri **small non coding RNA (sncRNA)** che regolano la traduzione e stabilità del mRNA. Per questo il **genoma dei procarioti è più denso**.



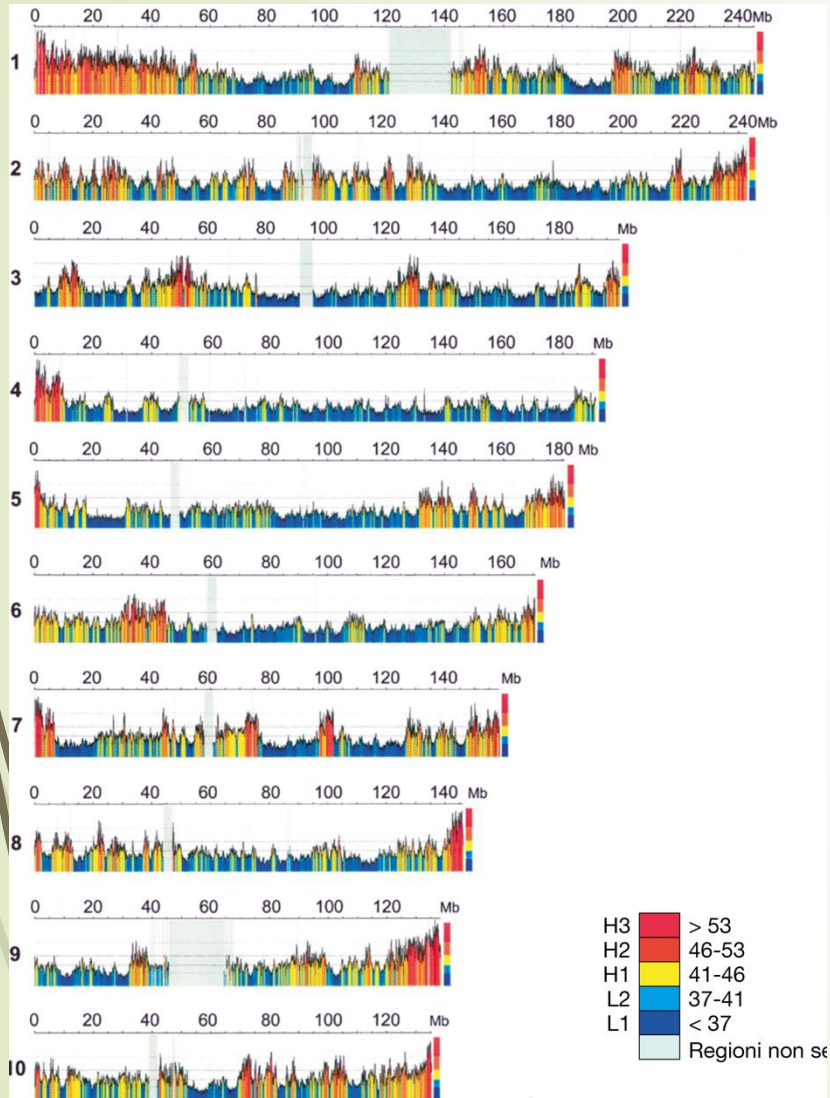
- **Altre** **sequenze** importanti negli **elementi** promuovono un altro punto di vista tra genomi (più avanti)
- In genere **presentano** **ripetute** in trovate neg





Genoma Eucariotico

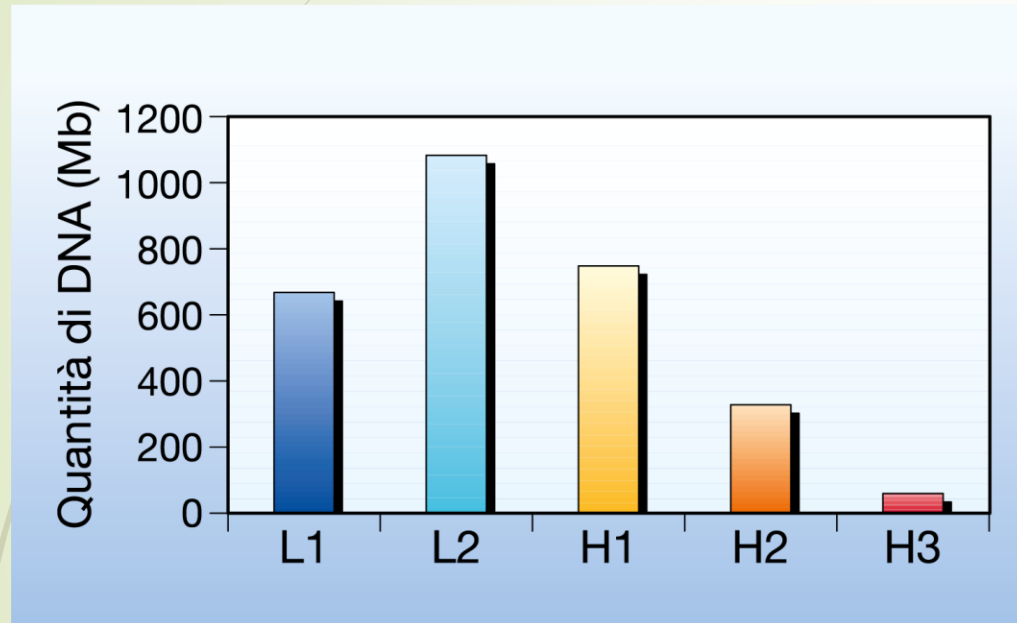
Genoma Eucariotico



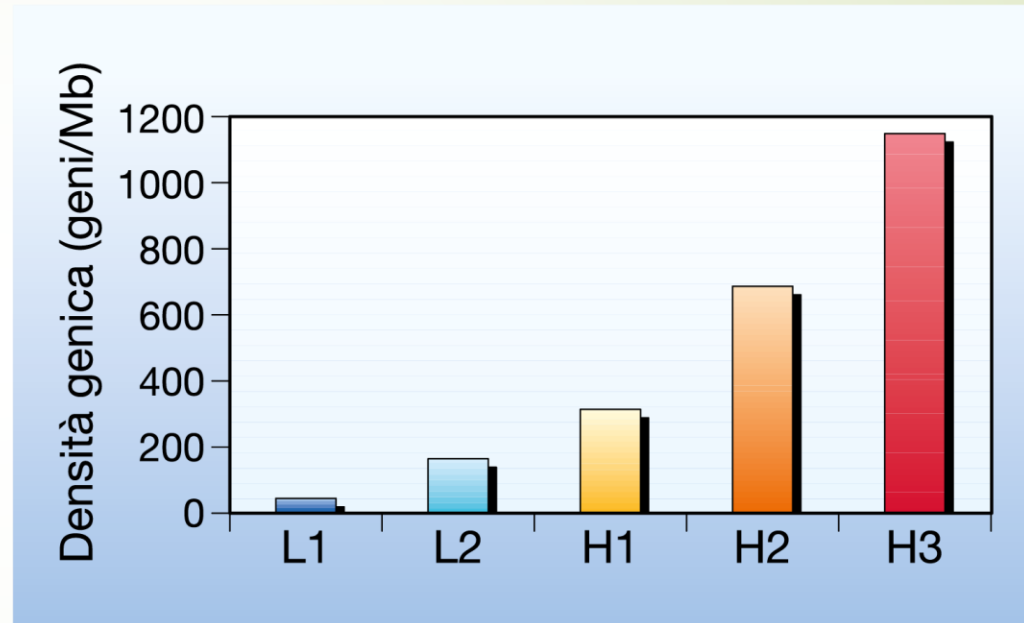
- I genomi eucariotici mostrano una **marcata simmetria composizionale a livello delle basi**, con un contenuto G+C meno variabile rispetto al genoma procariotico.
- In certi vertebrati come i mammiferi esistono **segmenti di DNA** distribuiti **a mosaico** caratterizzati da una specifica e **omogenea composizione in basi: ISOCORE**
- Questi ISOCORE sono denominati **L1, L2, H1, H2, H3**. Si **differenziano nel contenuto di G+C**.
- **L1 e L2** sono poveri in **G+C (33%)**, e costituiscono il 60%.
- **H1-H3** sono ricche in **G+C (59%)** e **contengono la maggior parte dei geni**. Sono caratterizzate da un **minore livello di metilazione**, la cromatina è in una conformazione più rilassata e contiene dei **geni costitutivamente espressi**.

Isocore: abbondanza e densità genica

A

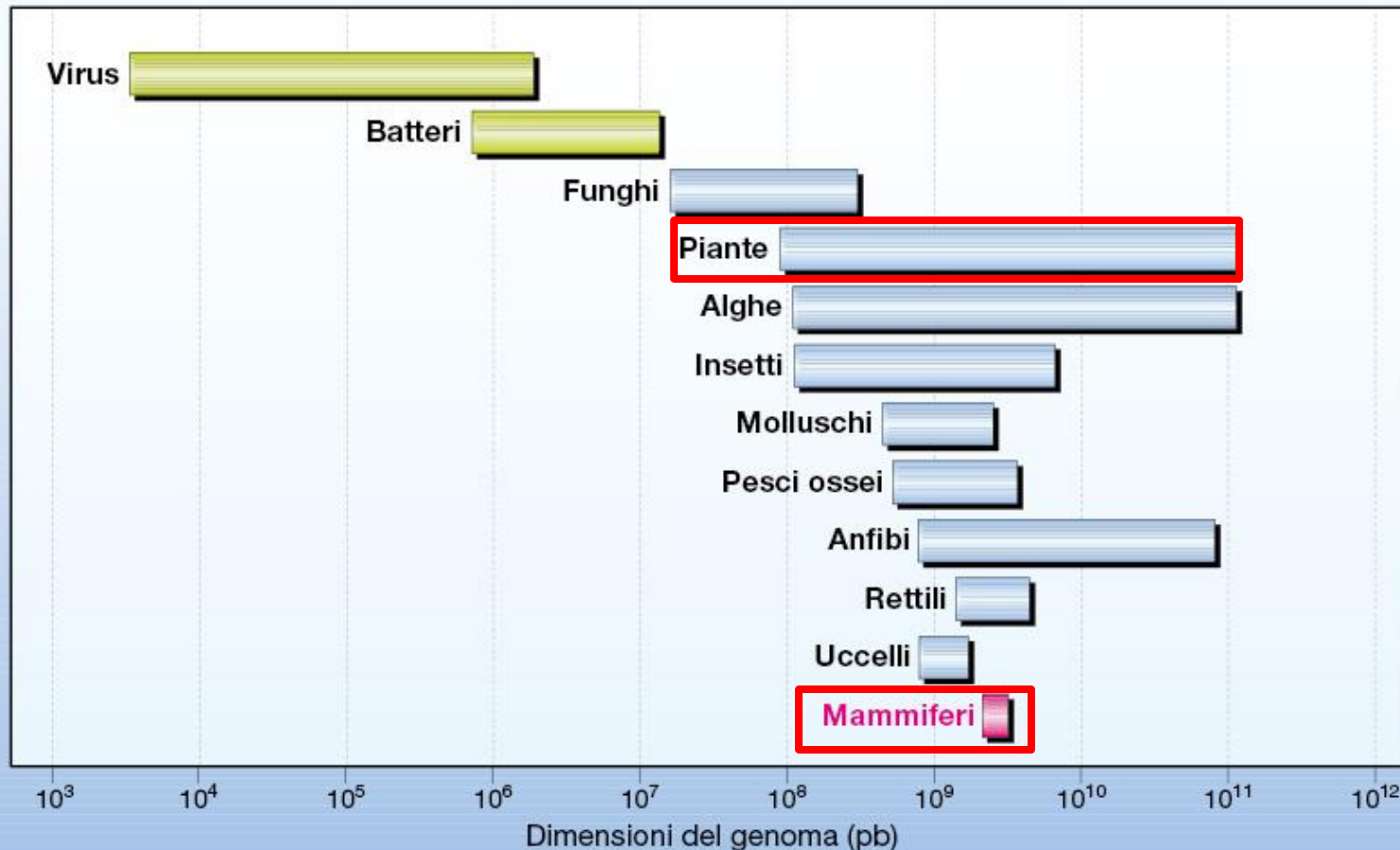


B



Genoma Eucariotico

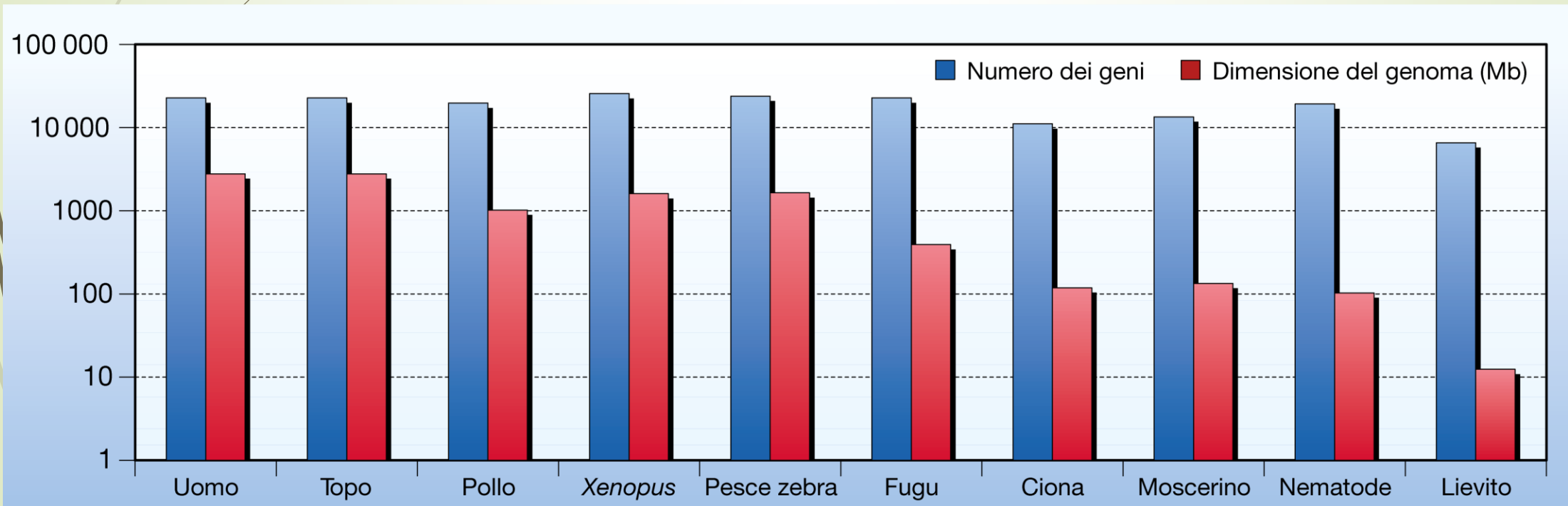
La dimensione della molecola di DNA non correla con la complessità del organismo: organismi apparentemente poco complessi presentano un genoma più lungo di quello di cellule umane (n° geni/lunghezza genoma).



Species	Size of genome (Mb)	Approximate number of genes
Eukaryotes		
<i>Arabidopsis thaliana</i> (plant)	125	25 500
<i>Caenorhabditis elegans</i> (nematode worm)	97	19 000
<i>Drosophila melanogaster</i> (fruit fly)	180	13 600
<i>Homo sapiens</i> (human)	3200	30 000
<i>Saccharomyces cerevisiae</i> (yeast)	12.1	5800
Bacteria		
<i>Escherichia coli</i> K12	4.64	4400
<i>Mycobacterium tuberculosis</i> H37Rv	4.41	4000
<i>Mycoplasma genitalium</i>	0.58	500
<i>Pseudomonas aeruginosa</i> PA01	6.26	5700
<i>Streptococcus pneumoniae</i>	2.16	2300
<i>Vibrio cholerae</i> El Tor N16961	4.03	4000
<i>Yersinia pestis</i> CO92	4.65	4100
Archaea		
<i>Archaeoglobus fulgidus</i>	2.18	2500
<i>Methanococcus jannaschii</i>	1.66	1750

Genoma Eucariotico

- La dimensione della molecola di DNA non correla con la complessità del organismo.
- La mancanza di correlazione tra la complessità genetica/morfologica di un organismo e le dimensioni del suo genoma è definita Paradosso del valore C.



Anatomia del Genoma ed Implicazioni evolutive

N° proteine umane: 80000-100000



Quanti geni?

1 mRNA → 1 proteina

80000-100000?

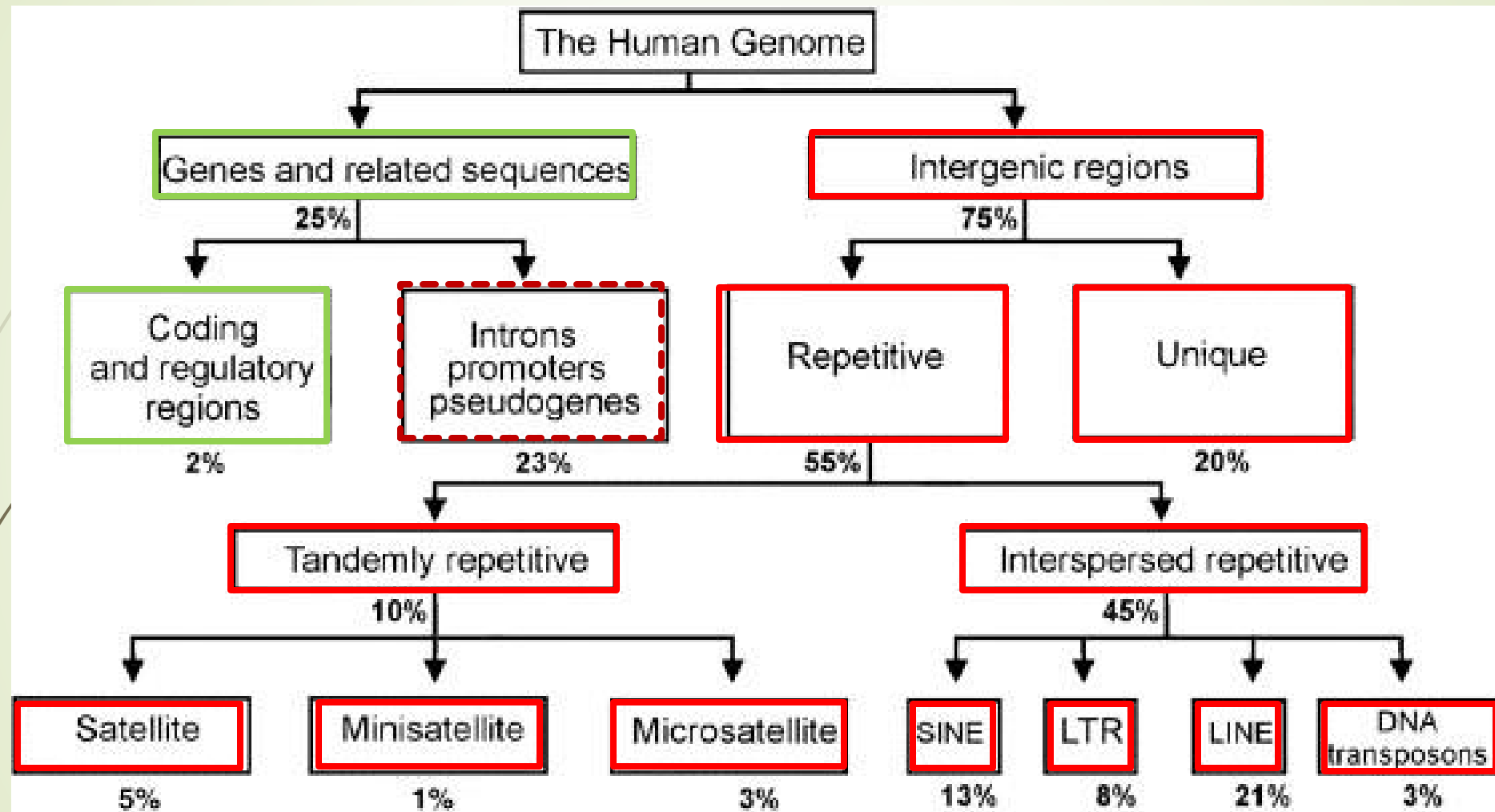


Progetto Genoma umano
ha rilevato l'esistenza di solo
23000-25000 geni che
codificano per proteine

Tabella 5.2 Dimensioni e relativo numero stimato di geni dei genomi di alcuni organismi modello.

Genoma	Dimensioni		Numero di geni codificanti proteine
	Mpb	Lunghezza (cm)	
ΦX174 (piccolo fago)	0,005	0,00016	11
T7 (grande fago)	0,08	0,00259	150
Mimivirus (virus gigante)	1,2	0,03886	1000
<i>Escherichia coli</i> (procariote)	4,6	0,14895	4300
<i>Saccharomyces cerevisiae</i> (lievito)	13,5	0,43714	6000
<i>Caenorhabditis elegans</i> (nematode)	100	3,23810	21700
<i>Arabidopsis thaliana</i>	120	3,88571	26000
<i>Arabidopsis thaliana</i> (cloroplasto)	0,155	0,00502	87
<i>Drosophila melanogaster</i>	165	5,34286	12000
<i>Homo sapiens</i>	3300	106,857	23000
<i>Homo sapiens</i> (mitocondrio)	0,016	0,00052	37

Genoma umano



Genoma Eucariotico

La dimensione della molecola di DNA è molto variabile incluso all'interno del gruppo tassonomico

- Questa differenza si può spiegare per la presenza di **regioni non codificanti**, che rappresentano l'90-98% del genoma:

- 1. Introni e pseudogeni
- 2. Regioni intergeniche

1. DNA ripetitivo (55%):

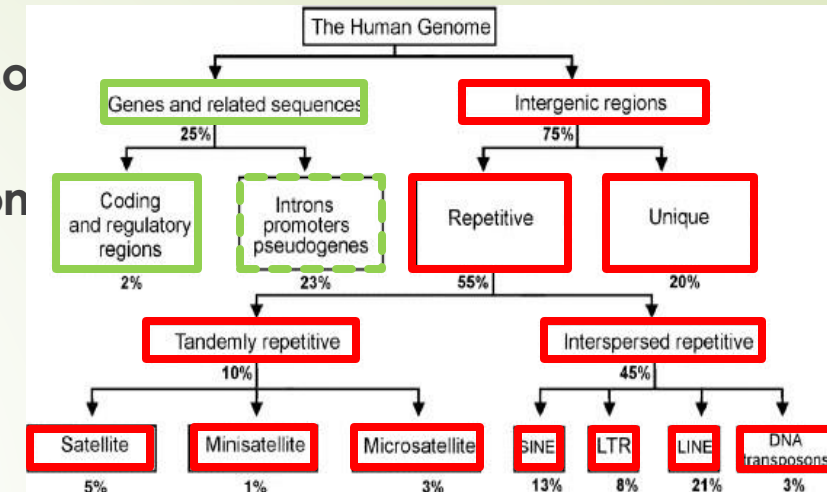
i. Genome wide or ripetizioni intersparse: elementi principalmente «parassiti» che risiedono nel DNA eucariotico però che non contribuiscono con le loro sequenze alla funzione del genoma. Questi elementi derivano da elementi trasponibili:

- Trasposoni non necessitano di un intermediario RNA
- LTR retrotrasposoni, LINEs e SINEs: si traspongono via un intermediario di RNA

ii. Ripetizioni in tandem: DNA satellite, minisatellite e microsatellite

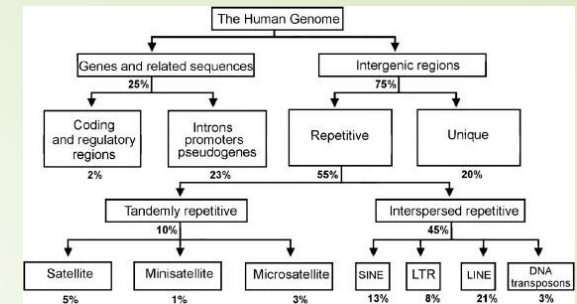
2. Sequenze uniche (20%)

Le **regioni codificanti o esoni** comprendono intorno al **2% del genoma umano**

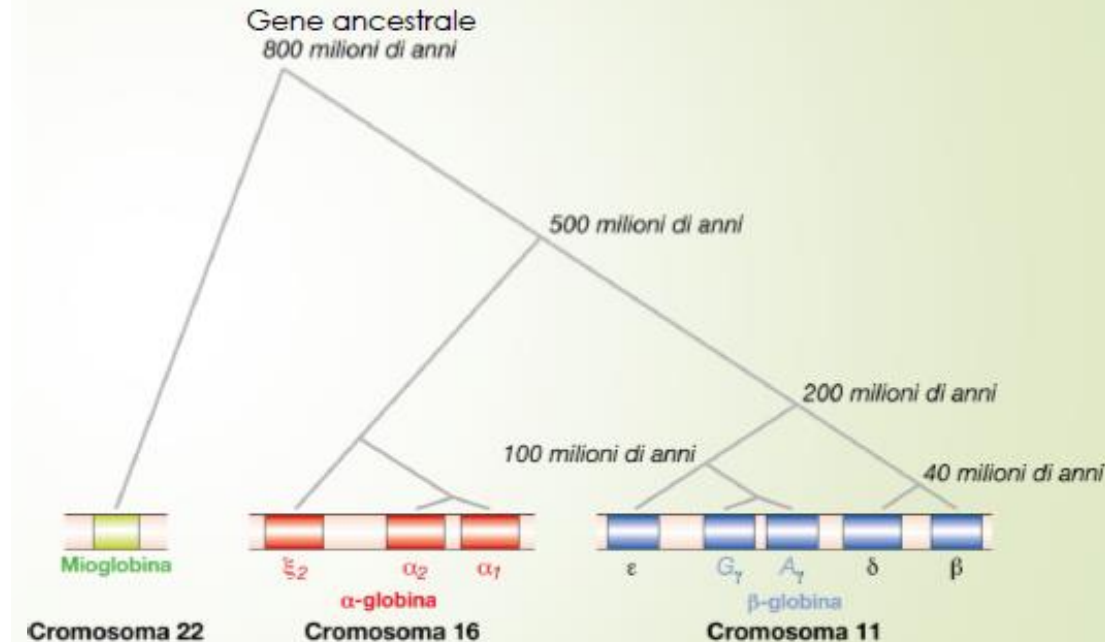


Geni e sequenze correlate

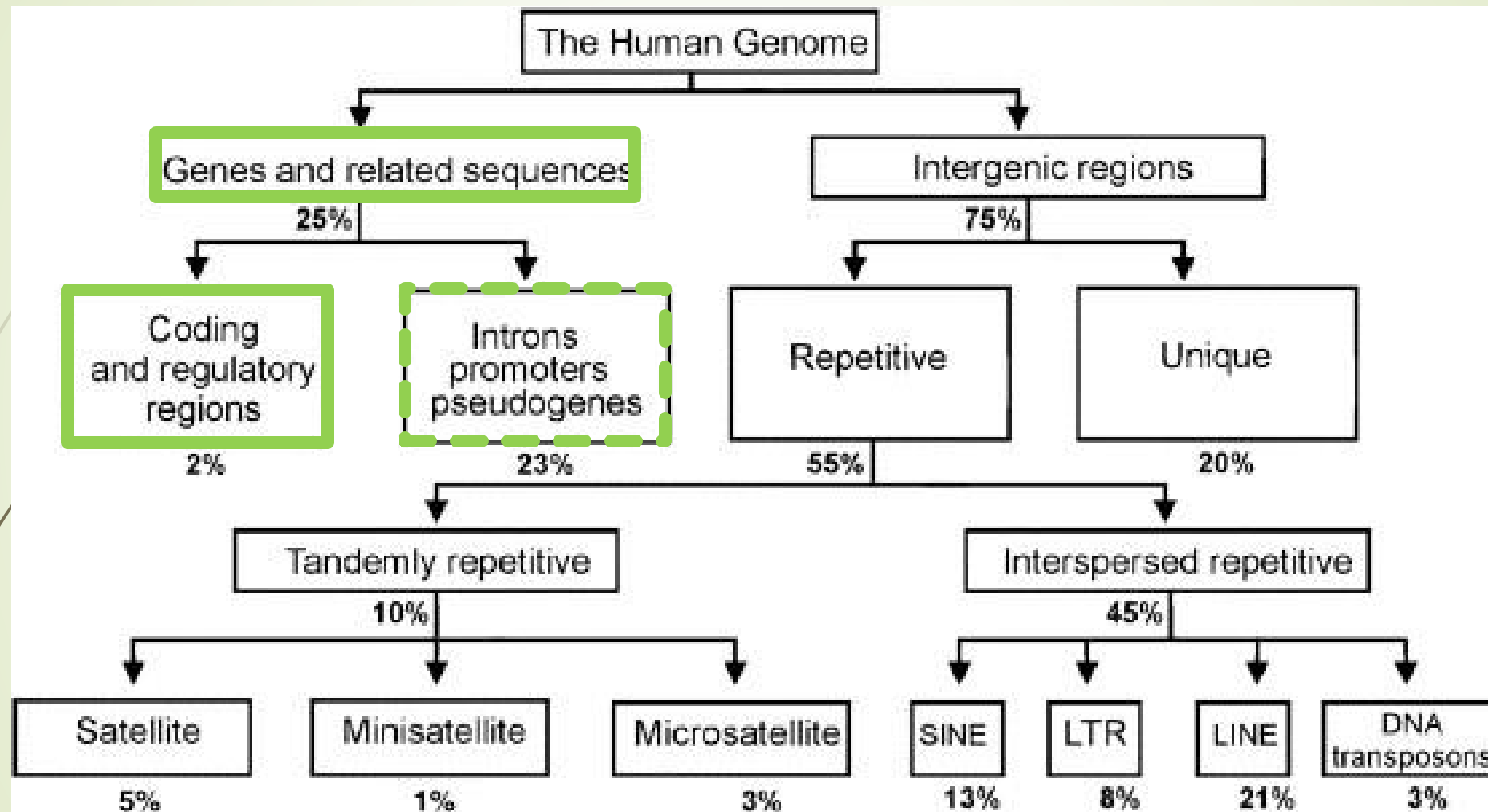
- I **geni** sono la **parte del genoma non ripetuto che codifica per proteine**.
- I geni si trovano in singola copia o in più copie correlate. Questi geni correlati si chiamano **Famiglie geniche**.
- Le famiglie geniche codificano per proteine o molecole di RNA con funzioni simili.
- Una **famiglia genica** proviene dalla **duplicazione di uno stesso gene ancestrale**.
- Alcune famiglie possono essere molto numerose (i geni per recettori dell'olfatto sono circa 1000 nei mammiferi).
- In alcuni casi la duplicazione genica serve semplicemente ad aumentare velocemente i livelli di una certa proteina: **geni reiterati (es: geni per l'istoni, rRNA)**



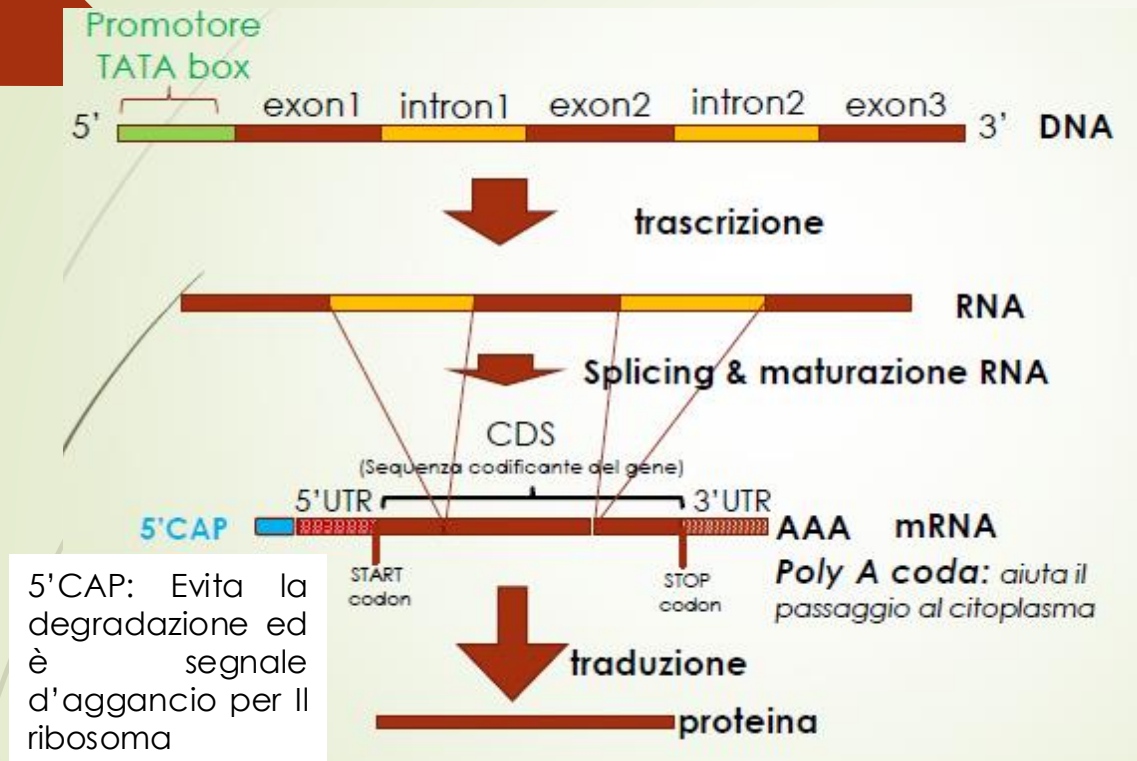
Famiglia genica globine umane



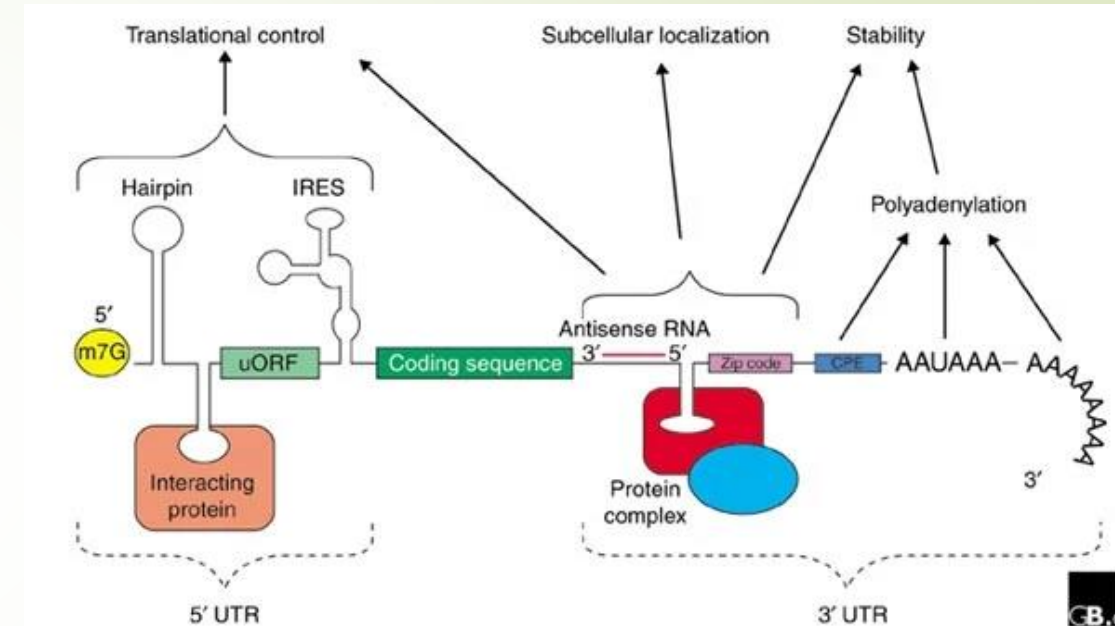
Genetic sequence correlate



Geni Eucariotici: Promotore, Esoni ed Introni



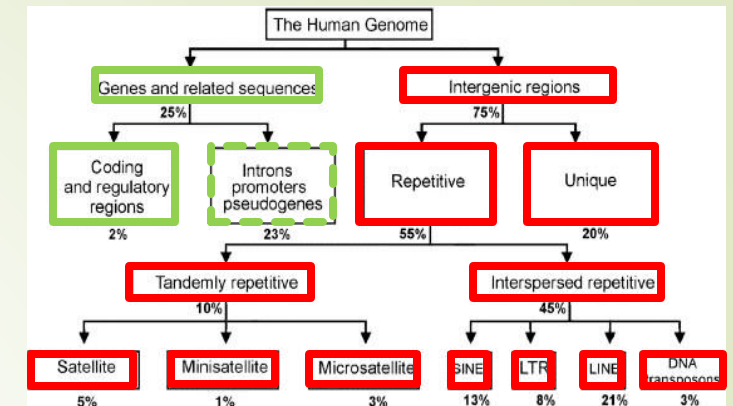
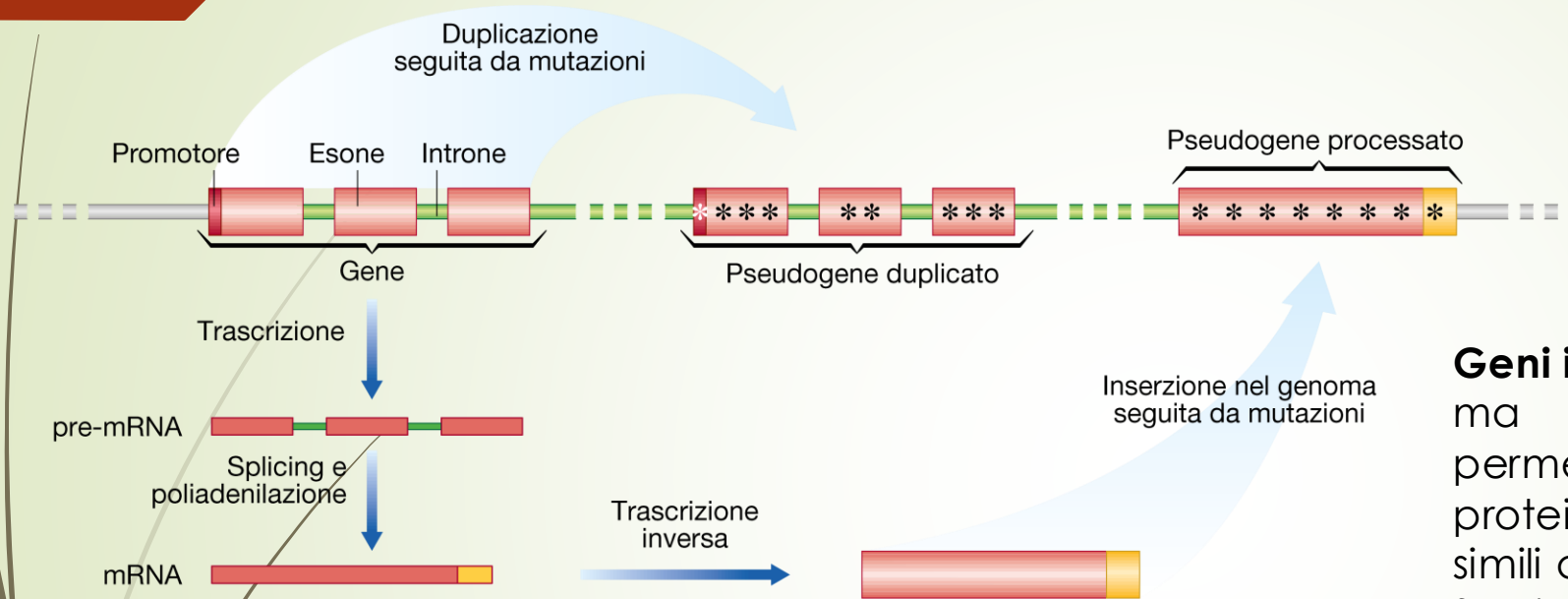
Promotore: regione genica (nel DNA) riconosciuta dai fattori di trascrizione e RNA polimerasi. Il più comune è il TATA box, che si trova da -35 a -20 basi a monte del sito di inizio della trascrizione (+1)



5'UTR: Untranslated region 5' o sequenza leader. Si trova nel mRNA a monte della CDS (coding sequence). Regola l'efficienza della traduzione.

3'UTR: Untranslated region 3' o sequenza trailer. Si trova nel mRNA a valle della CDS. Da stabilità e aiuta al trasporto.

Falsi geni o pseudogeni



Geni inattivi che assomigliano a geni funzionali ma contengono mutazioni che non permettono la espressione di una esprimono la proteina funzionale anche se derivano e sono simili ad un altro gene che esprime la proteina funzionale. Non presentano gli elementi necessari per la loro trascrizione o traduzione.

Possono originarsi in 2 modi:

- Duplicazione
- Retrotrasposizione (processati)

Diventano pseudogeni se presentano:

- uno stop-codon prematuro
- Estese delezioni o inserzioni
- Alterata fase di lettura o frameshift
- Assenza di promotore funzionale

Quando un gene non viene espresso o non è funzionale comincia ad accumulare mutazioni (***)

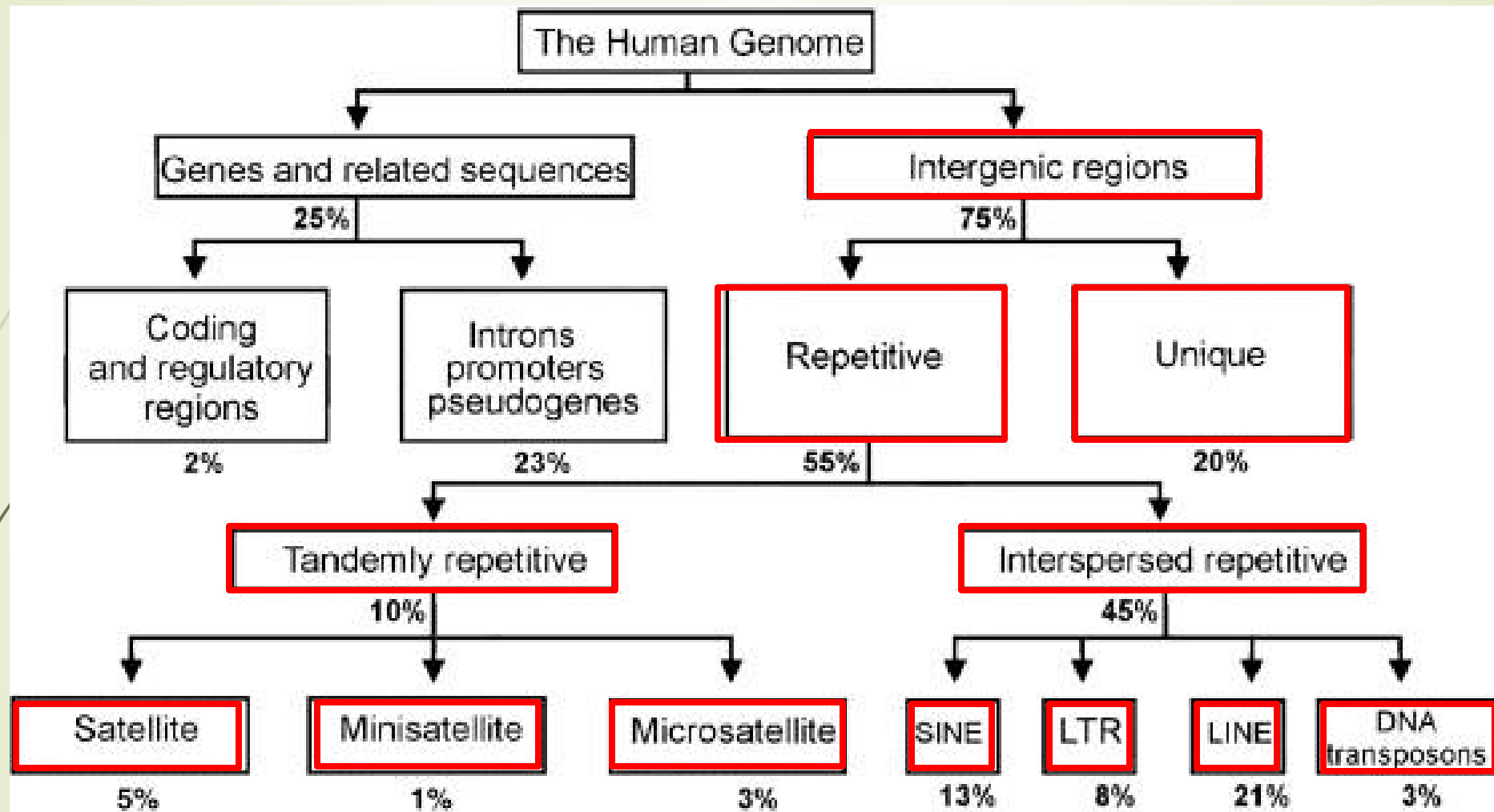


Le mutazioni che si accumulano all'interno dei geni duplicati possono portare alla formazione di :

- geni inattivi, come accade per gli **pseudogeni**,
- geni diversi da quello originario, codificanti per proteine con strutture e funzioni diverse. I geni che derivano da quest'evoluzione divergente, sono detti geni **paraloghi**, ad esempio i geni umani delle globine alfa e beta, i quali derivano da un unico gene ancestrale che ha subito duplicazione genica.

I geni paraloghi si distinguono dai **geni ortologhi**, che sono invece geni omologhi, presenti in specie diverse ma correlate, che codificano per proteine con strutture e funzioni simili. Anche i geni ortologhi originano da un unico gene ancestrale ma si sono separati non in seguito ad un evento di duplicazione ma per la separazione delle specie (speciazione) avvenuta nel corso dell'evoluzione.

Genoma umano: Regioni intergeniche



Genoma umano

Solo 1.5 - 2% del genoma contiene l'informazione per la sintesi delle proteine
Più **del 50% sono sequenze intersparse e ripetizioni in tandem**

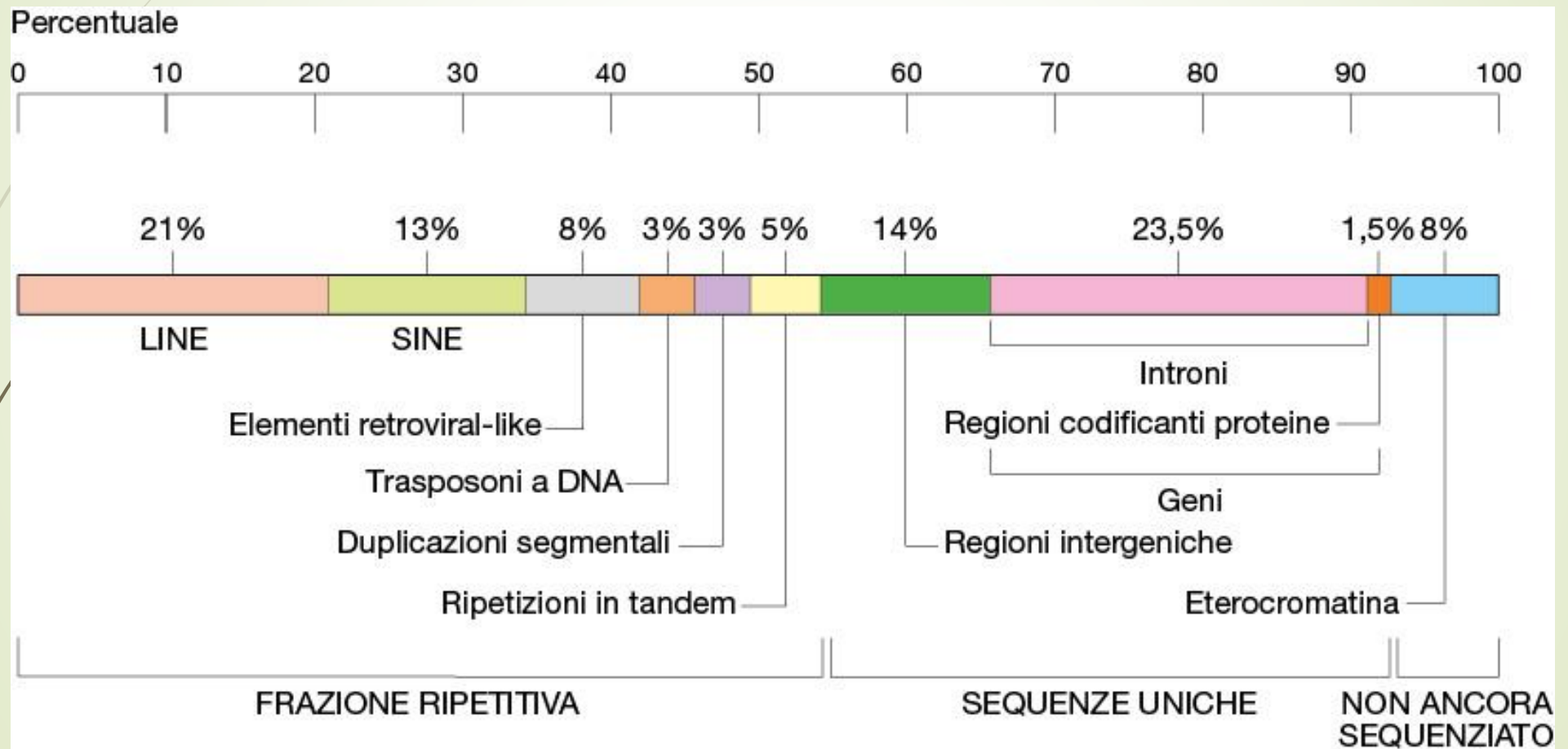
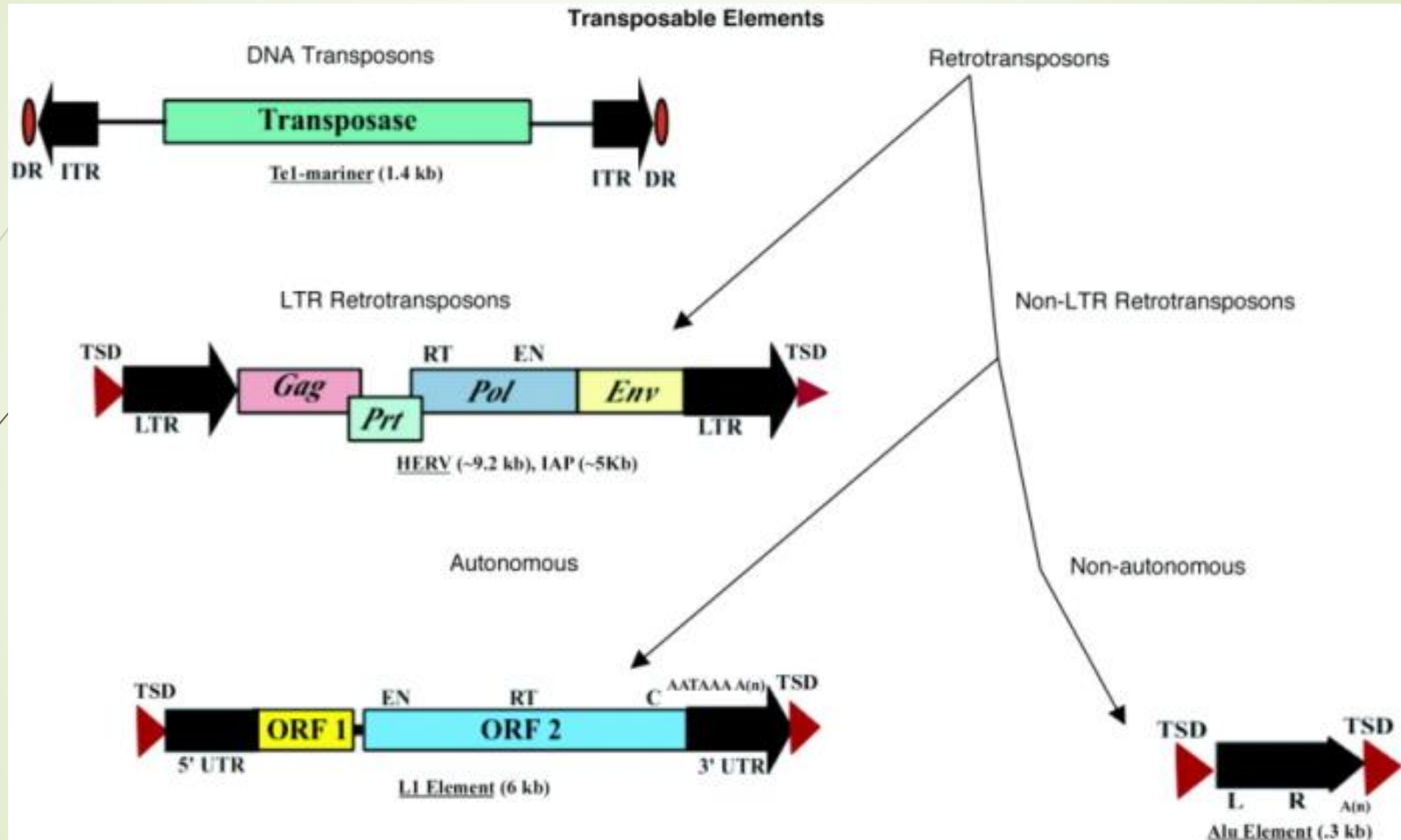


Fig.5.9

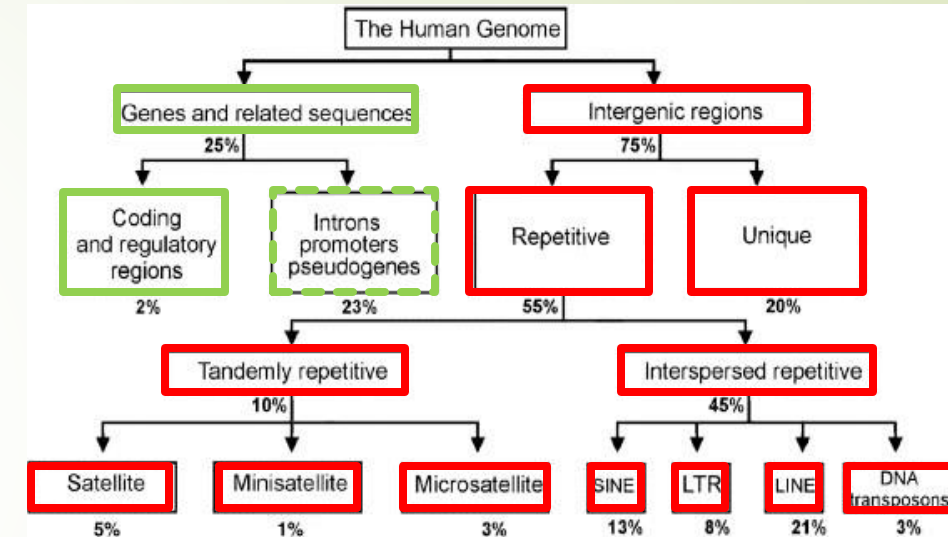
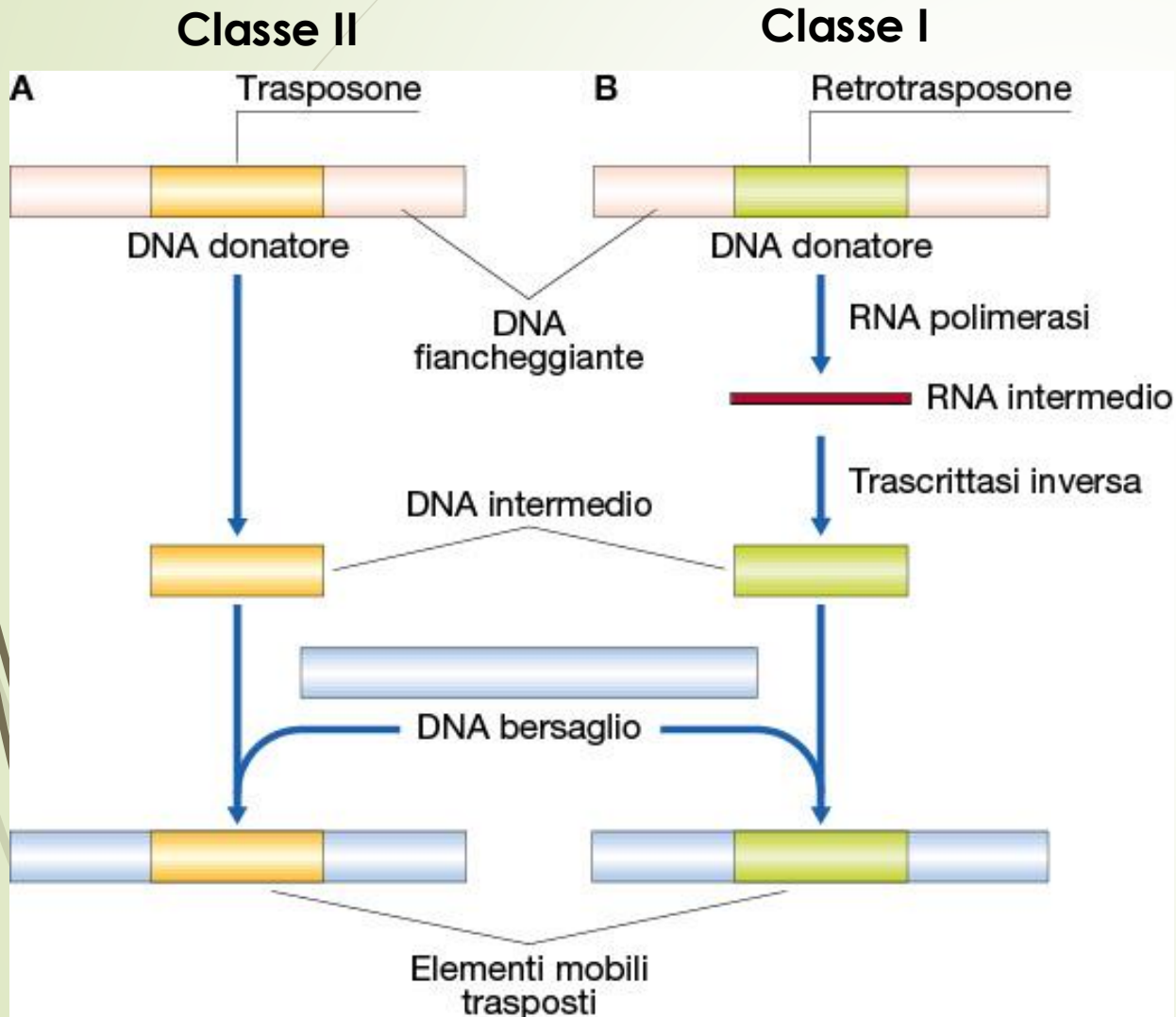
Elementi Trasponibili (Tes; Transposable Elements)



(ORF: Open Reading Frame, sequenza nucleotidica che contiene la sequenza ordinata di codoni che verrà letta dal ribosoma dando luogo a una proteina funzionale)

DNA ripetitivo intersparso o Genoma wide-

Sono elementi mobili o trasponibili (jumping genes) in grado di spostarsi nel genoma, distribuite in modo apparentemente casuale distanti fra di loro.



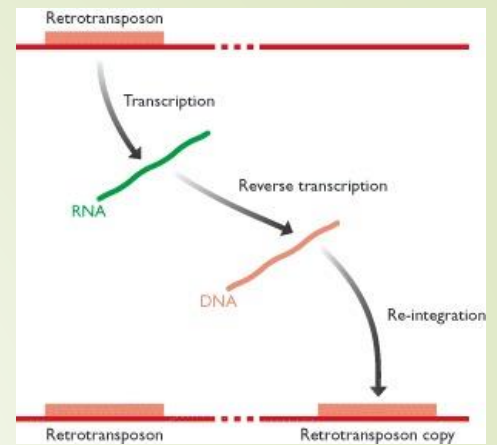
✓ Sono ubiquitari

✓ Contribuiscono alla plasticità del genoma (in modo attivo e passivo)

DNA ripetitivo intersparso o Genome wide-

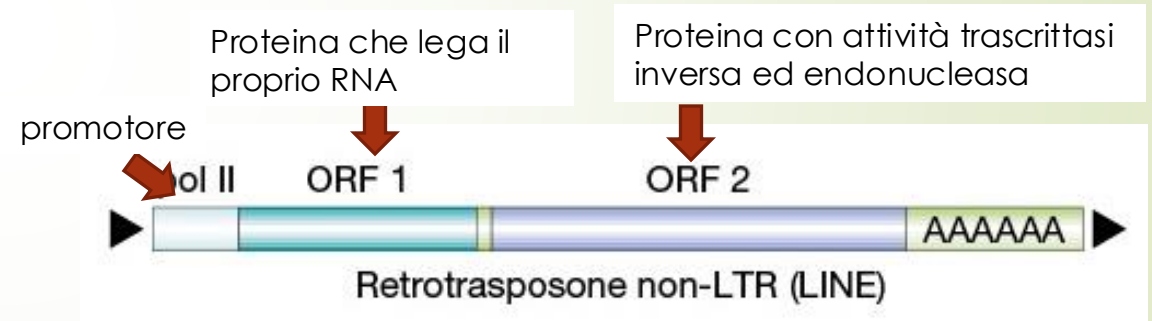
Classe I o Retrotrasposoni

► Usano un intermediario di RNA

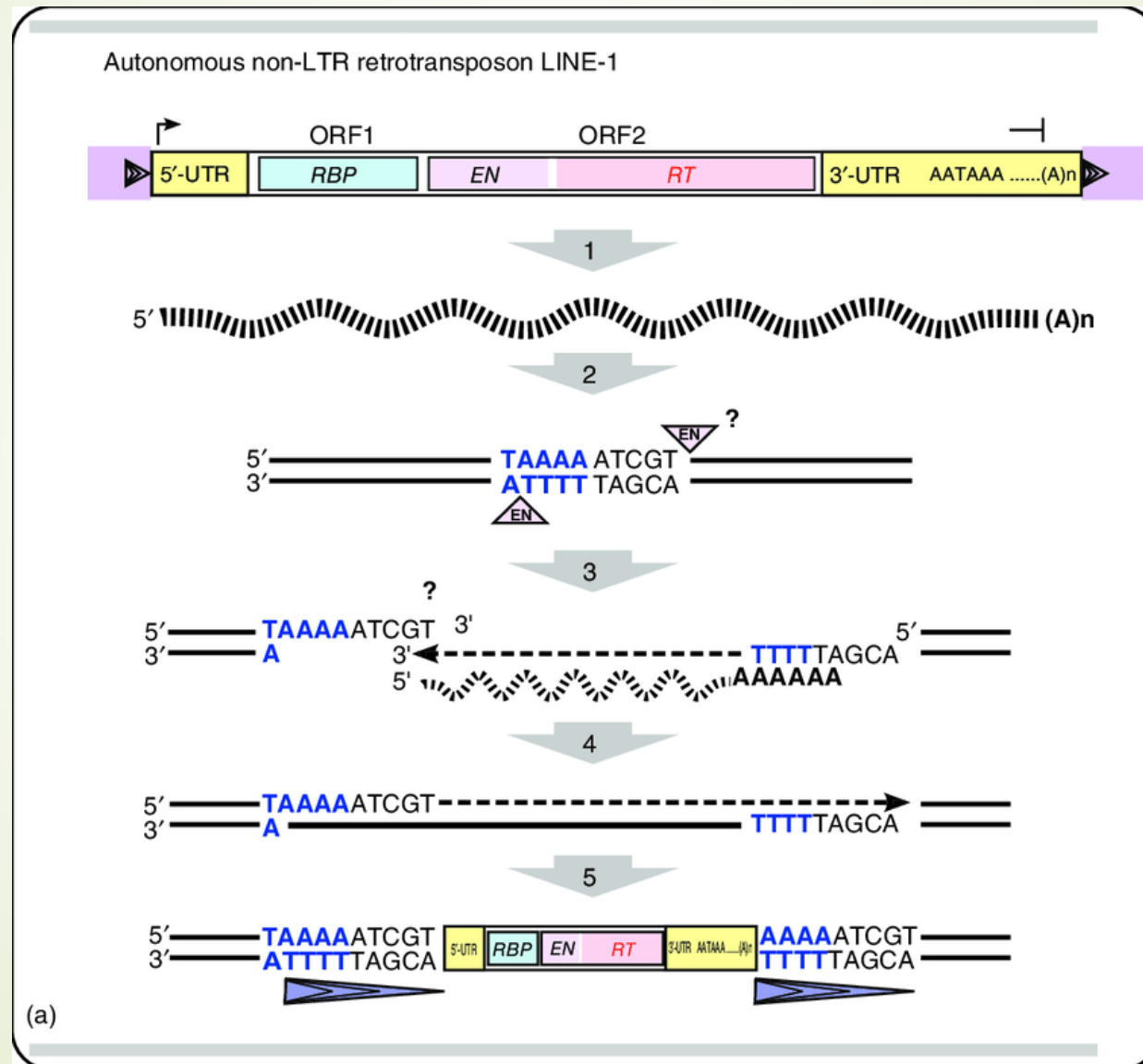


LINEs: (long interspersed nuclear elements, 5000pb).

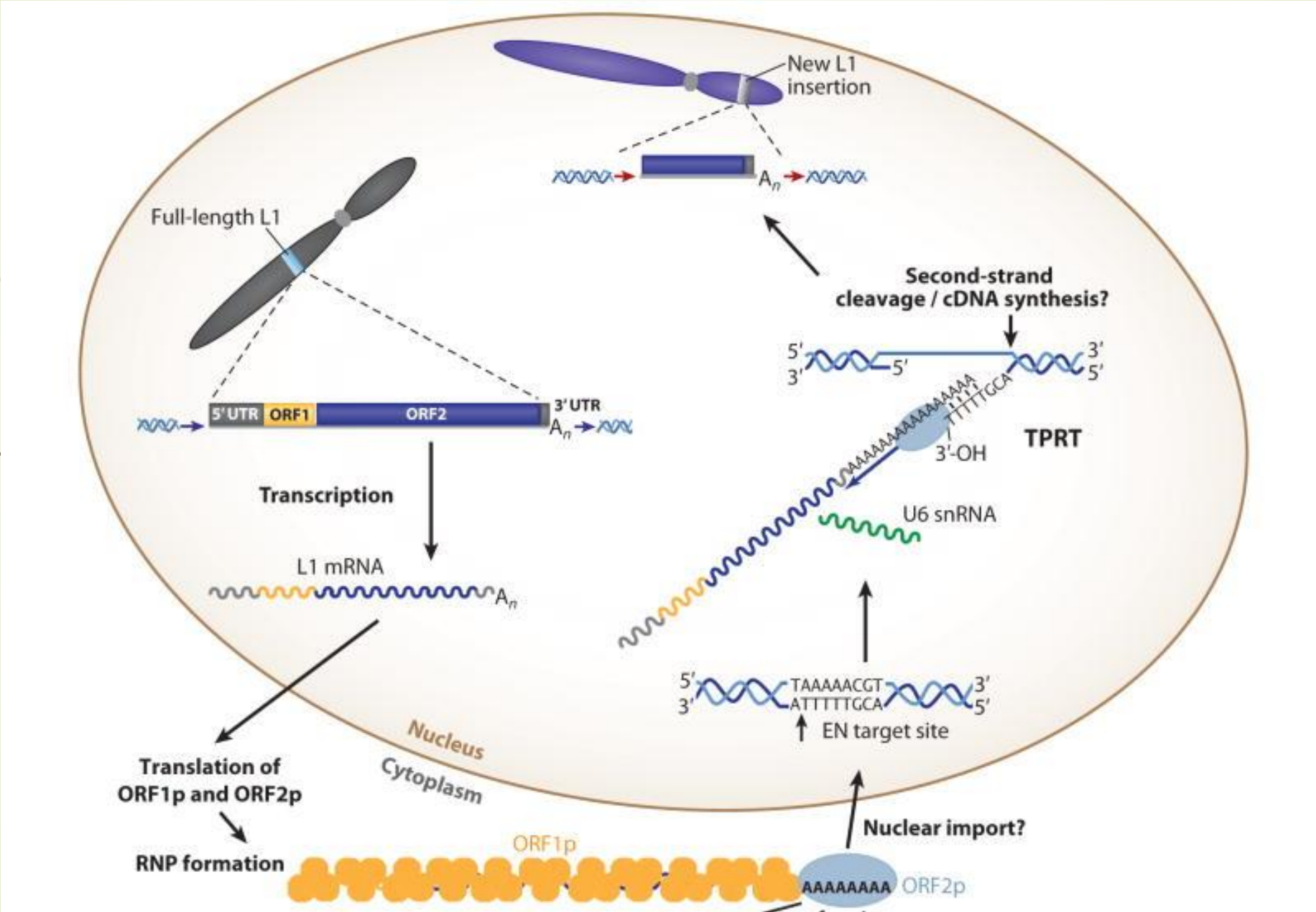
- Sono il **17-21%** del genoma umano.
- Contengono un promotore per la RNA pol II (5'UTR), 1 o 2 ORF, e una sequenza polyA (3'UTR).
- La famiglia più abbondante è **LINE-1 (L1)**.
- Solo 50 elementi L1 possono essere trascritti e tradotti dando luogo a proteine quali una endonucleasi che taglia il DNA e una trascrittasi inversa che sintetizza un cDNA a partire di un RNA.
- Autonomo.
- Meccanismo: target primed reverse transcription (TPRT)



LINES: target primed reverse transcription (TPRT)



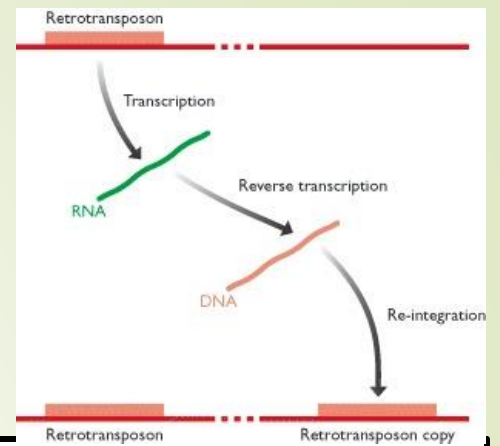
LINE-1 retrotransposition cycle.



DNA ripetitivo intersparso o Genome wide-

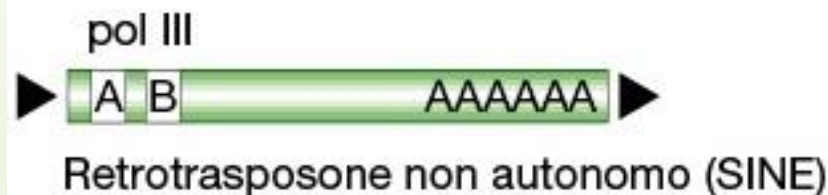
Classe I o Retrotrasposoni

► Usano un intermediario di RNA

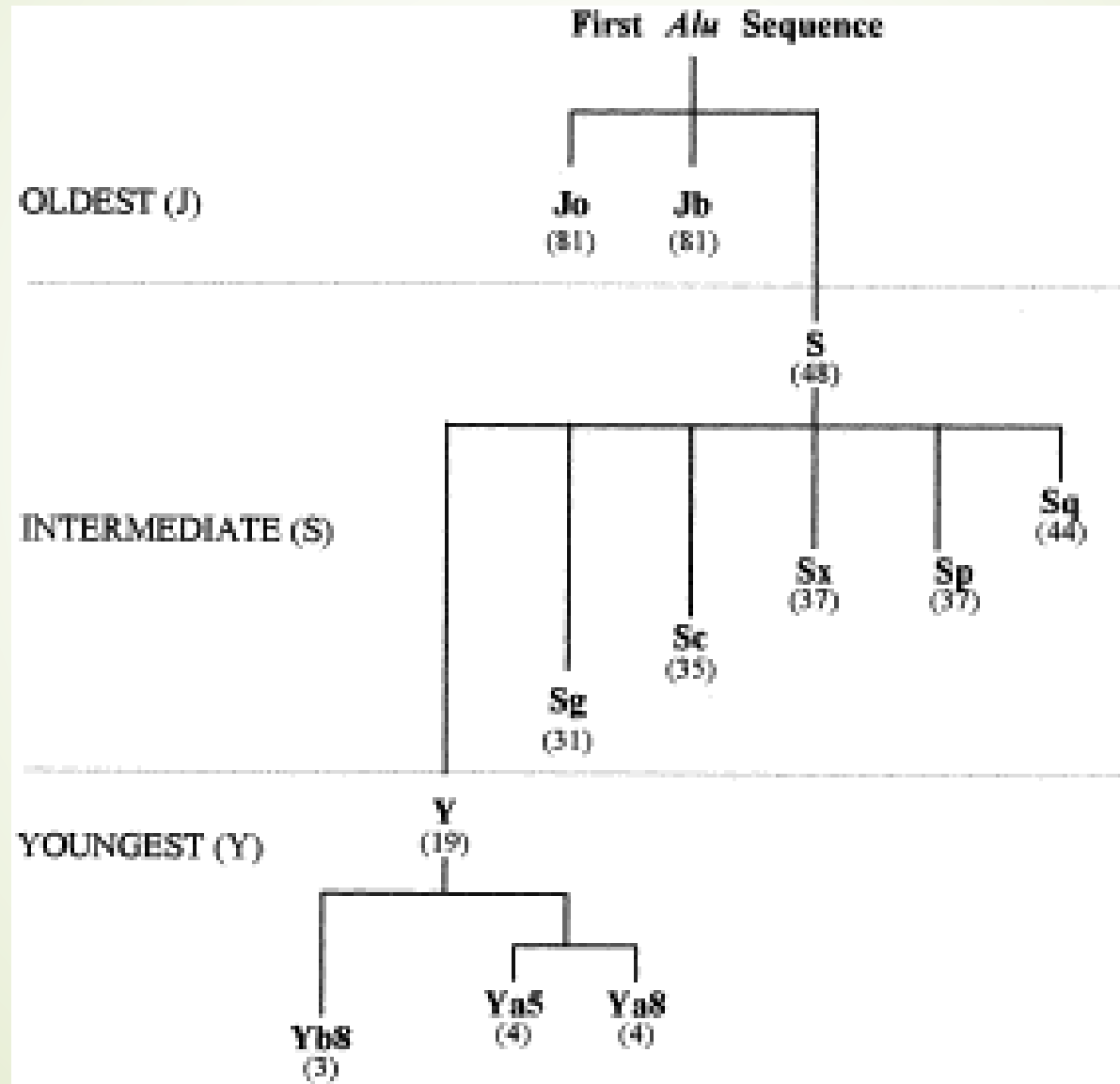


SINEs: (short interspersed nuclear elements, 80-400 pb, presenti 10^5 copie per genoma).

- Sono il **10-13 %** del genoma.
- I più abbondanti sono gli elementi **Alu**, sequenze di 260 pb che contengono una sequenza riconosciuta dal enzima di restrizione AluI.
- Non codificano per alcuna molecola funzionale e **dipendono da retrotrasposoni autonomi** (L1: trascrittasi inversa) per essere trasponibili.
- Trascrivono basi livelli di RNAs che contribuiscono alla retrotrasposizione.
- Alu e' **fiancheggiato da sequenze ripetute di lunghezza variabile** formate per duplicazione delle sequenze dove si inseriscono
- presenti in umani e primati, in molti geni influenzando espressione genica.
- Anche se sono stati elementi molto attivi anteriormente durante l'evoluzione, attualmente solo pochi continuano a trasponersi.
- Si pensa che ci sia un nuovo element Alu ogni 200 nascite
- **Non autonomo**



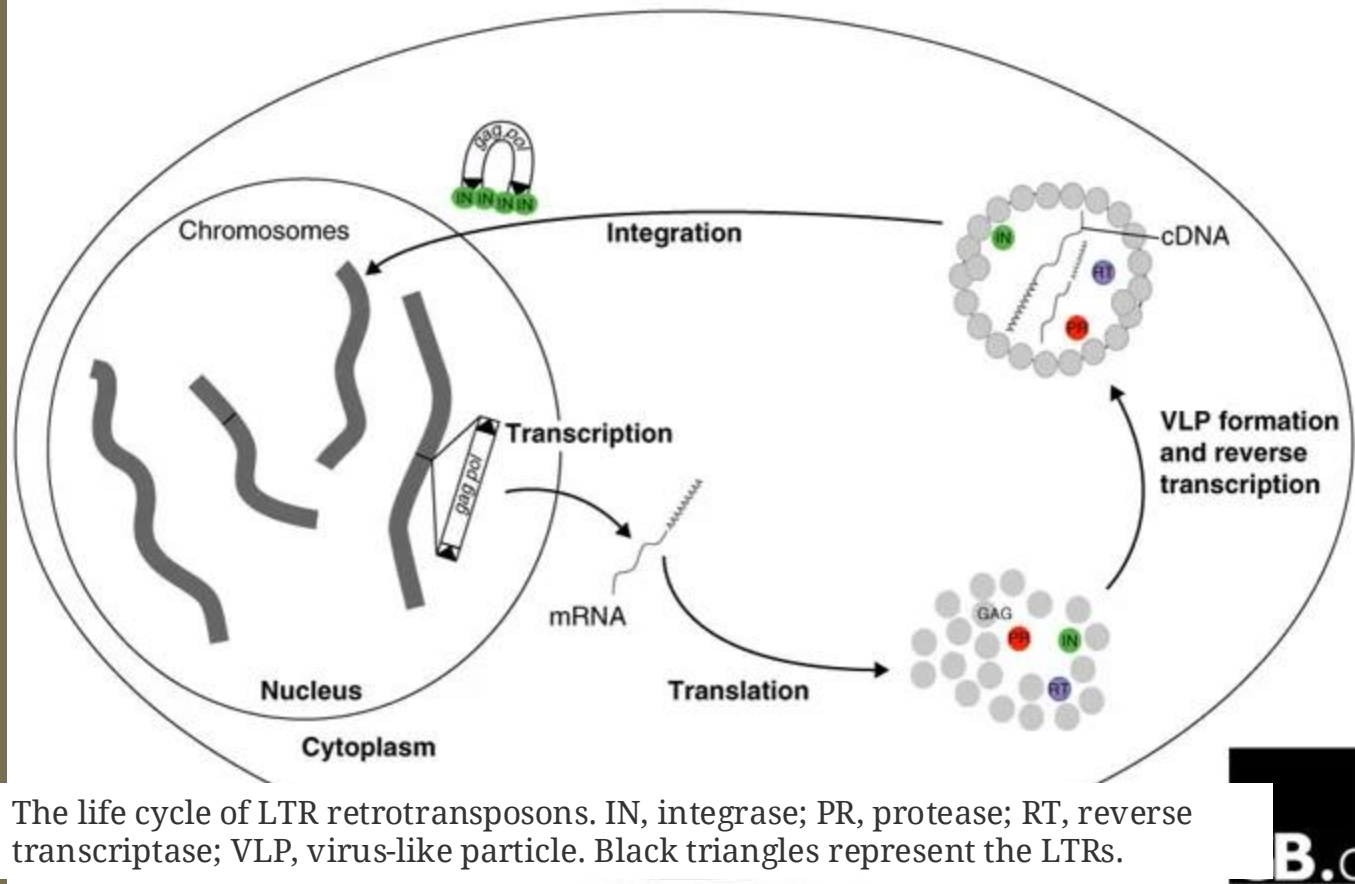
Alu sequences and evolution



DNA ripetitivo intersparso o Genome wide-

Classe I o Retrotrasposoni

► Usano un intermediario di RNA

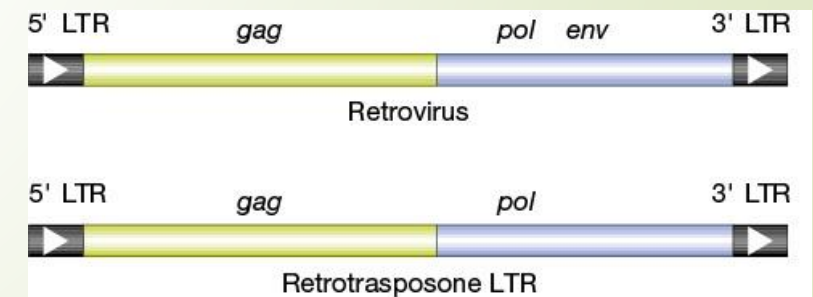


LTR element: (Long terminal repeat elements).

Sono un **8%** del genoma

Contengono sequenze di DNA ripetute di lunghezza variabile all'estremità (Long terminal repeats, LTR) che fiancheggiano una regione codificante.

Sono originate dopo la sintesi di un intermediario RNA (RNA pol II cellulare) e l'azione di una trascrittasi inversa che da luogo a un cDNA, che verra' poi integrato via integrasi LTR specifiche nel cromosoma della cellula ospite.



Pol: Enzymatic activity including reverse transcriptase and integrase

gag: Proteine strutturali

The life cycle of LTR retrotransposons. IN, integrase; PR, protease; RT, reverse transcriptase; VLP, virus-like particle. Black triangles represent the LTRs.

Classe I

Classe II

1. Long interspersed elements (LINE) (autonomous)		LINE-1 LINE-2 LINE-3	~600,000 copie ~370,000 copie ~44,000 copie
2. Short interspersed elements (SINE) (non-autonomous)		Alu family MIR MIR3	~1,200,000 copie ~450,000 copie ~85,000 copie
3. Retrovirus-like elements (LTR transposons)		Human endogenous retroviral sequences (HERV) (several classes; autonomous and non-autonomous)	~240,000 copie
4. DNA transposons		Several classes (autonomous and non-autonomous)	~300,000 copie

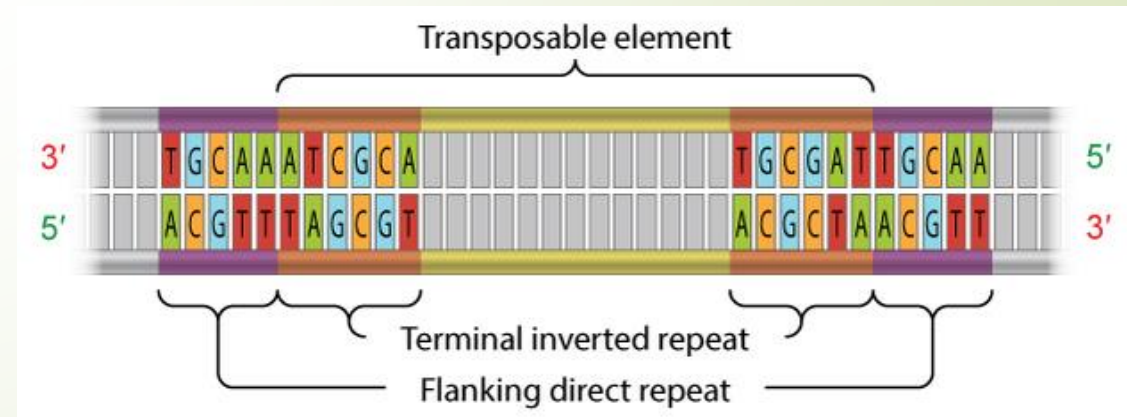
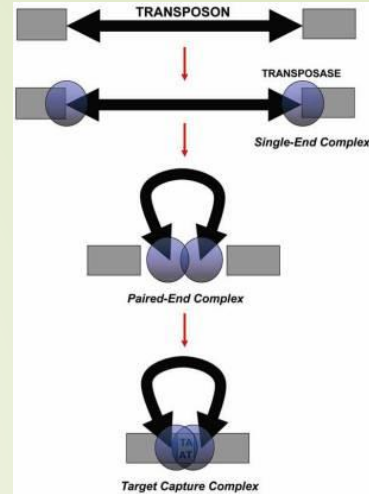
B. Major types of interspersed repetitive DNA in the human genome

DNA ripetitivo intersparso o Genome wide-

Classe II o trasposoni a DNA

➤ Non usano un intermediario di RNA e sono autonomi

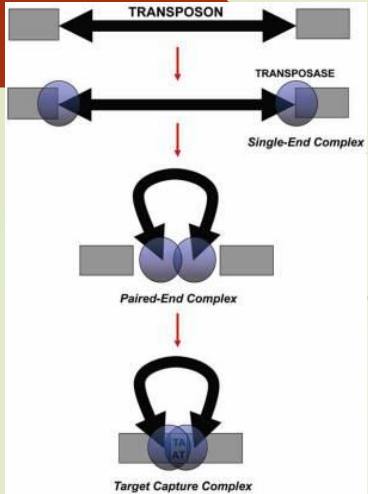
- Consistono in sequenze di DNA che contengono l'informazione che codifica per un enzima **transposasi** (che taglia e sposta il DNA) con delle sequenze fiancheggiatrici chiamate **Terminal Inverted Repeats (TIRs, 9-40 basi)**.
- Le TIR hanno sequenze ripetute e invertite che si complementano e sono riconosciute dalla trasposasa



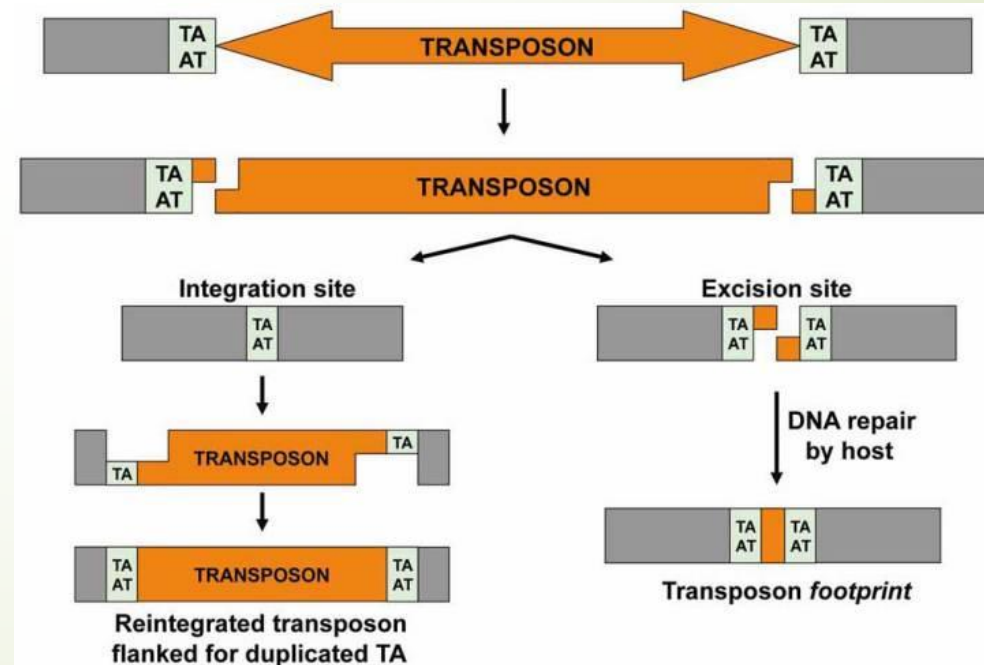
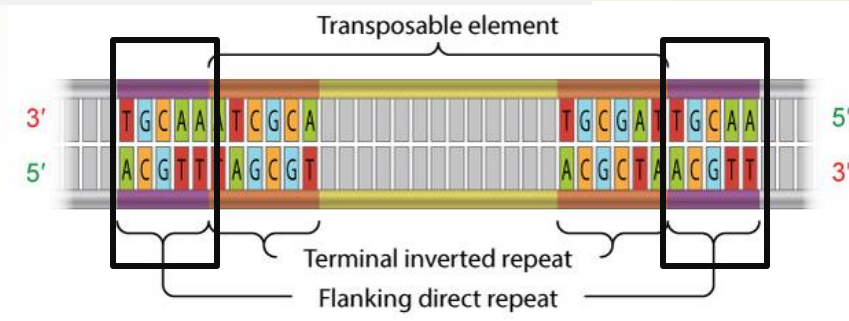
DNA ripetitivo intersparso o Genome wide-

Classe II o trasposoni a DNA

➡ Non usano un intermediario di RNA e sono autonomi



- La trasposasi riconosce le TIRs portando alla escissione del DNA trasposone e permettendo l'inserzione in un'altra posizione del genoma.
- Conseguenza della inserzione il sito d'inserzione si duplica.



DNA ripetitivo intersparso o Genome wide-

Classe II o trasposoni a DNA

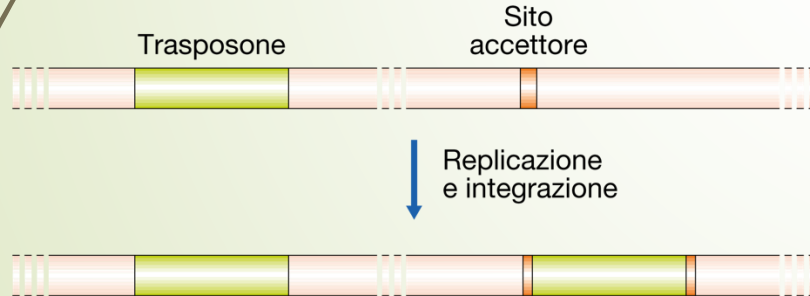
➡ Non usano un intermediario di RNA



Meccanismo di propagazione

REPLICATIVO

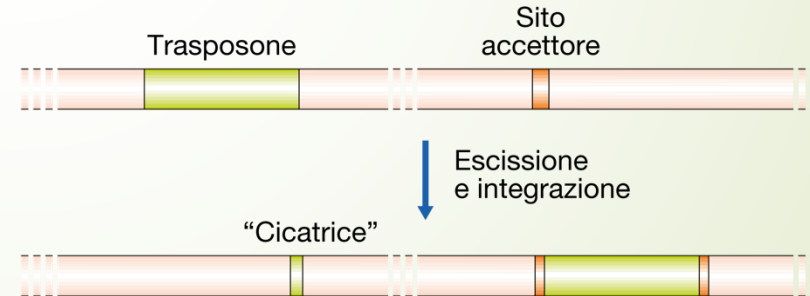
una nuova copia del trasposone si integra nel genoma in un'altra posizione. Il numero di elementi trasposti raddoppia.



Meccanismo di propagazione

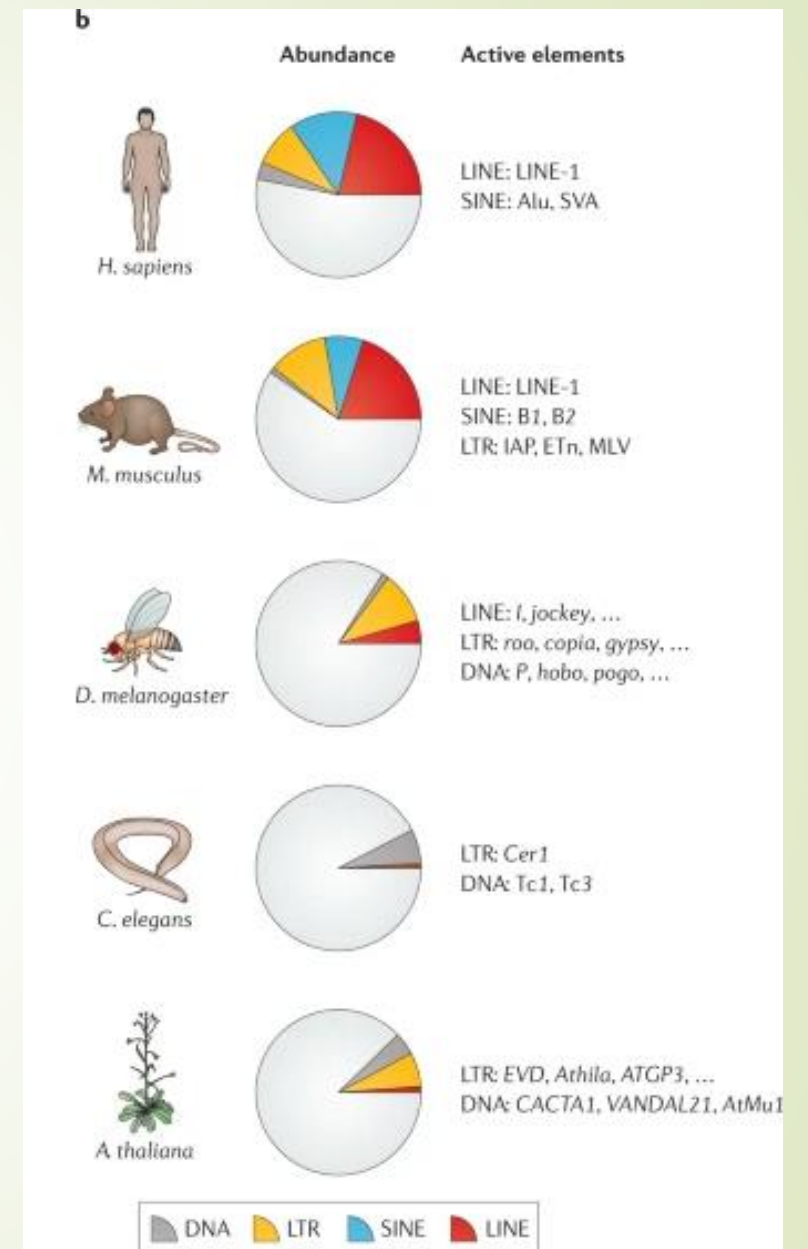
CONSERVATIVO

La trasposasi codificata dal trasposone taglia il trasposone permettendo la trasposizione ad un'altra regione del genoma (taglia e cuci). Il numero di copie rimane costante



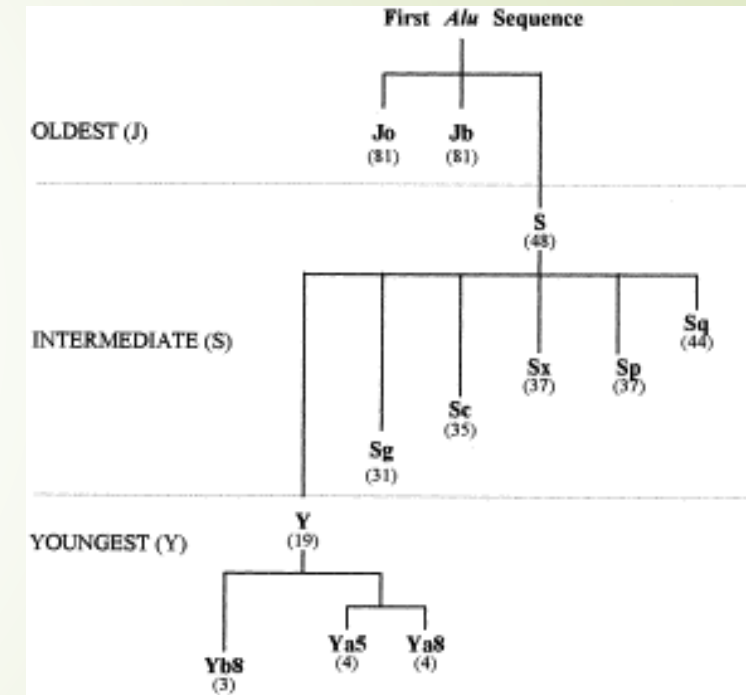
<https://www.youtube.com/watch?v=CroyUMRpbxg>

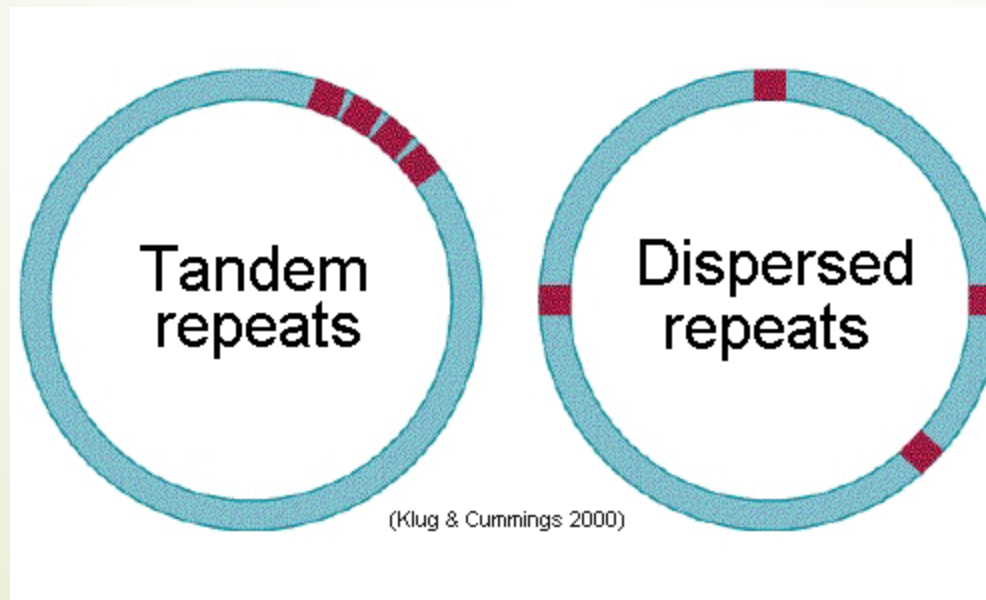
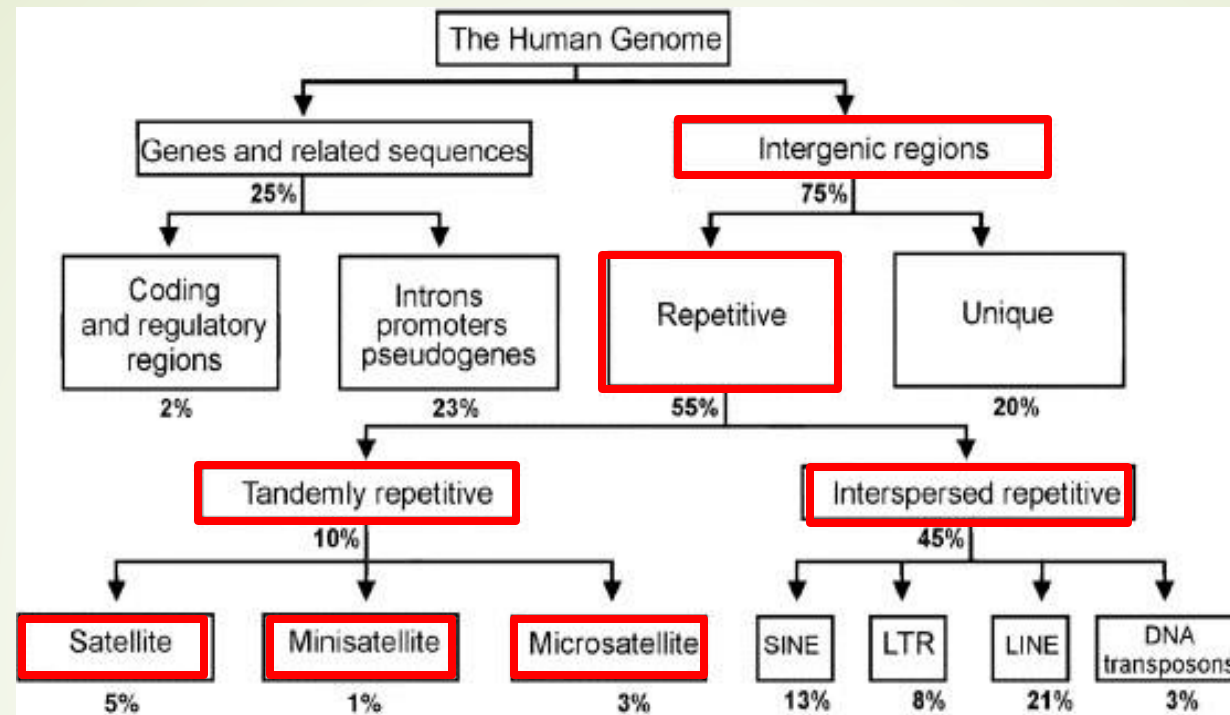
Type of repeat	Subtype	Approximate number of copies in the human genome
<u>SINEs</u>		1 558 000
	<u>Alu</u>	1 090 000
	MIR	393 000
	MIR3	75 000
<u>LINEs</u>		868 000
	<u>LINE-1</u>	516 000
	<u>LINE-2</u>	315 000
	<u>LINE-3</u>	37 000
<u>LTR elements</u>		443 000
	<u>ERV</u> class I	112 000
	<u>ERV</u> (K) class II	8000
	<u>ERV</u> (L) class III	83 000
	MaLR	240 000
<u>DNA transposons</u>		294 000
	hAT	195 000
	Tc-1	75 000
	PiggyBac	2000
	Unclassified	22 000



Rilevanza degli elementi mobili nel genoma

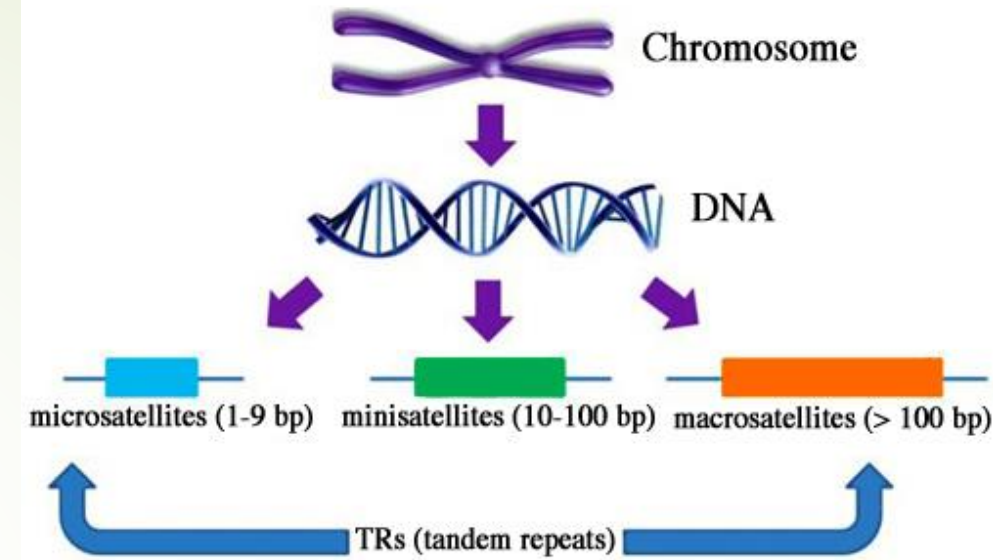
- **Transposoni e retrotrasposoni** sono **stabili** già che raramente spariscono dal genoma (non delezione completa).
- **L'integrazione** di un elemento mobile è **un evento random** e per questo è molto improbabile che 2 elementi mobili si integrino esattamente nella stessa posizione. Per questo **la loro presenza nello stesso locus in due individui diversi indica lo stesso ancestro.**
- Inoltre, il fatto che lo stato ancestrale del elemento mobile in un locus è di assenza, **la comparsa di un elemento mobile in un locus è utile in studi evoluzionistici, filogenetici e nella genetica delle popolazioni.**



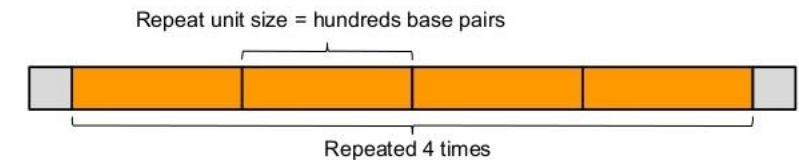


DNA ripetitivo in tandem

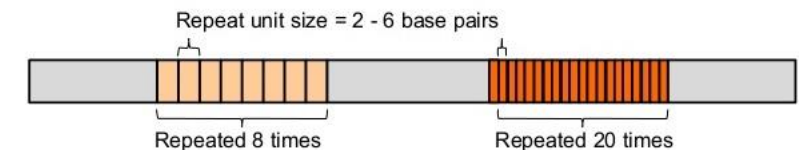
- **DNA satellite** o altamente ripetitivo:
 - ripetizioni in tandem di una breve sequenza
 - Lunghezza variabile della sequenza (5-100 bp), in tratti di fino a 100Mbp
 - Non sono trascritti
 - Si trovano nelle regioni eterocromatiche del cromosoma
- **DNA microsatellite e minisatellite**, o mediamente ripetitivo:
 - Ripetute un numero limitato di volte
 - **Mini:** 11-100 pb ripetuti in tandem più volte fino a 20 kpb
 - **Micro:** 2-10 pb ripetuti in tandem 10-30 volte. Sono elementi **SSR (simple sequence repeats)** o **STR (short tandem repeats)**



Minisatellite: Variable Number Tandem Repeats (VNTR)

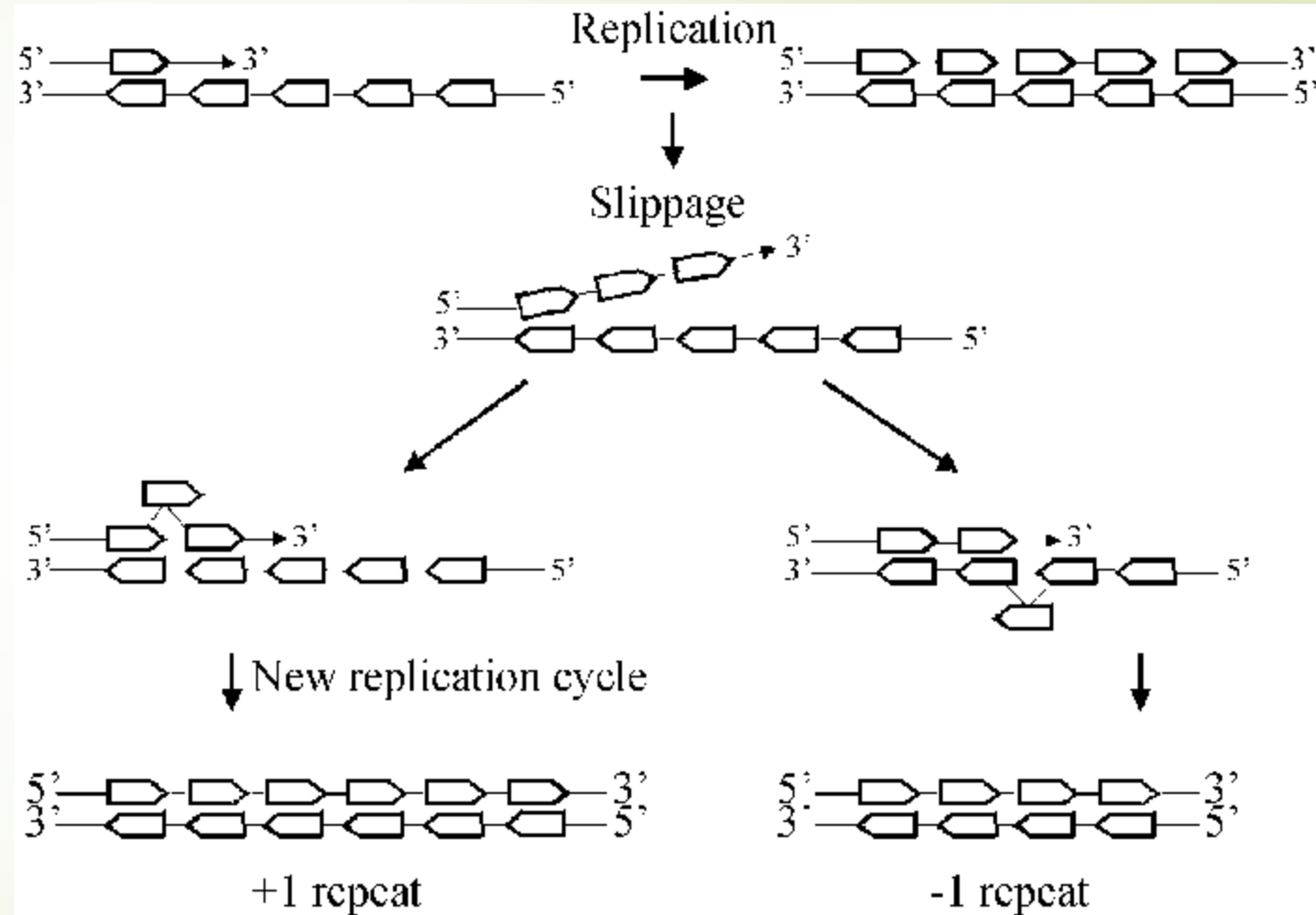


Microsatellite: Short Tandem Repeats



SCIVOLAMENTO o slippage della replicazione

- Si pensa che **durante la replicazione** la catena nascente si **dissocia e rialinea al di fuori della sua posizione originale**, alterando il numero di ripetizioni della catena nascente
- La **probabilità del evento** potrebbe dipendere della lunghezza del microsatellite. Alleli con più ripetizioni sono più instabili di quelli più corti.
- Questo polimorfismo le rende **utili come marcatori all'individualità genetica nel test del DNA**.



DNA microsatellite (STR, short tandem repeats) e Test del DNA

Gene: tratto di DNA che contiene un'informazione genetica

Locus: localizzazione di un gene sui cromosomi

Alleli: forme alternative dello stesso gene

Un **gene** con diversi alleli si dice che è *polimorfico*

Il **profilo genetico (DNA fingerprint)** di un individuo implica la determinazione degli alleli presenti in determinati loci altamente polimorfici caratterizzate da un diverso numero di ripetizioni:
DNA microsatellite o sequenze STR



- Il numero di ripetizioni delle sequenze STR varia da individuo a individuo
- Ogni locus STR presenta un numero variabile di ripetizioni



Profilo Genetico

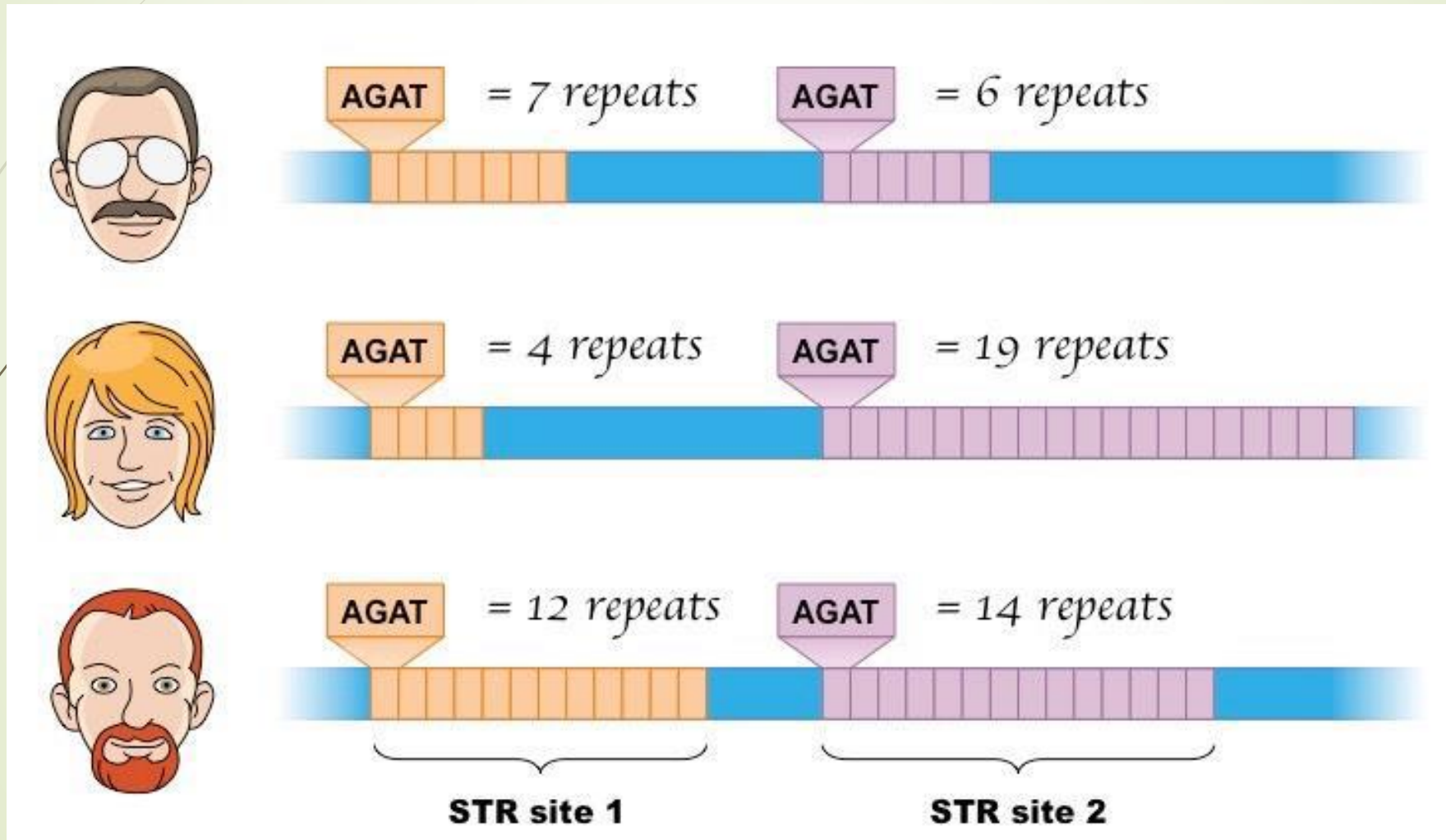
- Se ottiene tramite l'analisi di diversi **marcatori molecolari microsatelliti** che si trovano in diversi loci nel DNA.
- Ciascun **marcatore microsatellite** è **polimorfico** (cioè ci sono tanti alleli diversi per ogni singolo locus marcatore nella specie) e l'insieme dei genotipi (allele/allele) dei diversi microsatelliti analizzati in un soggetto (= profilo genetico) consente di ottenere una sorta di “microchip molecolare” o “impronta digitale genetica” che, con altissima probabilità, non troverà un eguale nella stessa specie.

Regione STR

GCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCAAGTGTATGTGCATACACGTG
CACACACACACACACACACAGAGGGGTGTGCACATGTGCATGCACACTCCAAGAGACAGT
GCCTAGTAAAGTGTCTCAGCACCATCTGCAGCAAACAGGTTCTGCAAAAACCAATCCCAACTG
ATGTTCCACAGTGACACTGT

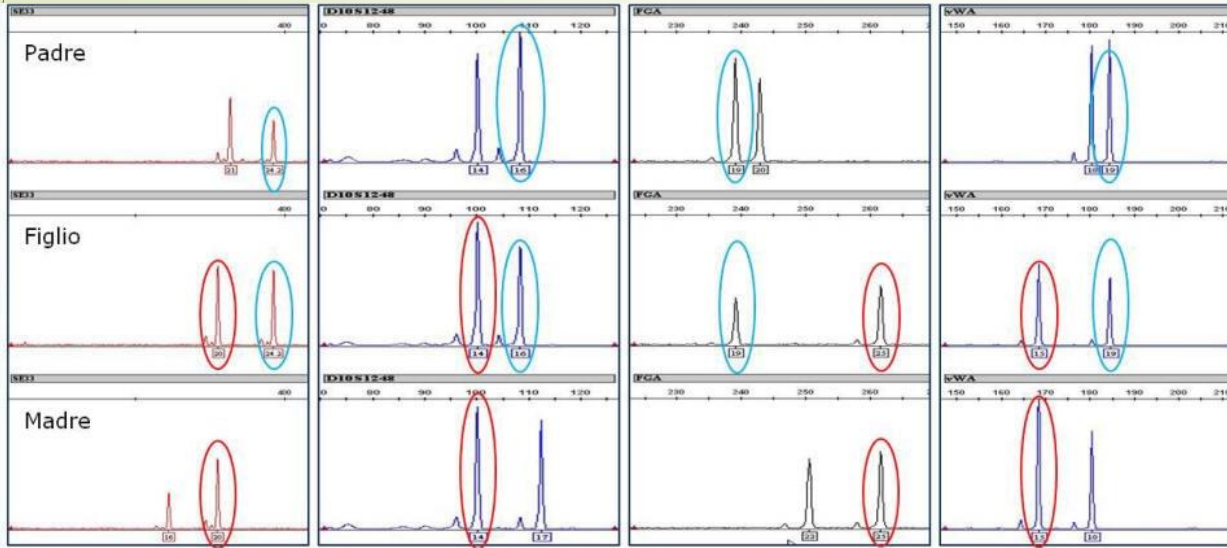
GCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCAAGTGTATGTGCATACACGTG
CACACACACACACACACACACACACACACACACACACAGAGGGGTGTGCACATGTG
CATGCACACTCCAAGAGACAGTGCCTAGTAAAGTGTCTCAGCACCATCTGCAGCAAACAGGT
CTGCAAAAACCAATCCCAACTGATGTTCCACAGTGACACTGT

Profilo genetico

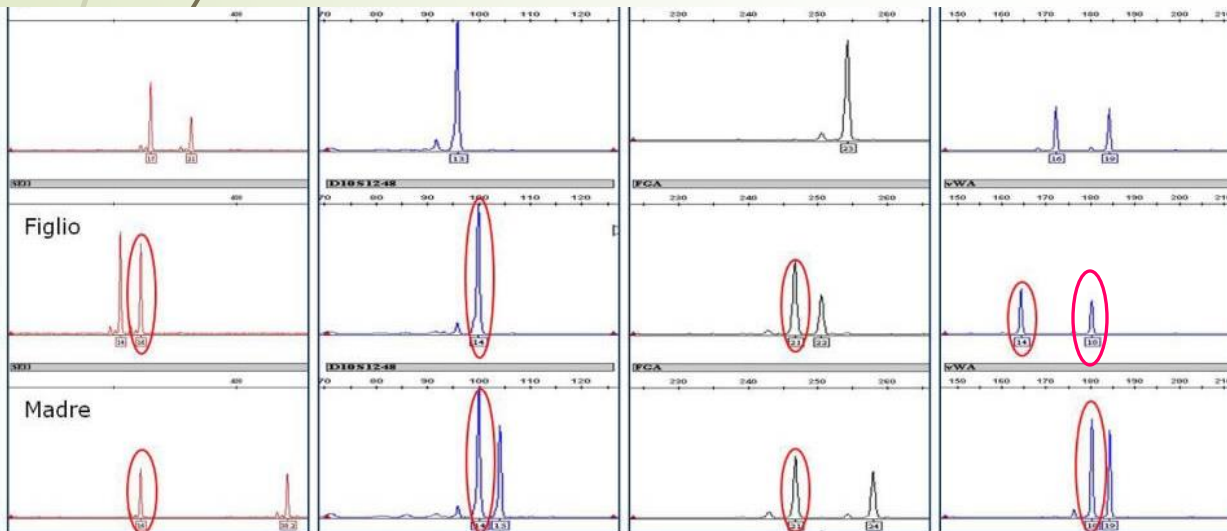


Profilo genetico

RICONOSCIMENTO



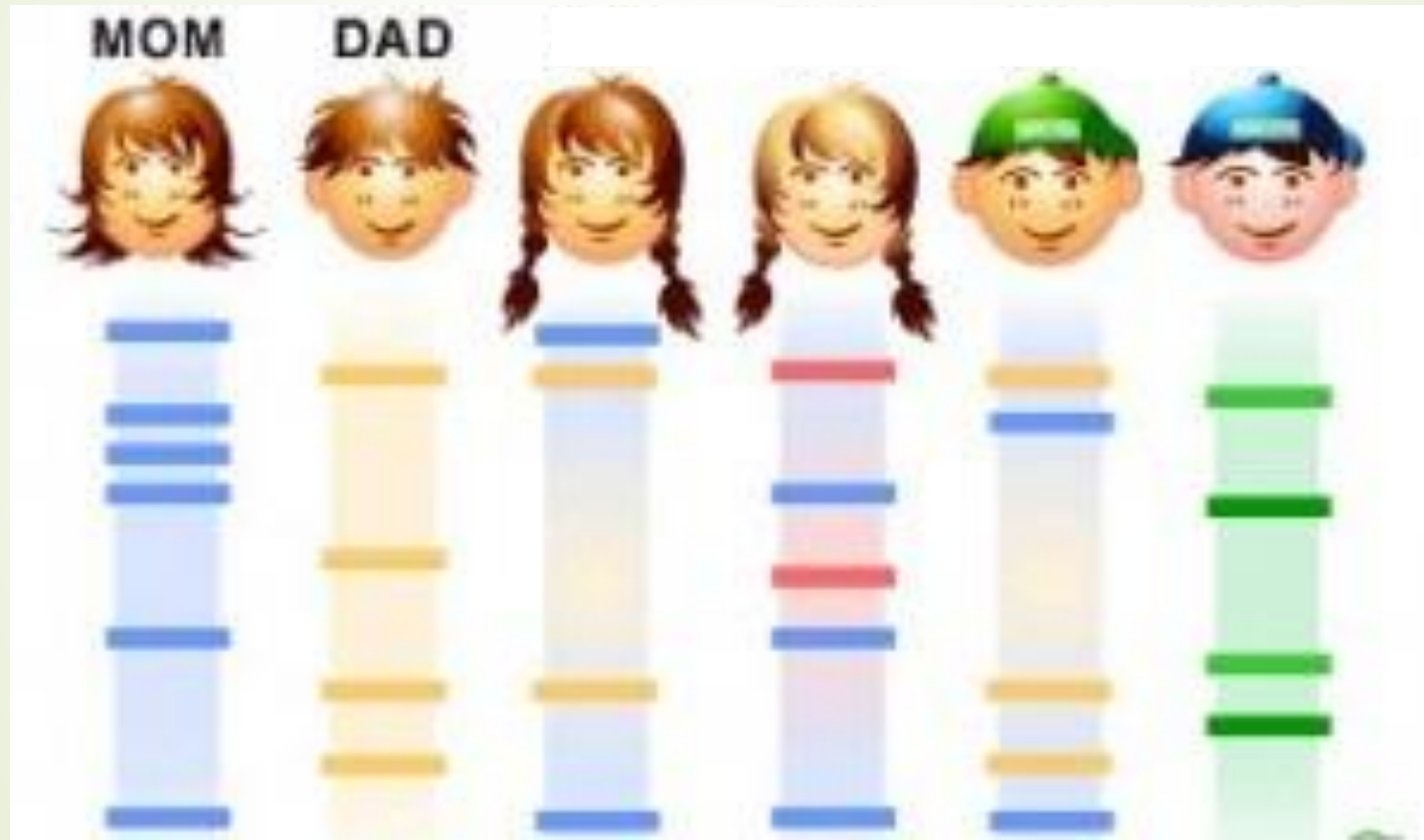
DISCONOSCIMENTO



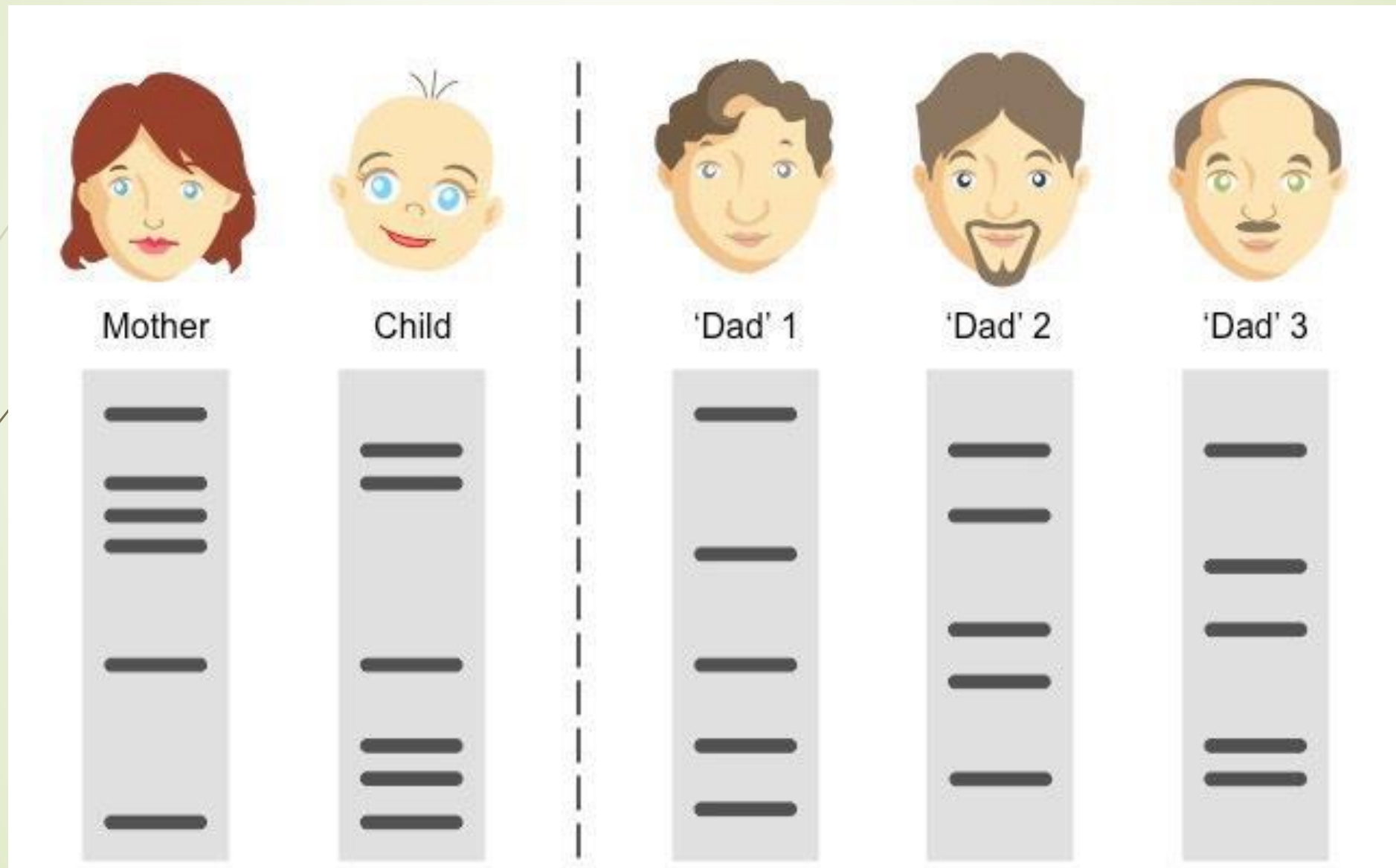
- Comporta la genotipizzazione di un numero variabile di regioni STR (9-15).
- Per questo bisogna **disegnare dei primer** che permettano amplificare queste regioni specifiche tramite una **reazione enzimatica di amplificazione del DNA** conosciuta come **Polymerase Chain reaction (PCR)**.
- Questa tecnica ci permette avere delle grandissime quantità di ogni sequenza che potrà poi essere sequenziato ed analizzata.

Analisi tramite sequenziamento o anche tramite gel di agarosio

Quali sono i figli nati da questa coppia?



Chi è il padre?





Sequenze ripetute all'interno dei geni: Patologie associate all'espansione anomala di triplette di nucleotidi

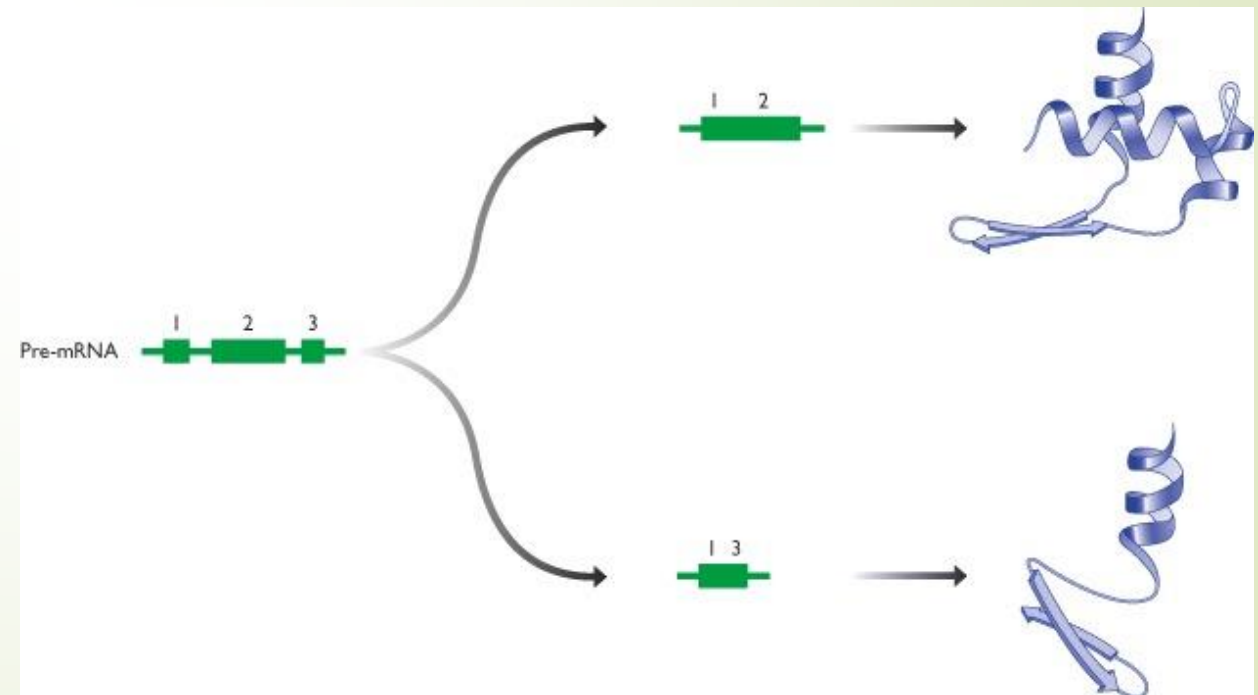
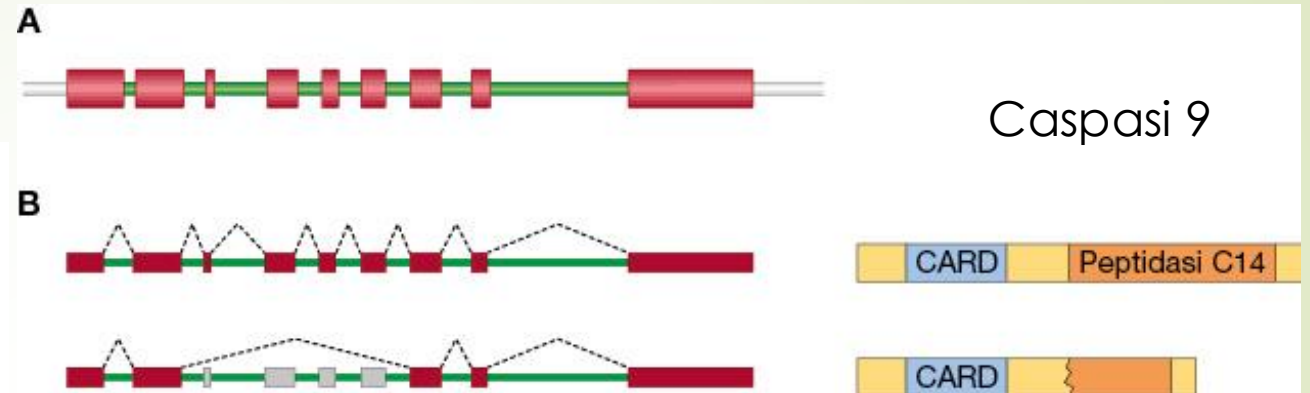
- Le sequenze ripetute di trinucleotide presenti nelle sequenze ripetute del genoma vengono trasmesse stabilmente.
- Al di sopra di un numero di ripetizioni la sequenza diventa molto instabile determinando un aumento delle ripetizioni nelle generazioni successive che può modificare l'espressione genica e la stabilità del mRNA dando luogo a patologie

Patologie associate all'espansione anomala di triplette di nucleotidi (DNA microsatellite)

Malattia	Gene mutante	Locus	Proteina	Ripetizione	Sequenza ripetuta		Sito di ripetizione
					Normale	Malattia	
Sindrome dell'X fragile	FMR1 (FRAXA)	Xq27.3	FMR-1 (FMRP)	CGG	6–53	60–200 (pre) > 230	5'UTR
Sindrome dell'XE fragile	FMR2 (FRAXE)	Xq28	FMR-2	GCC	6–35	60–200 (pre) > 200	5'UTR
Atassia di Friedreich	X25	9q13-21.1	Fratassina	GAA	7–34	34–80 (pre) > 100	Introne 1
Distrofia miotonica	DMPK	19q13	DMPK (Myotonic Dystrophy Protein Kinase)	CGT	5–37	50–migliaia	3'UTR
Atrofia muscolare spinale e bulbare (malattia di Kennedy)	AR	Xq13-21	Recettore degli androgeni (AR)	CAG	9–36	38–62	Regione codificante
Malattia di Huntington	HD	4p16.3	Huntington	CAG	6–35	49–88	Regione codificante
Atrofia dentato-rubro-pallidoluisiana (sindrome di Haw River)	DRPLA	12p13.31	Atrofina-1	CAG	6–35	49–88	Regione codificante

Anatomia del Genoma ed Implicazioni evolutive

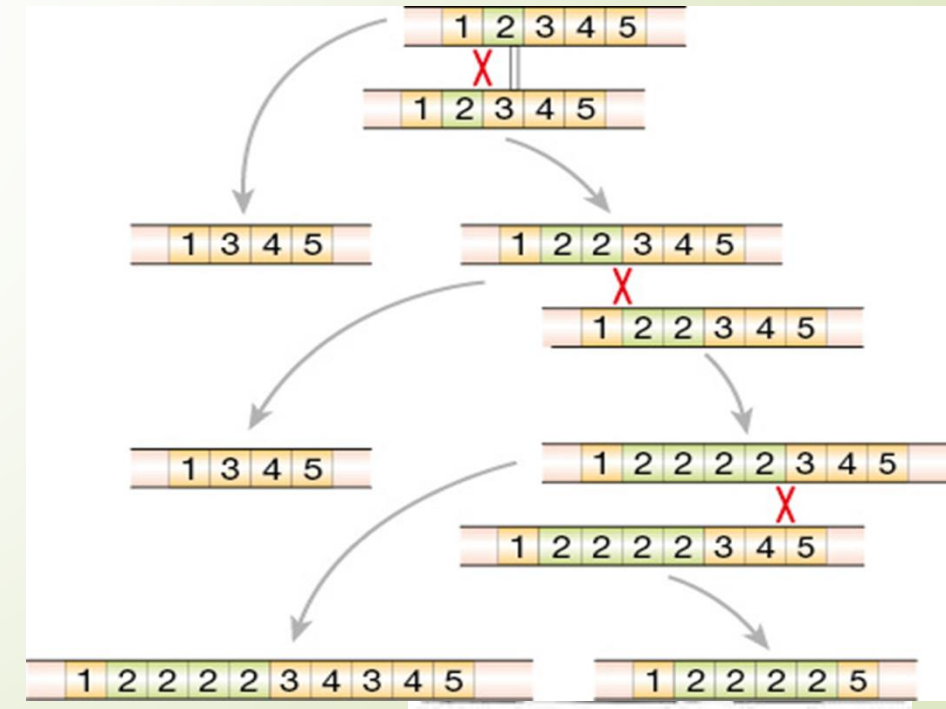
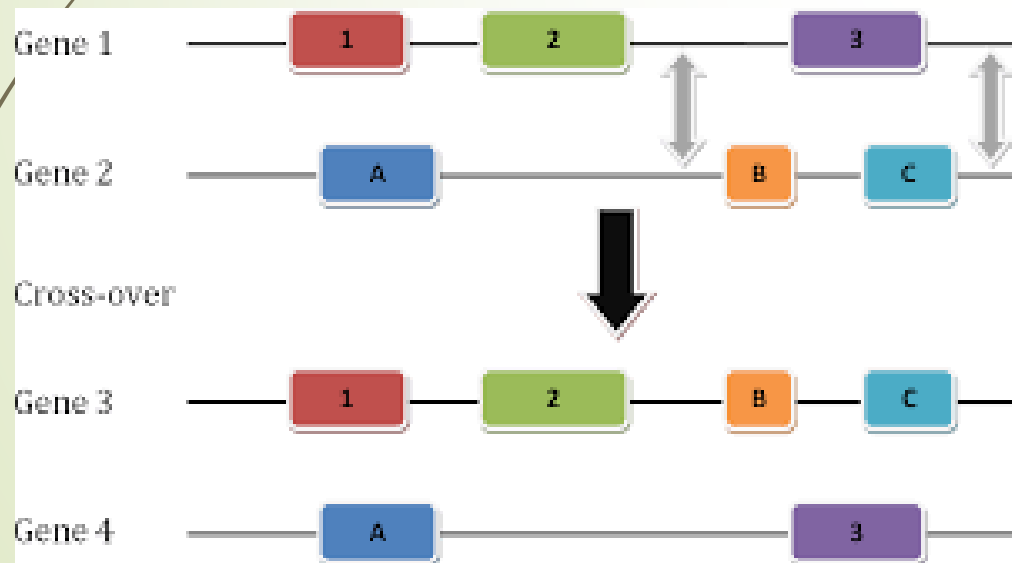
- La **comparsa degli introni**, sequenze apparentemente inutili, nel genoma eucariotico ha comportato un vantaggio evolutivo associato alla variabilità indotta dallo **splicing alternativo** dagli introni nel processo di maturazione del RNA.
- **Splicing alternativo**: processo mediante il quale gli esoni di un pre-mRNA possono essere assemblati in diverse combinazioni dipendendo della eliminazione degli introni che gli separano.
- Ogni mRNA dà luogo a proteine diverse che possono avere anche delle funzioni opposte
- Il processo di splicing è finemente regolato ed alterazioni in questo processo portano alla comparsa di malattie genetiche chiamate Spliceopatie.



Generazione di nuovi geni: Exon-shuffling

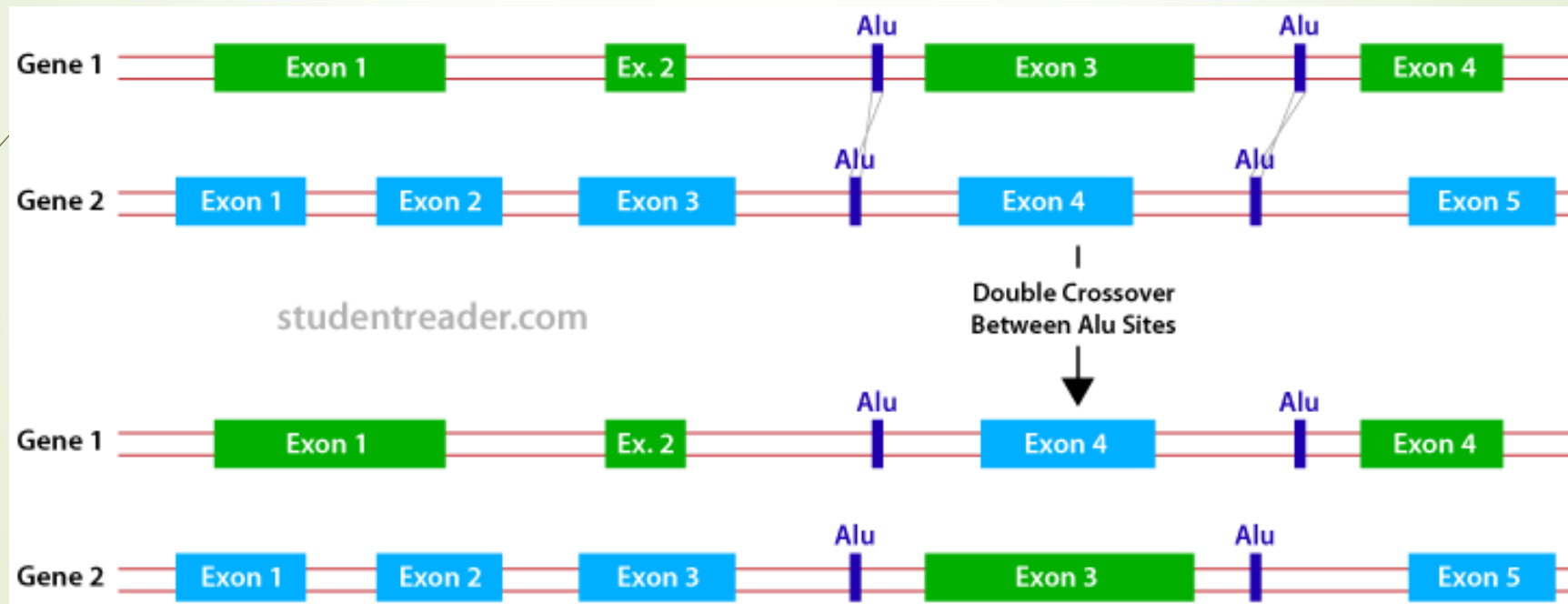
La presenza di geni a sequenza discontinua (esoni + introni) portò a Walter Gilbert a proporre l'esistenza di un processo chiamato **Riassortimento degli esoni** o **Exon-shuffling** come meccanismo implicato nella sintesi di nuove proteine, evento avvenuto durante l'evoluzione.

- Questo processo comporta la **combinazione di esoni provenienti da geni diversi formando una nuova combinazione di esoni** rispetto a quella di partenza.
- Il riassortimento avviene in seguito a **crossing over diseguale** (meccanismo osservato nei cromosomi come conseguenza del appaiamento non preciso tra cromosomi omologhi durante la ricombinazione meiotica)



Generazione di nuovi geni: Esonizzazione

- Nuove varianti geniche possono essere originate mediante un altro processo denominato **Esonizzazione**.
- **L'esonizzazione** porta all'inserimento di elemento mobili all'interno degli introni e la loro successiva attivazione per l'acquisizione di nuovi siti di splicing.



Poche volte queste modifiche sono mantenute nel tempo

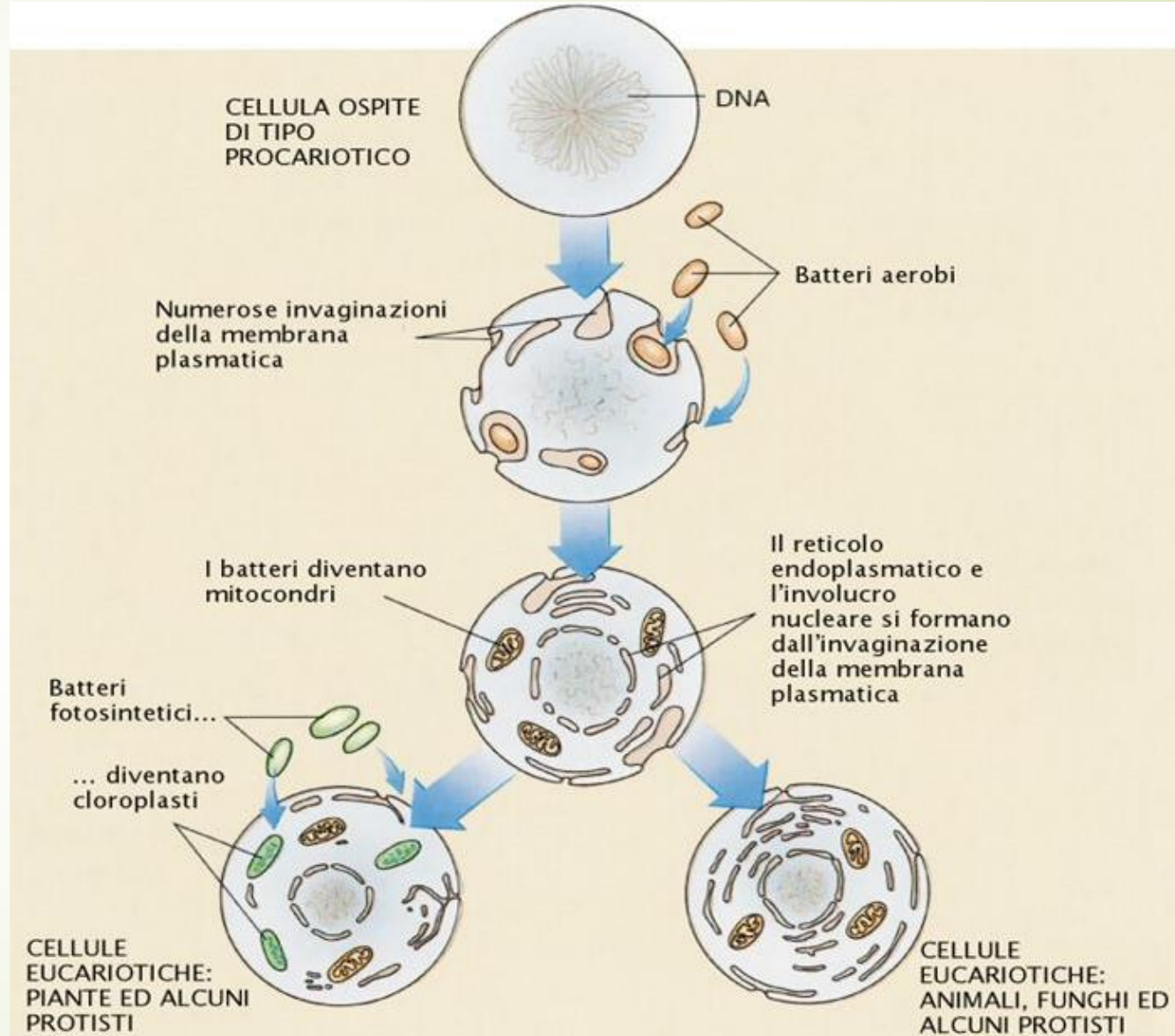
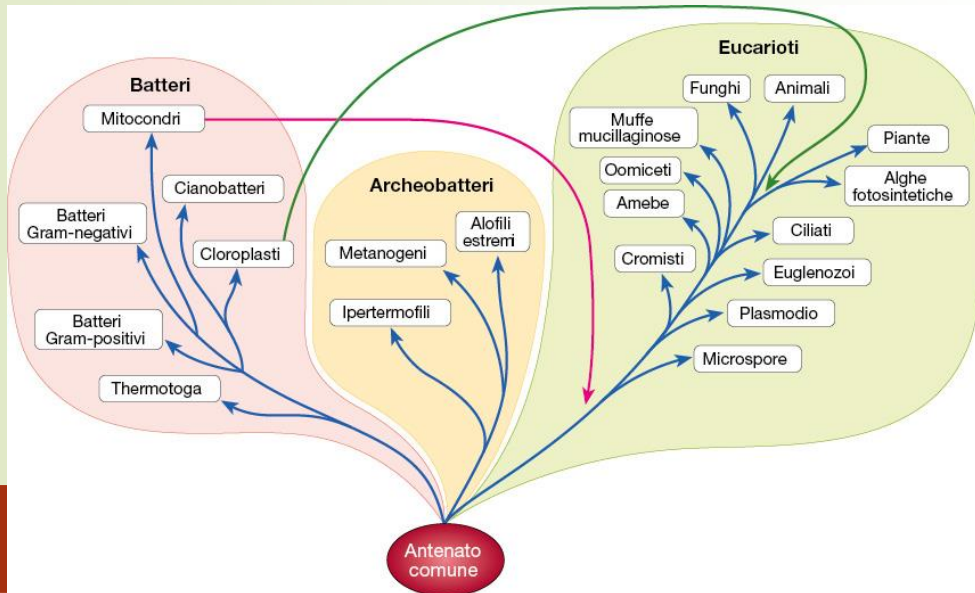
Genoma mitocondriale



Genoma mitocondriale

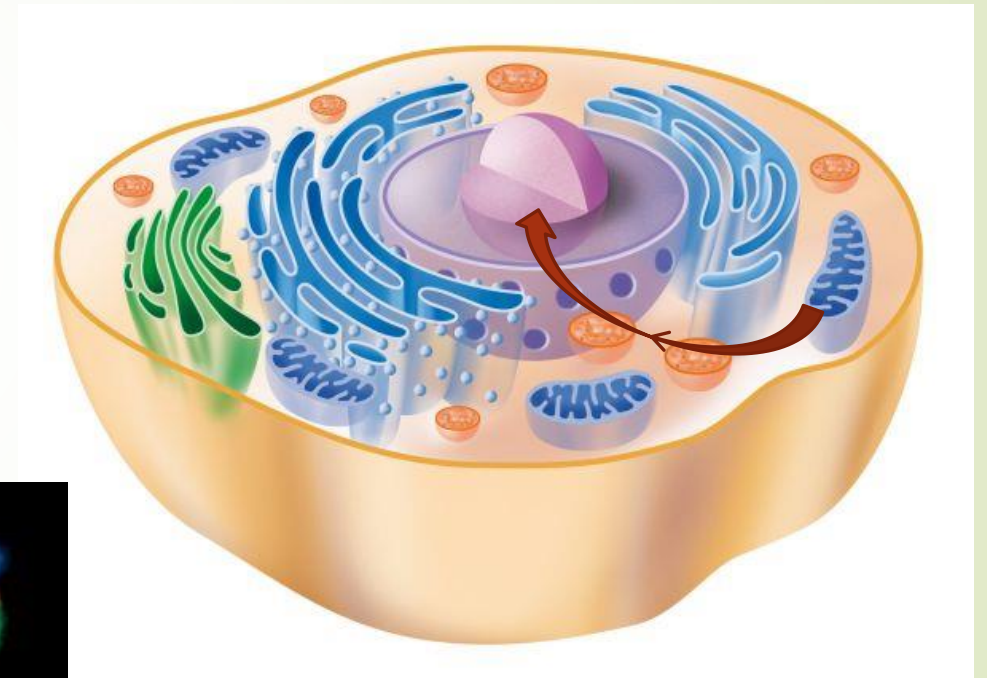
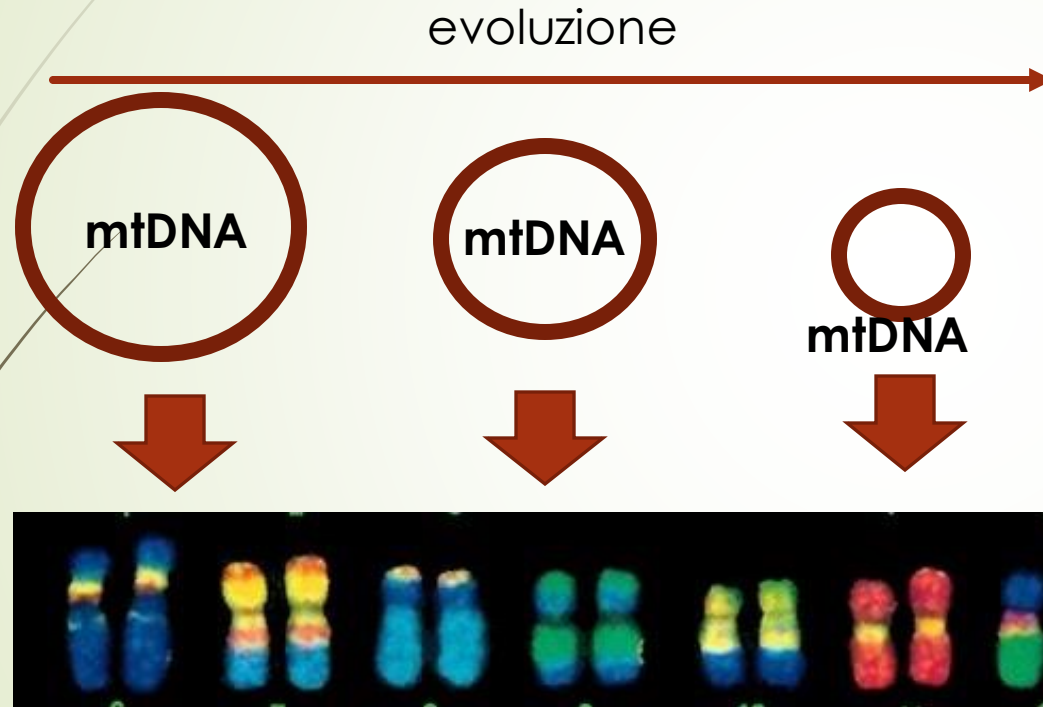
- I mitocondri sono organelli delle cellule eucariotiche con il proprio genoma, indicatore del loro origine batterico

Teoria Endosimbiontica



Genoma mitocondriale

- Nell'evoluzione il genoma mitocondriale (batterico) è diventato sempre più piccolo dovuto al trasferimento di geni al genoma nucleare

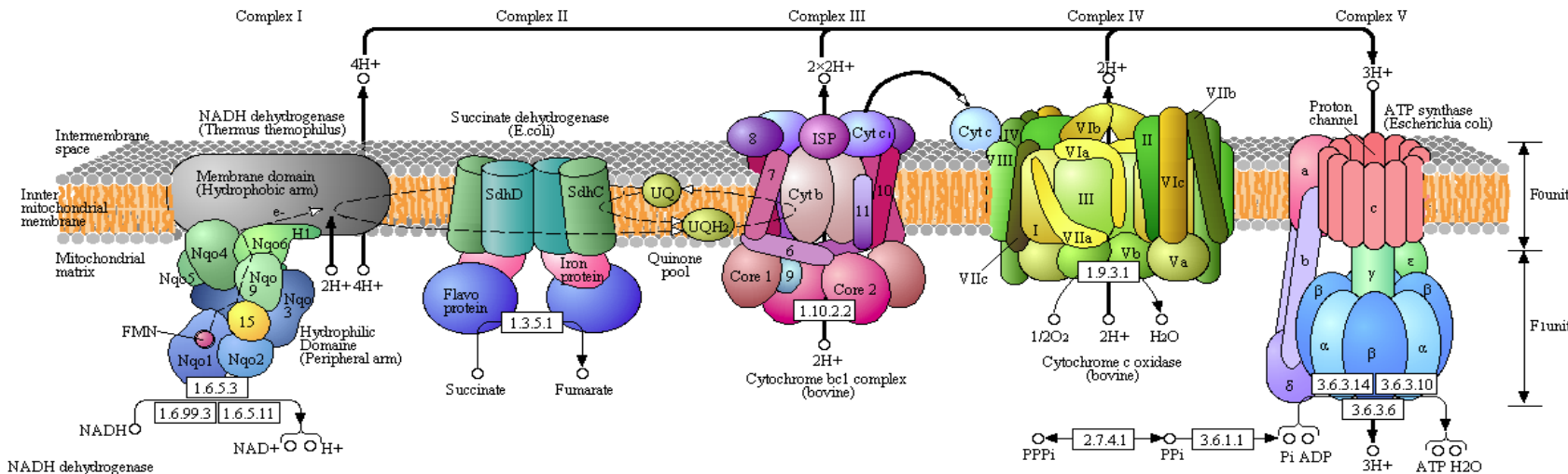


Perché non sono stati trasferiti tutti i geni mitocondriali al genoma nucleare?
Perché è stato conservato il mtDNA?

Perché non sono stati trasferiti tutti i geni mitocondriali al genoma nucleare?

Perché è stato conservato il mtDNA?

OXIDATIVE PHOSPHORYLATION



NADH dehydrogenase

E ND1 ND2 ND3 ND4 ND4L ND5 ND6

E Ndufs1 Ndufs2 Ndufs3 Ndufs4 Ndufs5 Ndufs6 Ndufs7 Ndufs8 Ndufv1 Ndufv2 Ndufv3

B/A NuoA NuoB NuoC NuoD NuoE NuoF NuoG NuoH NuoI NuoJ NuoK NuoL NuoM NuoN

B/A NdhC NdhK NdhJ NdhH NdhA NdhI NdhG NdhE NdhF NdhD NdhB NdhL NdhM NdhN HoxE HoxF HoxU

E Ndufa1 Ndufa2 Ndufa3 Ndufa4 Ndufa5 Ndufa6 Ndufa7 Ndufa8 Ndufa9 Ndufa10 Ndufab1 Ndufa11 Ndufa12 Ndufa13

E Ndubf1 Ndubf2 Ndubf3 Ndubf4 Ndubf5 Ndubf6 Ndubf7 Ndubf8 Ndubf9 Ndubf10 Ndubf11 Ndubf1 Ndubf2

Succinate dehydrogenase / Fumarate reductase

E SDHC SDHD SDHA SDHB

B/A SdhC SdhD SdhA SdhB
FrdA FrdB FrdC FrdD

Cytochrome c reductase

E/B/A ISP Cytb Cyt1

E COR1 QCR2 QCR6 QCR7 QCR8 QCR9 QCR10

Cytochrome c oxidase

E COX10 COX3 COX1 COX2 COX4 COX5A COX5B COX6A COX6B COX6C COX7A COX7B COX7C COX8

B/A CyoE CyoD CyoC CyoB CyoA

CoxD CoxC CoxA CoxB
QoxD QoxC QoxB QoxA

Cytochrome c oxidase, cbb3-type

B I II IV III

Cytochrome bd complex

B/A CydA CydB

F-type ATPase (Bacteria)

alpha	beta	gamma	delta	epsilon
a	b	c		

F-type ATPase (Eukaryotes)

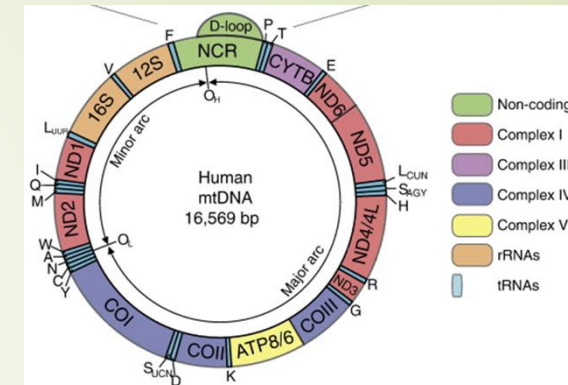
alpha	beta	gamma	delta	epsilon
OSCP	a	b	c	d
f	g	f6/f1	j	k
				8

V/A-type ATPase (Bacteria, Archaeas)

A	B	C	D	E	F	G/H
I	K					

V-type ATPase (Eukaryotes)

A	B	C	D	E	F	G	H
a	c	d	e	S1			



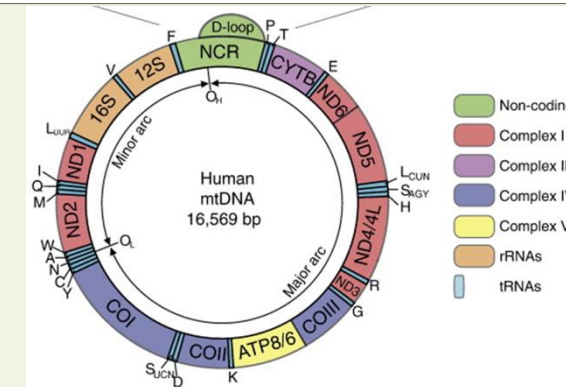
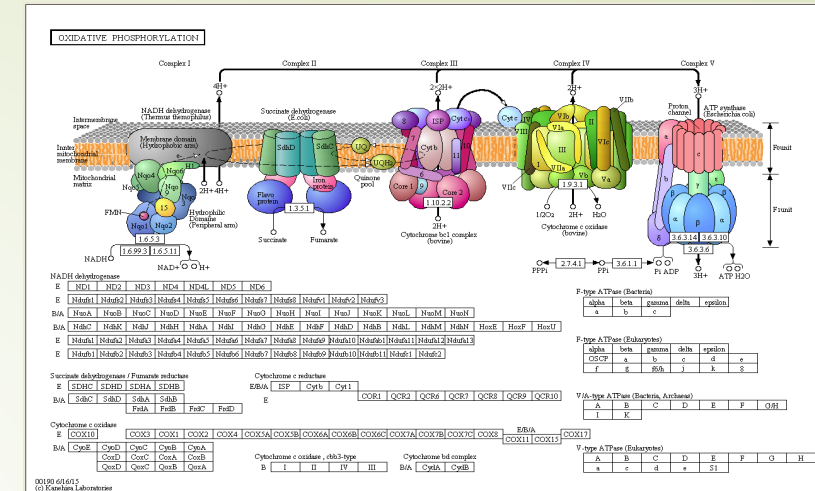
Perché sono state conservate in una molecola autonoma all'interno dei mitocondri?

- ▶ Sono poche proteine codificate dal mtDNA ma molto importanti per il funzionamento

Ipotesi:

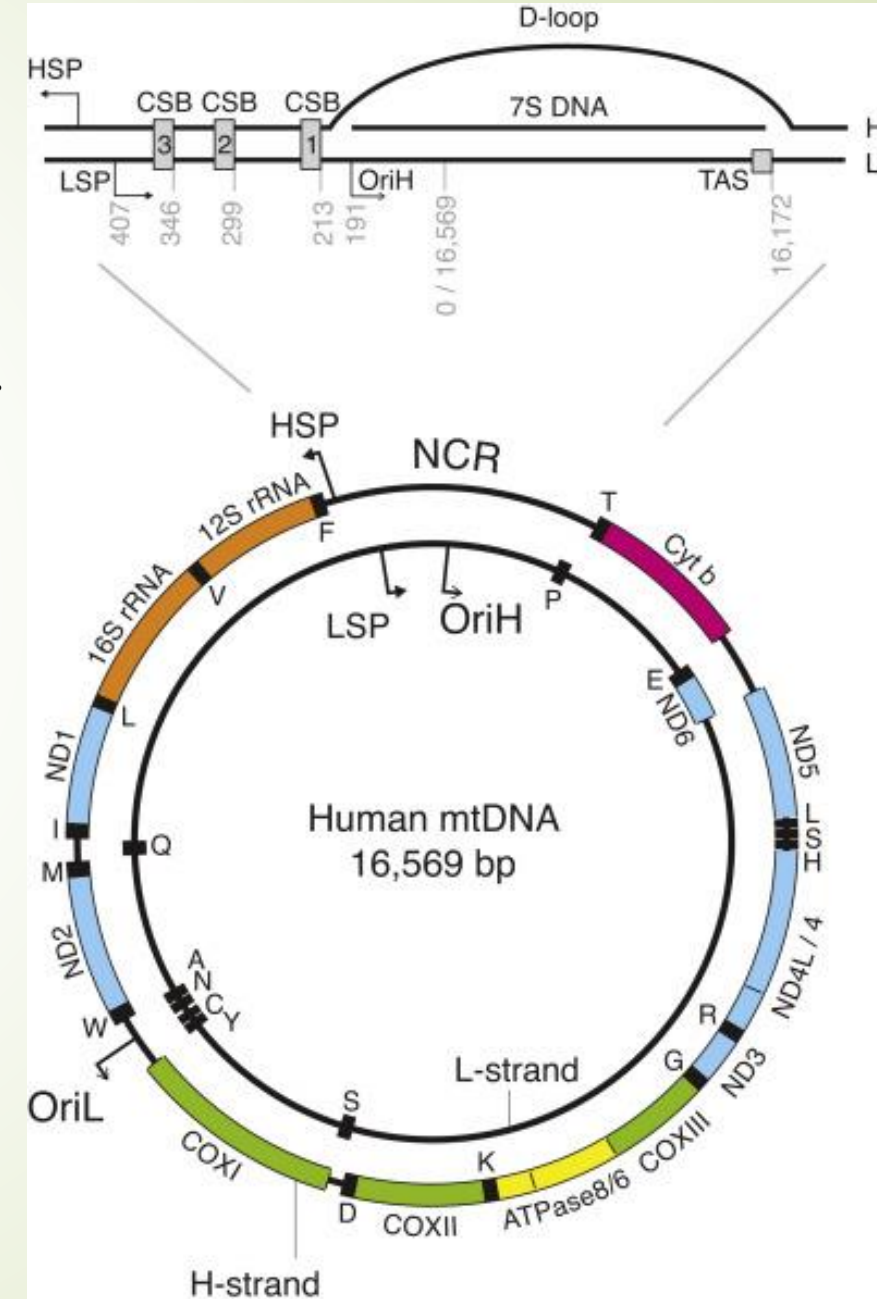
- ✓ Sono più protette dal punto di vista delle modificazioni di sequenza
- ✓ Sono molto idrofobiche e grandi, il loro trasporto all'interno dei mitocondri sarebbe troppo costoso e complicato

Si ipotizza che il genoma nucleare e mitocondriale comunicano in modo da regolare l'espressione delle proteine che portano a formare i complessi della catena respiratoria.



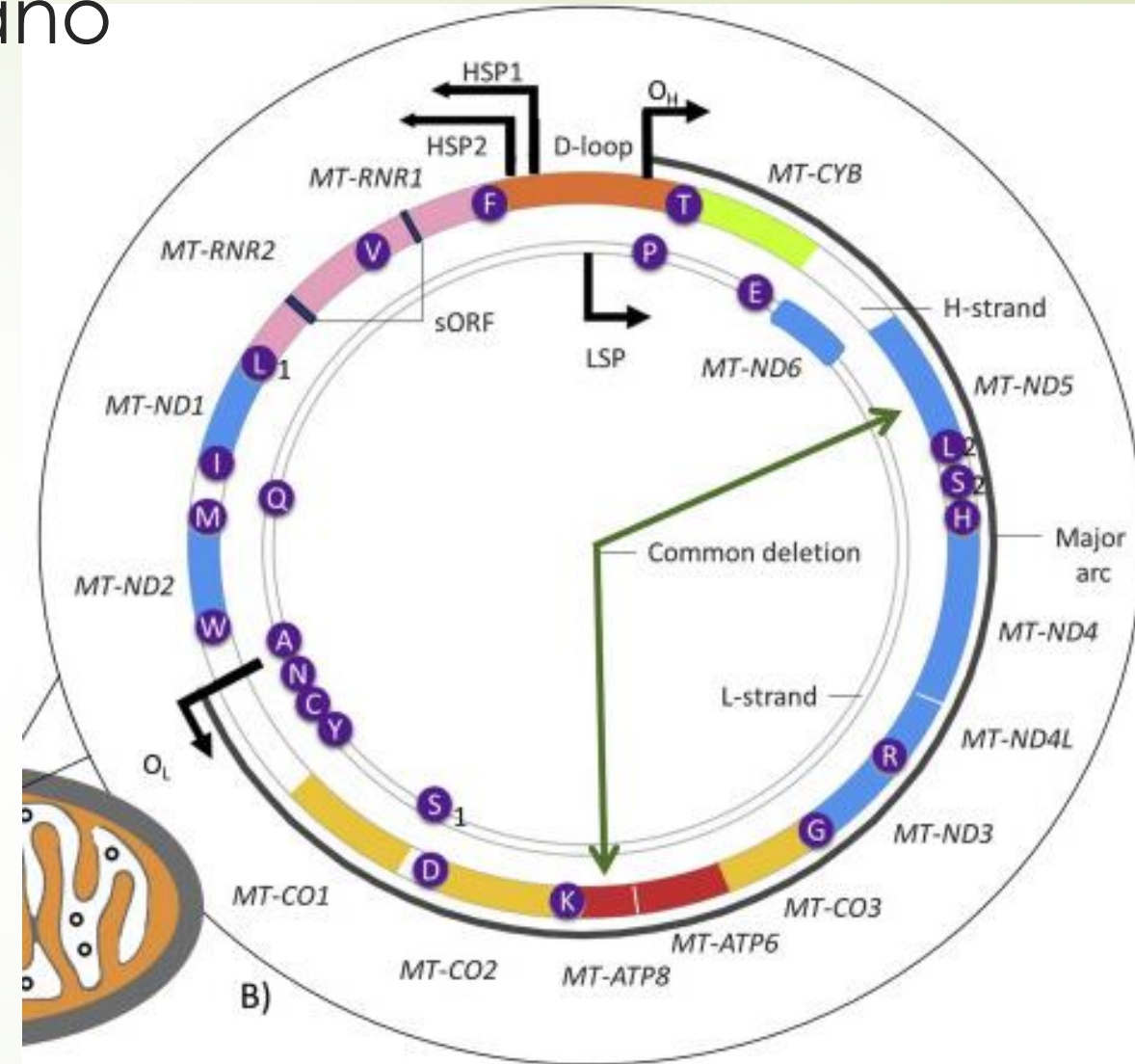
Genoma mitocondriale umano

- Singola molecola di DNA circolare a doppio filamento di 16.5 kbp
- Ogni mitocondrio contiene da 2-10 molecole, per ogni cellula centinaia di molecole di DNA organizzate in nucleoidi.
- Non contiene introni (solo in alcuni organismi presentano introni all'interno di alcuni geni)
- Non contiene regioni non codificanti ad eccezione della **No Coding Region (NCR)** e delle **sequenze intergeniche** di lunghezza non superior ad 8pb
- Codifica per **37 geni**:
 - 13 proteine parte dei complessi I, III, IV e V della catena respiratoria
 - 22 tRNA
 - 2 rRNA



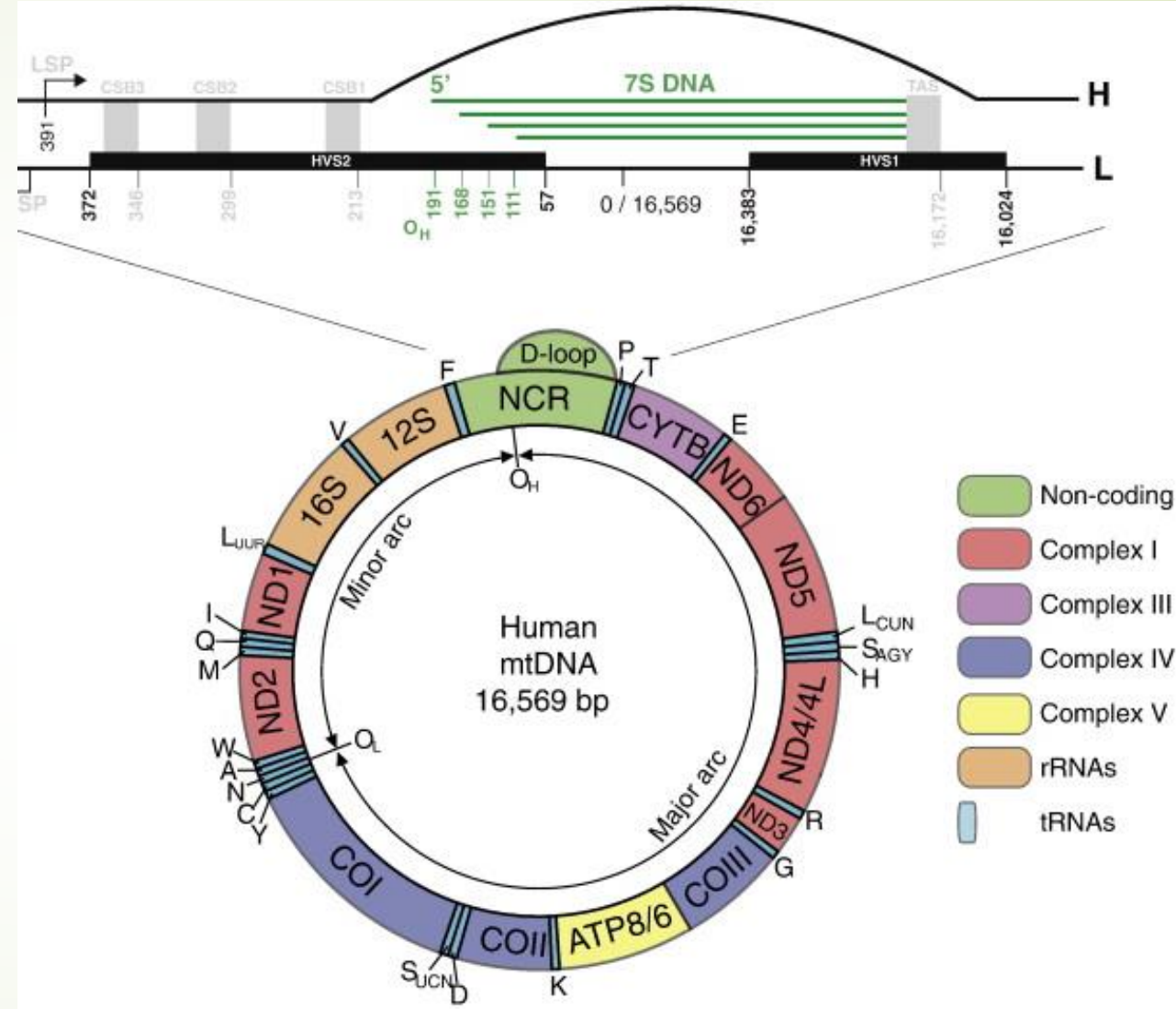
Genoma mitocondriale umano

- ▶ I **due filamenti di DNA (L e H)** differiscono nella composizione in G che conferisce una diversa densità quando separate in un **gradiente di cloruro di cesio**. Per questo motivo ricevono il nome di:
 - ▶ Light strand (L): filamento leggero
 - ▶ Heavy strand (H): filamento pesante (G)
- ▶ L'informazione è distribuita in modo non uniforme tra i due filamenti del DNA: Il filamento L codifica per 8 tRNA e la subunità ND6 del C-I; il filamento H codifica per i restanti geni.
- ▶ La **trascrizione** avviene in forma di **mRNA policistronico** a partire da principalmente 2 promotori a livello della regione non codificante. Un terzo promotore (HSP1) trascrive per 2 tRNA.
 - ▶ LSP (light strand promoter)
 - ▶ HSP1 e HSP2 (heavy strand promoters)



Genoma mitocondriale umano: NCR

- La **regione non codificante** (NCR) del mtDNA si estende 1.1 kbp tra il tRNA-Phe e tRNA-Pro
- Contiene i promotori per la trascrizione della catena pesante (**HSP**) e leggera (**LSP**), e l'origine di replicazione della catena pesante (**OriH**)
- Spesso in cellule metabolicamente attive una parte della regione non codificante si trova in forma di triple elica occupata da una catena di DNA nota come 7S. Questa regione occupata dalla triple elica si chiama **D-Loop** e non sempre occupa tutta la regione NCR
- Normalmente si estende dal OriH fino a sequenze associate alla terminazione della replicazione (TAS).
- La funzione del D-Loop e 7S non è conosciuta ma si ipotizza possa essere il **risultato di una replicazione non portata avanti**.





Genoma mitocondriale

- La traduzione avviene con meno dei 32 tRNAs suggeriti dalla ipotesi del vacillamento. Questo è dovuto a che, per esempio, **usa un unico tRNA per riconoscere tutti i codoni con U in prima posizione. I 22 tRNA codificati dal genoma mitocondriale sono sufficienti per la traduzione delle 13 proteine codificate all'interno del mtDNA.**
- Lo studio delle sequenze proteiche ha rivelato variazioni nel uso di codoni fra diverse specie



Genoma mitocondriale e mutazioni: Patologie mitocondriali

- Le **patologie mitocondriali** possono essere causate tanto **di mutazioni a livello del genoma nucleare come mitocondriale**
- Il sospetto che in qualche modo comunicano viene dal fatto che mutazioni in proteine mitocondriali in uno dei genomi ha una certa influenza su l'altro genoma.
- Il genoma mitocondriale **non è protetto** e ha **un alto tasso di mutazione** dovuto a prossimità alle fonti di produzione di specie reattive dell'ossigeno (ROS) anche se i mitocondri hanno compensato con la presenza di più difese antiossidanti. La probabilità di mutazione può essere più alta semplicemente per il fatto che non ci sono sequenze non codificanti, colpendo esclusivamente geni.
- Se tutti i geni nel mtDNA sono importanti, mutazioni in questi geni hanno di solito delle conseguenze patologiche evidenti più o meno drammatiche.
- Eteroplasmia

Patologie legate al mtDNA

- Mutazioni puntuali
- Delezioni di regioni della molecola
- Deplezione del mtDNA: una riduzione del numero di molecole del DNA



Conseguenza finale:
deficienza bioenergetica più o meno drammatica

Difficile diagnosi
Test genetici per mutazioni

