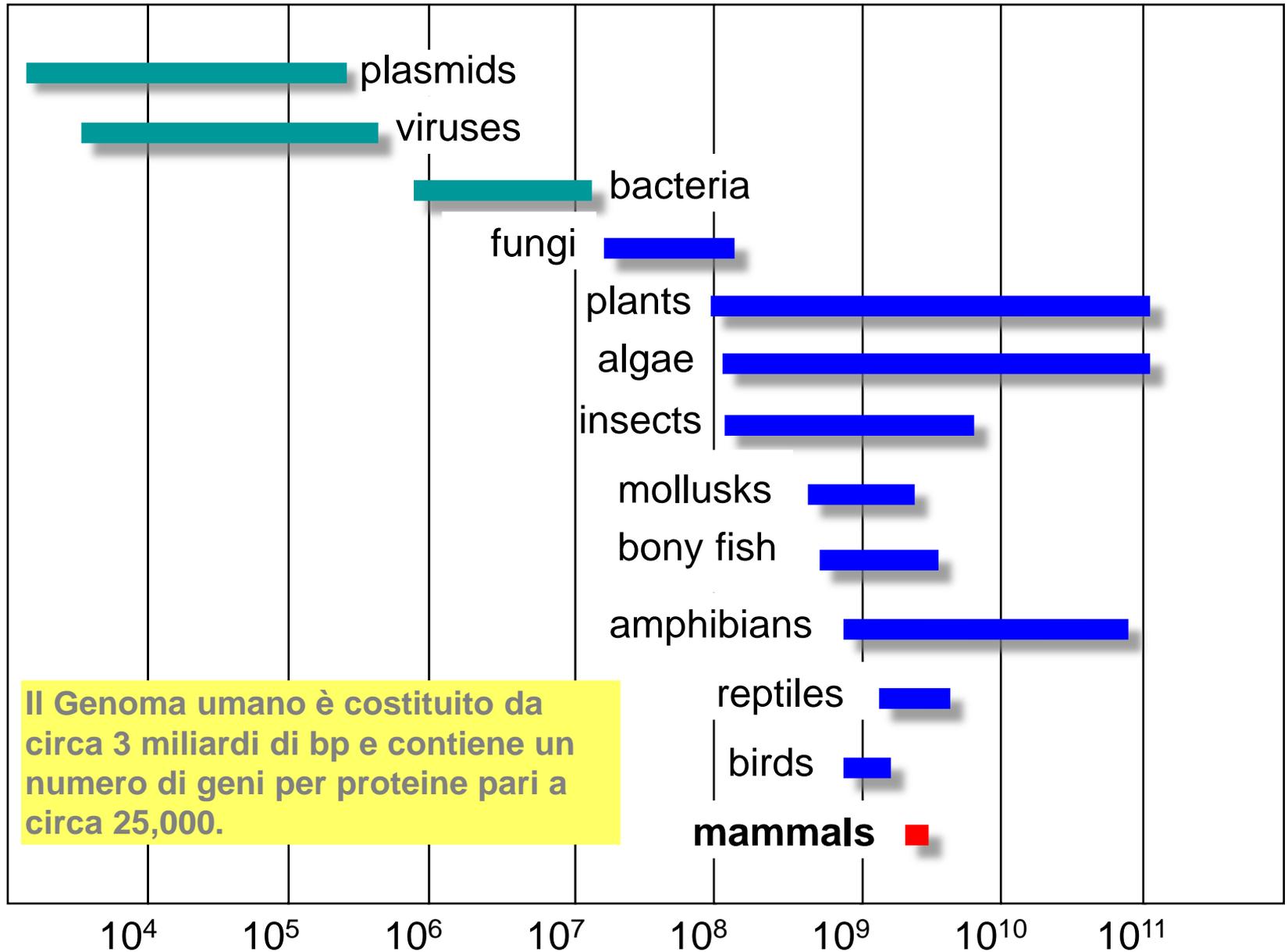


# Organizzazione genoma Umano

# Dimensioni di Genomi



Il contenuto di DNA per cellula aploide è detto valore C e si può esprimere in peso o in coppie di basi

- Il valore C è una costante per ciascuna specie vivente e la caratterizza.

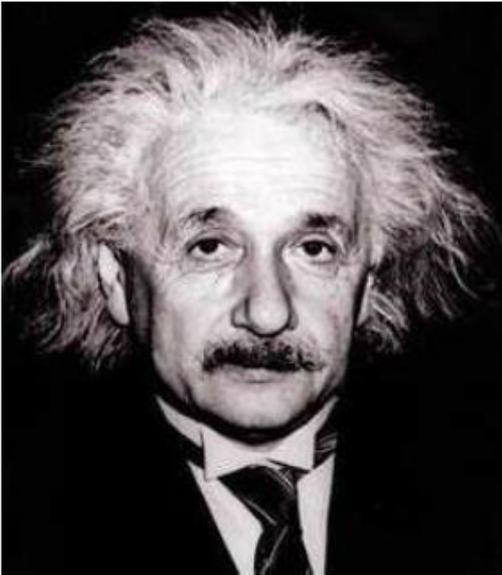
# PARADOSSO DEL VALORE C

NON C'E' CORRELAZIONE TRA LA DIMENSIONE DI UN GENOMA E LA COMPLESSITA' BIOLOGICA DI UN ORGANISMO

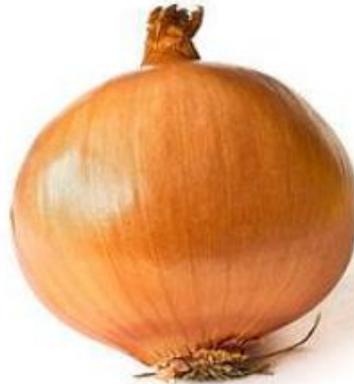
NON C'E' CORRELAZIONE TRA NUMERO DI GENI E LA COMPLESSITA' DI UN ORGANISMO

## Paradosso del valore C:

la complessità di un organismo non correla con il contenuto di DNA del suo genoma (1)



$3.4 \times 10^9$  bp  
*Homo sapiens*

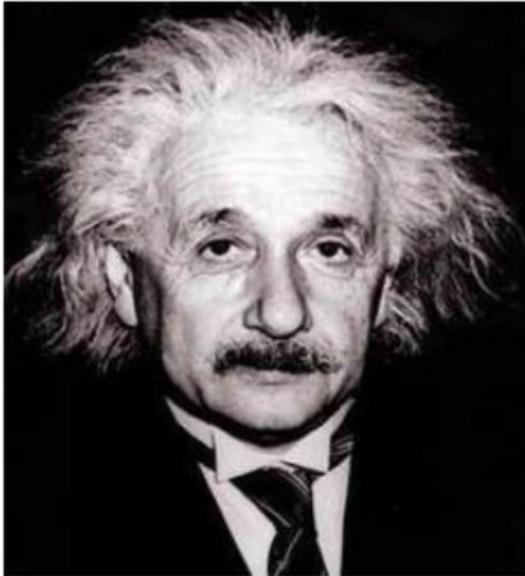


$1.5 \times 10^{10}$  bp  
*Allium cepa*



$6.8 \times 10^{11}$  bp  
*Amoeba dubia*

non c'è correlazione tra numero di geni e complessità di un organismo (2)



~21000 geni



~60000 geni

# Sequence complexity is not the same as length

- **Complexity** is the number of base pairs of unique, i.e. nonrepeating DNA.
- Example: consider 1000 bp DNA.
  - 500 bp is **sequence a**, present in a single copy.
  - 500 bp is **sequence b** (100 bp) repeated **5X**

a                    b   b   b   b   b



$$L = \text{length} = 1000 \text{ bp} = a + 5b$$

$$N = \text{complexity} = 600 \text{ bp} = a + b$$

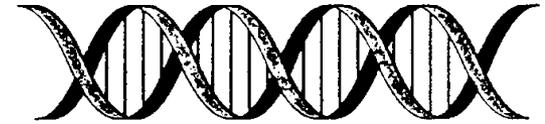
Lo studio della cinetica di riassociazione del DNA fornisce indicazioni su:

Dimensioni genoma

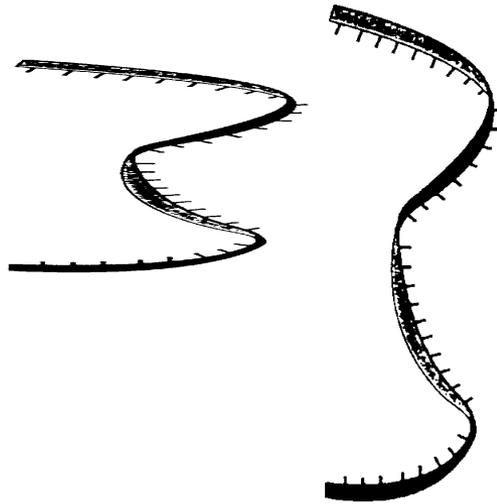
Complessità del genoma

# DNA RIASSOCIAZIONE (RINATURAZIONE)

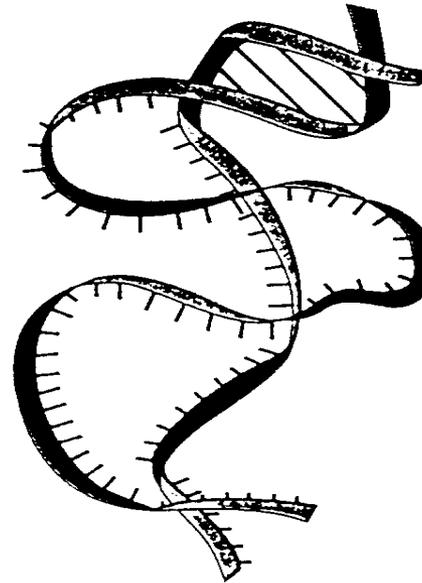
DNA a doppio-filamento



DNA denaturato in  
singolo-filamento



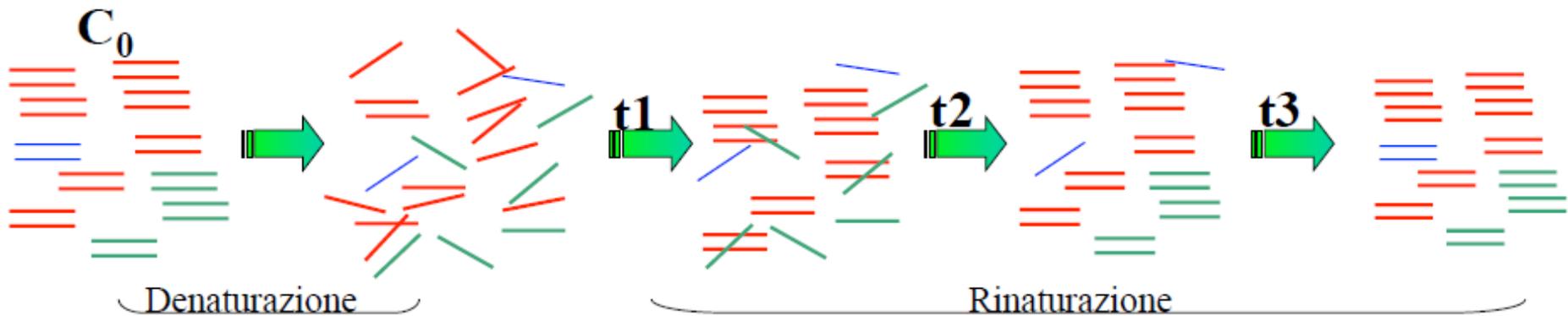
Più veloce,  
reazione  
“zippering” per  
formare lunghe  
molecole di  
DNA a doppio  
filamento



Fase lenta, processo di  
secondo-ordine a  
velocità limitante,  
per trovare sequenze di  
basi complementari



# Cinetica di riassociazione-complexità del genoma



Sequenze presenti in piu' copie (**rosse**, **verdi**)  
riassociano prima rispetto a sequenze presenti in  
**singola copia** (blu).

# Fattori che influenzano la riassociazione

**Concentrazione:** la probabilita' di incontro di due sequenze complementari aumenta con la **concentrazione** del DNA

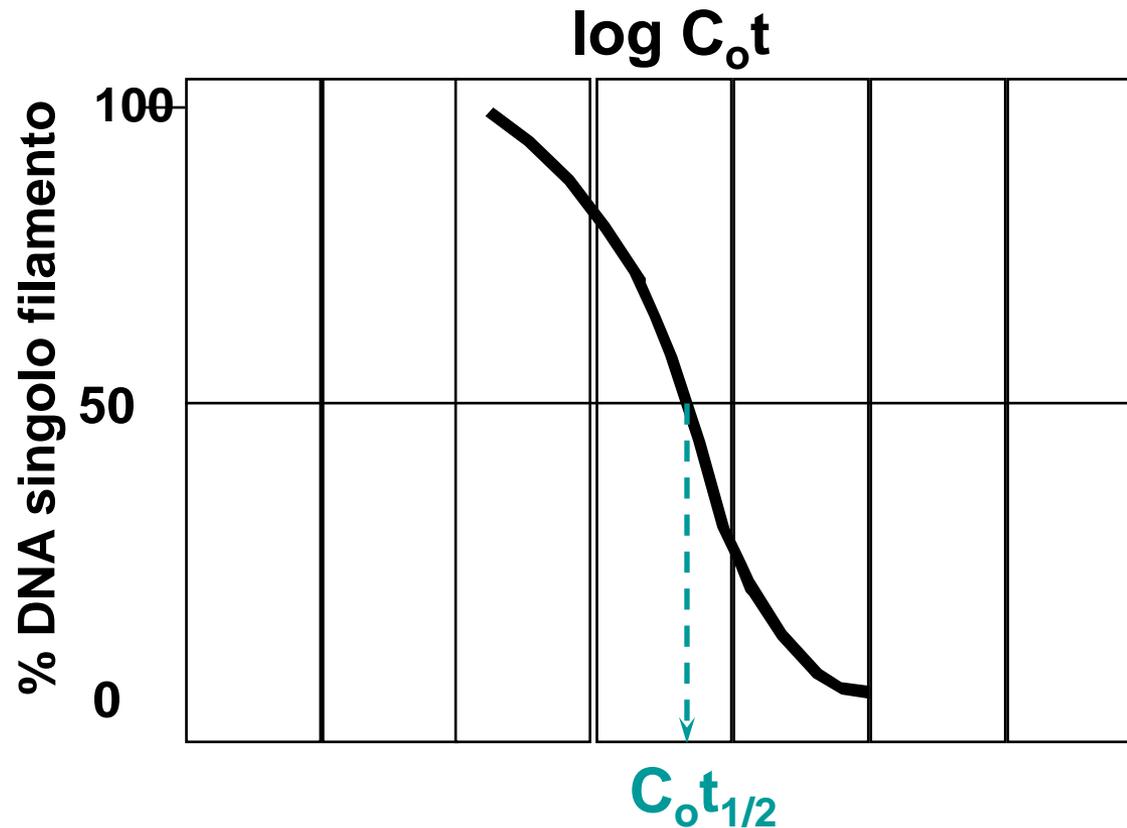
**Tempo:** piu' **il tempo** di reazione e' lungo, piu' e' alta la probabilita' che le sequenze complementari si incontrino.

Queste relazioni vengono descritte nella curva **Cot**

# Tipica cinetica di riassociazione di un DNA che contiene sequenze uniche non ripetute

$C_0 t_{1/2} = C_0$  = concentrazione di DNA

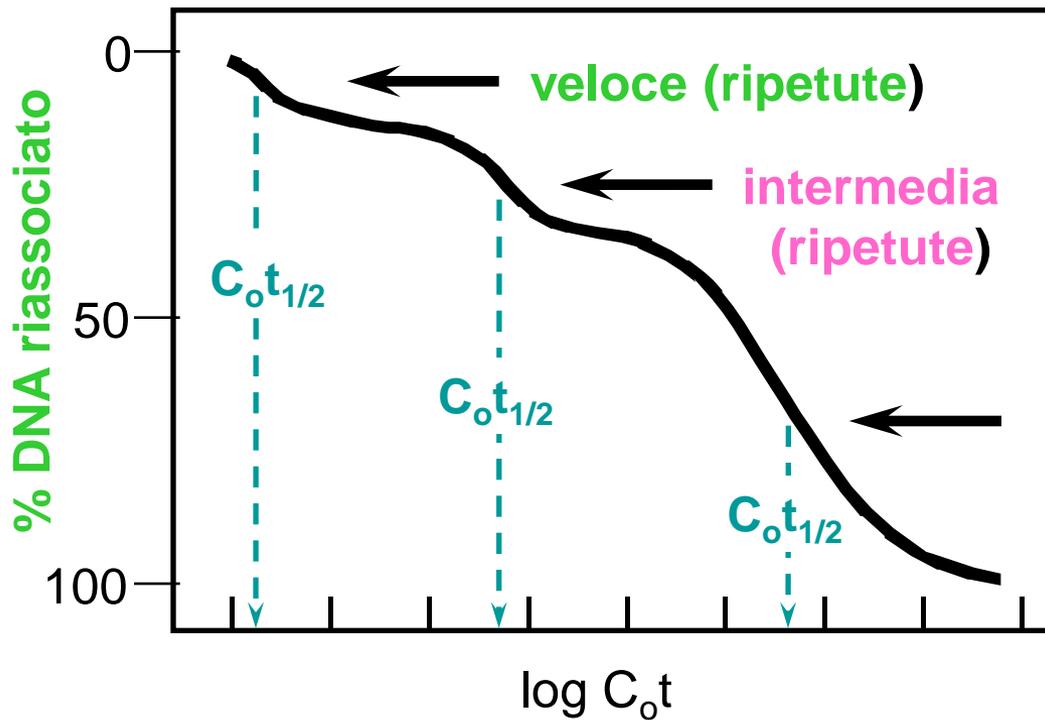
$t_{1/2}$  = tempo di metà reazione



A metà della reazione (cioè a  $t_{1/2}$ ) il 50% è riassociato.

# Cinetiche di riassociazione del DNA genomico umano

$C_0 t_{1/2}$   $C_0$  = concentrazione DNA  
 $t_{1/2}$  = tempo di metà reazione



Frazioni cinetiche:

veloce

intermedia

lenta

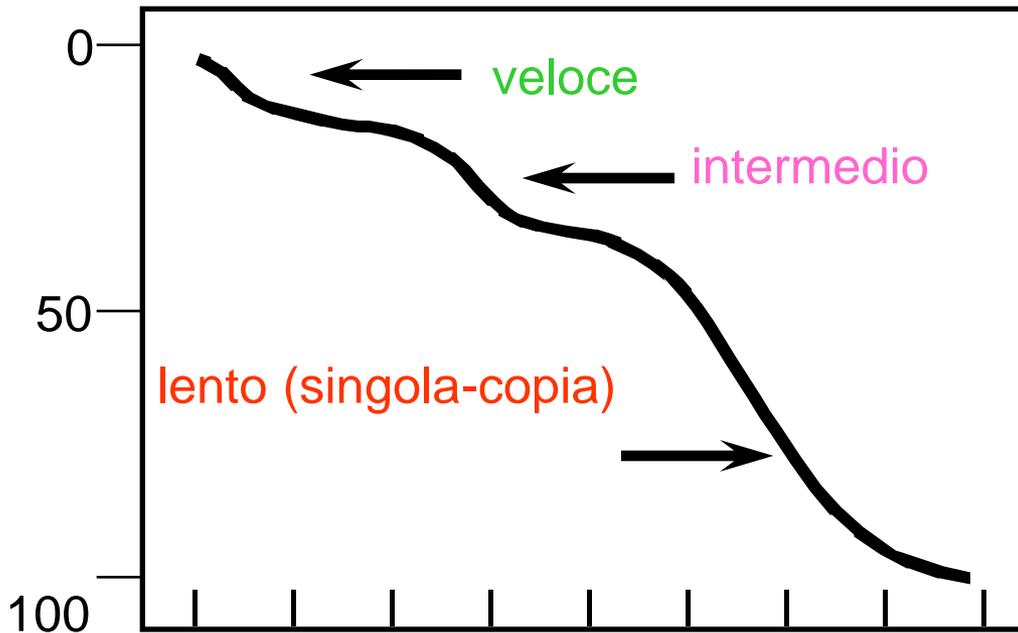
lenta (singola-copia)

Singola-copia (unica o poche copie) <50% Include la maggior parte dei geni

Sequenze moderatamente ripetute

Sequenze altamente ripetute

>50%



# Proprietà del genoma umano

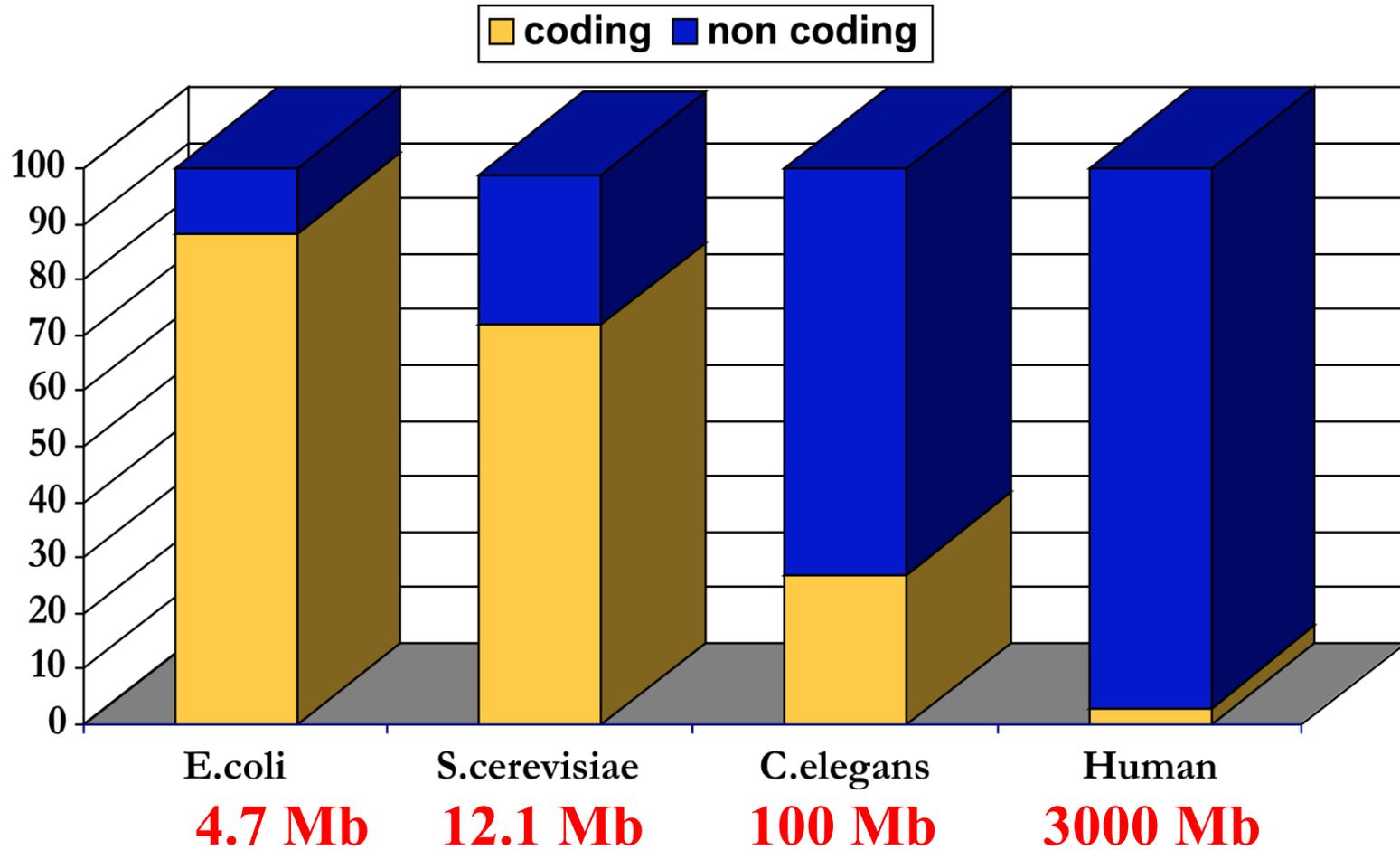
## Nucleare

- il genoma umano aploide ha  $\sim 3 \times 10^9$  bp di DNA
- il genoma umano contiene  $\sim 20000$ - $25000$  geni
- 

## Mitocondriale

- genoma circolare di  $\sim 17,000$  bp
- contiene 37 geni: 13 geni per proteine
- 22 geni per tRNA
- 2 geni rRNA
- i geni sono privi di introni.
-

# Porzioni codificanti e non codificanti dei genomi eucariotici



La **porzione codificante proteine** rappresenta **l'1-2% del genoma umano**

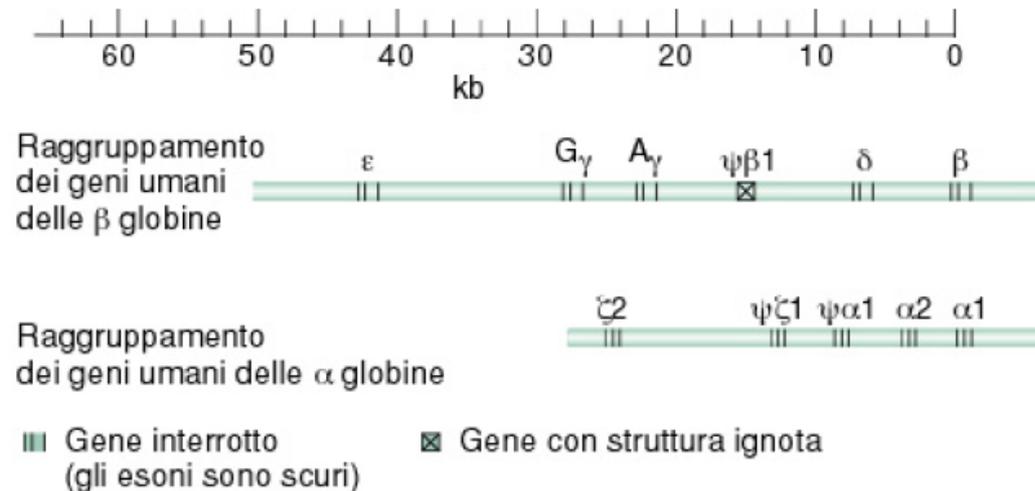
# Porzione codificante del genoma umano

La porzione codificante proteine rappresenta l'1-2% del genoma

- **sequenze uniche**, in singola copia o numero molto limitato di copie
- **sequenze moderatamente ripetitive**, es geni per gli istoni

## Esempi di famiglie geniche

### Le due famiglie dei geni per le globine $\alpha$ e $\beta$



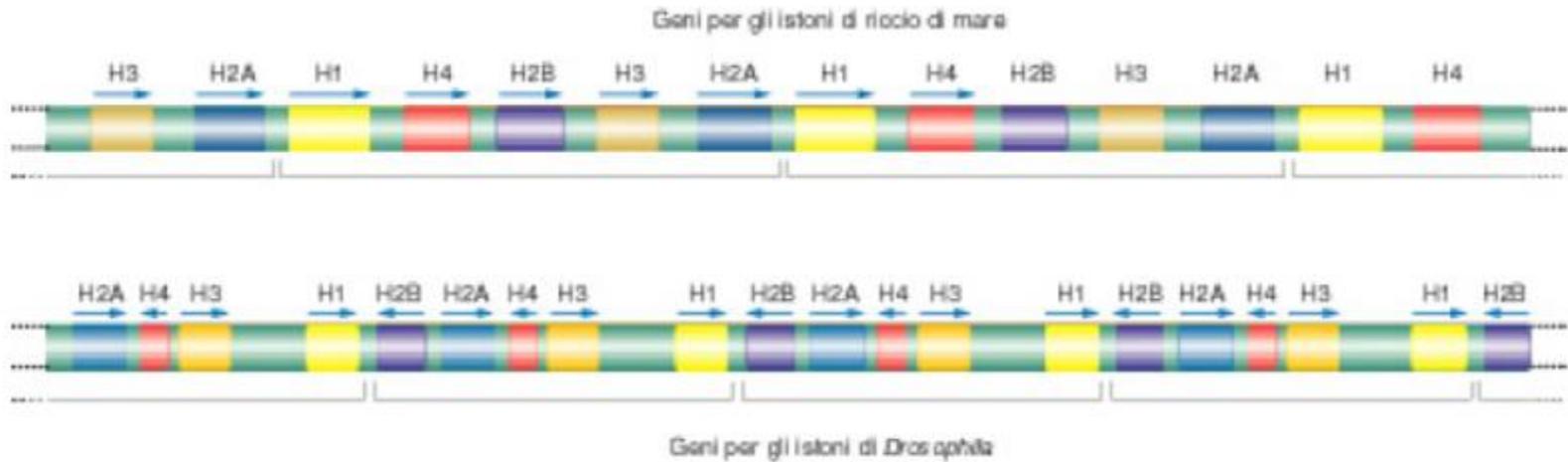
Altri esempi di famiglie geniche sono:

I geni delle actine (da 5 a 30)

I geni delle cheratine (circa 20)

I geni della regione variabile delle immunoglobuline (500)

## Sequenze moderatamente ripetitive



I geni delle proteine istoniche sono presenti nel genoma in centinaia di copie ripetute in tandem

# La struttura dei geni eucariotici

I geni eucariotici presentano una grande varietà di strutture e dimensioni.

Ad esempio nel **genoma umano**:

**Il più piccolo:**

tRNA<sup>GLU</sup> (69 bp)

**Il più grande:**

Distrofina (2.4 Mb, la sua trascrizione richiede circa 16h)

**Il numero di esoni può variare** da 1 (geni privi di introni come molti geni per ncRNA, interferoni, istoni, ribonucleasi, ecc.) sino a **363** (**Titina**).

Le **dimensioni degli esoni e degli introni** sono estremamente variabili. A fronte di esoni costituiti da pochi nucleotidi, l'esone più grande è presente nel gene per **ApoB** (7.6 kb). Anche le dimensioni degli introni possono variare da pochi nucleotidi fino a 800 kbp (gene **WWOX**).

**Numero di aa** Le proteine possono contenere pochi aa (piccoli ormoni) sino a molte migliaia (**Titina**, 38.138 aa).

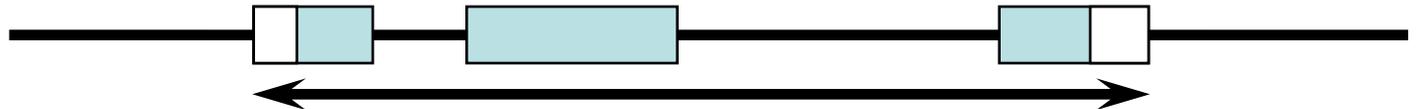
(esone-introne-esone)<sub>n</sub> struttura di diversi geni

istone



**totale = 400 bp; 1 solo esone = 400 bp**

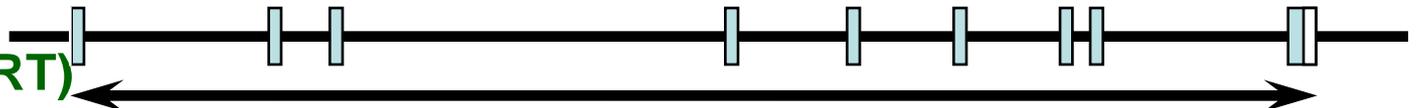
$\beta$ -globina



**totale = 1,660 bp; 3 esoni = 990 bp; 2 introni**

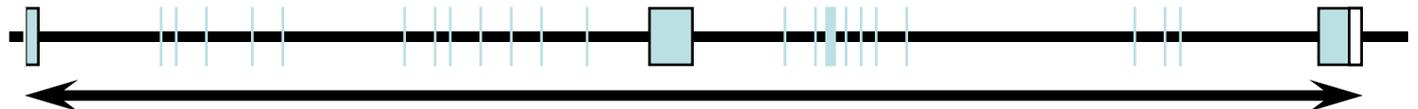
HGPRT (HPRT)

ipoxantina-guanina  
fosforibosil  
transferasi



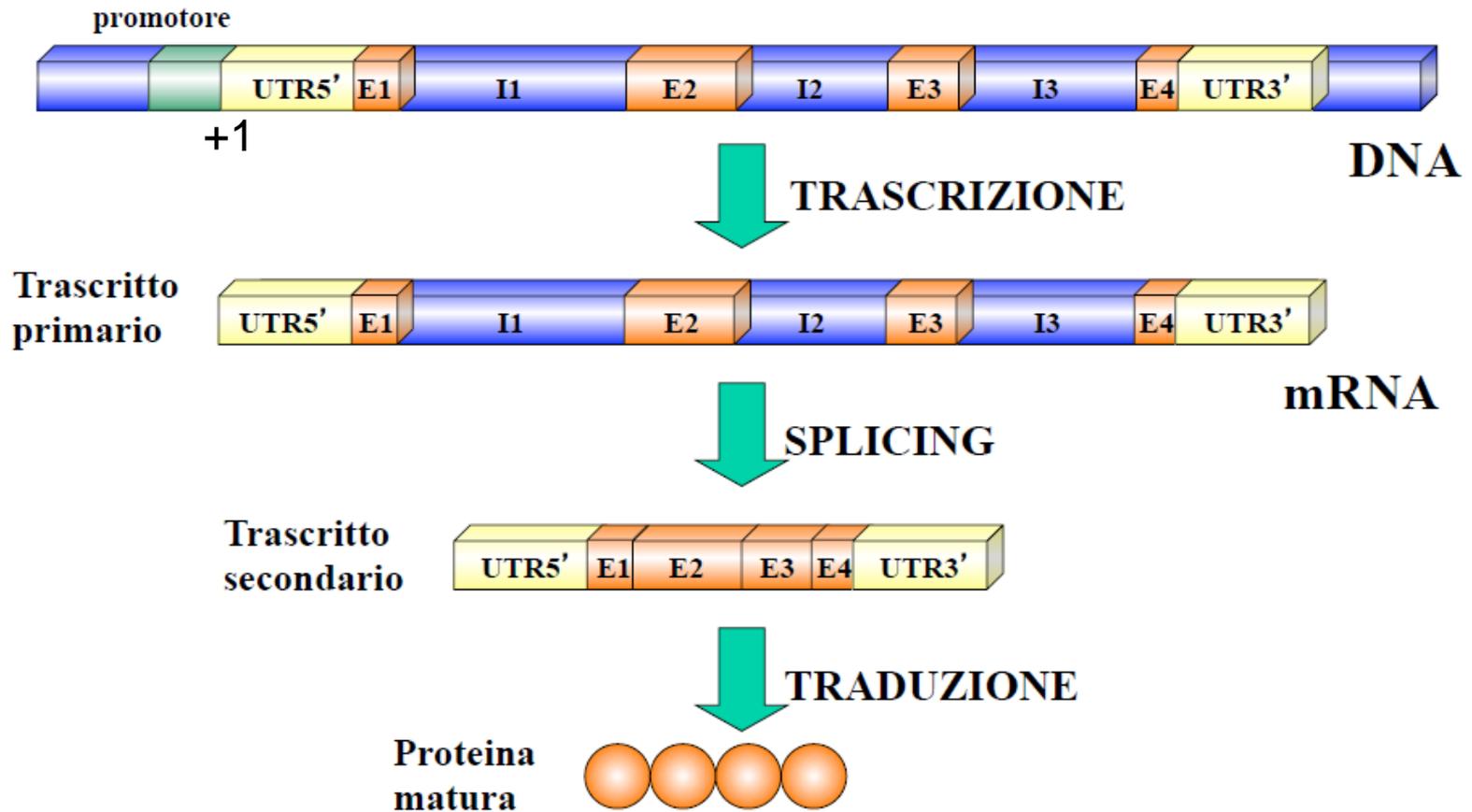
**totale = 42,830 bp; 9 esoni = 1263 bp**

factor VIII



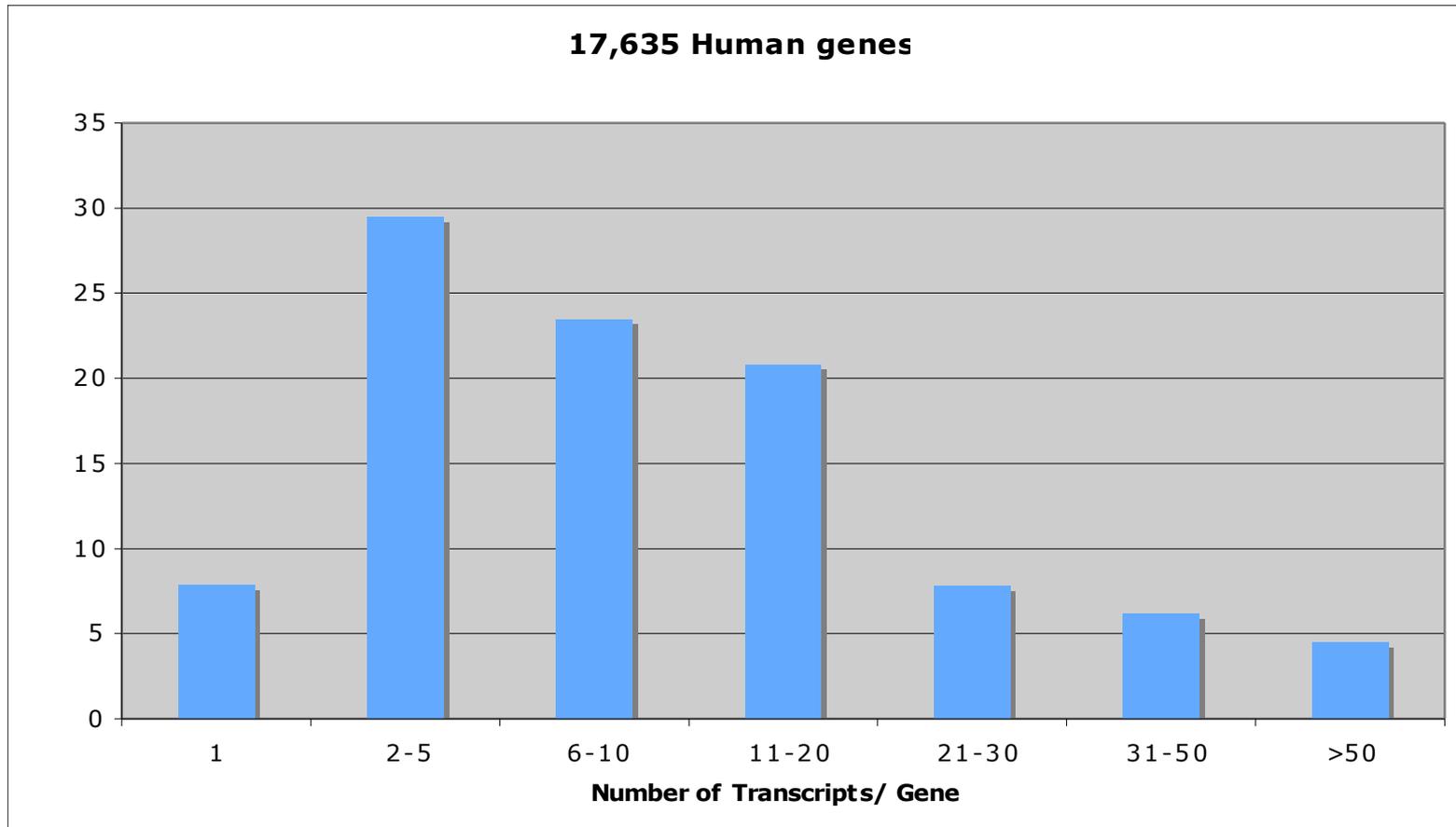
**totale = ~186,000 bp; molti esoni = ~9,000 bp**

# Struttura del Gene



# Splicing Alternativo

Oltre il 90% dei geni umani è in grado di esprimere più di un trascritto (ed è quindi soggetto a **splicing alternativo**). Le diverse **isoforme di splicing** possono avere specificità a livello di tessuto, di condizione fisiologica o patologica.



(Pesole et al., dati non pubblicati)

# Genoma nucleare umano

- 23 paia di cromosomi
- 3.000.000.000 paia di basi, circa 20000-25000 geni

La **porzione codificante** rappresenta l'**1-2% del genoma** (geni codificanti per proteine).

La **porzione non-codificante** è costituita da **sequenze uniche e sequenze ripetute**.

Piu' del 50% del genoma e' costituito da **sequenze ripetute** con funzione sconosciuta

1) **ripetizioni intersperse, disperse** (LINEs, SINEs, trasposoni a DNA)

2) **ripetizioni in tandem** (blocchi ripetuti in tandem dei centromeri e dei telomeri, micro- e mini satelliti).

Alcuni cromosomi contengono anche il 90% di DNA **non codificante**

# 1) SEQUENZE RIPETUTE INTERSPERSE (DISPERSE)

Sequenze moderatamente ripetitive



**Trasposoni o elementi genetici mobili:** sequenze in grado di spostarsi all'interno di un genoma, producendo molte copie di se stesse

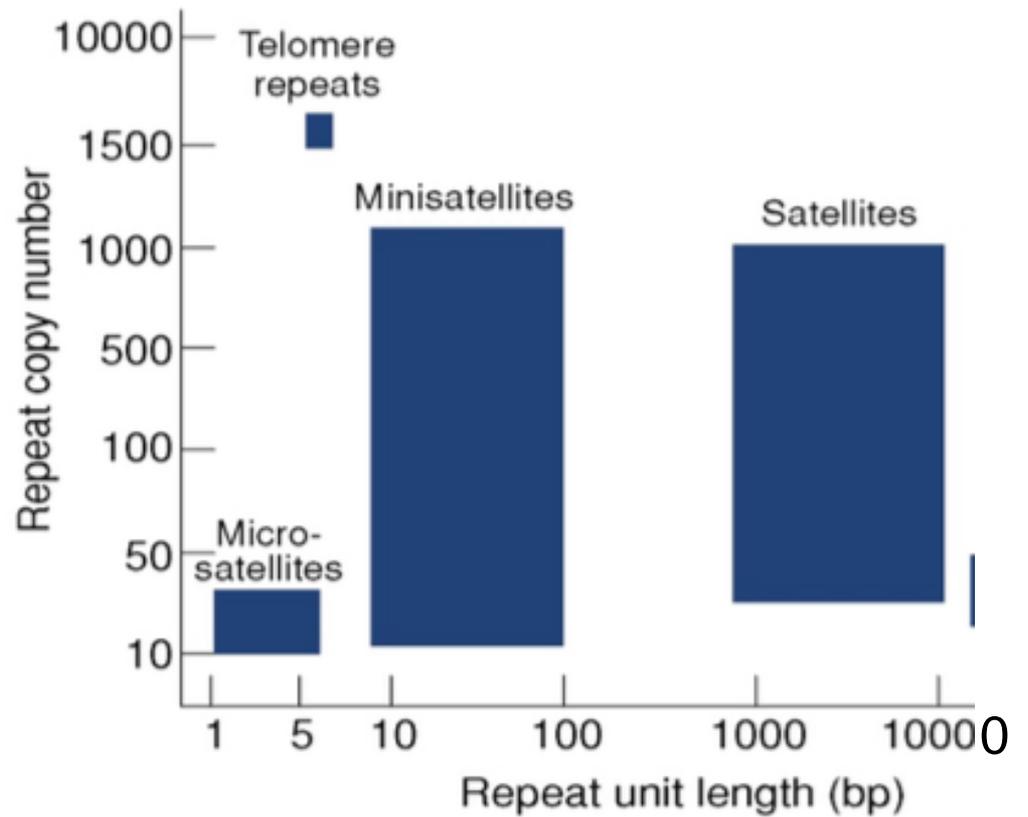
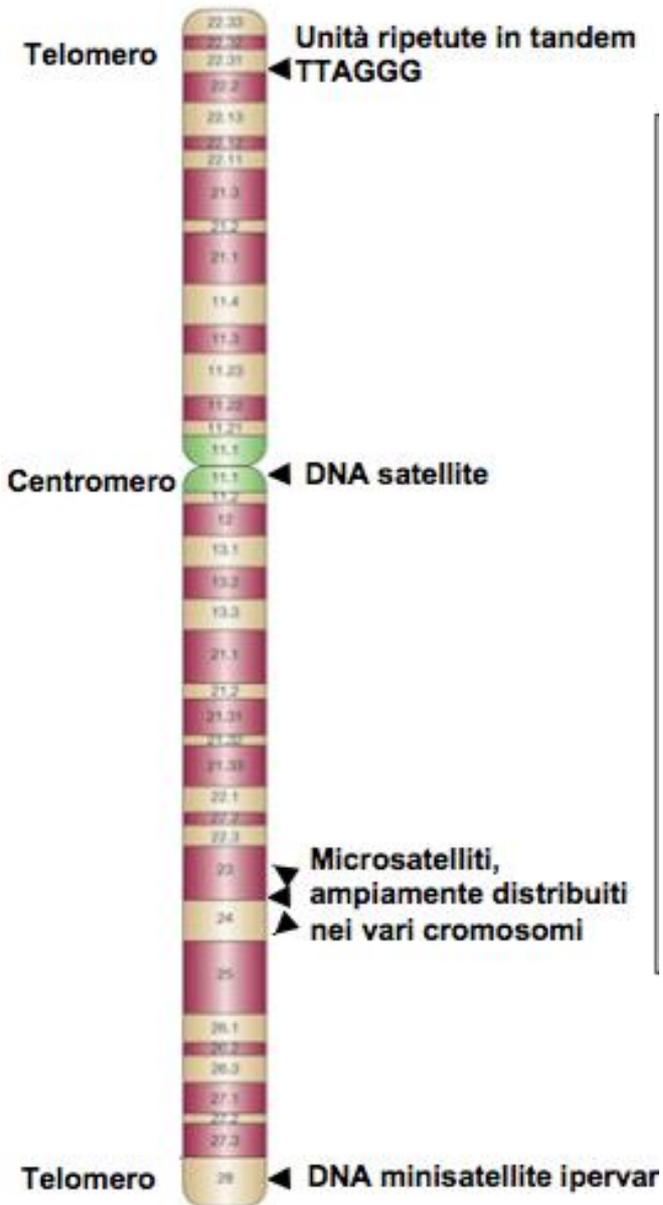
-Trasposoni a DNA

-Retrotrasposoni: sono elementi trasposti per mezzo di un intermedio di RNA

-**LINE:** Long Interspersed Elements. Lunghezza sequenza circa 6000-8000 pb, numero di copie circa 500000

- **SINE:** Short Interspersed Elements. Lunghezza circa 100-300 pb, numero di copie circa 1500000 (sequenze Alu)

## 2) RIPETIZIONI IN TANDEM



# SEQUENZE RIPETUTE IN TANDEM

## sequenze altamente ripetute

### DNA minisatellite: VNTR Variable Number Tandem Repeats

Lunghezza della sequenza ripetuta: 5-100 bp

Numero di ripetizioni: da decine a migliaia di copie

Frequenza nel genoma: circa 100 loci in tutto il genoma umano

I minisatelliti tendono ad essere instabili, **il numero di copie può variare**, l'alta variabilità fa che sia altissima la probabilità che singoli individui abbiano alleli diversi (**numero diverso di ripetizioni**) ad un dato locus. Poiché **polimorfici** sono utilizzati per il **DNA fingerprinting**

### DNA microsatellite: STR Short Tandem Repeats

Lunghezza della sequenza ripetuta : 1-4 bp

Numero di ripetizioni : 10-20

Frequenza nel genoma : circa 100mila loci, uno ogni 30000 bp

Nella maggior parte dei casi i loci microsatelliti sono al di fuori dei geni. Espansioni di triplette codificanti sono responsabili di alcune patologie (es Distrofia Miotonica, corea di Huntington)

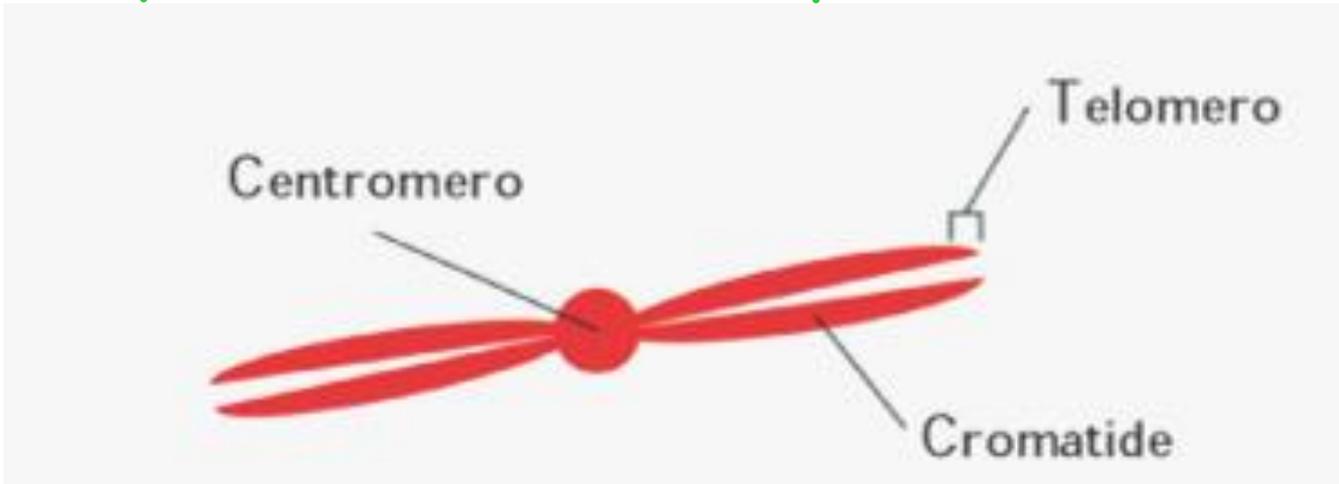
## Sequenze altamente ripetute in tandem

Il **DNA telomerico** consiste di alcune migliaia di copie di un motivo ripetuto in tandem, ricco in T e G, che nell'uomo e in altri vertebrati è 5'-TTAGGG-3'.

Il DNA telomerico identifica le estremità dei cromosomi, impedisce l'accorciamento dei cromosomi ad ogni evento di replicazione e la fusione dei cromosomi



## Sequenze altamente ripetute in tandem



Nel cromosoma in metafase è visibile una specie di strozzatura (il centromero) contenente sequenze altamente ripetute e una struttura proteica che funge da punto di attacco dei microtubuli del fuso durante mitosi e meiosi.

Nei primati (uomo compreso), il **centromero** è costituito da una sequenza di 171 pb ripetuta in tandem circa 10000 volte

## Cosa ancora non sappiamo.....

- Esatto numero dei geni, localizzazione e funzione
- Regolazione dei geni
- Tipi di DNA non-codificante, distribuzione, che tipo di informazione contengono e relativa funzione
- Struttura e funzione di molte proteine
- Correlazione tra variazioni di sequenza tra singoli individui con la salute e la malattia (predizione della suscettibilita' a specifiche malattie a secondo della variabilita' individuale)
- Geni coinvolti nelle malattie complesse