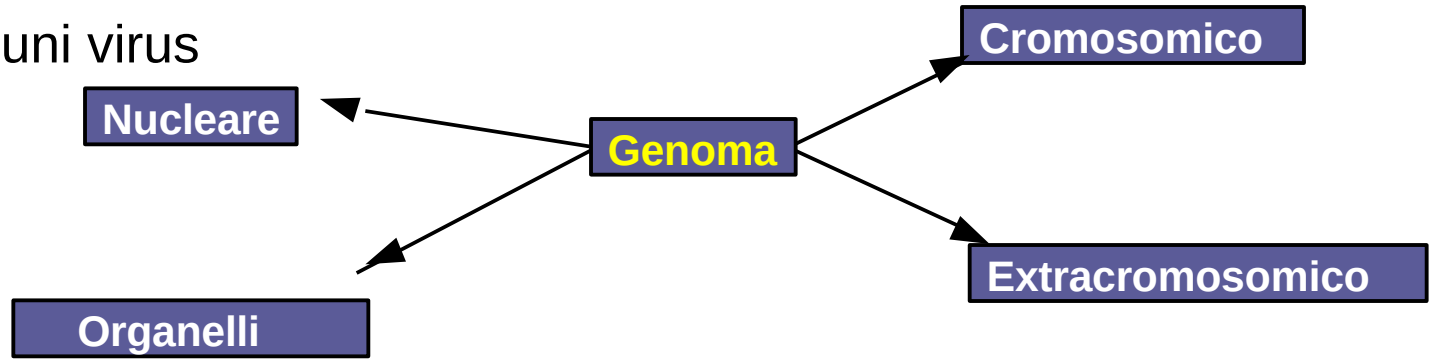
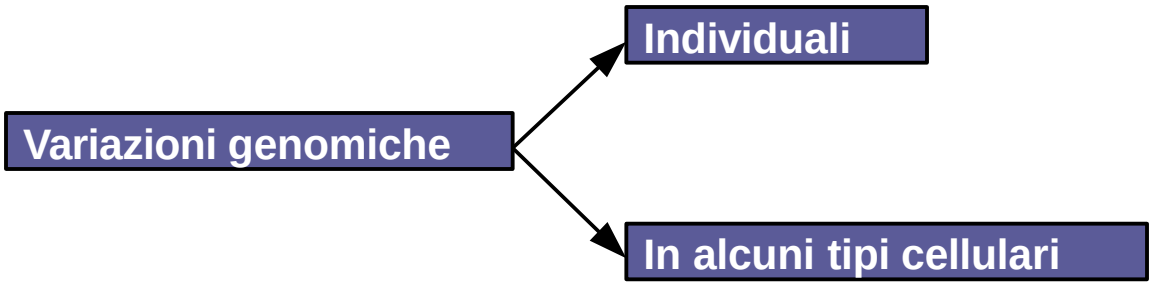
21.
L’organizzazione del genoma:
Genomica

Cos'e un Genoma?

- Il genoma il materiale genetico completo di un organismo, ovvero **tutto il DNA contenuto del nucleo di ogni cellula**:
- **insieme di tutte le informazioni biologiche necessarie alla costruzione e al mantenimento di un organismo vivente**.

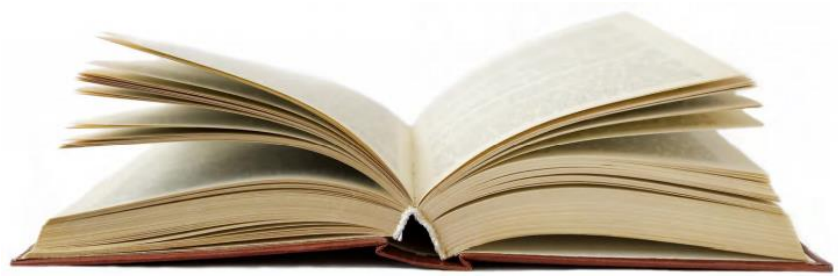
- **Genomi a DNA**
  - Organismi unicellulari e pluricellulari
- **Genomi a RNA**
  - Alcuni virus





Genoma: il libro della vita

- Ogni genoma **contiene tutte le informazioni** necessarie per
  - costruire e mantenere quell'organismo,
  - è quindi una sorta di **manuale di istruzioni**



- in cui sono contenute tutte le informazioni che regolano lo sviluppo ed il funzionamento di un individuo



L’avvento dell’era Genomica

- Gli sviluppi delle tecniche di indagine bio-molecolare rendono oggi possibile acquisire le informazioni relative al flusso dell’informazione genetica a ritmi fino a poco fa inimmaginabili (es. determinazione delle sequenze di DNA, studio del profilo di trascrizione, determinazione della struttura delle proteine, ecc.)
- Questo ha prodotto un drastico cambio di prospettiva e di orizzonti nella ricerca biologica:

  - Genoma: 1920 Hans Winkler, Botanico
    - **gen "gene" + (chromos)om "chromosome**”
    - Si riferiva all’insieme dei cromosomi, intesi come vettori dei fattori ereditari.

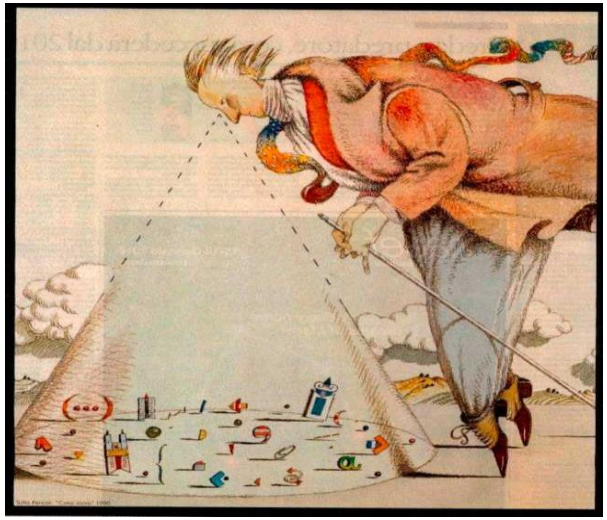
  - **Genomica: 1**979 Victor McKusic, Genetista

Dal G E N E al G E N O M A

Il complesso delle unità elementari di un dato sistema (es. rizoma)

OMICS

- Lo scopo di tale approccio olistico (o sistemico) è quello di poter **comprendere**, operando con approcci integrativi, **un siste-ma complesso**
- (ovvero un gruppo di entità, interconnesse, che formano un insieme integrato),
- **considerato come l'insieme di molecole biologiche** che lo compongono (i.e. geni, trascritti, proteine, metaboliti)



OMICS

- L’avvento delle “omics” coincide con uno
- spettacolare miglioramento tecnologico in diversi ambiti

- **biologia molecolare**



- **Informatica**



- **Internet**



Genomica vs. Genetica

- **Genetica:** scienza che studia i geni, l'ereditarietà e la variabilità degli organismi
- **Genomica:** disciplina che studia i profili genetici su ampia scala per il genoma di una data specie

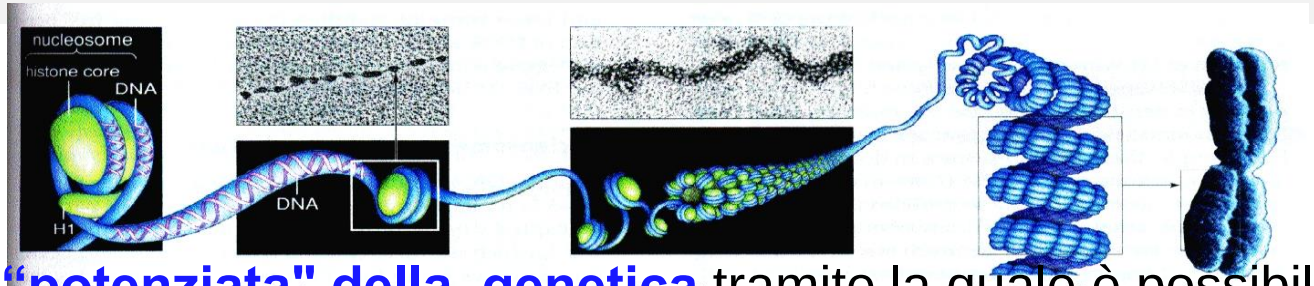
Quindi:

- Mentre
- **la genetica** guarda ai singoli geni, uno alla volta
- la **genomica** cerca di fotografare l'immagine complessiva, esaminando l'insieme dei geni (e delle caratteristiche genomiche) **come un sistema completo**

COS'E' LA GENOMICA?

↳ Lo **studio sistematico dei genomi**, a partire dalla determinazione delle sequenze nucleotidiche del/dei cromosoma/i, che rappresenta la quantità totale di una cellula o di un organismo di un organismo

G - A - T - C ←



- Una **versione "potenziata" della genetica** tramite la quale è possibile studiare, simultaneamente, tutti i geni di un organismo vivente allo stesso momento. (NIGMS, NIH)

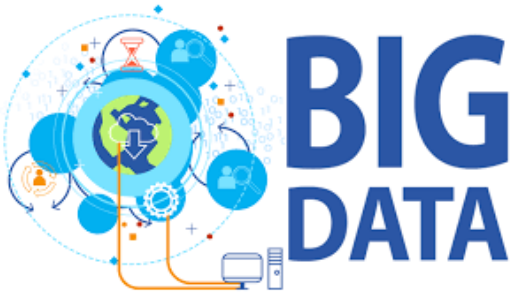
- La conoscenza del genoma risulta quindi **utile in ogni campo della biologia**
- l'avvento di metodi per il sequenziamento del DNA ha accelerato significativamente la ricerca.
- In medicina, ad esempio, il sequenziamento è usato per
  - identificare e diagnosticare malattie ereditarie
  - sviluppare nuovi trattamenti.
- In modo simile, il genoma degli agenti patogeni può portare allo sviluppo di medicine

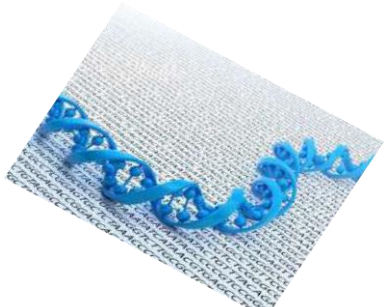
Genomica e bioinformatica

- L'introduzione dei **computer e metodi computazionali** nei laboratori di biologia molecolare è **uno dei fattori chiave nello sviluppo della genomica.**



- Lo sviluppo dell'automazione in laboratorio e di tecnologie di analisi in grado di produrre grandi quantità di dati da un singolo esperimento ha dato spinta allo **sviluppo della Bioinformatica.**





Sviluppi correlati

- **Bioinformatica e biologia computazionale** implicano l'uso di tecniche che includono:

- Matematica applicata,
- informatica,
- statistica,
- scienza dei computer,
- intelligenza artificiale,
- chimica e biochimica,

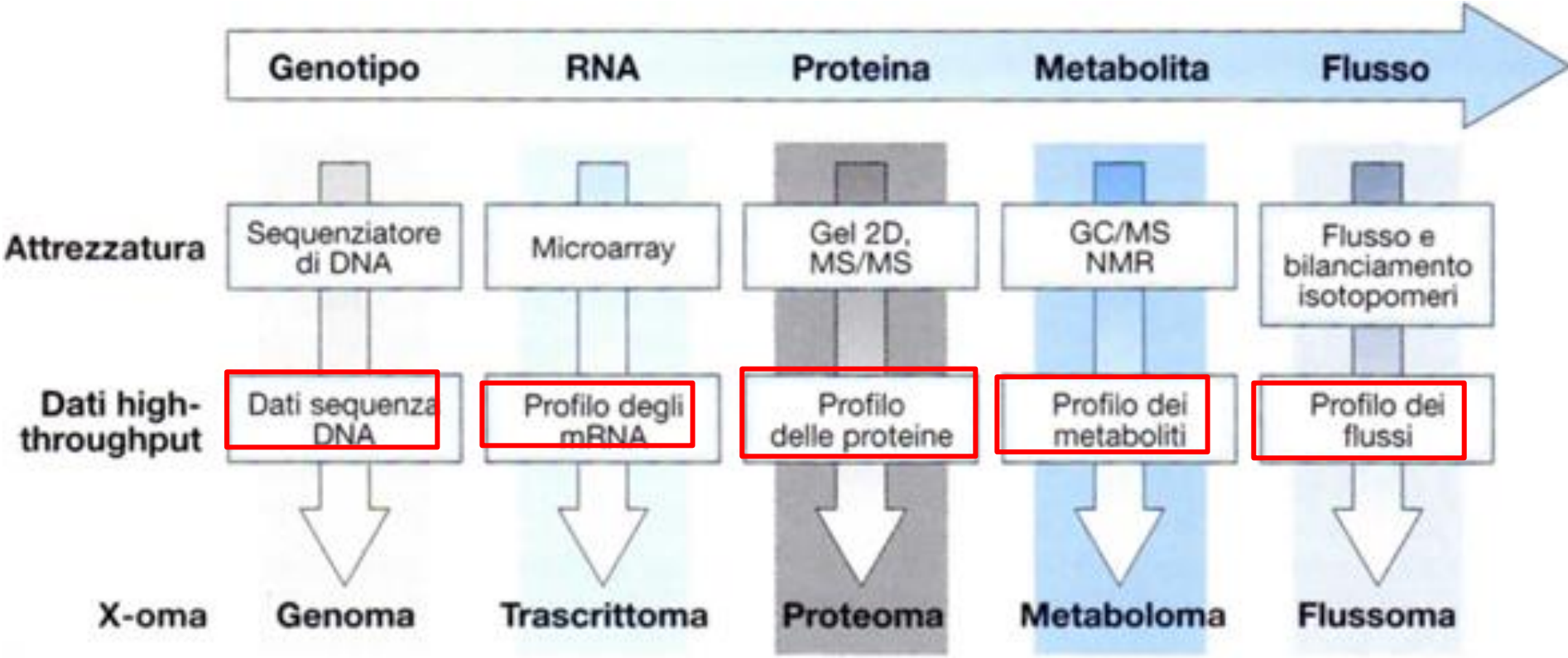
✓ **Questo al fine di risolvere problemi biologici generalmente a livello molecolare.**

Genomica: Obiettivi

- Compilare le sequenze genomiche degli organismi
- Stabilire la **localizzazione** di tutti i geni di un genoma
- Annotare l'insieme dei geni di un genoma
- **Stabilire la funzione** di tutti i geni di un genoma
- Generare profili di espressione genica per cellule sottoposte a condizioni diverse
- Identificare tutte le proteine che possono essere prodotte da un determinato genoma
- **Confrontare i geni e le proteine** tra gli organismi per stabilire **relazioni evolutive**

I metodi "omici" (system biotechnology)

- La domanda chiave circa il genoma è **quanti geni contiene** .
- Possiamo pensare circa il numero totale di geni diversi livelli, corrispondenti a successivi stadi di espressione genica:



GC: Gas Cromatografia
MS: Spettrometria Massa

Le tecnologie omiche

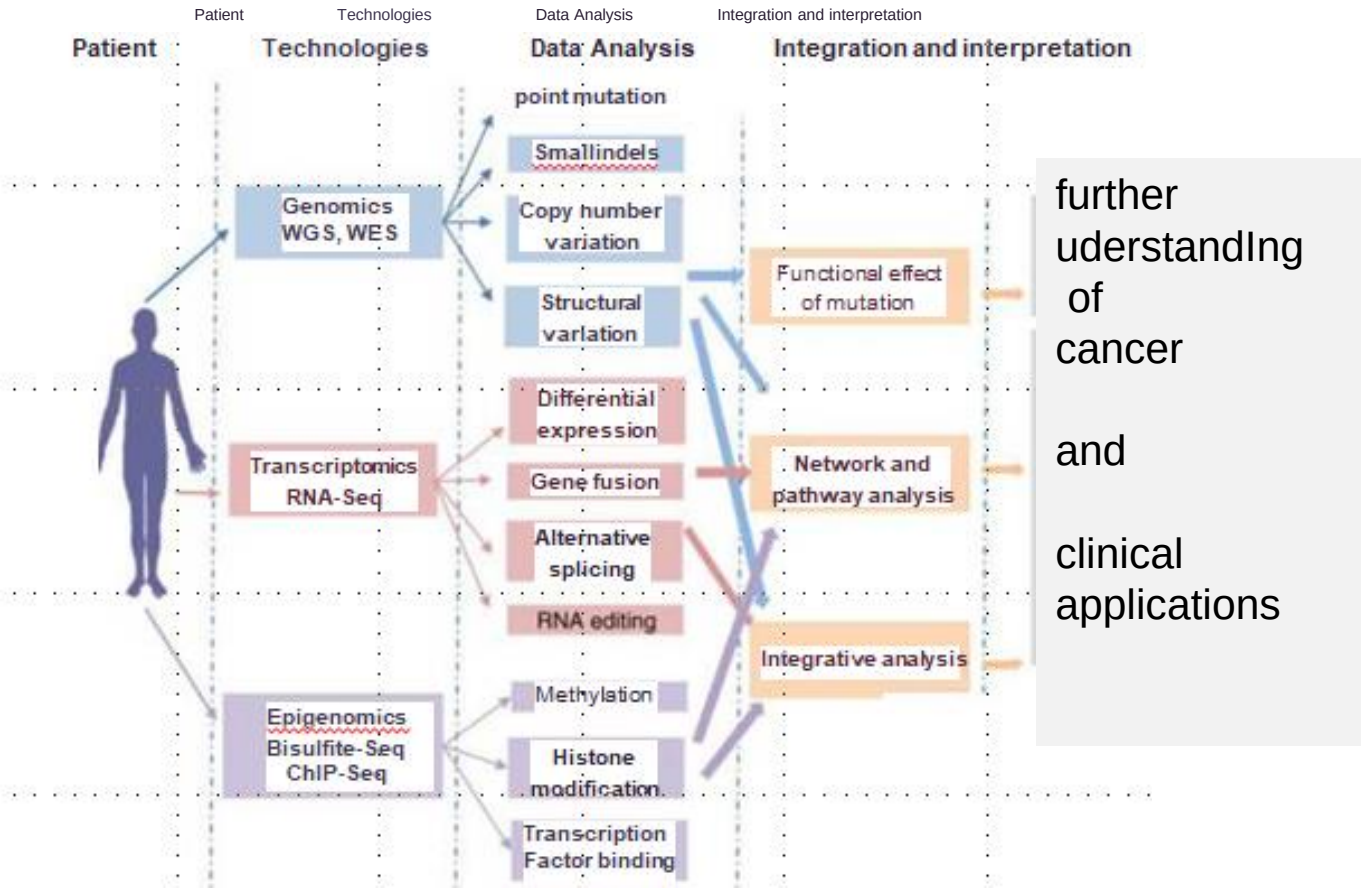
- Queste tecnologie ad esempio analizzano nel loro insieme:
  - **GENOMICA:** i geni contenuti nel DNA (genomica) e le loro molteplici funzioni (genomica funzionale);
  - **TRASCRITTOMICA:** un gran numero o di tutti i trascritti (trascrittoma);
  - **PROTEOMICA: il** complesso delle proteine, in particolare delle loro strutture e funzioni.;
  - **METABOLOMICA:** tutti i metaboliti di un organismo biologico, che sono i prodotti finali del suo metabolismo.
  - **EPIGENOMICA:** le modifiche epigenetiche presenti nel materiale genetico di una cellula (l'epigenoma).
- Tra gli altri obiettivi delle tecnologie omiche vi è anche quello di studiare le
  - connessioni
  - interazioni reciproche
    - tra pool di **molecole biologiche (interattomica**)
    - tra queste molecole e
      - microrganismi intestinali (**microbiomica),**
      - patogeni **(infettivomica),**
      - cibi e/o i nutrienti (**nutrigenomica)**
      - ambiente in generale (**exposomica).**

Tecnologie omiche: tecnologie di applicazione mediche

- Hanno un **potenziale di applicazione molto vasto**

- va dall'aumento della comprensione di svariati processi fisiologici e fisiopatologici,

- al loro impiego nello screening, nella diagnosi e nella valutazione della storia naturale, della risposta alla terapia e/o della prognosi di diverse patologie sia del bambino così come dell'adulto



IoT e Big Data per la medicina di precisione

- **IoT:** rete di dispositivi connessi che possono comunicare tra loro e fornire dati agli utenti tramite Internet.



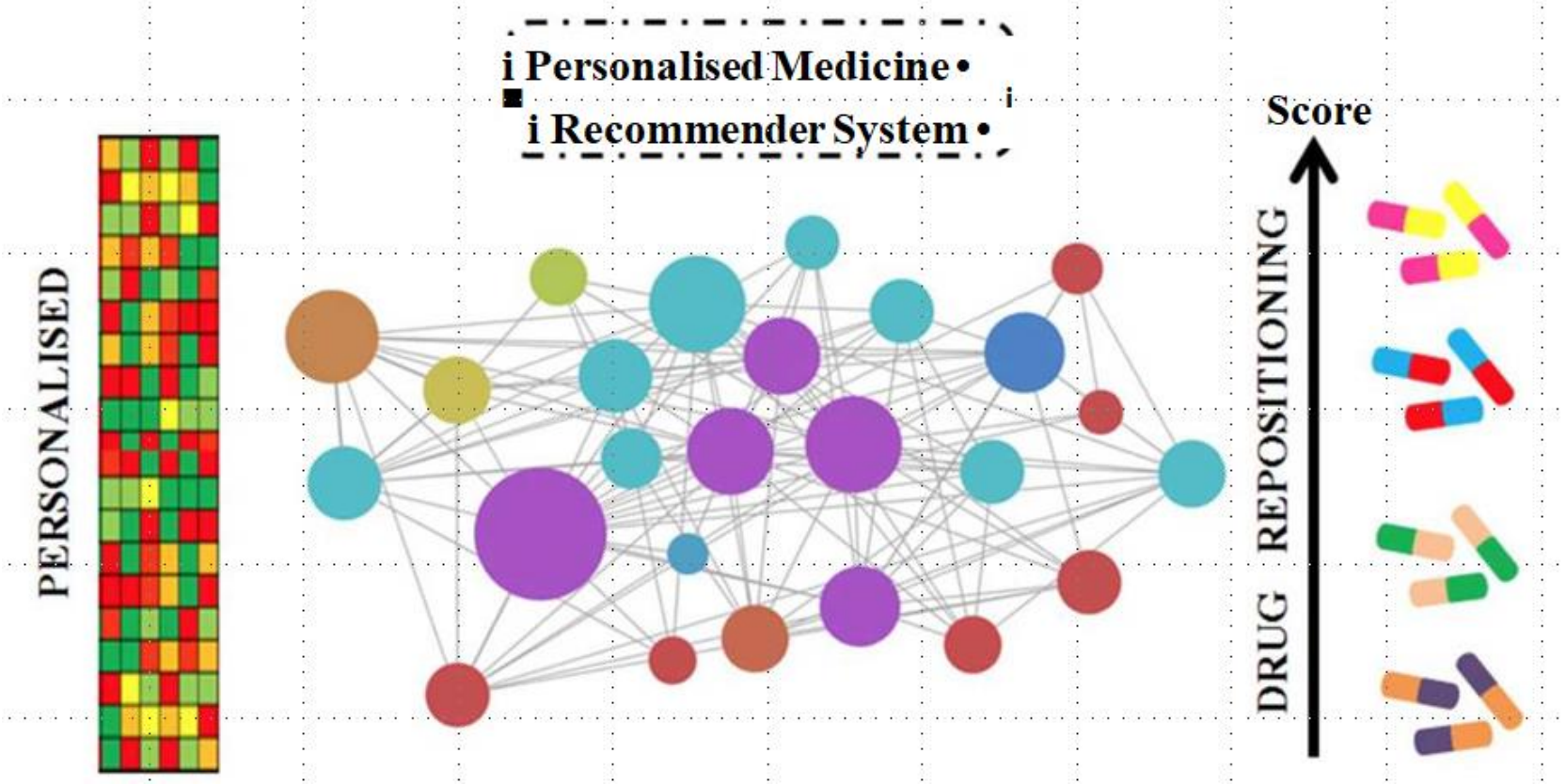
Per ogni **singolo individuo** sarà possibile mettere a punto **strategie di prevenzione** e identificare il miglior percorso terapeutico.

Grazie alla generazione di una **massa enorme di dati** e a nuove tecniche di data mining, le **informazioni genetiche** saranno correlabili alle **malattie**.

Si **prevede che a breve il sequenziamento dell'intero genoma di ogni neonato sarà fattibile ed economicamente sostenibile**

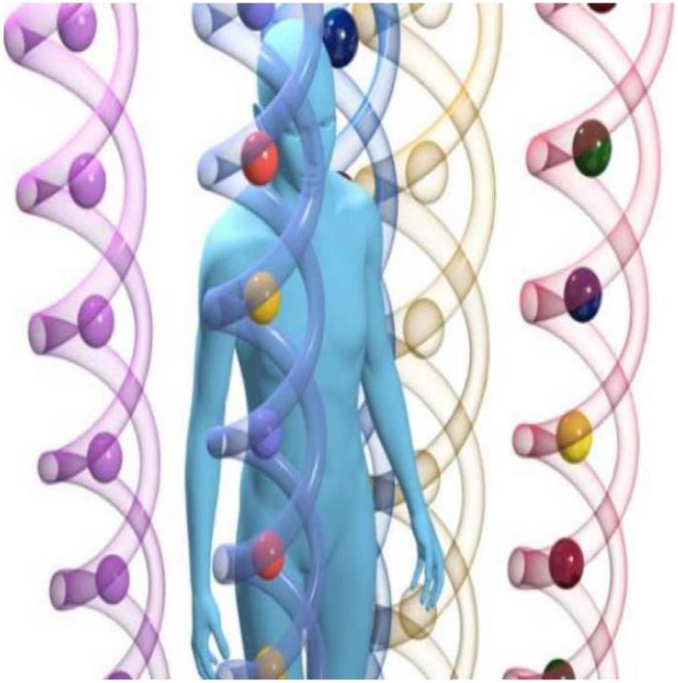
❑ Le tecnologie omiche permetteranno in futuro di comprendere e prevedere

- perché la stessa malattia si presenta in individui differenti con diversa sintomatologia clinica e severità,
- perché risponde diversamente al trattamento terapeutico.
- Il futuro è nella '**medicina di precisione**' o 'm**edicina personalizzata',**
- ciò permetterà di personalizzare le terapie, i protocolli diagnostici e la prognosi del malato.



GENETICA, MEDICINA PREDITTIVA
MEDICINA PERSONALIZZATA

- **Con il termine di Medicina Personalizzata o medicina di precisione s'intende oggi il modello di medicina caratterizzato dalla possibilità di :**

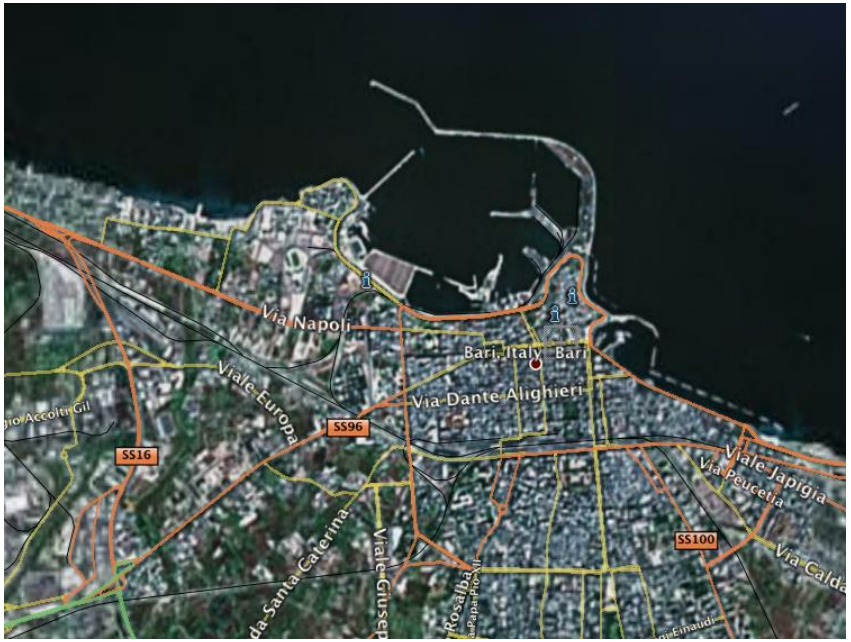


- **identificare la suscettibilità di una singola persona alle malattie comuni,**
- **misurarne il livello di rischio, personalizzare la terapia in base alla costituzione genetica del paziente**
- **offrire nuove opzioni terapeutiche basate sulla interazione dei farmaci con nuovi bersagli molecolari**
- **Ciò al fine di assicurare il miglior esito possibile in termini di salute.**

- **In ambito oncologico, i test vengono spesso utilizzati a due livelli per:**
  - **stabilire terapie ad hoc**
  - **valutare il rischio individuale di ammalarsi.**
- **Oggi si stanno sviluppando strategie di medicina personalizzata ormai in tutti i campi, nessuno escluso, dalle malattie cosiddette complesse (cardiovascolari, neurodegenerative, immunologhe, tumori), alle diete.**
- **La medicina personalizzata sta rivoluzionando la medicina pratica**
  - **nel giro di 10 anni, ognuno avrà il genoma sequenziato come parte della propria cartella clinica e quindi il farmaco appropriato.**

Suddivisioni della Genomica





❑ Genomica Strutturale.

➢ **Studio della struttura del genoma,**

- identificazione di :
- geni e dei loro prodotti di espressione,
- elementi regolatori
- altre entità informazionali.

❑ Genomica Funzionale.

- **Studio delle funzioni dei geni**,
- delle loro interazioni (pathways metabolici)
- dei meccanismi che ne regolano l'espressione.

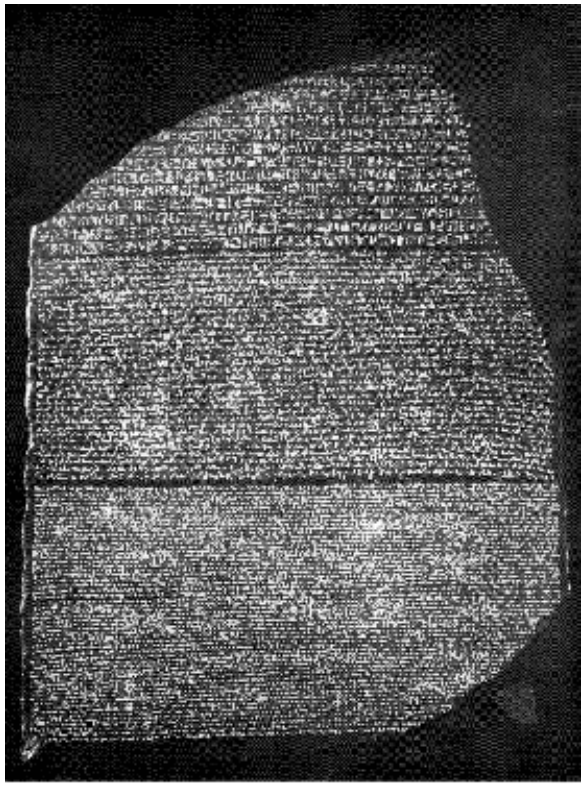
Genomica Mutazionale

- **Studio dei genomi in termini di mutazioni** che sono presenti in DNA o genoma di un individuo.
- Uno degli aspetti della genomica funzionale.
- E' anche indicata come determinazione genomica.
- **Scopo:** determinare la funzione del gene o sequenze anonime

Genomica Comparata

- Gli obiettivi sono volti a **confrontare caratteristiche genomiche tra specie differenti.**
- *e.g*. per una migliore comprensione delle relazioni evolutive.
- Determinare la funzione di ogni genoma: *e.g* studiando i geni in organismi modello
- Lievito e uomo.

La Genomica Comparata



Stele di Rosetta (British Museum)

- Gli studi **Genomica Strutturale** e **Funzionale** traggono grande vantaggio **dall'analisi comparativa** dei genomi e dei loro prodotti di espressione.
- Il confronto di entità "omologhe" ci aiuta ad interpretare l'informazione genica.

La Genomica Comparata: obiettivi

"Nothing in Biology makes sense except in the light of evolution"



Theodosius Grigorevich Dobzhansky
(1900 - 1975)

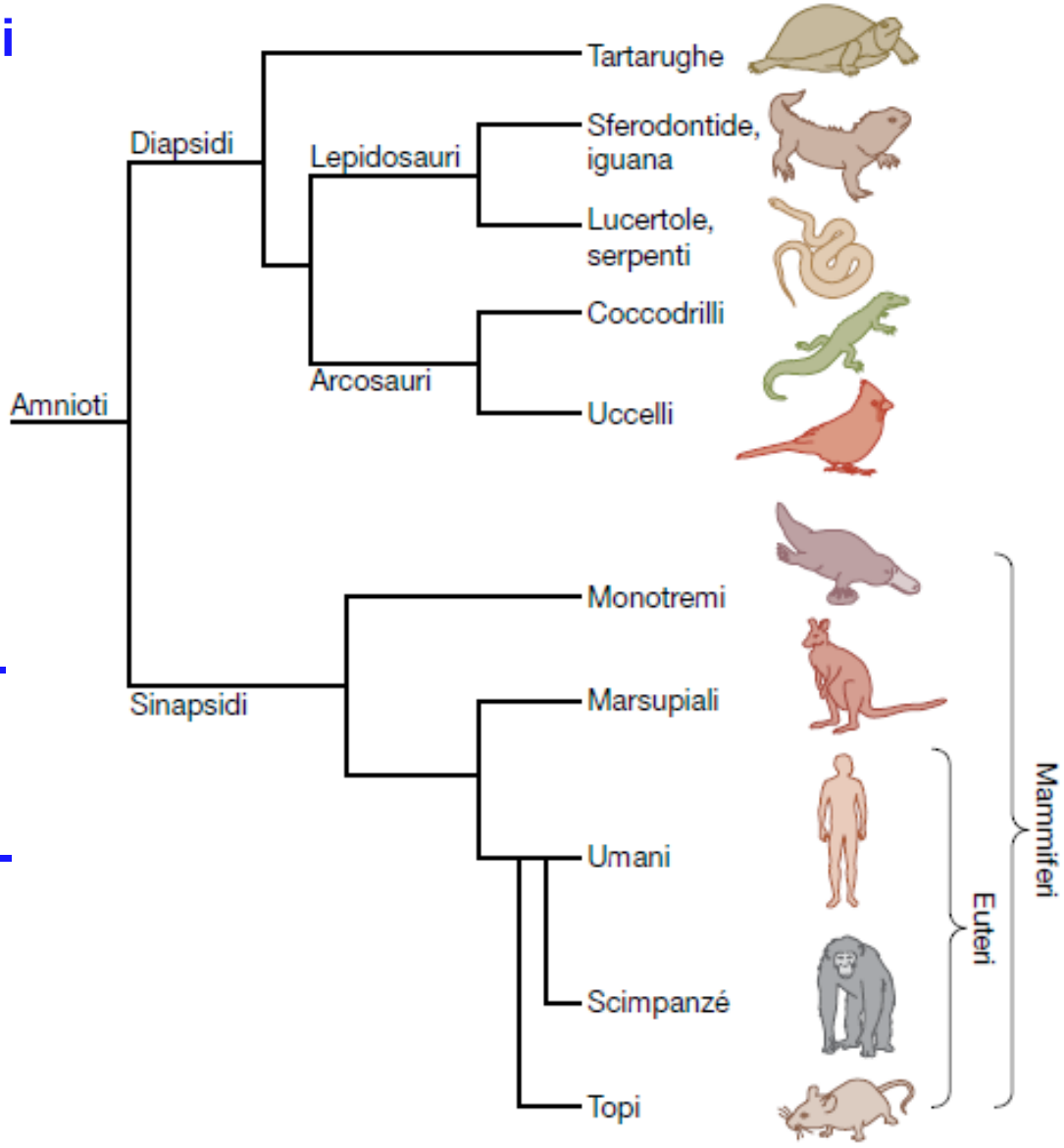
- La conservazione delle sequenze ci permette di **osservare gli effetti dell'evoluzione**.
- Ciò che è **conservato,** ovvero preservato nel corso dell'evoluzione, è molto probabile che abbia una **precisa funzione biologica**.
- La conservazione può essere a livello della sequenza (nucleotidica o proteica), della struttura, dell'espressione, ecc.
- Analogamente, possiamo **attribuire una medesima funzione a geni** (o altre entità biologiche) che risultino **simili e conservati nel corso dell'evoluzione.**

Inferenza filogenetica

- **Filogenesi dei mammiferi e di altri amnioti**
- L'albero filogenetico rappresenta le relazioni evolutive presenti tra i tre gruppi principali di mammiferi (monotremi, marsupiali ed euteri) e altri amnioti, inclusi gli uccelli e vari rettili.

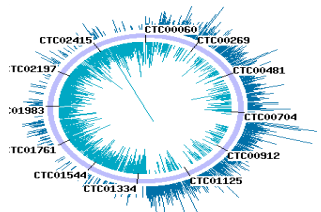
- **Mappando la presenza e l'assenza di geni in particolari gruppi e conoscendo la filogenesi si può dedurre la direzione delle modificazioni evolutive**

(perdita o amplificazione) in una particolare linea evolutiva.



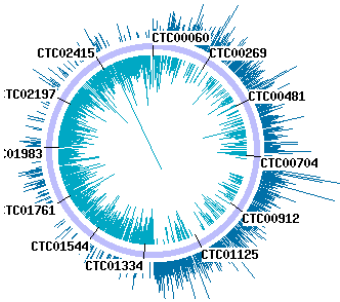
Benefici attesi da ricerche genomiche

- Medicina molecolare
- Genomica Microbica
- Valutazione del rischio
- Bioarcheologia, Antropologia, Evolutione e Migratione umana
- Identificazione del DNA (Forense)
- Agricoltura, miglioramento animale, e Bioprocessi



Mappaggio dei genomi

- Genomi possono essere mappati con:.
  - Linkage,
  - Polimorfismi di Restrizione,
  - Sequenze di DNA .



Sequenziamento del DNA



- **Sequenziamento:** determinazione di una sequenza nucleotidica per analizzare il **potenziale codificante proteine della sequenza del DNA.**

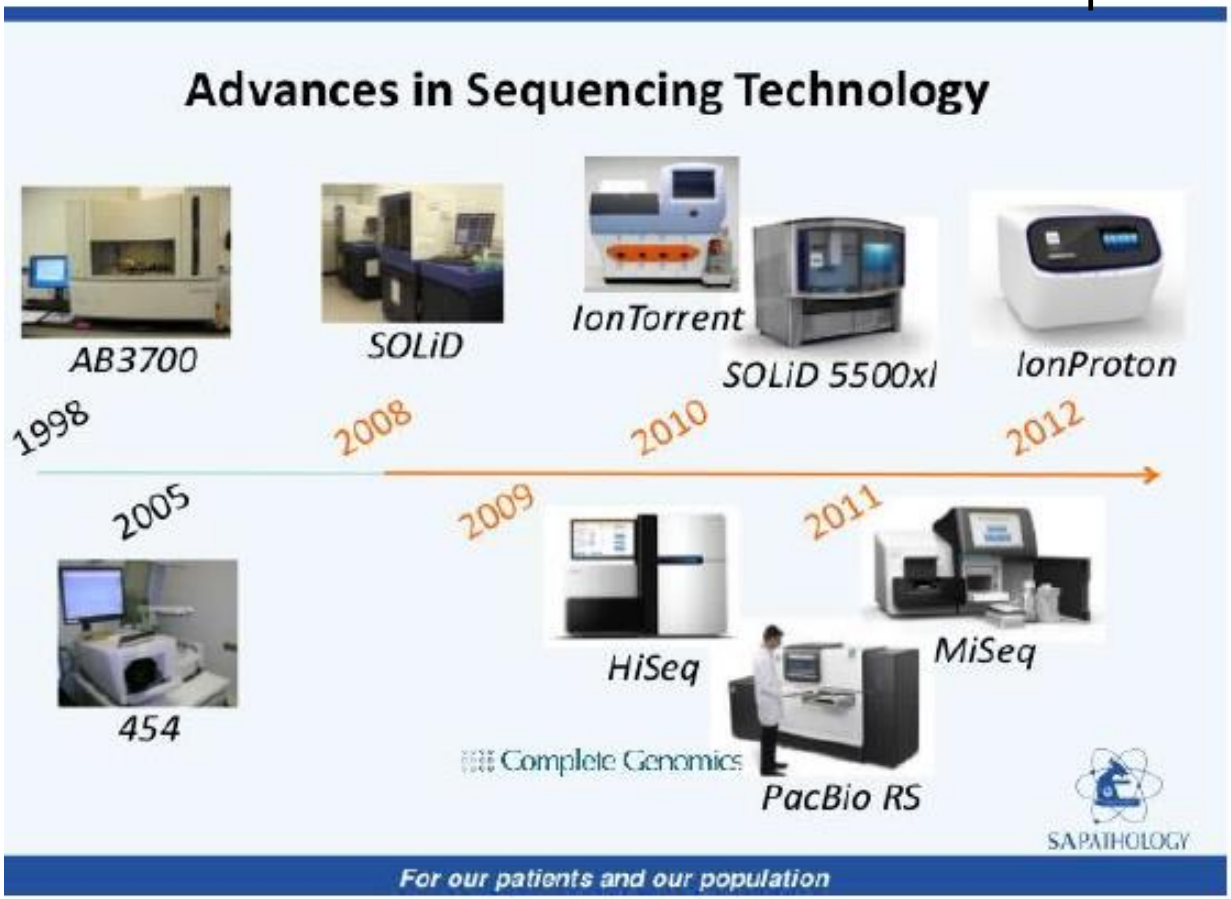
- **Il principio è ottenere una serie di frammenti sovrapposti di DNA che possono essere connessi in una mappa continua.**

Il sequenziamento … oggi

SEQUENZIAMENTO DEL DNA DI NUOVA GENERAZIONE

NGS: *Next Generation Sequences*

- Si basa sul sequenziamento di cluster clonali: milioni di analisi di sequenziamento di DNA contemporaneamente sullo stesso campione e su più campioni diversi

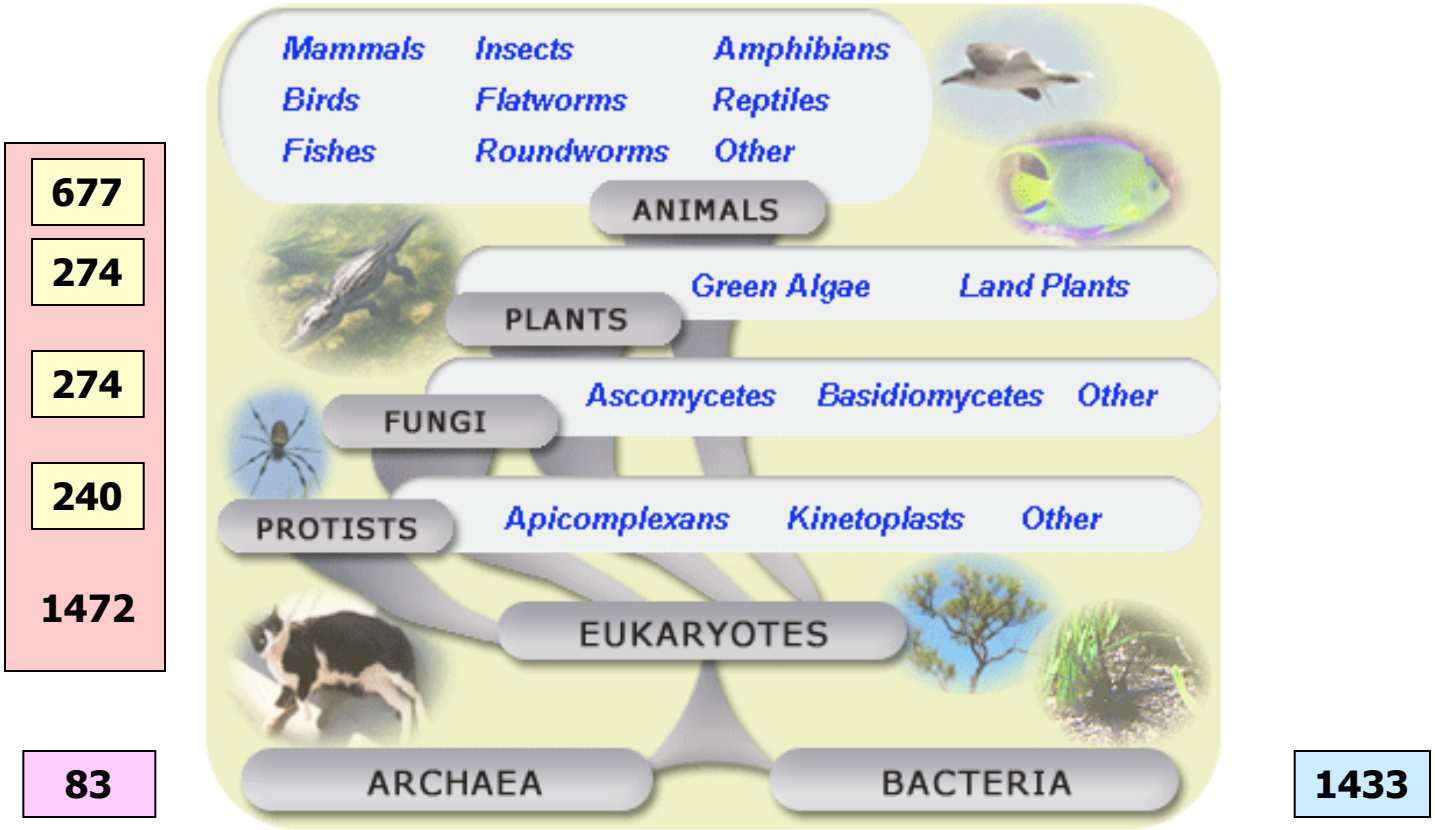


- Nel ventennio trascorso dal completamento del Progetto Genoma Umano, significativi progressi nelle tecnologie di sequenziamento e analisi del genoma hanno notevolmente **ridotto il costo del sequenziamento.**
- I ricercatori possono ora sequenziare più genomi e quindi creare set di dati più grandi da analizzare alla ricerca del loro 'ago nel pagliaio.

$3,000,000,000	2003 Human Genome Project	
$20,000,000	2006 1st individual genome	
$2,000,000	2007 1st NGS Genome	
$200,000	2008 1st 30x genome	
$10,000	2010 1st sub-10K genome	
$1,000	2014 1st $1,000 genome	
$100	2017 1st $100 genome	

Genome sequencing projects

- NCBI genome portal, as of 11/19, 2007



2018: Eukaryotes: 511 Bacteria: 5209 Archaea: 296

Il sequenziamento del genoma utilizza le metodiche che prevedono diverse fasi

1: compilazione – consiste nella ripetizione per numerose volte dell'intero processo per riempire le lacune e correggere gli errori dovuti alla natura empirica e casuale del processo

2: annotazione – consiste nell'identificazione di presunti geni, riconoscibili come "open reading frames" (ORF), nel caso di codificazione di polipeptidi

In un tratto di DNA sequenziato per la ricerca di ORF sono possibili 3 tentativi di lettura, sfalsati di un nucleotide, per ciascuno dei due filamenti complementari.

Oggi sono disponibili software sofisticati che minimizzano gli errori, tenendo conto di caratteristiche note del genoma (p. es. abbondanza di "isole CpG"- 5'CG3' in prossimità di geni)

Le lacune, oltre agli errori di metodo, sono dovute a regioni difficilmente sequenziabili: DNA satellite e geni ripetuti in tandem

Le ORF iniziano con codoni di inizio traduzione (sul filamento trascritto 3'TAC5', su quello complementare 5'ATG3') e , dopo numerose triplette, terminano con codoni di stop.

3: verifica delle ORF – consiste nella ricerca in prossimità delle ORF di altre sequenze funzionali (promotori; negli eucarioti siti di splicing, di poliadenilazione etc.).

Solo il 5% del genoma umano consiste nei suoi circa 26.000 geni; il 10% consiste in DNA satellite, il 50% in elementi trasponibili interspersi.

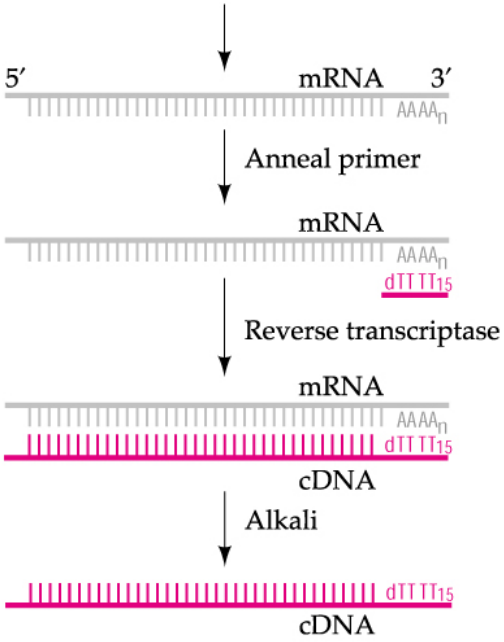
Genome sequences: overview

- **Genome Assembly** (noum): current version of genome sequence
  - Shotgun Sequencing
  - Genetic Maps

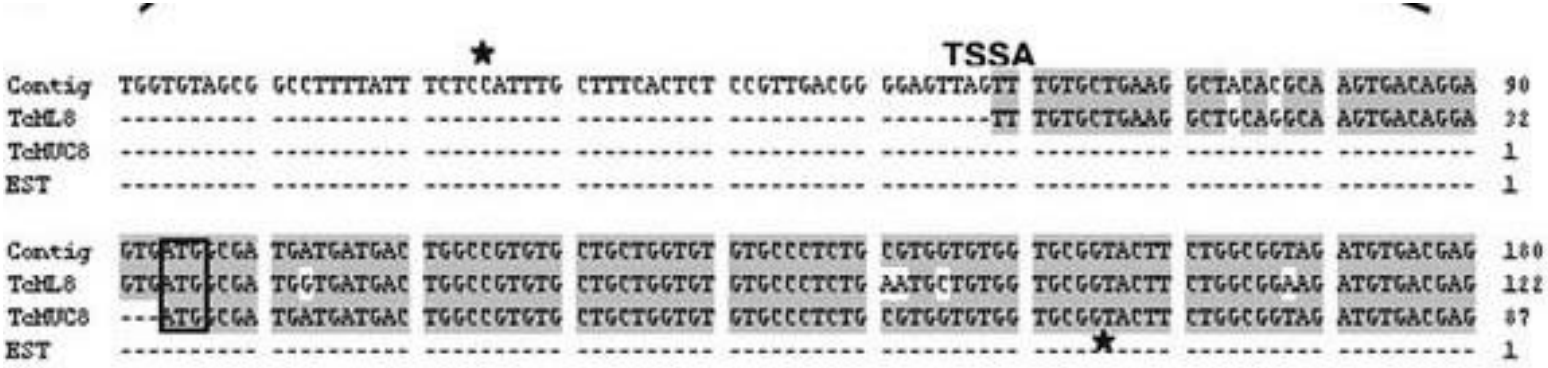
- **Gene annotation** (process): identifying and describing important sequences in genome
  - *Ab initio*
  - Homology-based Annotation

Come si identificano i geni?

1. **Metodi biomolecolari**: Isolando gli mRNA, sintetizzando un'elica di DNA complementare con la trascrittasi inversa e poi sequenziando il DNA copia



2. **Metodi bioinformatici:** Cercando Open Reading Frames (ORF), cioè regioni che iniziano con un'AUG e terminano, dopo un numero di basi multiplo di 3, con un codone di stop



Come si identificano i geni?

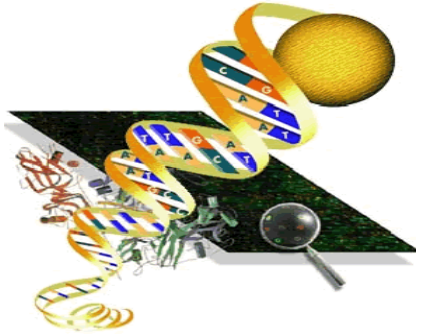
- Riconoscimento dei geni
- Predizioni dei siti funzionali
- Annotazione funzionale dei prodotti genici
- Studio dell’evoluzione del genoma

Metodi per l'identificazione di geni codificanti proteine

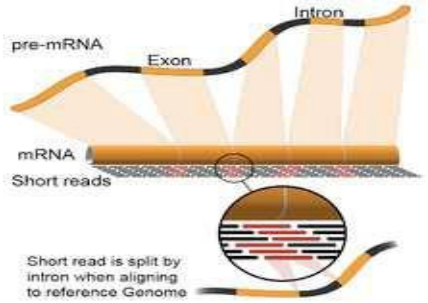
- Metodi basati sull'allineamento delle evidenze sperimentali.
- Metodi basati sulla predizione genica *ab initio.*
- Metodi basati sulla predizione genica *ab initio* guidata da evidenze sperimentali.
- Metodi basati sul confronto tra genomi

Metodi basati sull'allineamento delle evidenze sperimentali

❑ Si possono utilizzare utilizzare diverse evidenze sperimentali, che opportunamente elaborate e allineate al genoma permettono di identificare le regioni codificanti proteine:



- **cDNA full-length**: sequenze di RNA maturi (mRNA) retro-trascritti a cDNA, quindi completo di UTR e CDS.
- **EST** (Expressed Sequence Tags): brevi frammenti parziali, tra 400-800 bp, di mRNA retrotrascritti a cDNA.
- **Proteine omologhe**: sequenze aa corrispondenti a proteine omologhe di organismi evolutivamente vicini.
- **Tilling arrays**: microarray con sonde equamente spaziate su tutto il genoma, permettono l'identificazione di regione espresse mediante l'ibridazione di campione marcati.



- **MPSS: Massively Parallel Signature Sequencing**: piattaforma che analizza il livello di espressione e **identifica una regione di 17-20 bp degli mRNA tramite sequenziamento**.
- **RNA-seq:** frammenti di cDNA di lunghezza tra 50-150 bp che derivano dal sequenziamento shotgun di un intero trascrittoma.

Principali programmi per la predizione genica

Predittore	Predizione ab initio	Predizione di geni eucarioti	Training in locale per nuovi genomi	Utilizzo di EST e Proteine per la predizione	Utilizzo di RNA-Seq per la predizione	Predizione degli UTR	Predizione dei trascritti alternativi
Augustus	SI	SI	SI	SI	SI	SI	SI
Snap	SI	SI	SI	NO	NO	NO	NO
GeneMark-ES	SI	SI	NO	NO	NO	NO	NO
GeneID	SI	SI	SI	SI	SI	SI	SI
FGenesh	SI	SI	SI	NO	NO	NO	NO
Genescan	SI	SI	NO	SI	SI	SI	NO
MZEF	SI	SI	NO	NO	NO	NO	NO
mGene.NGS	SI	SI	SI	SI	SI	SI	NO
Contrast	SI	SI	SI	SI	NO	SI	NO
GrailExp	SI	SI	NO	SI	NO	SI	NO
TwinScan/N-Scan	SI	SI	SI	SI	NO	NO	SI

Mappaggio di geni “noti”

BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool, o strumento di ricerca di allineamento locale)

- **algoritmo** usato per comparare le informazioni contenute nelle strutture biologiche primarie, come ad esempio le sequenze proteiche o le sequenze nucleotidiche delle molecole di DNA.
- Una ricerca BLAST permette al ricercatore di confrontare una sequenza di interesse con un database di sequenze già conosciute, e di identificare tra queste ultime quelle che presentano delle somiglianze con la sequenza di interesse.

  - **BLASTx: //www.ncbi.nlm.nih.gov/BLAST/**
  - Molta informazione quando un genoma ‘vicino’ è stato studiato in precedenza.
  - **Problemi**: scelta di valori di soglia, mappaggio ‘fine’ delle posizioni d’inizio in altri casi. Soluzioni non esatte.

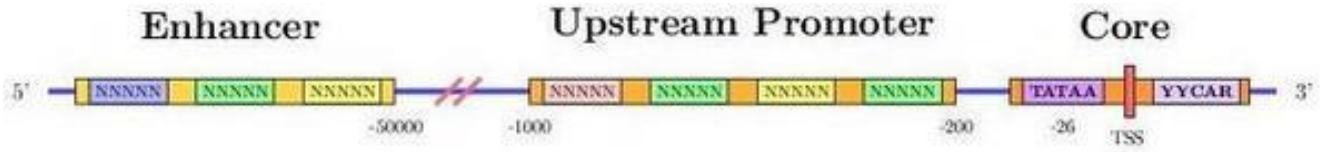
Mappaggio esatto delle posizioni d'inizio dei geni.

- **Procariotici:** La MET iniziale è preceduta da un sito di legame per il ribosoma (chiamato box di Shine-Dalgarno)

- Estensione dell'allineamento nucleotidico con regioni ortologhe appartenenti ad un genoma correlato evolutivamente.

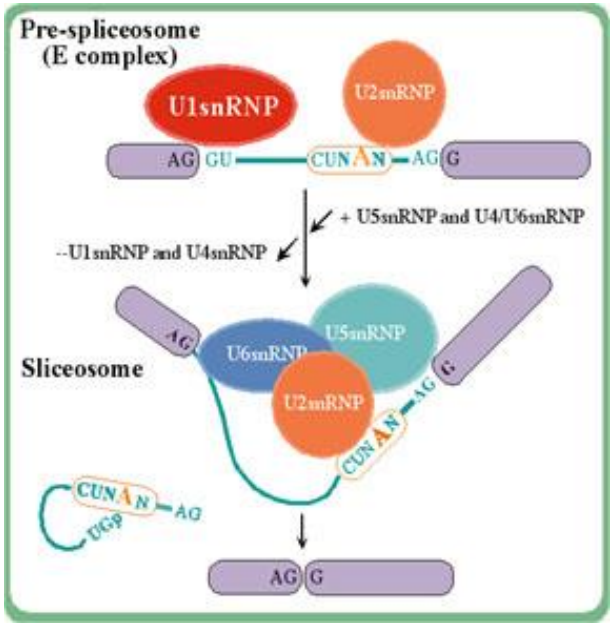
Riconoscimento dei geni II: Eucarioti

- **Preceduto da un promotore e seguito da un sito di poli-adenilazione (le sequenze caratteristiche poco definite)**



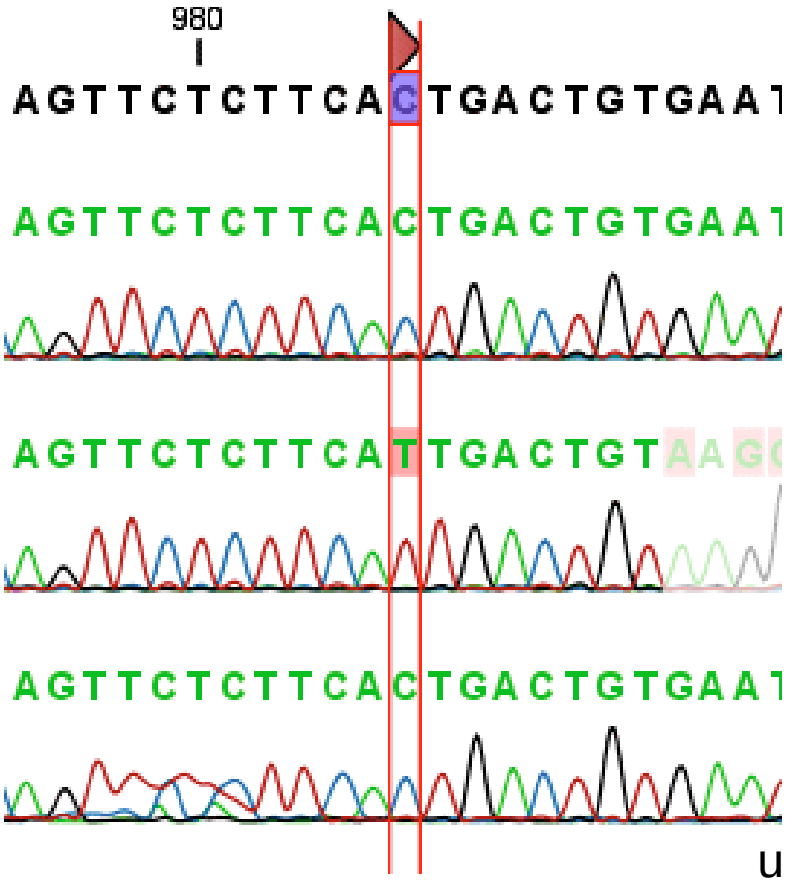
- Gli introni sono fiancheggiati dalle due coppie **AG** e **GT**, ma non ci sono limiti riguardo la distanza fra di loro.

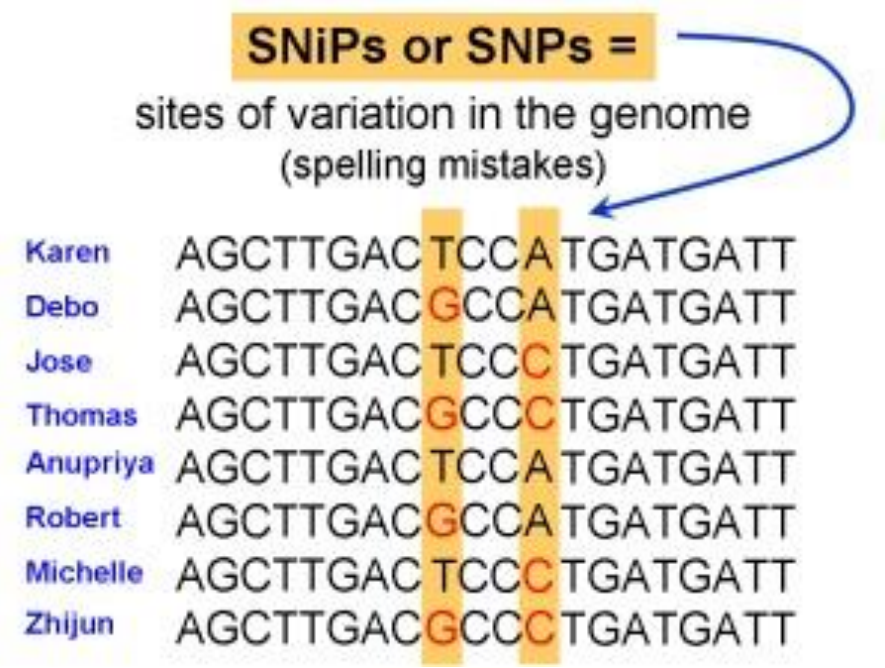
- L'introne può contenere un numero qualunque di basi senza caratteristiche ovvie.



- Negli esoni ci aspettiamo una frequenza di codoni compatibile con i codon usage dell'organismo al quale appartengono
- Effetti di contesto: All'interno degli esoni, alcuni codoni tendono ad essere più spesso adiacenti ad altri;
- la frequenza di esanucleotidi (pe. 5'-TTAGGG-3') è uno dei parametri per calcolare la probabilità che una sequenza sia codificante

Allineando e confrontando più genomi si individuano polimorfismi di singolo nucleotide, o SNPs (Single-Nucleotide Polymorphisms)





Aplotipo:

un insieme di alleli marcatori a loci strettamenti associati che tendono ad essere ereditati in blocco ; consentono di identificare un singolo genotipo in maniera univoca

Genome size

Units: bp: base pair

Common multiples are: 1 kb = 10^3 bp

1 Mb = 10^6 bp

1 Gb = 10^9 bp

Bacterial genomes are typically expressed in **Mb**

Length conversion

Dickerson *et al* (1982) *Science*, 216:475-485.

1 bp ≈ 0.33 nm

1 kb ≈ 0.33 µm

1 Mb ≈ 0.33 mm

1 Gb ≈ 0.33 m

Bacterial genomes are typically in the mm range (and therefore 1000x bigger than the typical bacterial size).

Mass conversion (1 pg = 10^{-12} g)

Dolezel *et al* (2003) *Cytometry*, 51A:127-128. Number of base pairs = mass in pg × 0.978 10^9

1 kb ≈ 10^{-6} pg

1 Mb ≈ 10^{-3} pg

1 Gb ≈ 1 pg

Bacterial genomes are typically in the **10^{-3} pg** range (femtogram).

Dimensione dei genomi

- In confronto ad altri

Quadro generale

❑ **Virus, organelles**
✓ Tiny genomes (kb)
✓ High gene density
✓ Bacteriophages: 10-100 genes

❑ **"Bacteria"**
✓Small genomes (Mb)
High gene density
✓*E. coli*: ~ 5000 genes

❑ **Eucarya**
✓Large genomes (Gb)
✓ Low gene density
✓*Homo Sapiens*: ~ 25000 genes

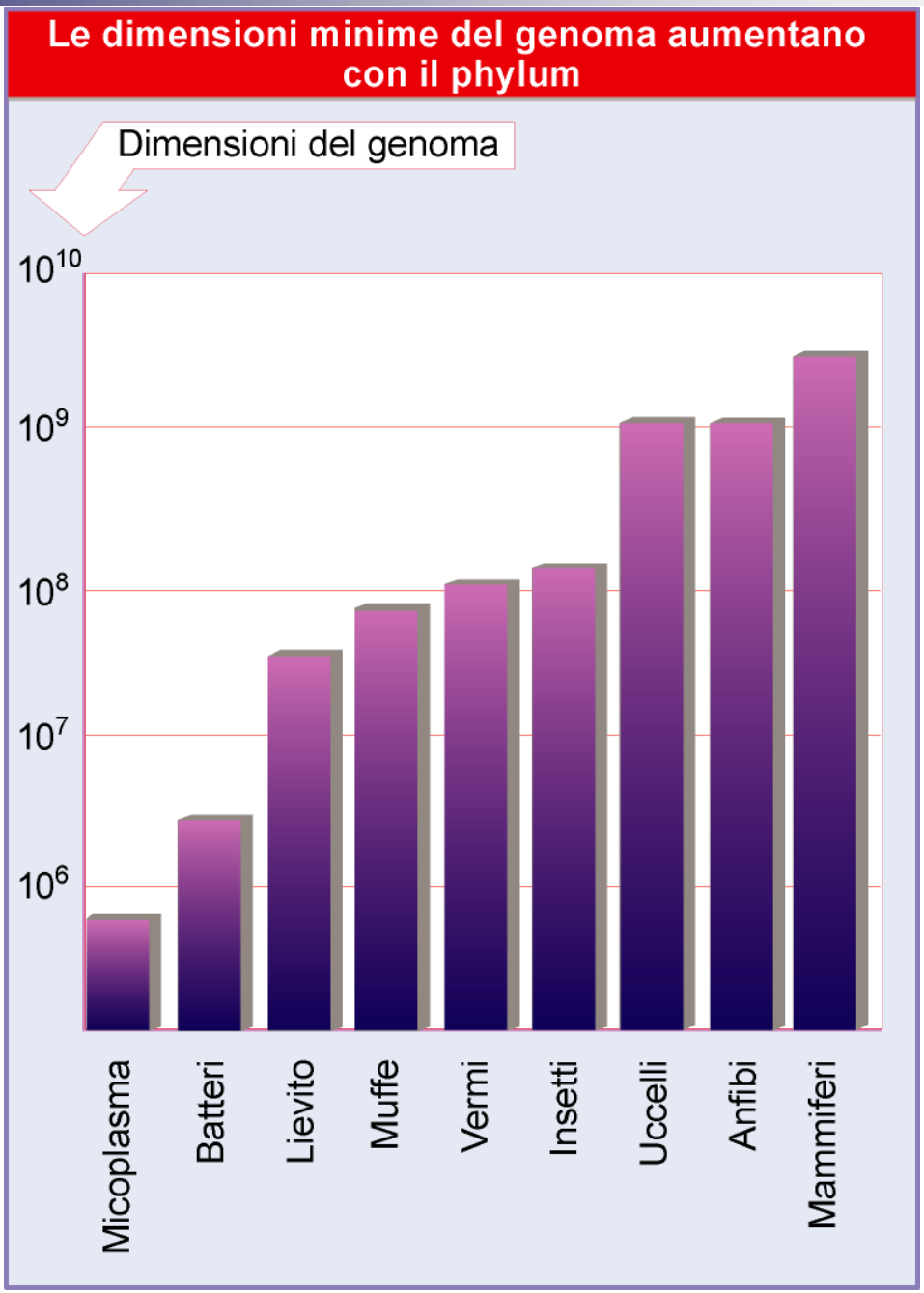
Organizzazione dei genomi

- **Procarioti**
  - La maggior parte del genoma è codificante
  - Piccola parte di sequenze non-codificanti sono sequenze regolatorie
- **Eucarioti**
  - La maggior parte del genoma è non-codificante (98%)
    - Sequenze regolative
    - Introni
    - DNA ripetitivo

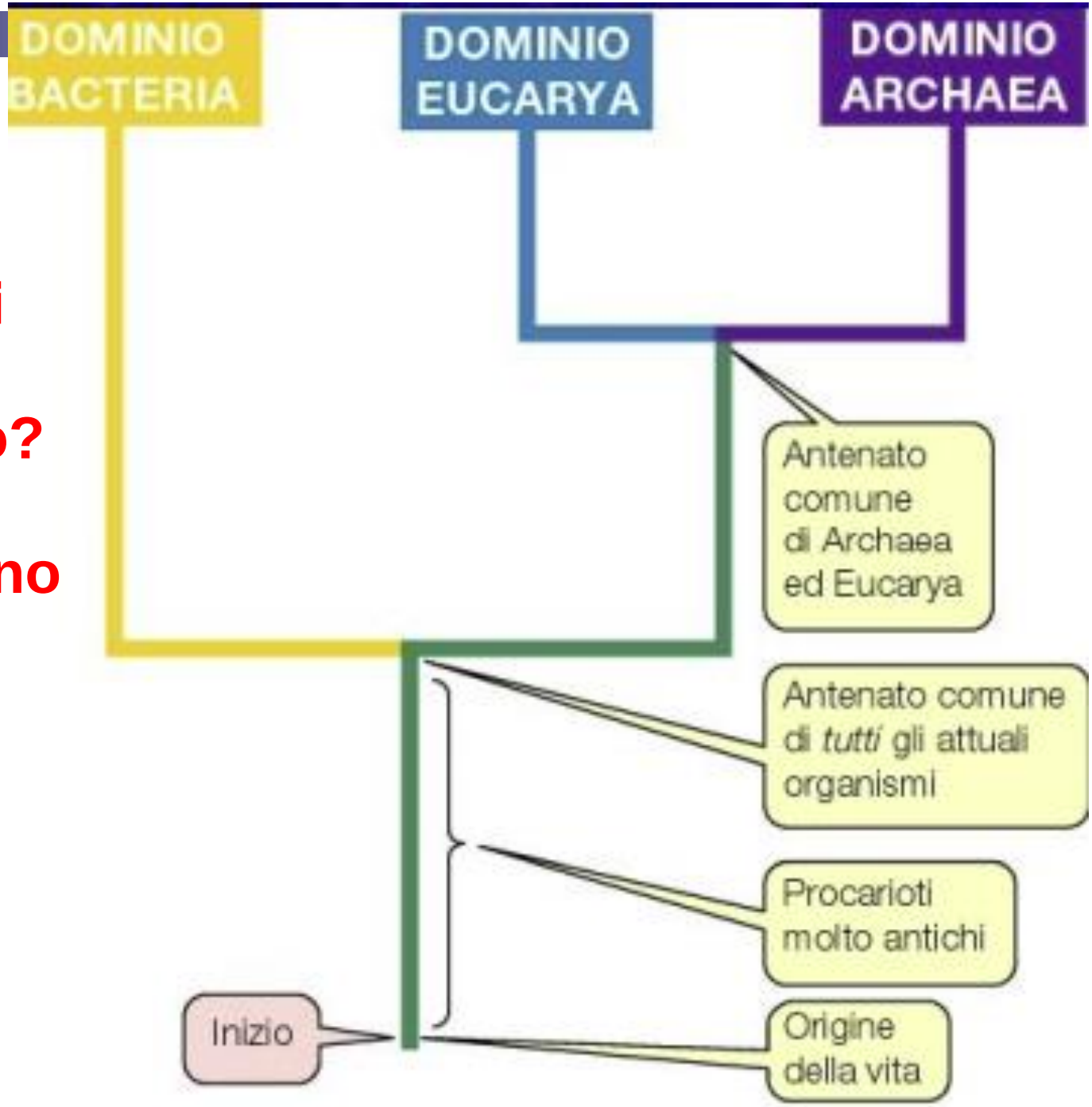
Dimensioni di genomi utili

Phylum	Specie	Genoma (bp)
Alghe	*Pyrenomas salina*	$6{,}6 \times 10^5$
Micoplasma	*M. pneumoniae*	$1{,}0 \times 10^6$
Batterio	*E. coli*	$4{,}2 \times 10^6$
Lievito	*S. cerevisiae*	$1{,}3 \times 10^7$
Muffa mucillaginosa	*D. discoideum*	$5{,}4 \times 10^7$
Nematode	*C. elegans*	$8{,}0 \times 10^7$
Insetto	*D. melanogaster*	$1{,}4 \times 10^8$
Uccello	*G. domesticus*	$1{,}2 \times 10^9$
Anfibio	*X. laevis*	$3{,}1 \times 10^9$
Mammifero	*H. sapiens*	$3{,}3 \times 10^9$

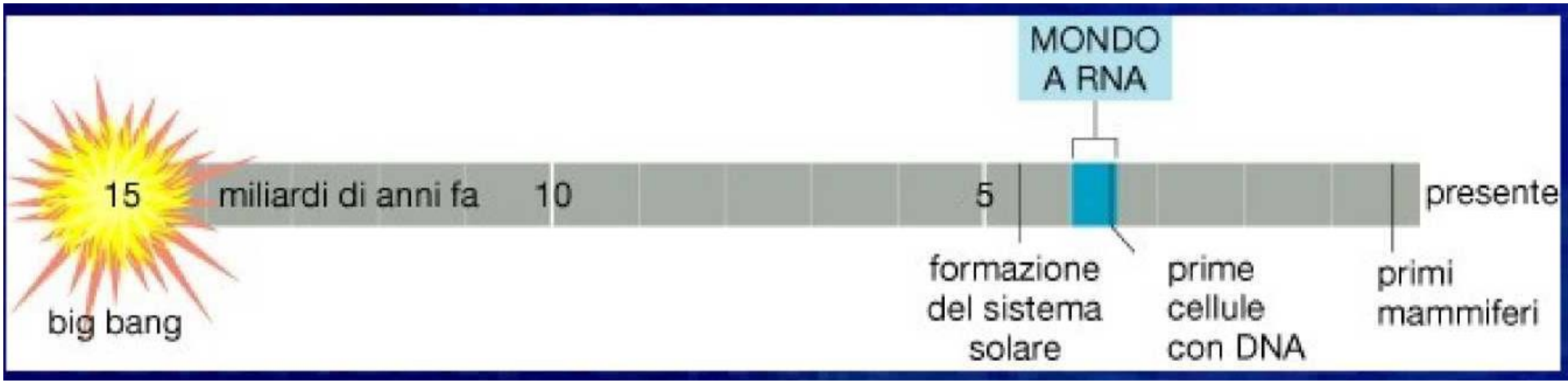
Le dimensioni minime del genoma aumentano con il phylum



Qual è l'origine di tutto questo? Come si sono evoluti i genomi?



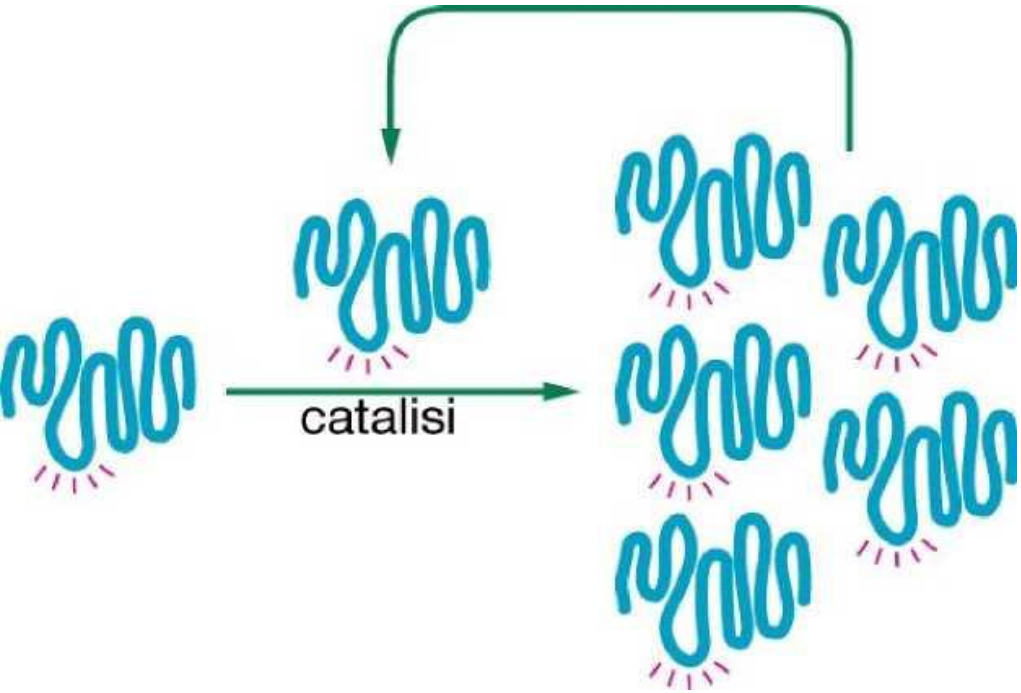
Origine ed evoluzione dei genomi

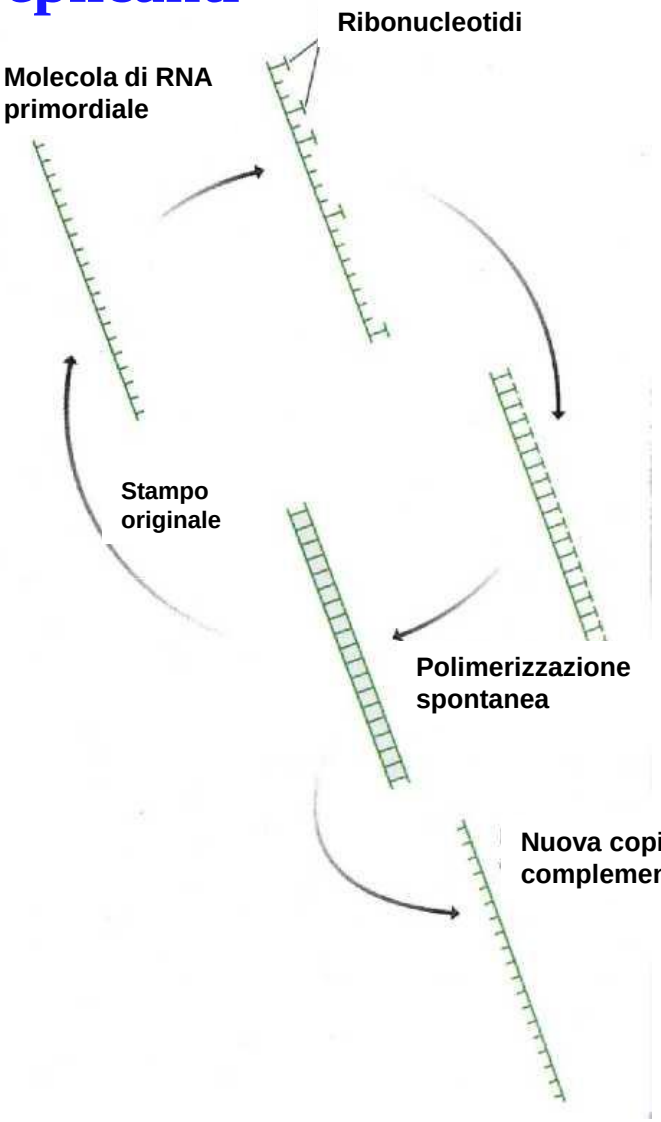


Origine ed evoluzione dei genomi

Mondo a RNA

Nascita di molecole autoreplicanti

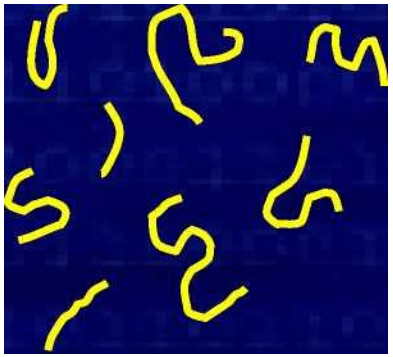


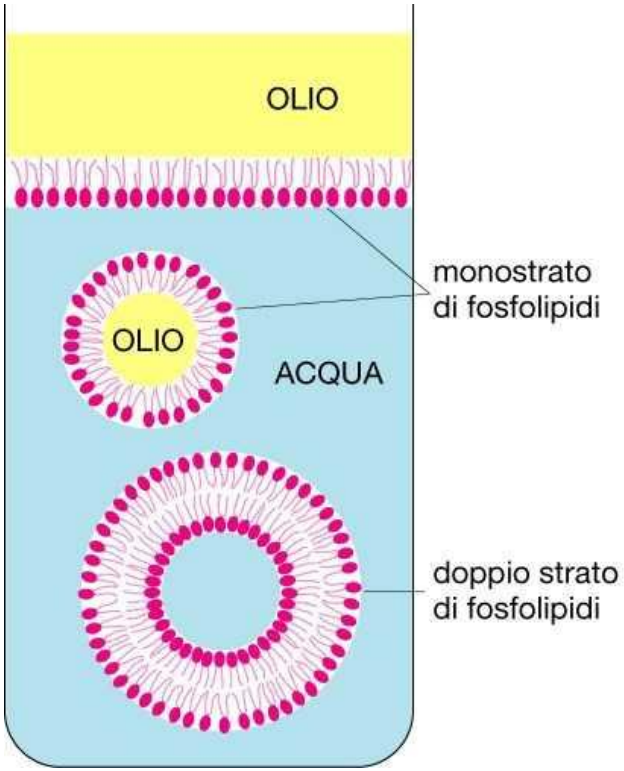


Origine ed evoluzione dei genomi

Mondo a RNA

Protogenomi a RNA





Compartimentalizzazione all'interno di membrane lipidiche

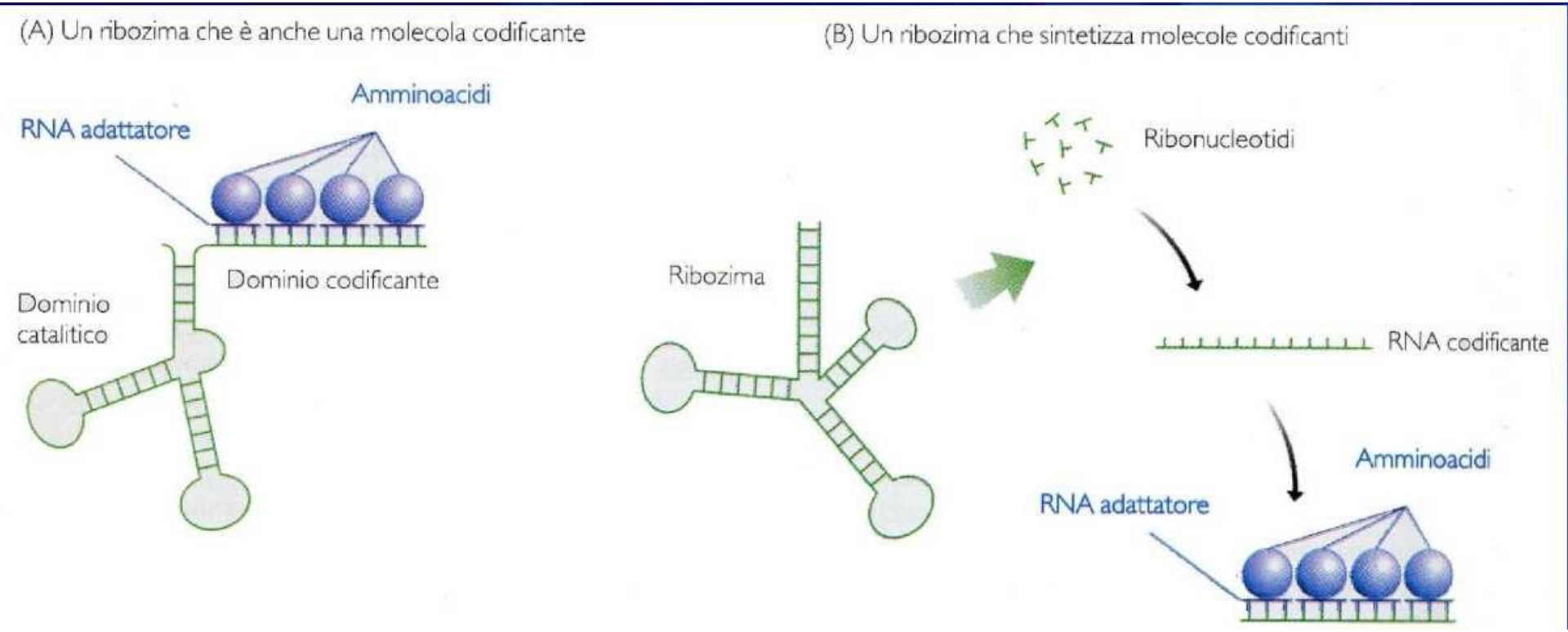


- Prime strutture di tipo cellulare

Origine ed evoluzione dei genomi

Come si è evoluto, il genoma a DNA?

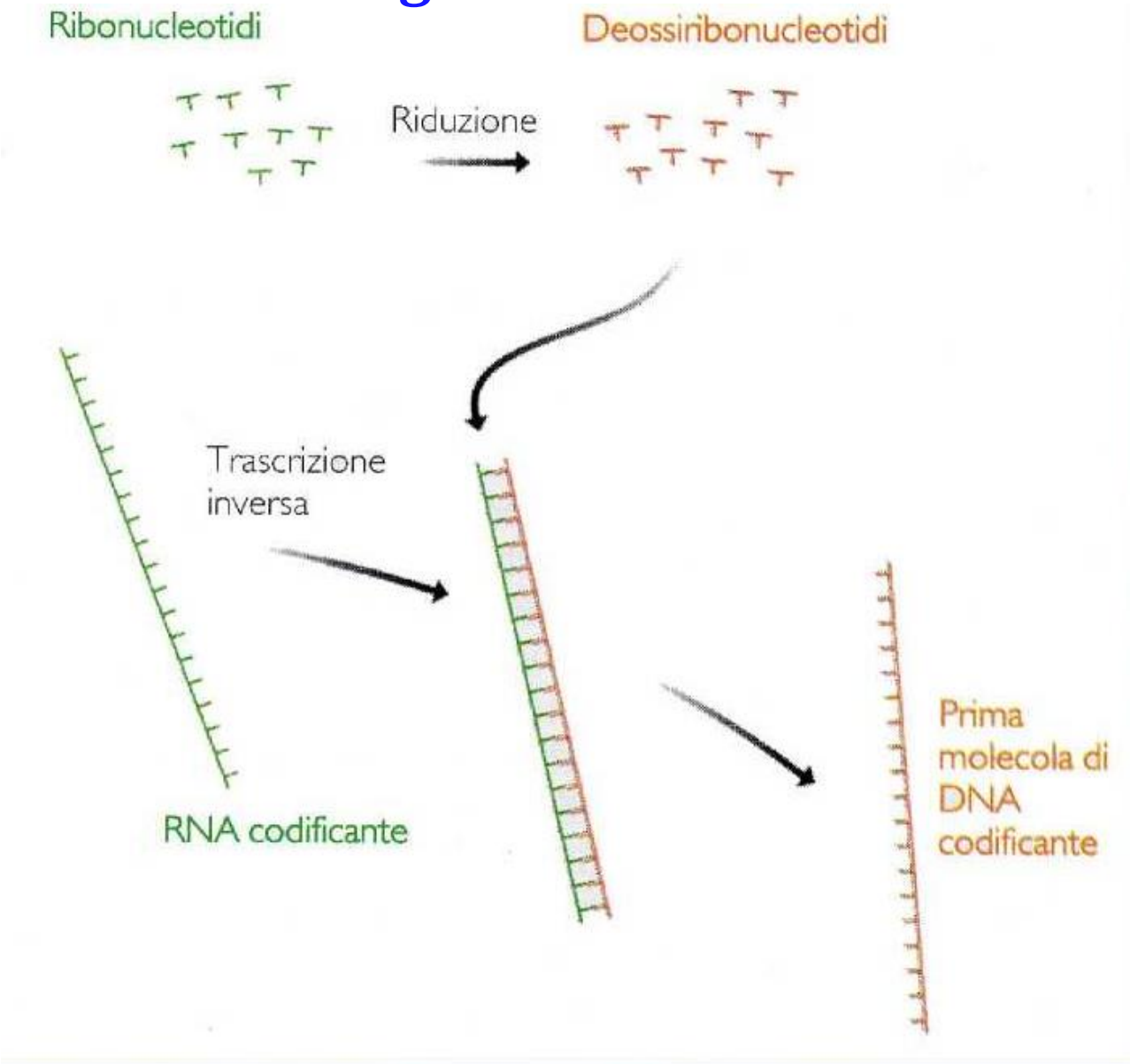
Nascita di enzimi proteici



Origine ed evoluzione dei genomi

Come si è evoluto il genoma a DNA?

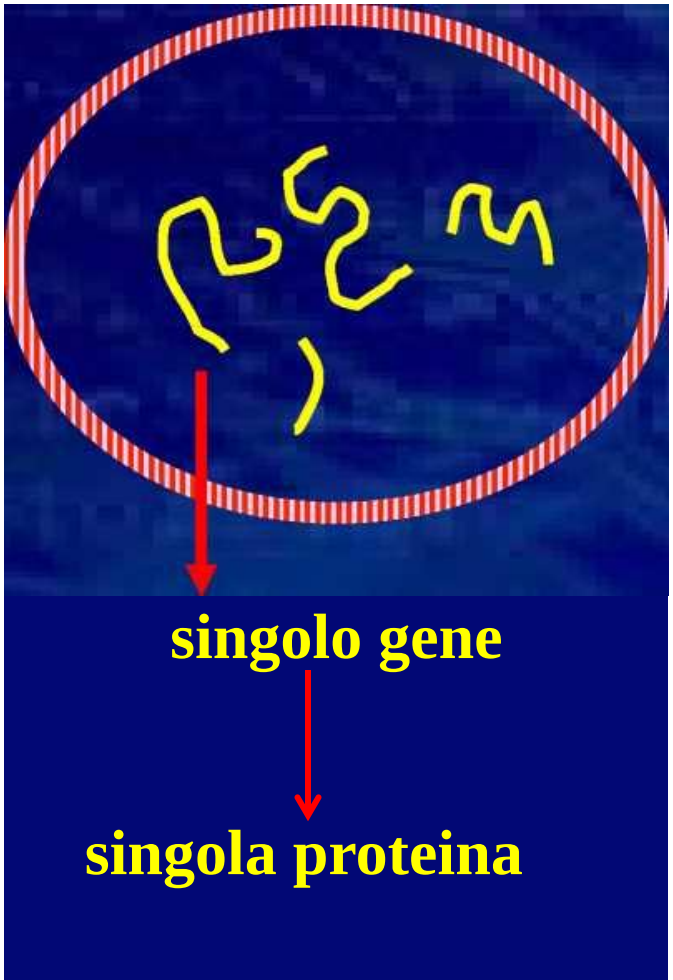
Trasferimento della **funzione codificante dall'RNA al DNA** (chimicamente piu' stabile)



Origine ed evoluzione dei genomi

➢ Primi Genomi a DNA (3,8 miliardi di anni fa)

Ogni molecola di DNA rappresenta un singolo gene che codifica per una singola proteina



Come differiscono i genomi

Similarità e differenze tra sequenze genomiche compaiono

1) a livello di singole basi;
2) a livello di geni;
3) in blocchi su scala maggiore;
4) a livello di interi genomi che hanno subito duplicazioni complete

Come si originano le novità evolutive

- Cambiamenti in geni già esistenti
  - Mutazioni puntiformi
  - Inserzioni/delezioni
  - Traslocazioni
  - Riarrangiamenti cromosomici

- Acquisizione di nuovi geni
  - *De novo*
  - Duplicazione
  - Trasferimento orizzontale
  - Ibridazione

DUPLICAZIONI

Specie	Geni duplicati (%)
Bacteria (batteri)	
Mycoplasma pneumoniae	44
Helicobacter pylori	17
Haemophilus influenzae	17
Archaea (archei)	
Archaeoglobus fulgidus	30
Eukarya (eucarioti)	
Saccharomyces cerevisiae	30
Caenorhabditis elegans	49
Drosophila melanogaster	41
Arabidopsis thaliana	65
Homo sapiens	38

Da: J. Zhang, *Evolution by gene duplication: an update*, «Trends Ecol. Evol.», 18 (2003), pp. 292-298.

DUPLICAZIONE GENICA: DESTINO DEI GENI DUPLICATI

1. Tutte le copie mantengono la stessa funzione.



2. Alcune copie scompaiono.



3. Alcune copie evolvono con una nuova funzione.



DUPLICAZIONE GENICA: DESTINO DEI GENI DUPLICATI

1.

1. Tutte le copie mantengono la stessa funzione.



Numbers of rRNA and tRNA genes per haploid genome in various organisms

Genome Source	Number of rRNA sets	Number of tRNA genes[a]	Approximate genome size (bp)
Human mitochondrion	1	22	2×10^4
Nicotiana tabacum chloroplast	2	37	2×10^5
Escherichia coli	7	~ 100	4×10^6
Neurospora crassa	~ 100	~ 2,600	2×10^7
Saccharomyces cerevisiae	~ 140	~ 360	5×10^7
Caenorhabditis elegans	~ 55	~ 300	8×10^7
Tetrahymena thermophila	1	~ 800[c] 2×10^8	
Drosophila melanogaster	120-240	590-900	2×10^8
Physarum polycephalum	80-280	~ 1,050	5×10^8
Euglena gracilis	800-1,000	~ 740 2×10^9	
Human	~ 300	~ 1,300	3×10^9
Rattus norvegicus	150-170	~ 6,500	3×10^9
Xenopus laevis	500-760	6,500-7,800	8×10^9

DUPLICAZIONE GENICA: DESTINO DEI GENI DUPLICATI

1. Tutte le copie mantengono la stessa funzione.



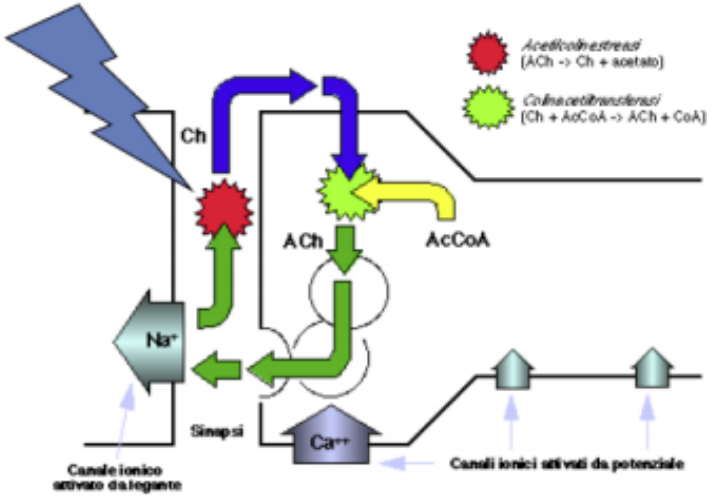
CN
V

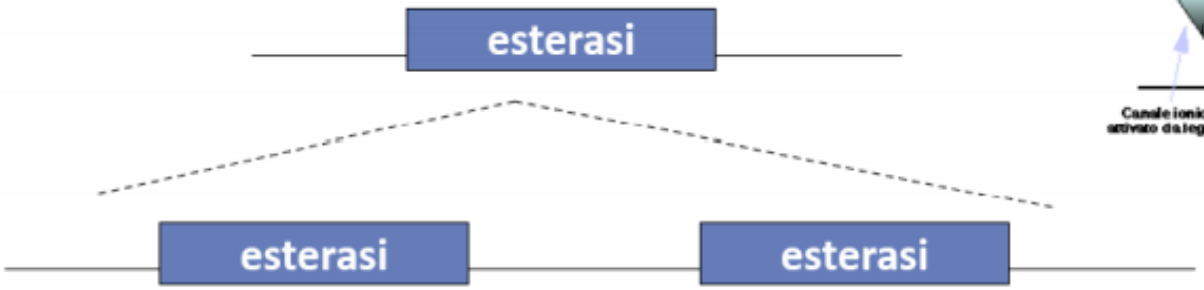
Resistenza ai pesticidi: **zanzare**



Organofosfati: neurotossine che colpiscono il SN degli insetti

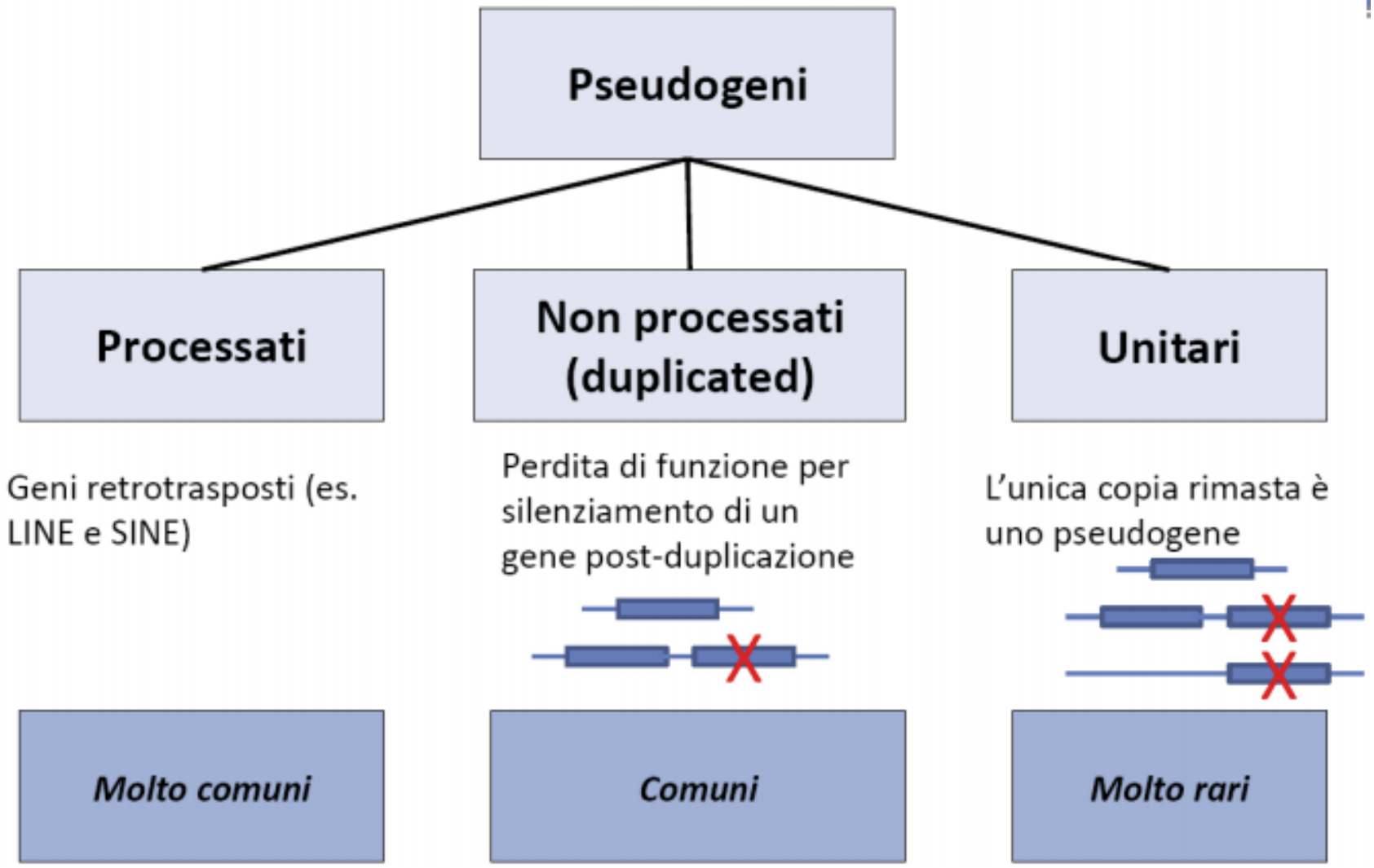
Resistenza:
Amplificazione del gene dell'Esterasi che blocca il pesticida prima dell'arrivo al terminale nervoso





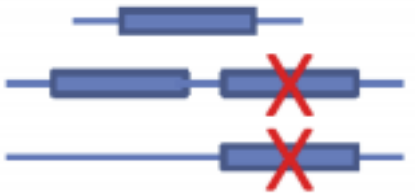
DUPLICAZIONE GENICA: DESTINO DEI GENI DUPLICATI

2. Alcune copie scompaiono.



DUPLICAZIONE GENICA: DESTINO DEI GENI DUPLICATI

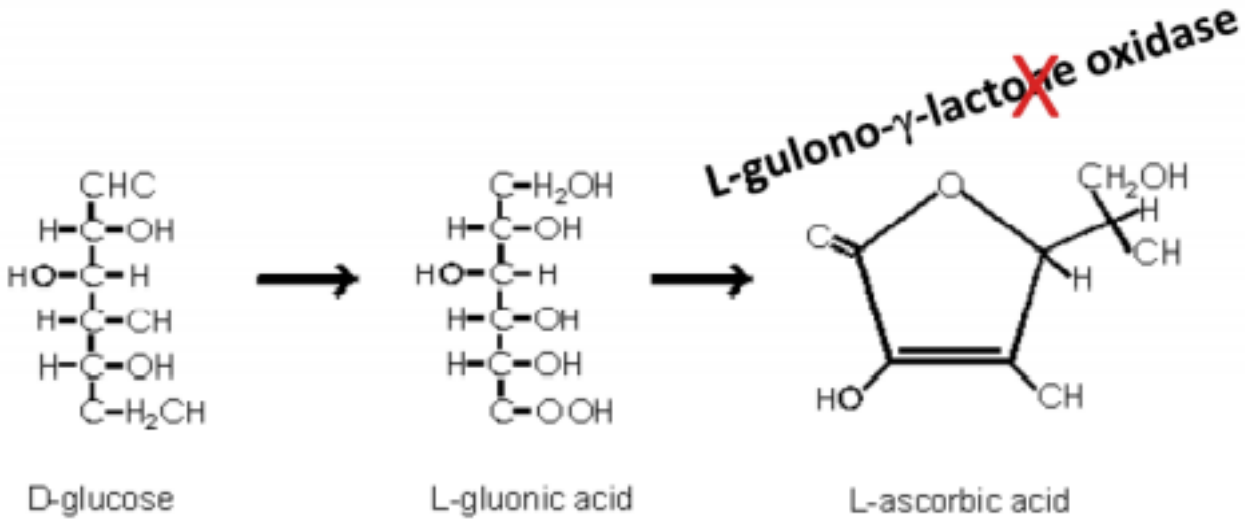
2. Alcune copie scompaiono. Pseudogeni unitari



Come si può fissare in una popolazione la perdita di un gene?
Per deriva se la selezione è rilassata

Ipoascorbemia o scrobuto = incapacità di sintetisi dell'acido ascorbico

Cavie
Uomo — si ammalano di scorbuto se non assumo acido ascorbico (vitamina C)
Trote

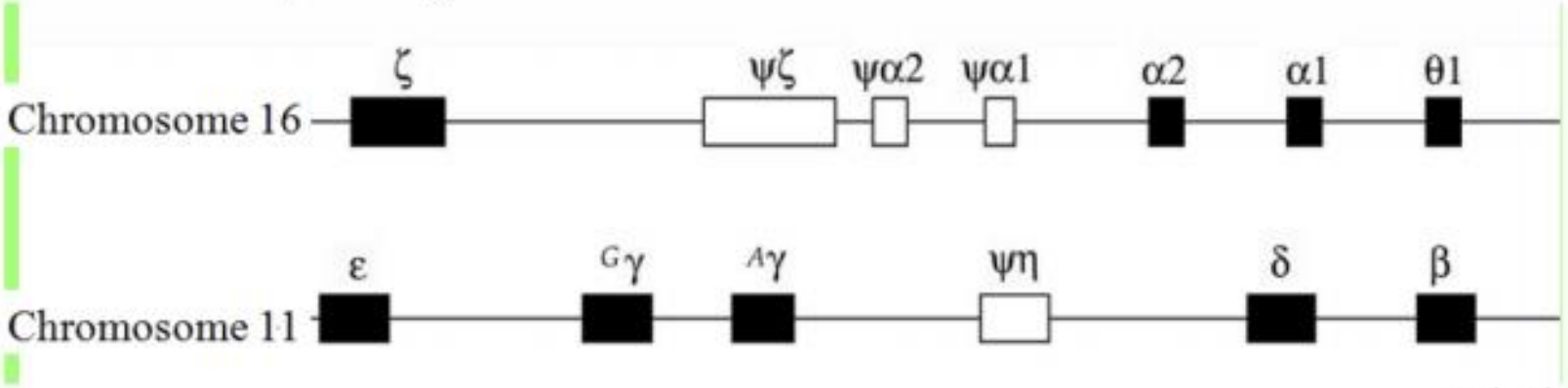


DUPLICAZIONE GENICA: DESTINO DEI GENI DUPLICATI

2. Alcune copie scompaiono. Pseudogeni non processati



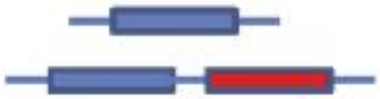
Molti esempi nelle globine



Pseudogene	TATA box	Init. codon	Frame shift	Premature stop	Essential amino acid	Splice GT/AG	Stop codon	Polyadenilation signal AATAAA
Human ψα1		+	+	+	+	+	+	+
Human ψζ1				+				
Mouse ψα3	+		+	+		+		
Mouse ψα4			+		+			
Mouse βh3	?	+	+	+	+	+	?	?
Goat $\psi\beta^x$	+		+	+	+	+	+	+
Goat $\psi\beta^z$	+		+	+	+	+	+	+
Rabbit ψβ2			+	+	+	+		

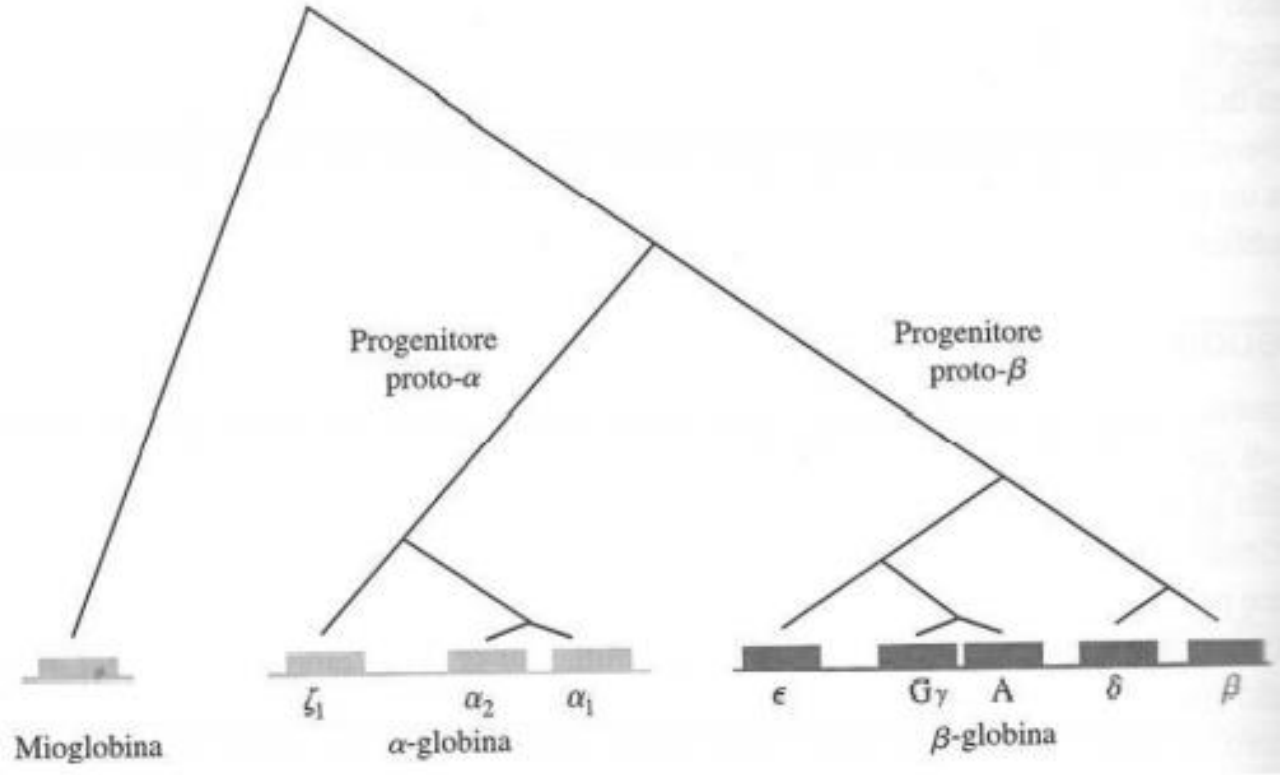
DUPLICAZIONE GENICA: DESTINO DEI GENI DUPLICATI

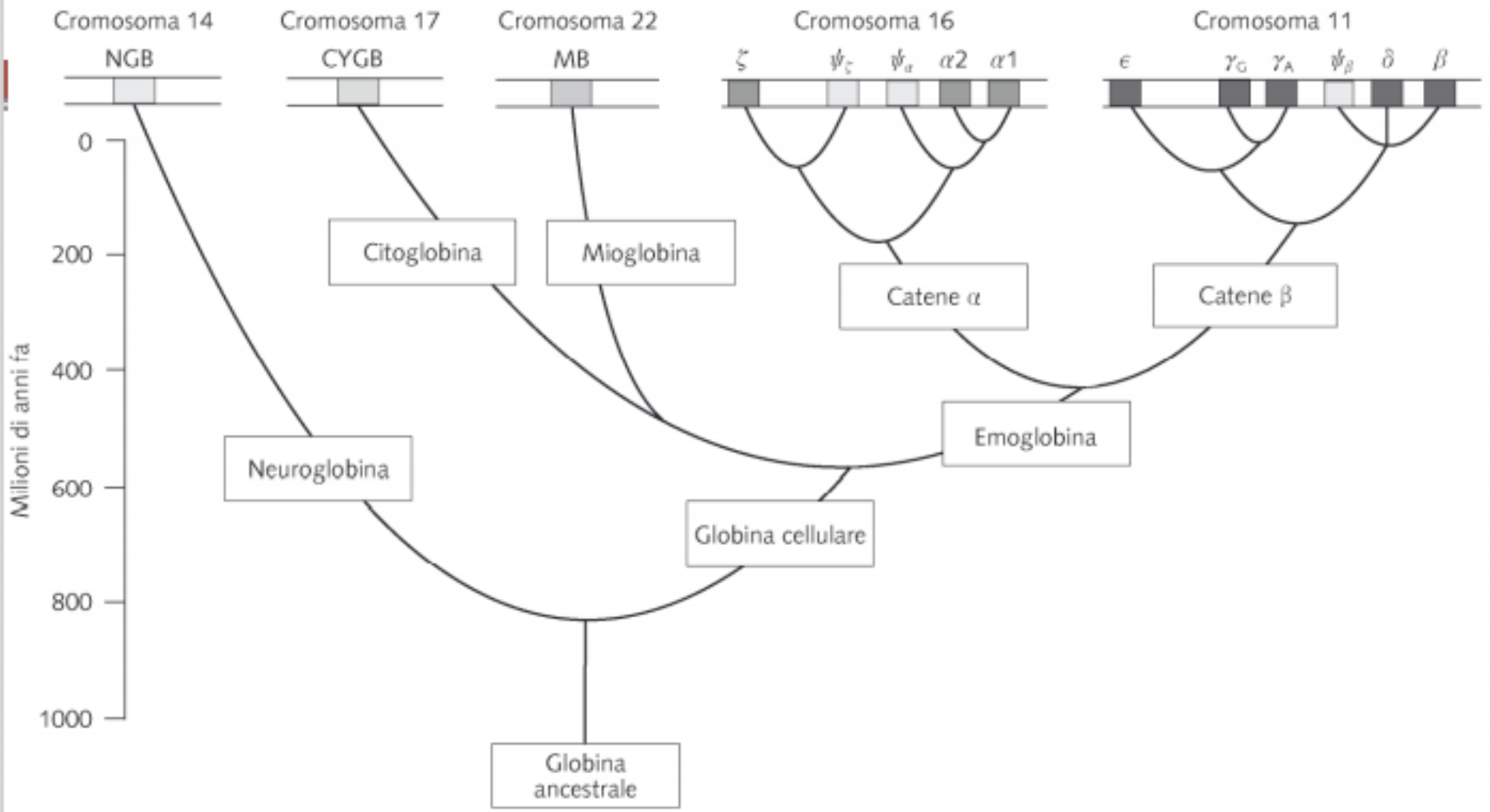
3. Alcune copie evolvono con una nuova funzione.

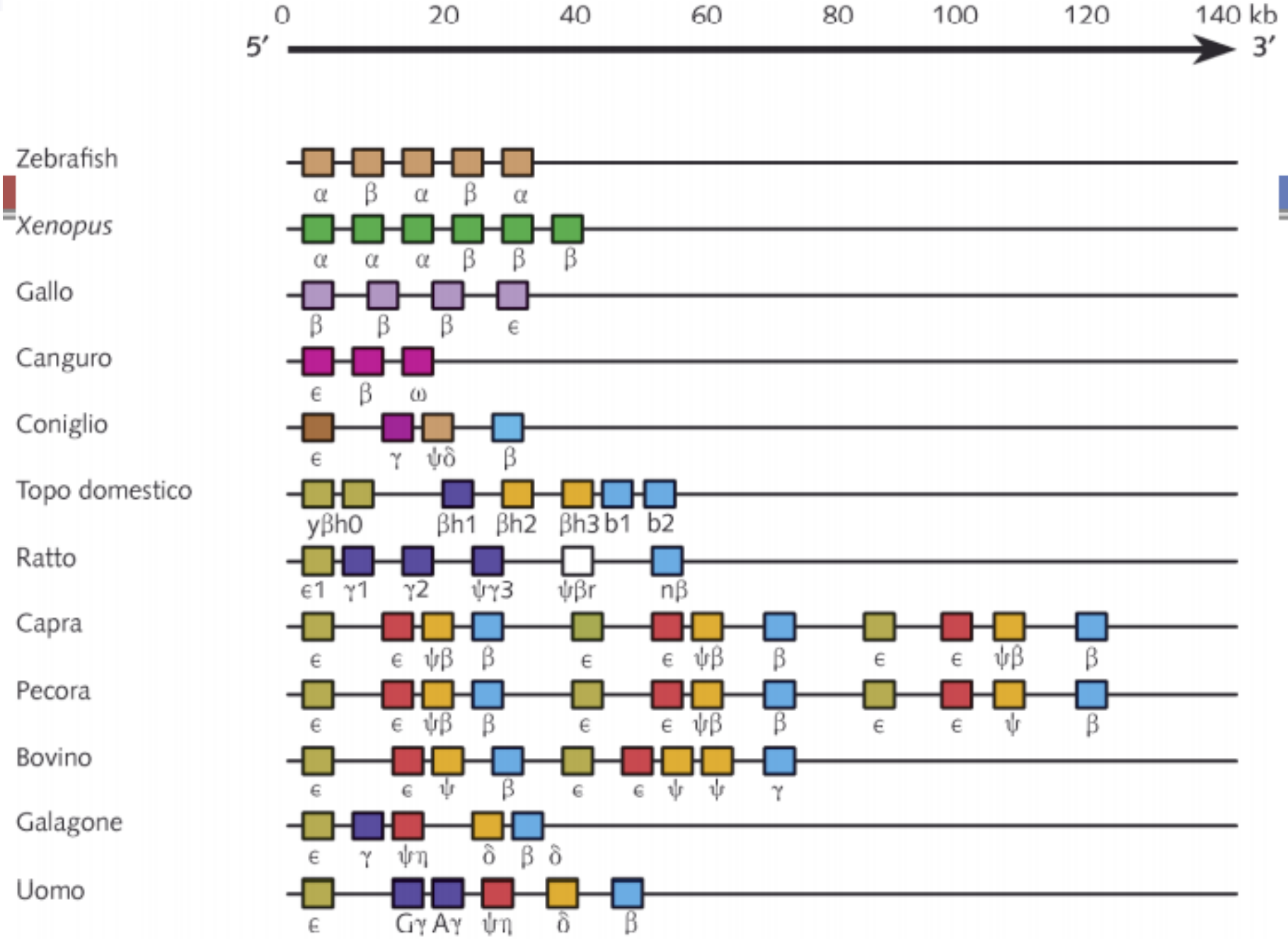


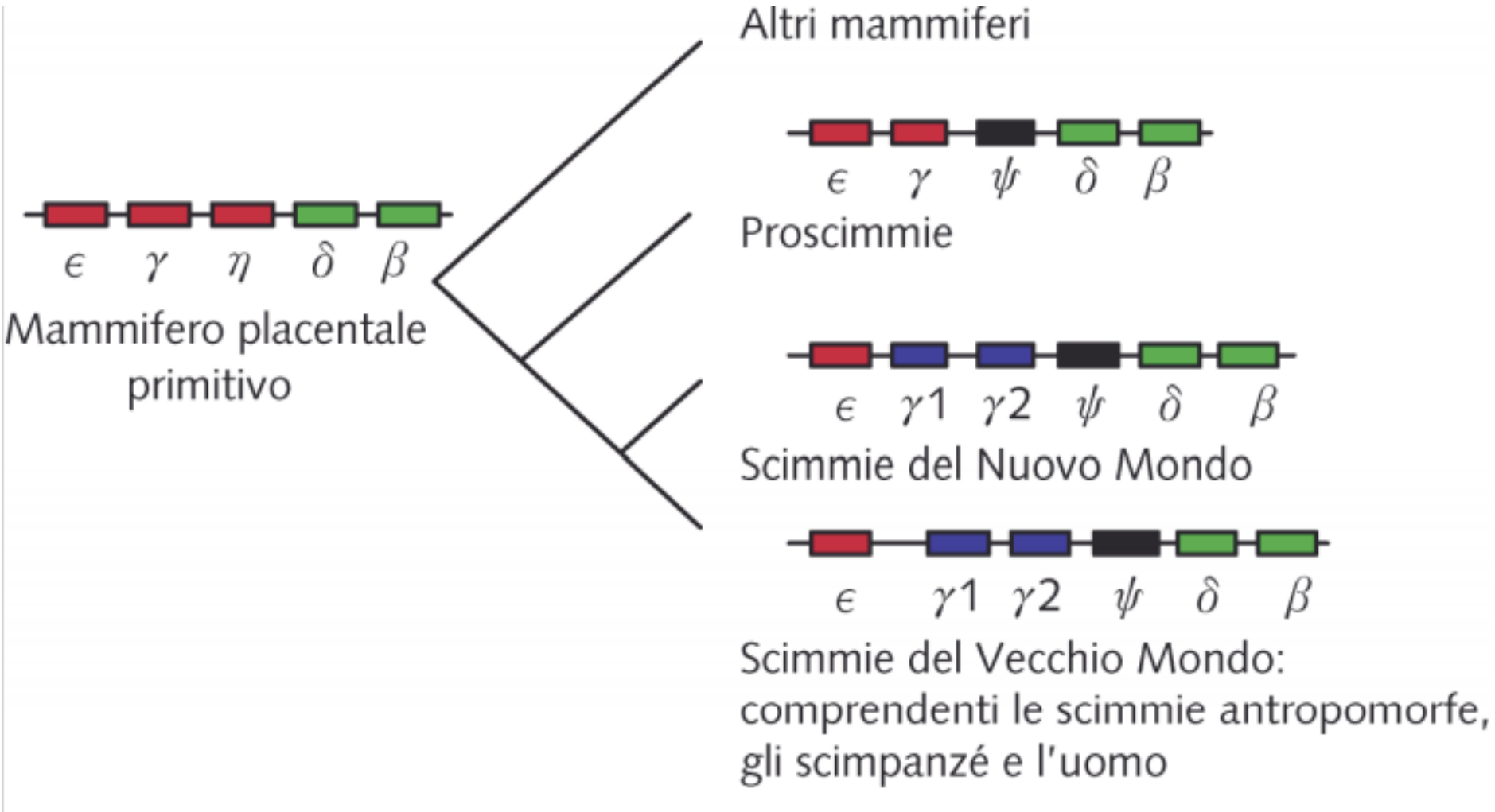
Rosso e blu sono PARALOGHI

Figura 7.9
Vi sono, per esempio nella nostra specie, diversi geni per le globine, probabilmente derivati da duplicazioni avvenute nel passato. Questo albero mostra la "storia" dei geni per le globine, che spiega la presenza dei paraloghi.

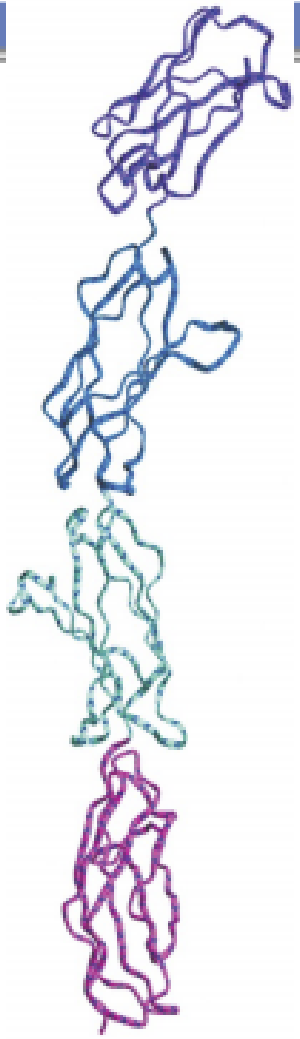



Cromosoma 14
NGB
Cromosoma 17
CYGB
Cromosoma 22
MB
Cromosoma 16
ζ ψζ ψα α2 α1
Cromosoma 11
ε γG γA ψβ δ β
0
200
400
600
800
1000
Milioni di anni fa
Citoglobina
Mioglobina
Catene α
Catene β
Emoglobina
Neuroglobina
Globina cellulare
Globina ancestrale


0
20
40
60
80
100
120
140 kb
5′
3′
Zebrafish
α β α β α
Xenopus
α α α β β β
Gallo
β β β ε
Canguro
ε β ω
Coniglio
ε γ ψδ β
Topo domestico
yβh0 βh1 βh2 βh3 b1 b2
Ratto
ε1 γ1 γ2 ψγ3 ψβr nβ
Capra
ε ε ψβ β ε ε ψβ β ε ε ψβ β
Pecora
ε ε ψβ β ε ε ψβ β ε ε ψ β
Bovino
ε ε ψ β ε ε ψ ψ γ
Galagone
ε γ ψη δ β δ
Uomo
ε Gγ Aγ ψη δ β


Altri mammiferi
ε γ ψ δ β
Proscimmie
ε γ1 γ2 ψ δ β
Scimmie del Nuovo Mondo
ε γ1 γ2 ψ δ β
Scimmie del Vecchio Mondo: comprendenti le scimmie antropomorfe, gli scimpanzé e l'uomo
ε γ η δ β
Mammifero placentale primitivo

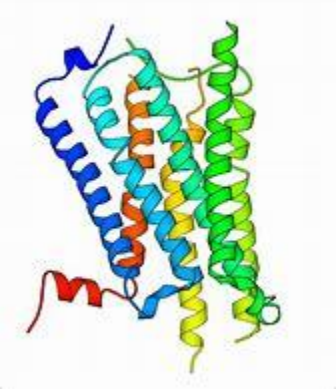
La duplicazione può influenzare singoli esoni

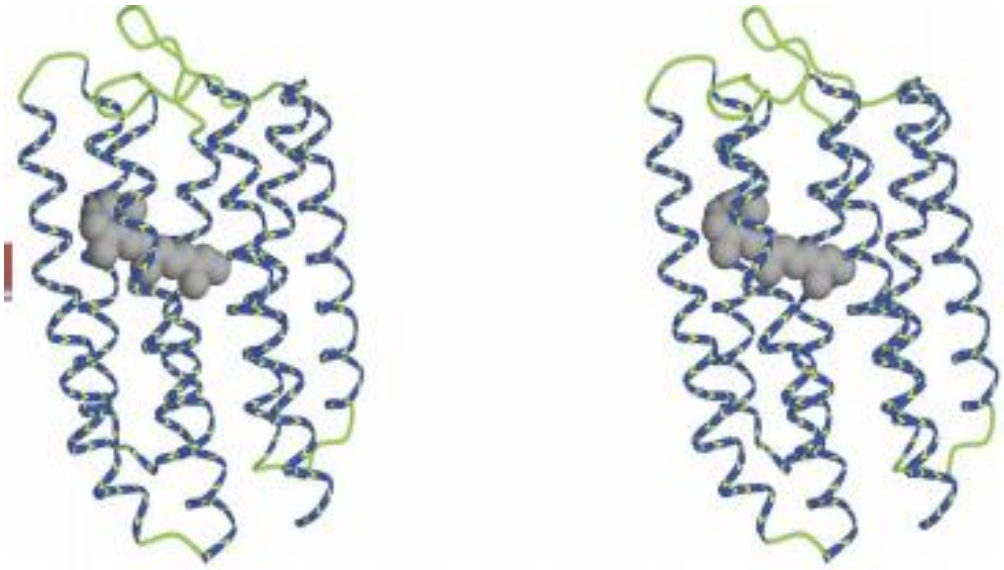


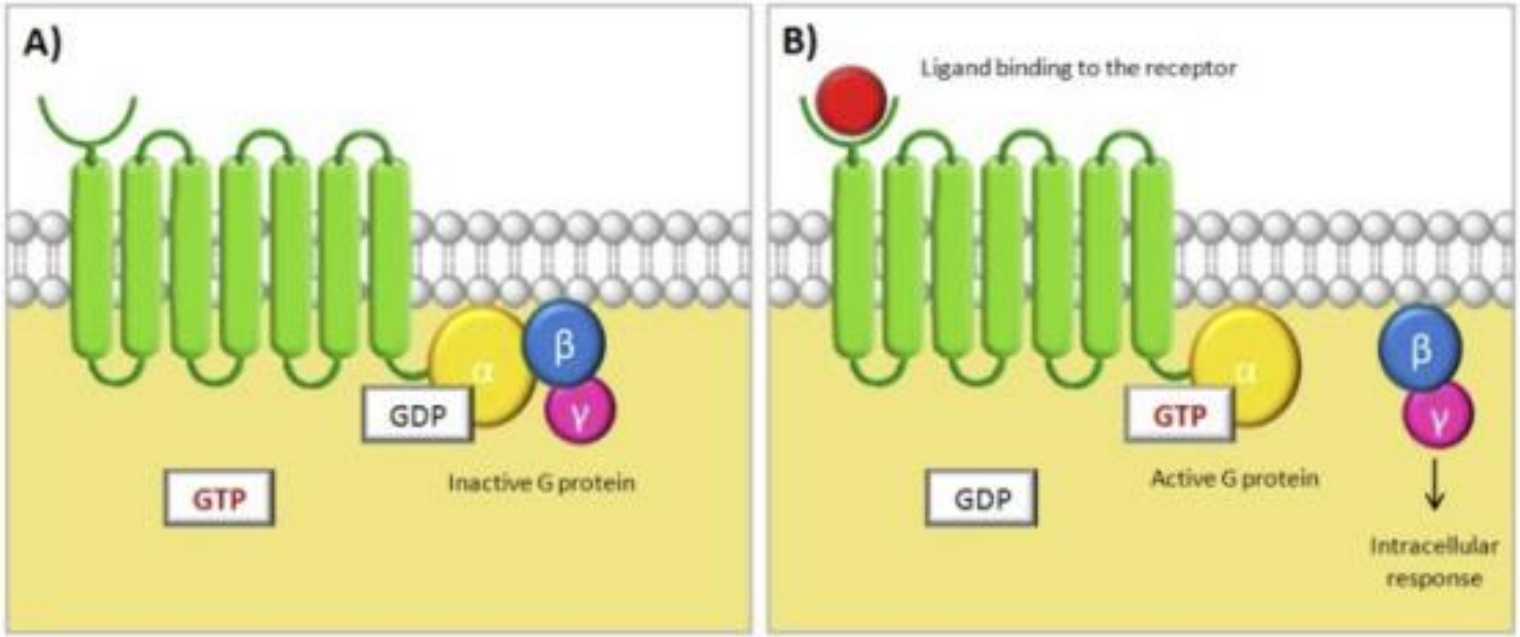
La fibronectina, una grande proteina extracellulare che interviene nell'adesione e nella migrazione cellulare, è una **proteina modulare** contenente più ripetizioni in tandem di tre tipi di dominio denominati FI, F2 e F3.
Codificati da singoli esoni

Espansione delle famiglie

- I recettori accoppiati a proteine C (GPCR) sono una grande famiglia dì proteine transmembrana che intervengono nella trasduzione dei segnali nelle cellule.
- Hanno in comune una struttura contenente 7eliche transmembrana, disposte in una topologia comune
- Questa famiglia ha la stessa età degli eucarioti
- è grande ed eterogenea.
- I genomi dei mammiferi contengono circa 1500-2000 GPCR, che costituiscono circa il 3 - 5% del genoma.
- Il genoma umano ha circa 1000 geni per i **recettori olfattivi**, di cui soltanto il 40% sono attivi.
- Il genoma murino ne ha circa 1300, di cui l'80% sono attivi.






A)
GDP
α
β
γ
GTP
Inactive G protein
B)
Ligand binding to the receptor
GTP
α
β
γ
GDP
Active G protein
Intracellular response

Duplicazione genica: destino dei geni duplicati

3. Alcune copie evolvono con una nuova funzione



Novità evolutiva nell'ornitorinco: capacità di produrre veleno

Vol 453|8 May 2008|doi:10.1038/nature06936 nature

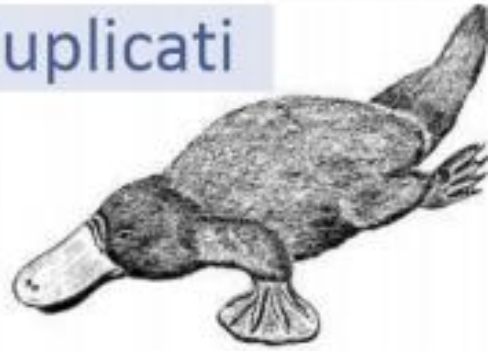
ARTICLES

Genome analysis of the platypus reveals unique signatures of evolution

A list of authors and their affiliations appears at the end of the paper

We present a draft genome sequence of the platypus, *Ornithorhynchus anatinus*. This monotreme exhibits a fascinating combination of reptilian and mammalian characters. For example, platypuses have a coat of fur adapted to an aquatic lifestyle; platypus females lactate, yet lay eggs; and males are equipped with venom similar to that of reptiles. Analysis of the first monotreme genome aligned these features with genetic innovations. We find that reptile and platypus venom proteins have been co-opted independently from the same gene families; milk protein genes are conserved despite platypuses laying eggs; and immune gene family expansions are directly related to platypus biology. Expansions of protein, non-protein-coding RNA and microRNA families, as well as repeat elements, are identified. Sequencing of this genome now provides a valuable resource for deep mammalian comparative analyses, as well as for monotreme biology and conservation.

Duplicazione genica: destino dei geni duplicati



3. Alcune copie evolvono con una nuova funzione.

Le proteine del veleno derivano da copie duplicate neofunzionalizzate di geni già presenti nel genoma.

Una duplicazione simile ma indipendente si osserva nei rettili (convergenza evolutiva!)

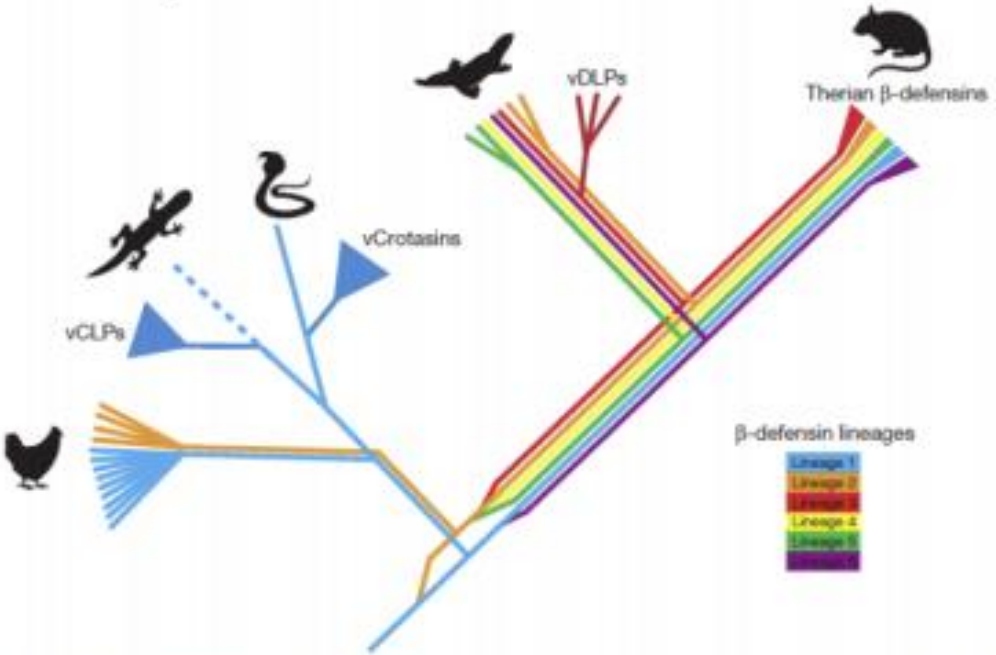
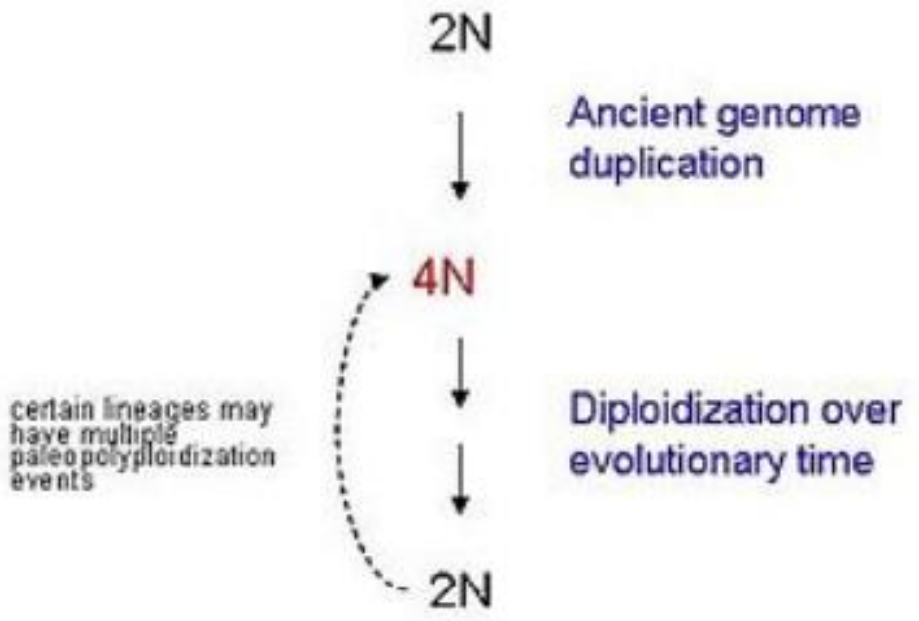


Figure 4 | The evolution of β-defensin peptides in platypus venom gland. The diagram illustrates separate gene duplications in different parts of the phylogeny for platypus venom defensin-like peptides (vDLPs), for lizard venom crotamine-like peptides (vCLPs) and for snake venom crotamines. These venom proteins have thus been co-opted from pre-existing non-toxin homologues independently in platypus and in lizards and snakes[45].

Duplicazioni su larga scala: intero genoma

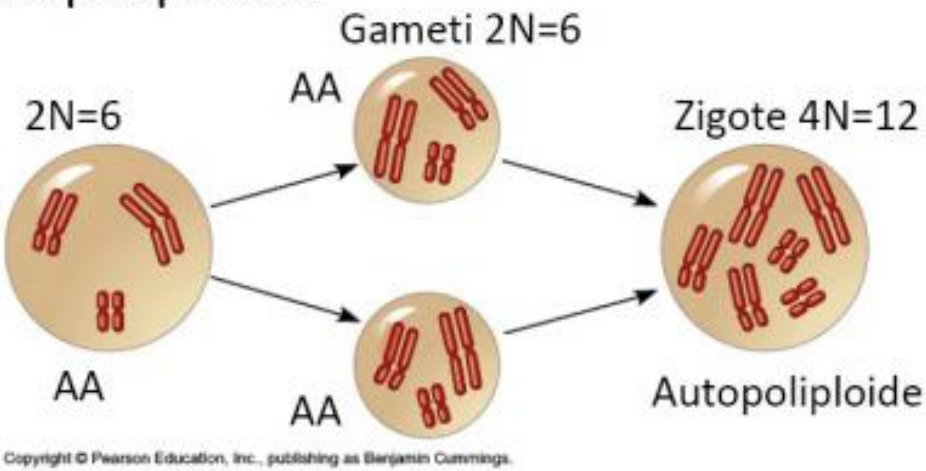
Paleoploidia

Serve come prova la presenza, nell'intero genoma, di coppie di geni omologhi che *si presentano nello stesso ordine (colinearità);* o l'evidenza di un «orologio molecolare» che indica uguali tempi di divergenza in molte coppie di omologhi.



Come si originano le novità evolutive: duplicazioni genomiche

Autopoliploidia

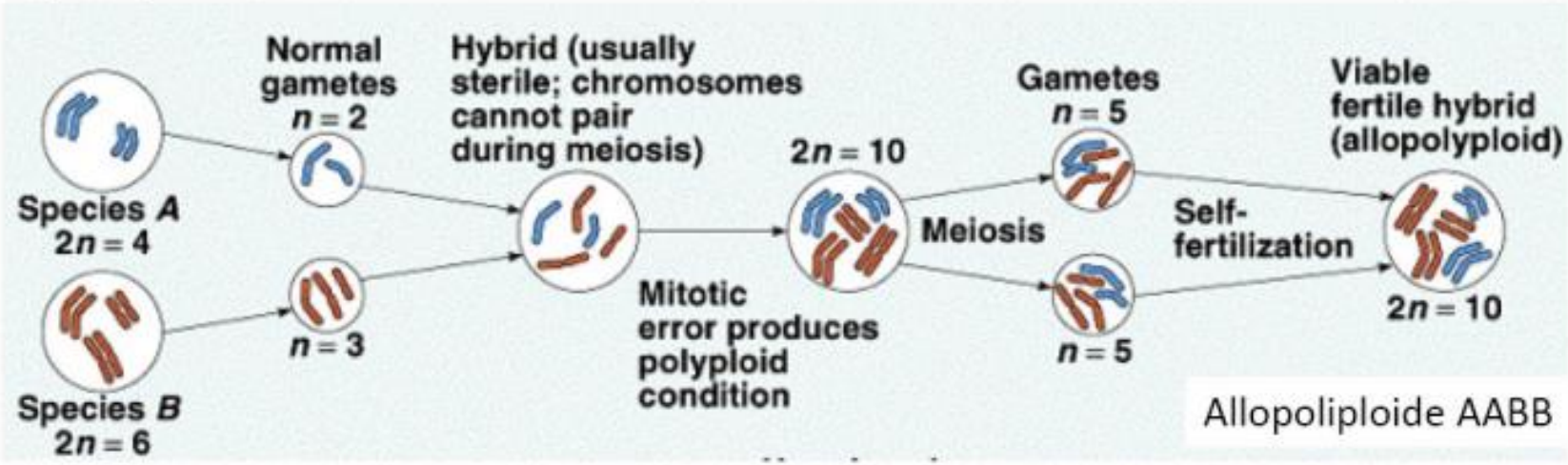


Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

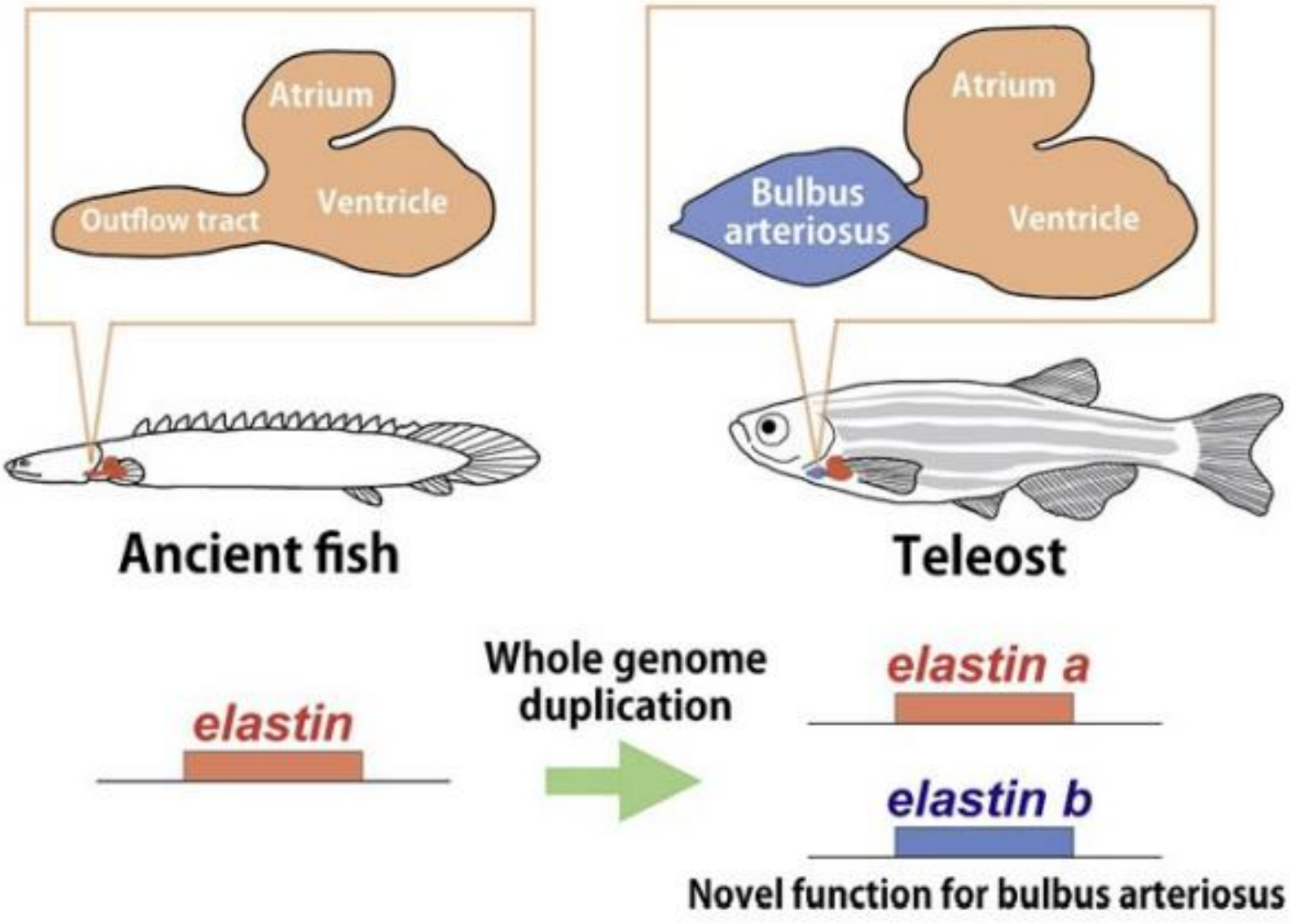
Solo l'incrocio con un altro autopoliploide darà prole fertile

Comune fra le piante

Allopoliploidia



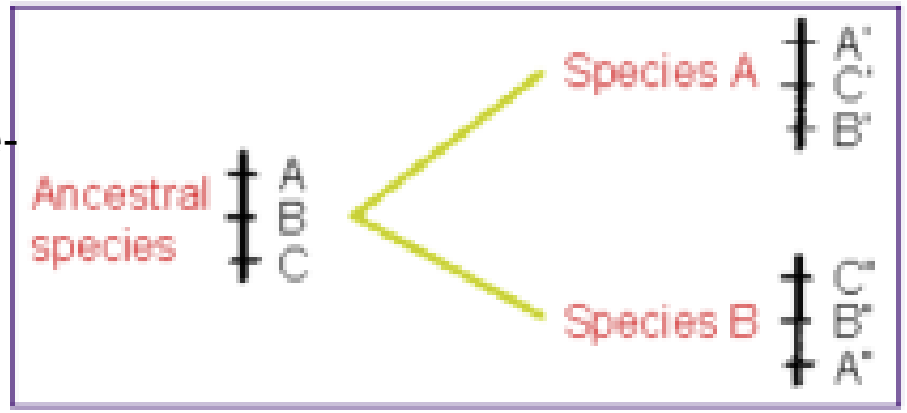
Duplicazioni su larga scala



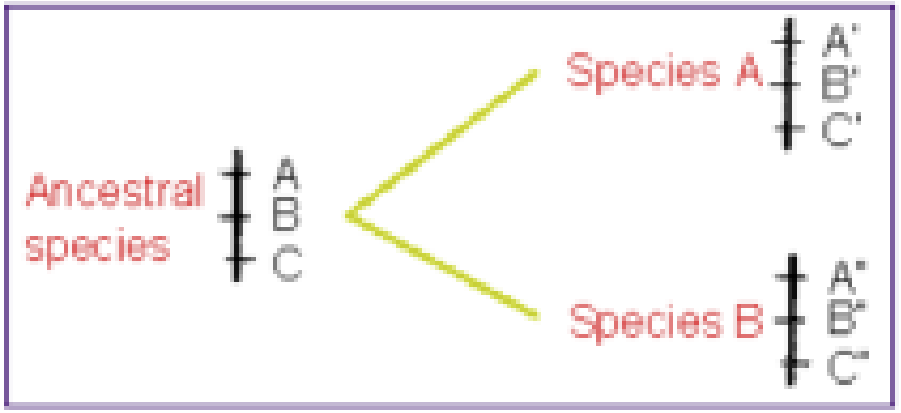
Sintenia

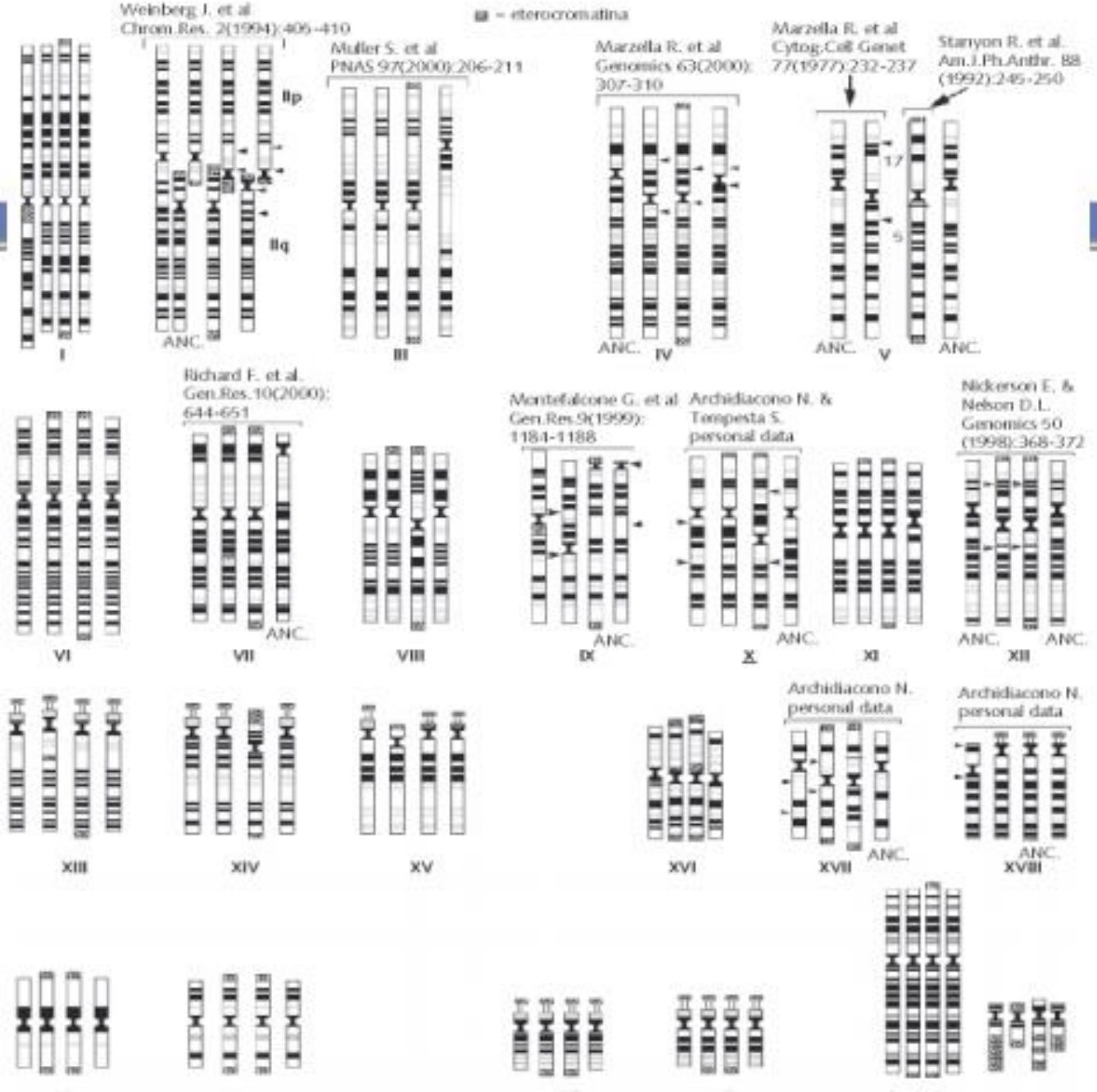
- **Blocchi di sintenia**: regioni del genoma dove l'ordine dei geni è conservato, come risultato della discendenza da un antenato comune
  - Microsintenie possono essere evidenziate con bandeggi cromosomici

➢ **Sintenia:** un set di loci in due specie differenti che è localizzato sullo stesso cromoso-Ma in ciacuna (non necessariamente nello stesso ordine



➢ **Collinerità:** un set di loci in due specie differenti che è localizzato sullo stesso cromosoma in ciascuna ed è conservato nello stesso ordine



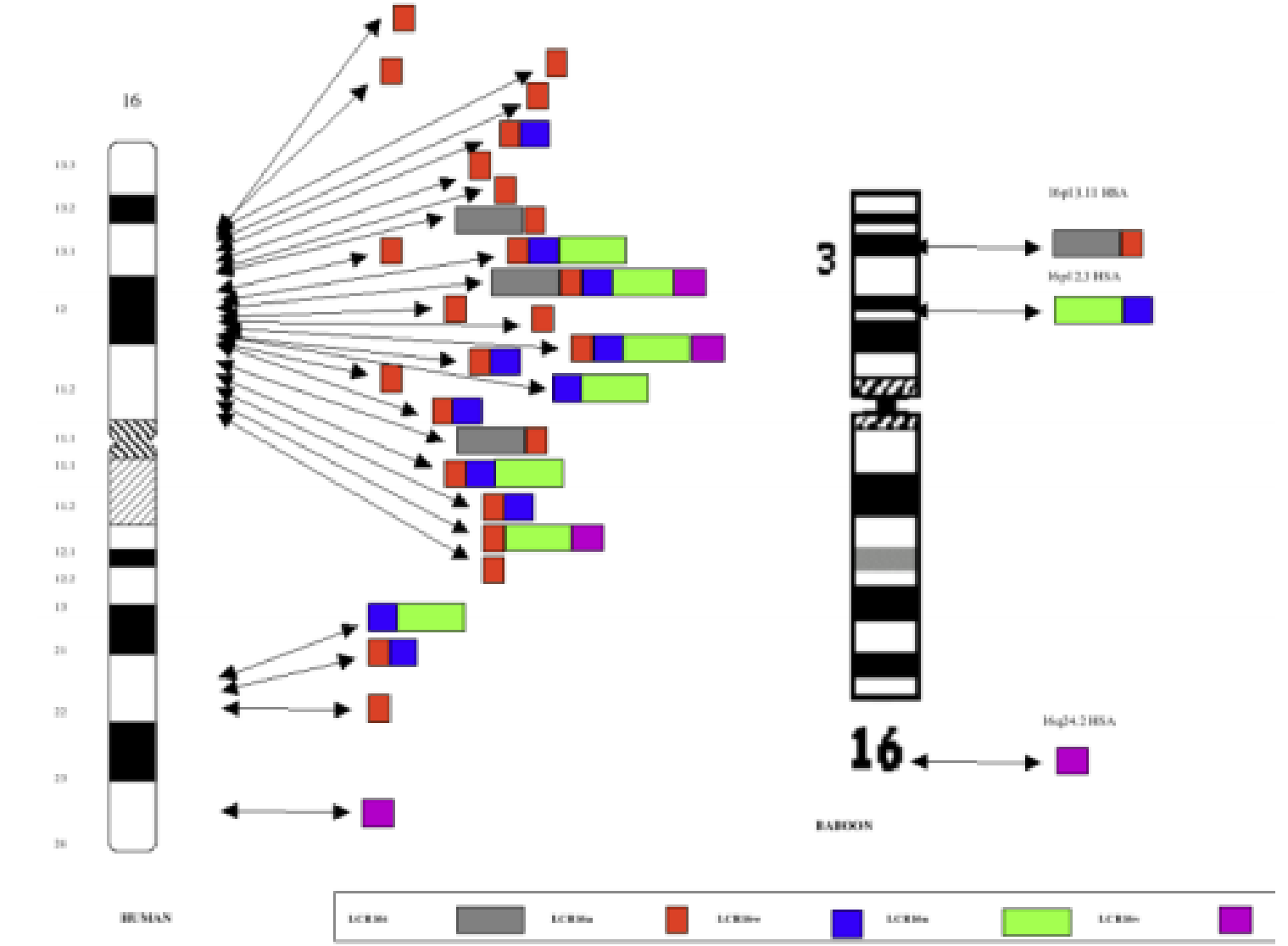

= eterocromatina
Weinberg J. et al
Chrom.Res. 2(1994):406-410
Muller S. et al
PNAS 97(2000):206-211
Marzella R. et al
Genomics 63(2000):
307-310
Marzella R. et al
Cytog.Cell Genet
77(1977):232-237
Stanyon R. et al.
Am.J.Ph.Anthr. 88
(1992):245-250
IIp
IIq
17
5
ANC.
I
III
IV
V
Richard F. et al.
Gen.Res.10(2000):
644-651
Montefalcone G. et al
Gen.Res.9(1999):
1184-1188
Archidiacono N. &
Tempesta S.
personal data
Nickerson E. &
Nelson D.L.
Genomics 50
(1998):368-372
VI
VII
VIII
IX
X
XI
XII
Archidiacono N.
personal data
Archidiacono N.
personal data
XIII
XIV
XV
XVI
XVII
XVIII

Duplicazioni su larga scala

I genomi di molte specie contengono duplicazioni di regioni multigeniche, la cui lunghezza varia da specie a specie.

- Le duplicazioni segmentali su grande scala sono una componente importante nella differenza tra genoma dell'uomo e genoma dello scimpanzé, interessando circa il 2,7% del genoma.

Duplicazioni su larga scala

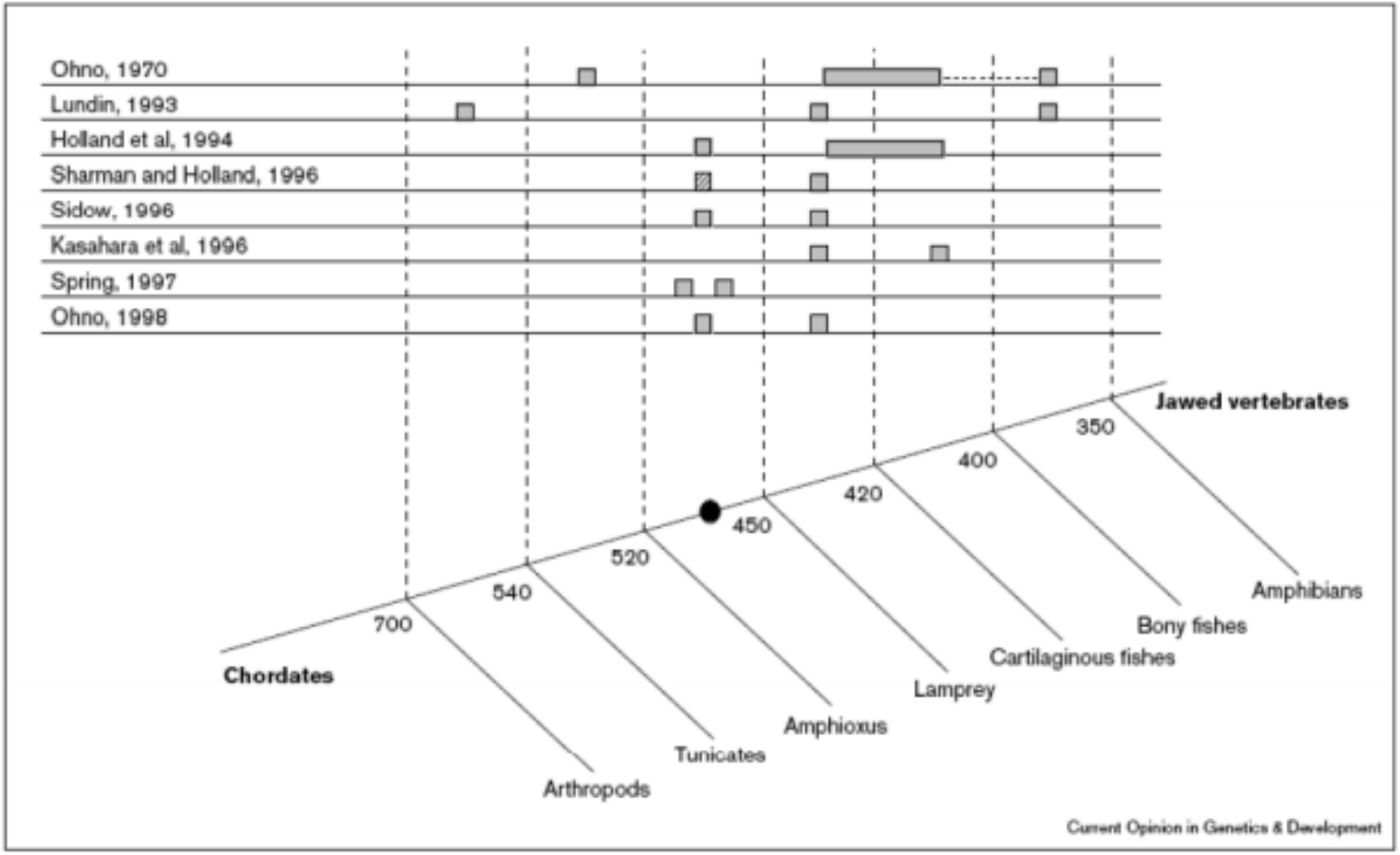


Yeast

Il genoma del lievito subì una duplicazione circa 10^8 anni fa. Gli effetti sono oscurati dai successivi riarrangiamenti cromosomici e dalla massiva perdita di materiale duplicato.
Comunque il genoma del lievito contiene 55 regioni duplicate, lunghe in media 55 kb, che coprono insieme il ~ 50% del genoma e comprendono 376 coppie di geni omologhi.

Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis Susumu Ohno



Da Skrabanek L, Wolfe KH. 199ε

Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis

1R: il genoma raddoppia, **2R**: il genoma quadruplica (aggiungiamo una fish-specific genome duplication: **3R** per la linea dei pesci)

Ancora molta discussione sul numero e la datazione delle duplicazioni

Si vedono delle tracce della 2R (esempio: geni Hox), ma altre sono cancellate da eventi successivi

Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis

Si vedono delle tracce della 2R (esempio: geni Hox)

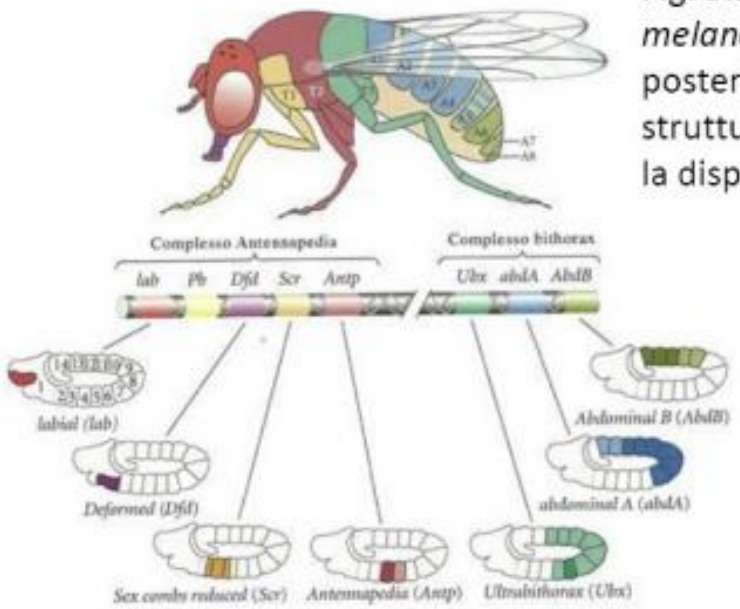
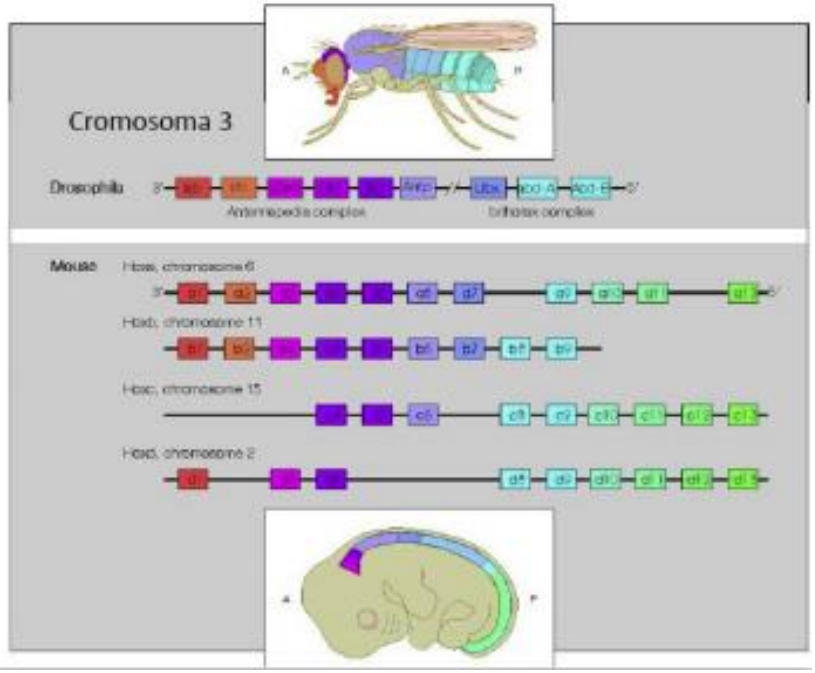


Fig. 9.10: L' espressione dei geni Hox nell'embrione di *D. melanogaster* (in basso) in specifici domini dell'asse antero-posteriore del corpo definisce la posizione di specifiche strutture lungo lo stesso asse nell'adulto (in alto). Al centro la disposizione dei geni Hox sul cromosoma



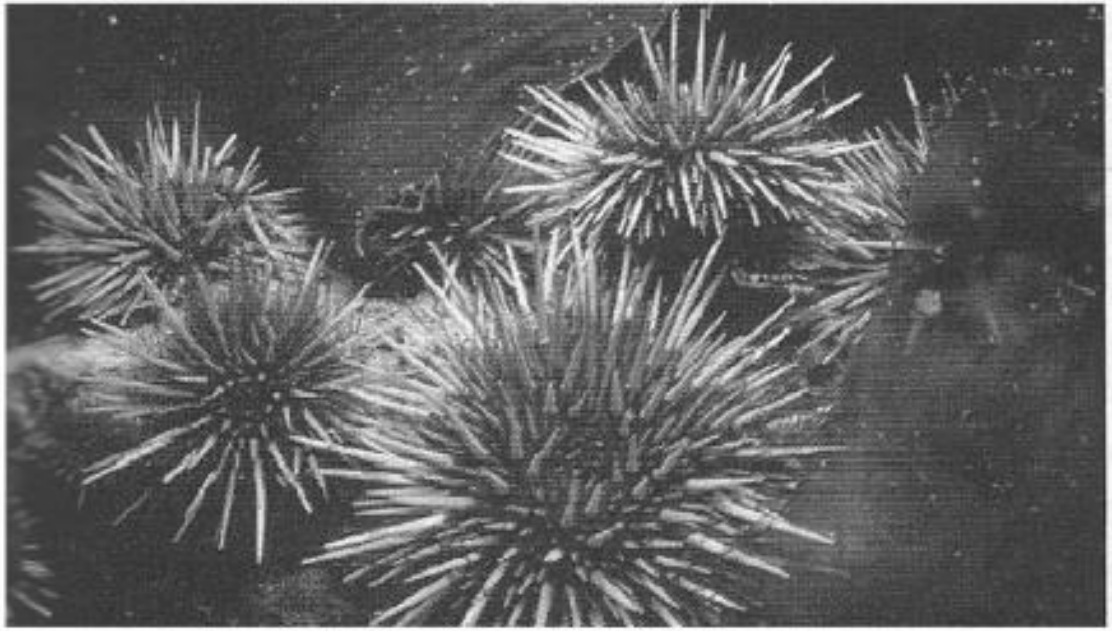
>Complessità
>Nr geni omeotici

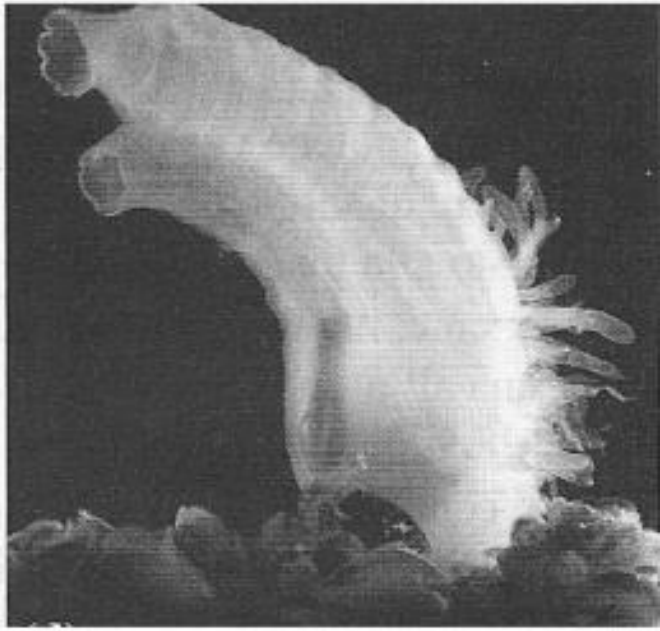
4 clusters nei vertebrati
Geni con similarità di sequenza con quelli di Drosophila

Come si originano le novità evolutive: duplicazioni genomiche

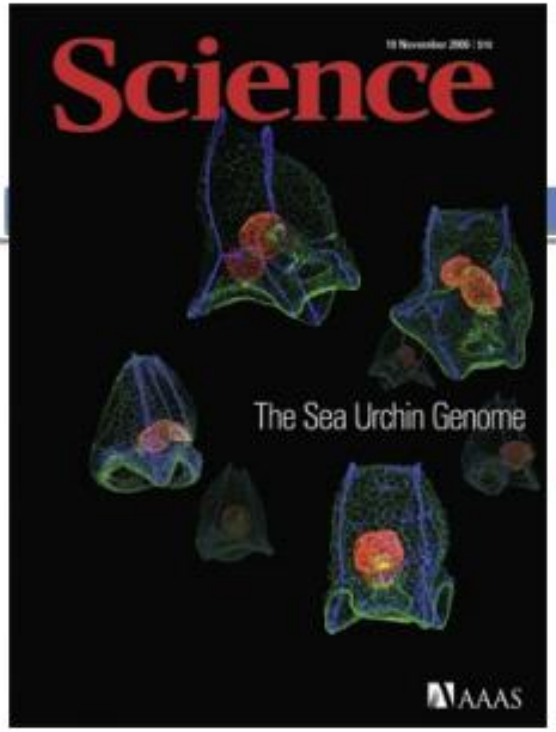
2R Hypothesis

Si vedono delle tracce della 2R (esempio: sequenziamento di genomi)





Il sequenziamento dei genomi di riccio di mare (*Strongylocentrotus purpuratus*, echinoderma) e dell'ascidia (*Ciona intestinalis*, tunicato) mostrano numerosi geni in **proporzione 1 a 4 rispetto agli omologhi vertebrati**





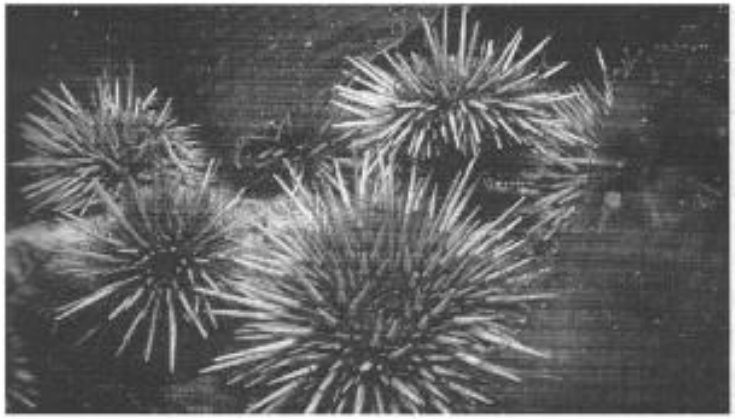
The sea urchin, *Strongylocentrotus purpuratus*, has been a premier model for understanding the gene networks that underlie development. Shown are sea urchin pluteus larvae stained to reveal the ectoderm (green), mesoderm

A group of about 30 tadpole-stage *Ciona intestinalis* embryos 12 to 14 hours after fertilization. The embryos were electroporated at the one-cell stage with a fusion gene containing the 5' cis-regulatory region of the *Brachyury* gene attached to a green fluorescent protein (GFP) reporter gene. The *Brachyury* regulatory DNA directs GFP expression in the

Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis

Tuttavia....



~ 23.000 geni



~ 24.000 geni

Nel complesso il riccio di mare ha tanti geni quanti l'uomo: **dopo le duplicazioni genomiche molte copie non sopravvivono come geni funzionanti**

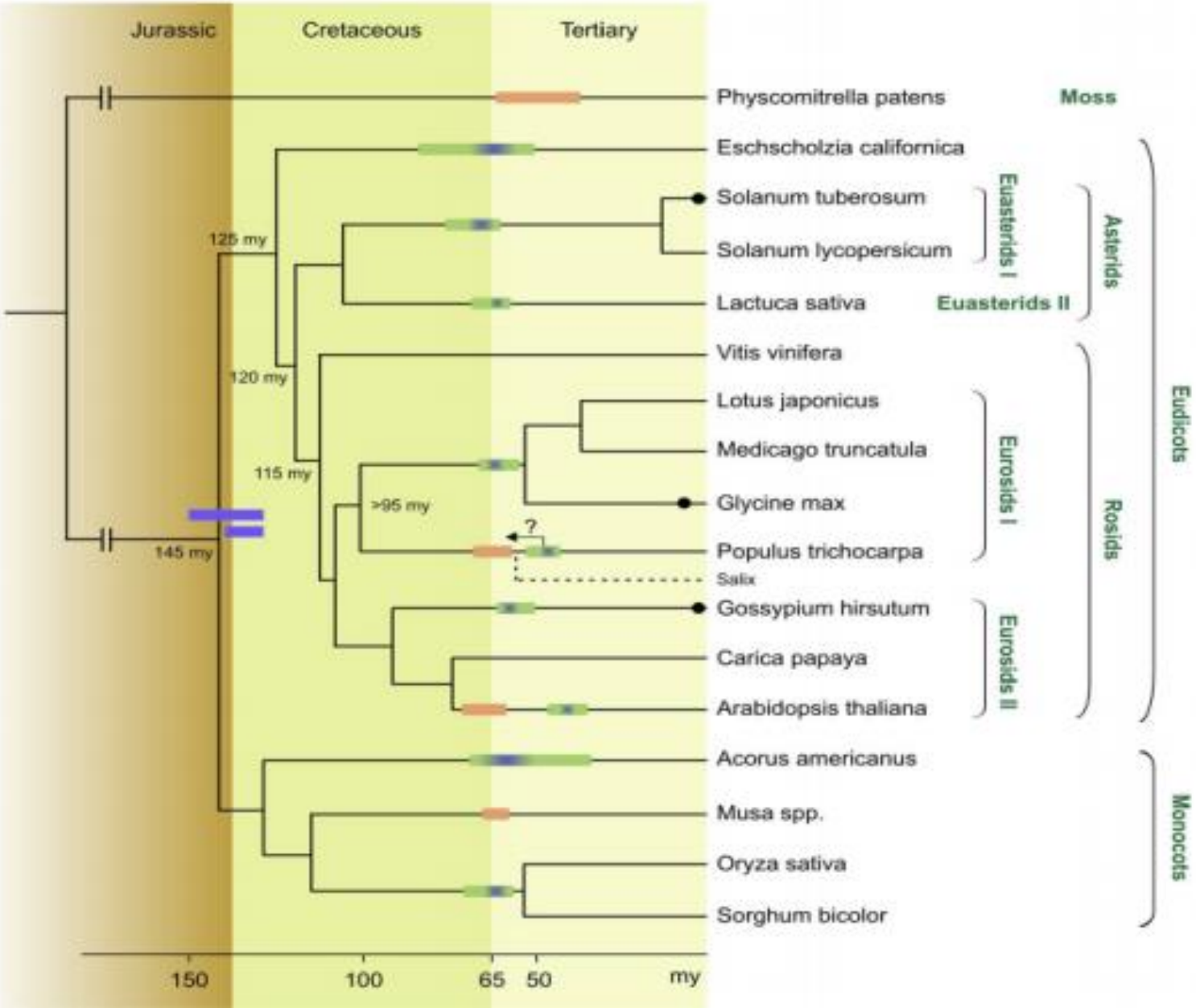
Nelle piante

Il sequenziamento di *A. thaliana* ha rilevato almeno 2 duplicazioni genomiche.
La maggior parte delle piante sono poliploidi.

In studi condotti sulla flora artica, si è osservato che la frazione di specie vegetali diploidi e poliploidi aumenta all'aumentare della latitudine.

Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event

Jeffrey A. Fawcett[a,b,1], Steven Maere[a,b,1], and
[a]Department of Plant Systems Biology, Flanders Institute
Ghent University, 9052 Gent, Belgium
Communicated by Marc C. E. Van Montagu, Ghent Univ

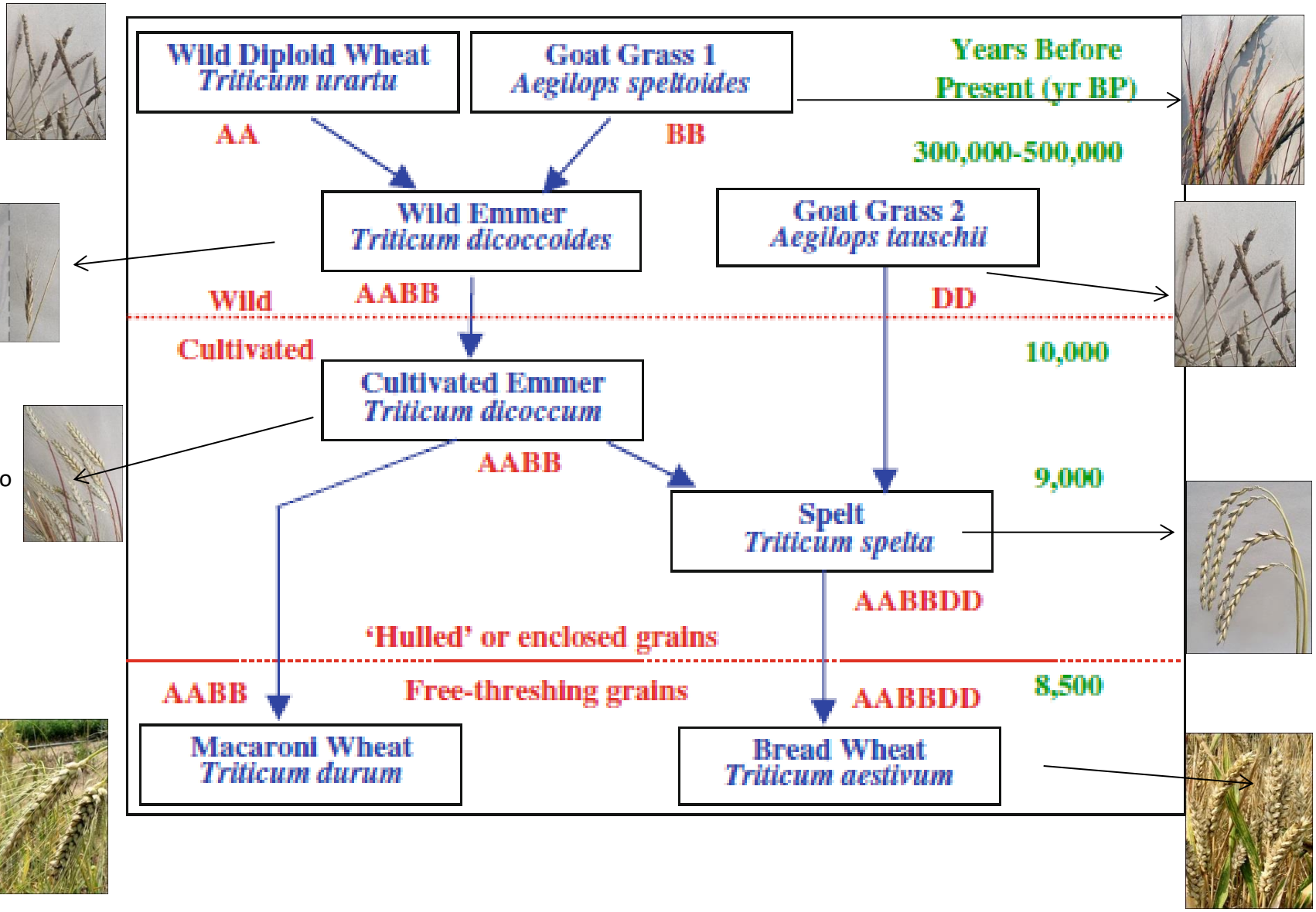


Poliploidia nel frumento

Triticum monococcum	farro piccolo, farro monococco, einkorn	AA
Triticum dicoccum	farro medio, farro dicocco	AABB
Triticum turgidum	grano duro, frumento duro	AABB
Triticum spelta	farro grande, granfarro, spelta	AABBDD
Triticum aestivum	grano tenero, frumento tenero	AABBDD
Triticosecale	triticale	AABBRR

legenda: A, genoma del frumento diploide originale o di una specie affine; B, genoma di una graminacea selvatica, *Aegilops speltoides*, o di una specie affine; D, genoma di un'altra graminacea selvatica, *T. tauschii*, o di una specie affine; R, genoma di segale *(Secale cereale)*.

Come si originano le novità evolutive: duplicazioni genomiche



POLIPLOIDIA NEI MAMMIFERI


Tympanoctomys barrerae

Nel 1999 viene identificato un mammifero poliploide, il topo rosso di Mendoza (Argentina), 4N= 102.

Il successo evolutivo di questa specie poliploide è legato al mantenimento di un unico set di cromosomi sessuali XY.

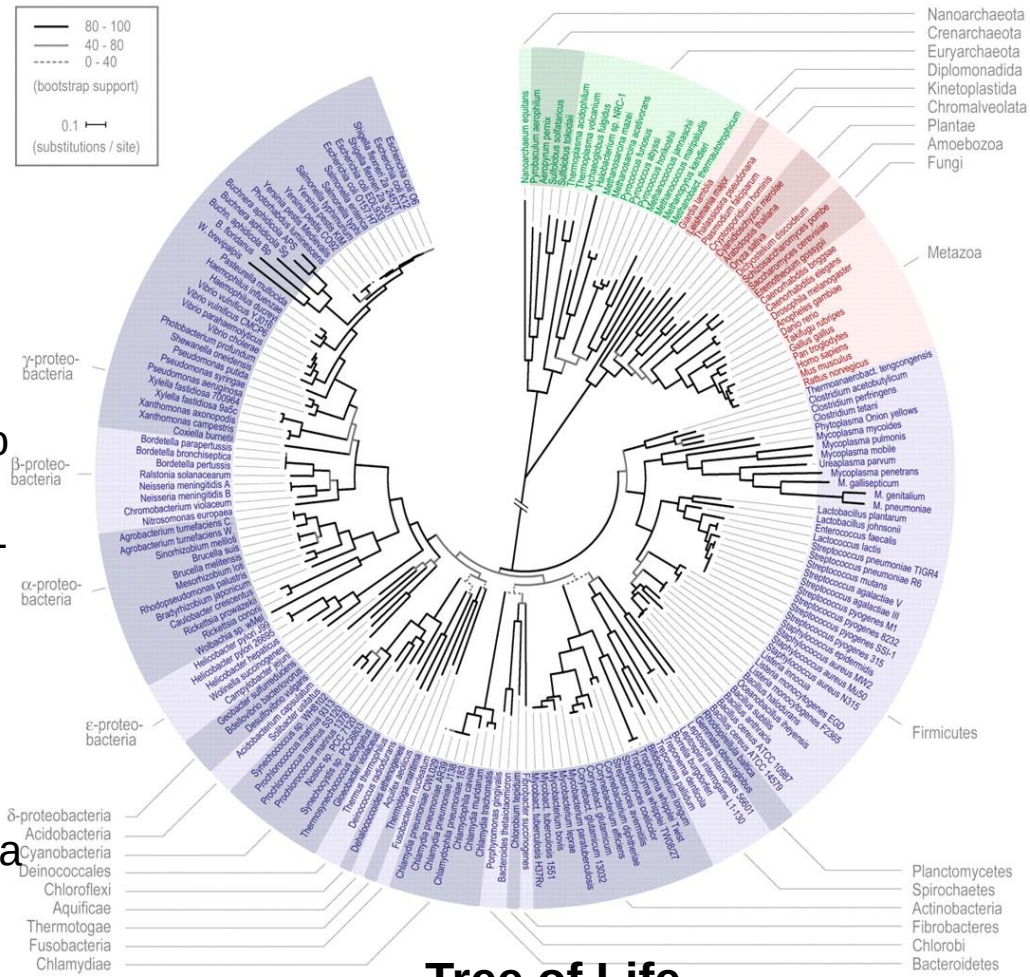


Genomi: albero della vita

- La classificazione molecolare suddivide i procarioti in due domini:
  - **Batteri** e **Archebatteri,**
  - **Eucarioti**

formano i tre rami principali dell'Albero delle Vita (**Tree of Life**).

- Gli Archea presentano una serie di caratteristiche e proprietà che li differenziano sostanzialmente dagli Eubacteria:
  - le proteine "informazionali" (cioè quelle implicate in replicazione, trascrizione e traduzione) sono più simili alle controparti eucariotiche che a quelle eubatteriche;
  - hanno proteine simili a istoni per l'impacchettamento del DNA;
  - differente struttura del promotore (TATA box a -30 rispetto al TSS come negli eucarioti, mentre nei eubatteri il TATA box è a -10);
  - RNA polimerasi simile in struttura a RNA pol II eucariotica;
  - la traduzione inizia per Met (formil-Met negli eubatteri);
  - alcuni geni per tRNA e rRNA possono contenere introni



Tree of Life

Blu. eubacteria, verde archaea, rosso Eukaryotes.

Da: Ciccarelli et al. (Science 311: 1283-1287)

I. IL GENOMA DEI PROCARIOTI

- Mondo microbico è il prodotto di 3.7 miliardi di anni di evoluzione
- I microrganismi hanno quindi esplorato :

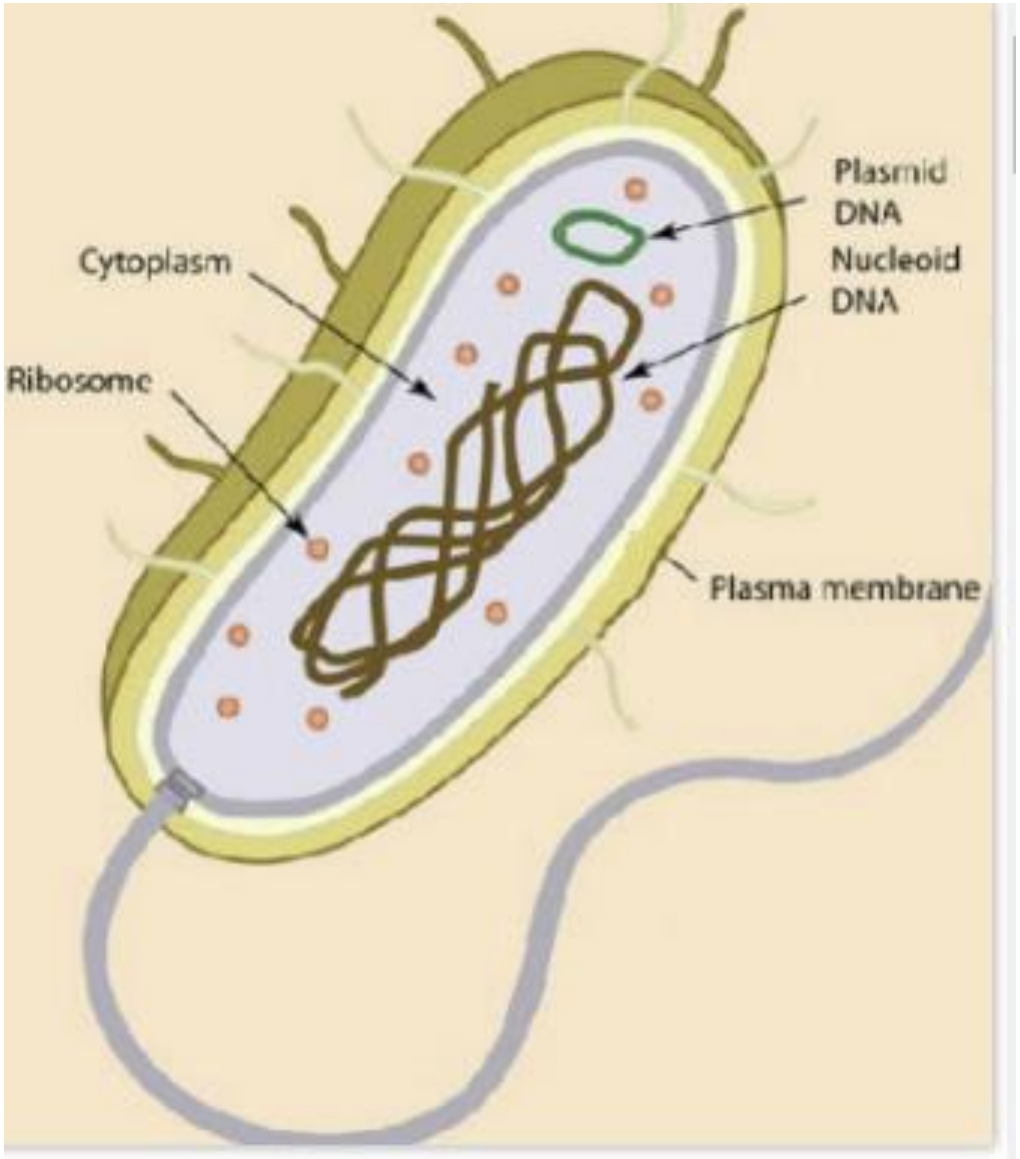
Una moltitudine di nicchie ecologiche
tutte le varie combinazioni di condizioni climatiche

- ENORME SFRUTTAMENTO DELLE CAPACITA' BIOCHIMICHE DI UNA CELLULA
- ENORME DIVERSITA' DI SEQUENZA

- Una delle maggiori differenze con gli eucarioti è costituta dalla **DIVERSITA' DELLA SEQUENZA PER LA STESSA FUNZIONE GENICA**
- E' proprio questa diversità di sequenza che fornisce diverse proprietà
  - **cinetica**
  - **regolazione**
  - **stabilità delle macromolecole** (proteine, enzimi etc) che sono importanti per lo sfruttamento globale degli elementi della terra

Procarioti

- Organismi unicelulari
  - Nessun organello legato alla membrana
    - es: nessun nucleo o mitocondri
  - DNA fluttuante libero
- I batteri sono procarioti



Quante specie di procarioti ci sono?

- Per gli animali e le piante viene considerata una specie, una popolazione di individui che può riprodursi dando origine ad individui riproducibili
- I batteri sono aploidi e si riproducono asessualmente .

Come si definisce quindi una specie?

- **Concetto di specie nei procarioti:**

• **due ceppi appartengono alla stessa specie** quando la % di ibridazione DNA -DNA è uguale o superiore al 70%

• due ceppi appartengono alla stessa specie quando le sequenze di RNA 16S hanno un'omologia del 97% (meno del 3% di diver-genza)

Specie ---Generi---Famiglie---Ordini---Classi---Phyla---- Dominio

Batteri Archea Eucarioti

Quante specie ci sono?

Stime indicano chr il numero di specie procariote oscilla da 10^5 to 10^7
Hammond (1995) in Microbial Diversity and Ecosystem Function 29–71.

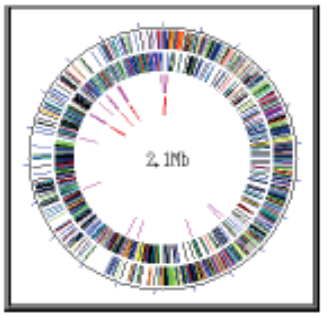
Qualche numero…..molto provvisorio Bergey's, Manual 2005

	Batteri	Archea	Totale
Dominio	1	1	2
Phyla	25	4	29
Classe	34	9	43
Ordine	78	13	91
Famiglie	230	23	243
Genere	1227(871)	79(69)	1306(941)
Specie	6740(5007)	217(289)	7029(5224)

* si sono identificati oltre 15 nuova Phyla tramite sequenziamento batteri NON COLTIVABILI



IL GENOMA DEI PROCARIOTI: organizzazione



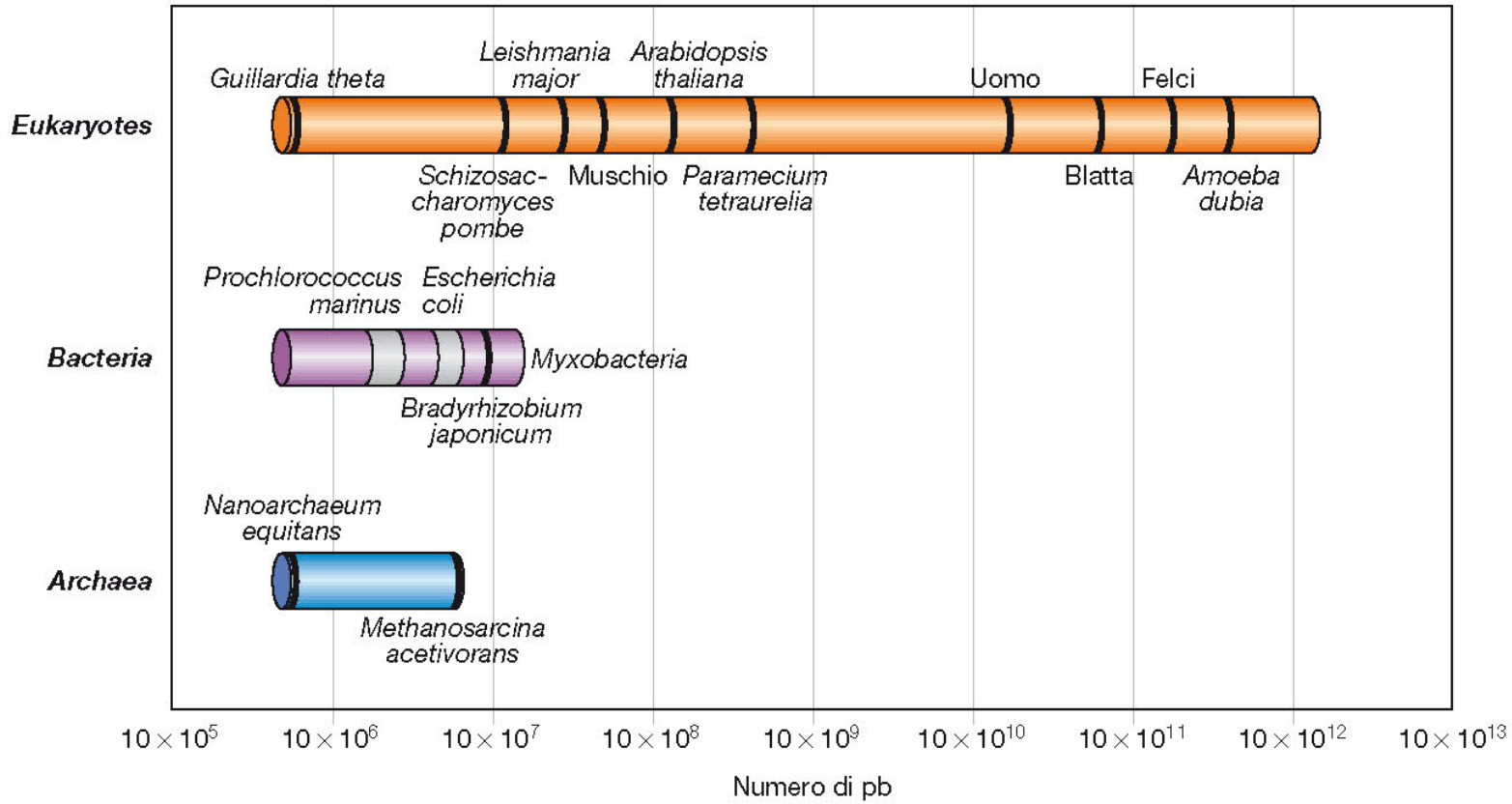
1. Non è provvisto di una membrana nucleare che lo separa dal citoplasma.
2. Ha un **unico genoma circolare** costituito da una doppia elica di DNA (0.5-10 Mbp, 500-10,000 geni)
3. *E. Coli* **= 4x10^6 bp** di cui il **99%** è costituito da sequenze codificanti o regolatrici e solo **l'1%** da sequenze ripetute non codificanti.
4. I **geni sono organizzati in operoni,** ovvero unità funzionali, in cui i geni correlati per attività (es. enzimi di una via metabolica) sono posti l'uno accanto all'altro sul cromosoma, trascritti in un unico mRNA e regolati in blocco

ORGANIZZAZIONE DEL GENOMA BATTERICO

Specie batterica	Numero di cromosomi e organizzazione	Numero di plasmidi
Agrobacterium tumefaciens	uno lineare + uno circolare	due circolari
Bacillus subtilis	uno circolare	
Bacillus thuringiensis	uno circolare	sei
Borrella burgdorferi	uno lineare	>17 circulari + lineari
Buchnera sp.	uno circolare	due circolari
Deinococcus radiodurans	due circolari	due circolari
Escherichia coli	uno circolare	
Vibrio cholerae	due circolari	

Cromosomi circolari e cromosomi multipli nei batteri

Quanti geni nei procarioti?



- Dall'anali dei genomi sequenziati **il numero dei geni**
- **Procarioti** (Batteri e Archea) varia **da 500 a 10.000**
- **Eucarioti** da **2.000 a 35.000.**
- **queste stime non includono i geni che sono trascritti in sRNAs**.
- A livello di dimensioni del genoma vediamo però che i batteri arrivano ad un massimo di 10 Mb

I genomi dei procarioti hanno alcune caratteristiche inattese

DIMENSIONE DEL GENOMA E NUMERO DI GENI IN ALCUNI PROCARIOTI

	Dimensione del genoma (Mb)	Numero di geni
Archaea		
Archaeoglobus fulgidis	2,17	2493
Methanococcus jammaschii	1,66	1738
Thermoplasma acidophilium	1,56	1509
Eubacteria		
E. coli	4,64	4397
Bacillus subtilis	4,21	4212
H. influenzae	1,83	1791
Aquifex aeolicus	1,55	1552
Rickettsia prowazekii	1,11	834
Mycoplasma pneumoniae	0,82	710
Mycoplasma genitalium	0,58	503

Genomi di Batteri sequenziati: esempi

Organismo	Dimensioni (in paia di basi)	ORF[p]	Commenti
Bacteria			
Mycoplasma genitalium	580 070	470	Il più piccolo genoma cellulare noto (➤ vol. 2A, cap. 16.21)
Mycoplasma pneumonite	816 394	677	Causa la polmonite (➤ vol. 2A, cap. 16.21)
Borrelia burgdorferi	910 725	853	È uno spirochete, possiede un genoma lineare e causa la malattia di Lyme (➤ vol. 2A, cap. 16.33 e vol. 2B, cap. 31.4)
Clamydia trachomatis	1 042 519	894	Parassita intracellulare obbligato, patogeno umano comune (➤ vol. 2A, cap. 16.27 e vol. 2B, cap. 30.13)
Rickettsia prowazekii	1 111 523	834	Parassita intracellulare obbligato, causa il tifo epidemico (➤ vol. 2A, cap. 16.13 e vol. 2B, cap. 31.3)
Treponema pallidum	1 138 006	1041	Spirochete, causa la sifilide (➤ vol. 2A, cap. 16.33 e vol. 2B, cap. 30.12)
Aquifex aeolicus	1 551 335	1544	Ipertermofilo, autotrofo (➤ vol. 2A, cap. 16.37)
Prochlorococcus marinus	1 657 990	1716	Il più abbondante fototrofo nell'oceano (➤ vol. 2A, capp. 16.26 e 20.6)
Helicobacter pylori	1 667 867	1590	Causa l'ulcera peptica (➤ vol. 2B, cap. 30.10)
Streptococcus pyogenes	1 852 442	1752	Patogeno, causa la faringite e la scarlattina (➤ vol. 2B, cap. 30.2)
Thermotoga maritima	1 860 725	1877	Ipertermofilo (➤ vol. 2A, cap. 16.36) *vedi* anche fig. 13.7
Chlorobium tepidum	2 154 946	2288	Batterio fototrofico modello (➤ vol. 2A, cap. 16.32)
Staphylococcus aureus	2 814 816	2593	Principale responsabile delle infezioni nocosomiali (➤ vol. 2A, cap. 16.19 e vol. 2B, cap. 29.7)
Deinococcus radiodurans	3 284 156	2185	Resistente alle radiazioni, contiene cromosomi multipli (➤ vol. 2A, cap. 16.34)
Synecocystis sp.	3 573 470	3168	Cianobatterio (➤ vol. 2A, cap. 16.25)
Bdellovibrio bacteriovorus	3 782 950	3584	Parassita di altri procarioti (➤ vol. 2A, cap. 16.14)
Geobacter sulfurreducens	3 814 139	3467	Modello per il biorisanamento (➤ vol. 2A, cap. 18.18)
Caulobacter crescentus	4 016 942	3767	Ciclo di vita complesso (➤ vol. 2A, cap. 16.16)
Bacillus subtilis	4 214 810	4100	Modello genetico per i Gram-positivi (➤ vol. 2A, cap. 16.20)
Mycobacterium tuberculosis	4 411 529	3924	Causa la tubercolosi (➤ vol. 2A, cap. 16.23 e vol. 2B, cap. 30.5)
Escherichia coli	4 639 221	4288	Modello genetico per i Gram-negativi (➤ vol. 2A, cap. 16.11)
Bacillus anthracis	5 227 293	5738	Patogeno, agente per la guerra biologica (➤ vol. 2A, cap. 16.20)
Rhodopseudomonas palustris	5 459 213	4836	Fototrofo anossigenico metabolicamente versatile (➤ vol. 2A, cap. 16.2)
Pseudomonas aeruginosa	6 264 403	5570	Patogeno opportunistico metabolicamente versatile (➤ vol. 2A, cap. 16.7)
Streptomyces coelicolor	8 667 507	7825	Possiede un cromosoma lineare, produce antibiotici (➤ vol. 2A, cap. 16.24)
Bradyrhizobium japonicum	9 105 828	8317	Fissatore di azoto, forma noduli nelle piante di soia (➤ vol. 2A, cap. 20.22)

- Il genoma di ***Mycoplasma genitalium*** è più piccolo di quello del virus Chlorella e del batterio-fago G

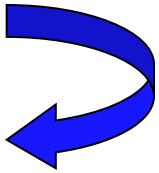
Genomi degli Archea sequenziati

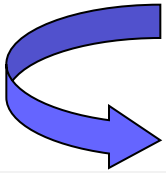
Archaea			
Nanoarchaeum equitans	490 885	552	Il più piccolo genoma conosciuto (➤ vol. 2A, cap. 17.11)
Thermoplasma acidophilum	1 564 905	1509	Termofilico, acidofilo (➤ vol. 2A, cap. 17.5)
Methanocaldococcus jannaschii	1 664 976	1738	Metanogeno (➤ vol. 2A, cap. 17.4)
Aeropyrum pernix	1 669 695	1841	Ipertermofilo (➤ vol. 2A, cap. 17.9)
Pyrococcus horikoshii	1 738 505	2061	Ipertermofilo (➤ vol. 2A, cap. 17.6)
Methanothermobacter thermoautotrophicus	1 751 377	1855	Metanogeno (➤ vol. 2A, cap. 17.4)
Archeoglobus fulgidus	2 178 400	2436	Ipertermofilo (➤ vol. 2A, cap. 17.7)
Halobacterium salinarum	2 571 010	2630	Alofilo estremo, contiene batteriorodopsina (➤ vol. 2A, cap. 17.3)
Sulfolobus solfataricus	2 992 245	2977	Ipertermofilo, chemiolitotrofo sulfureo (➤ vol. 2A, cap. 17.9)

Quali dati emergono dal sequenziamento dei genomi?

- I genomi procariotici variano da 0.5 Mb a 10 Mb

- Le dimensioni del genoma ed il contenuto genico riflettono la nicchia ambientale





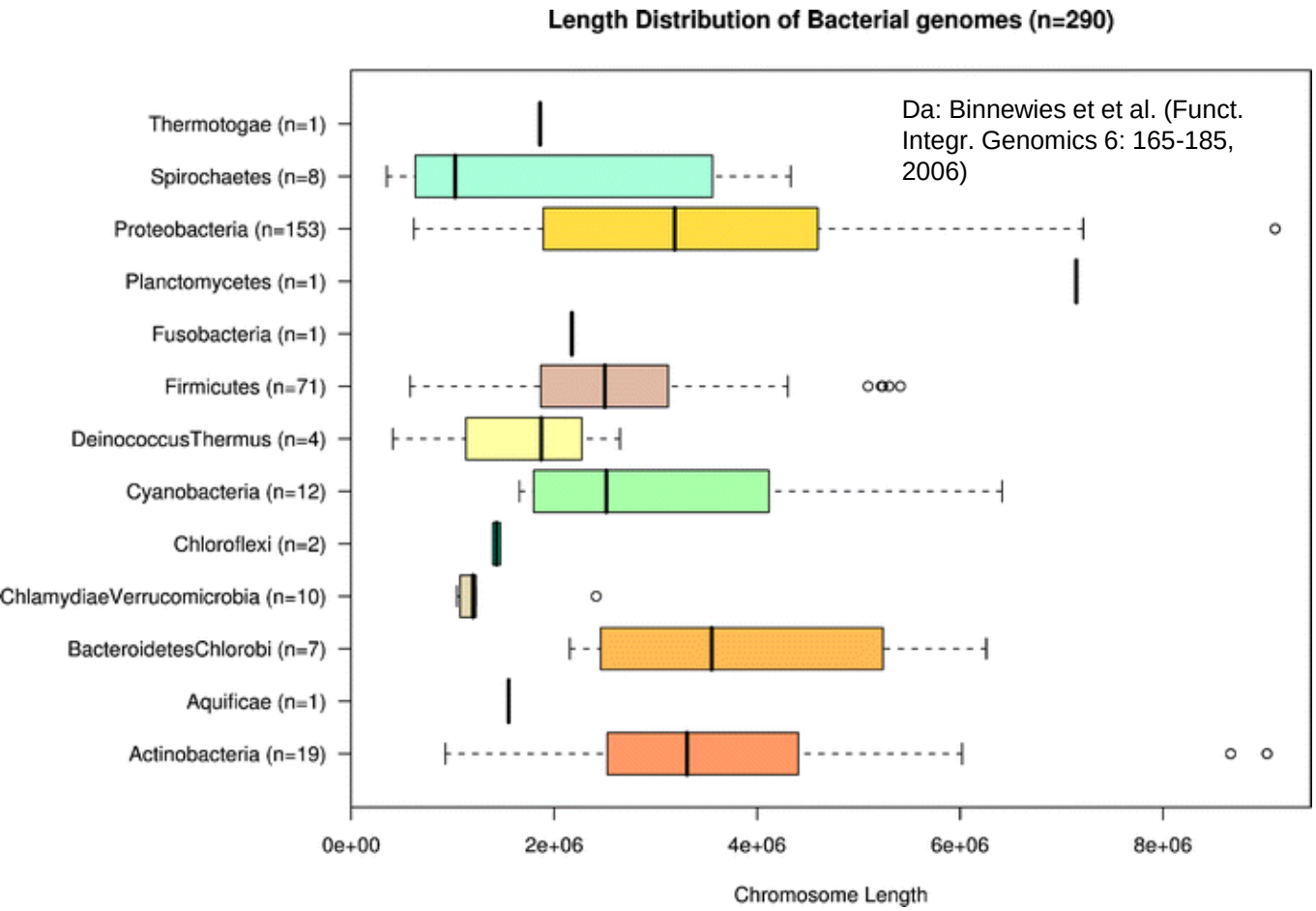
✓ I microrganismi con **genomi più GRANDI**
- **occupano ambienti** dove è necessaria una **maggiore variabilità**

✓ I microrganismi con genomi più **PICCOLI** sono
- **specializzati e**
- **ristretti** ad una determinata nicchia.

- Alcune specie presentano un elevato livello di variabilità all'interno della stessa specie
- questa variabilità è correlata alla capacità dell'organismo di adattarsi al proprio ambiente

I Genomi dei Procarioti: Dimensione

- **I genomi eubatterici mostrano una notevole variabilità nelle dimensioni, normalmente correlate al loro stile di vita (da 580 kb in *Mycoplasma genitalium* a 9,7 Mpb in *Burkholdera xenovorans*) .**



- **Simbionti intracellulari obbligati, patogeni (specializzati):**

genomi piccoli (media 1 Mb), molto stabili, con riarrangiamenti genomici e innovazioni geniche scarse o nulle.

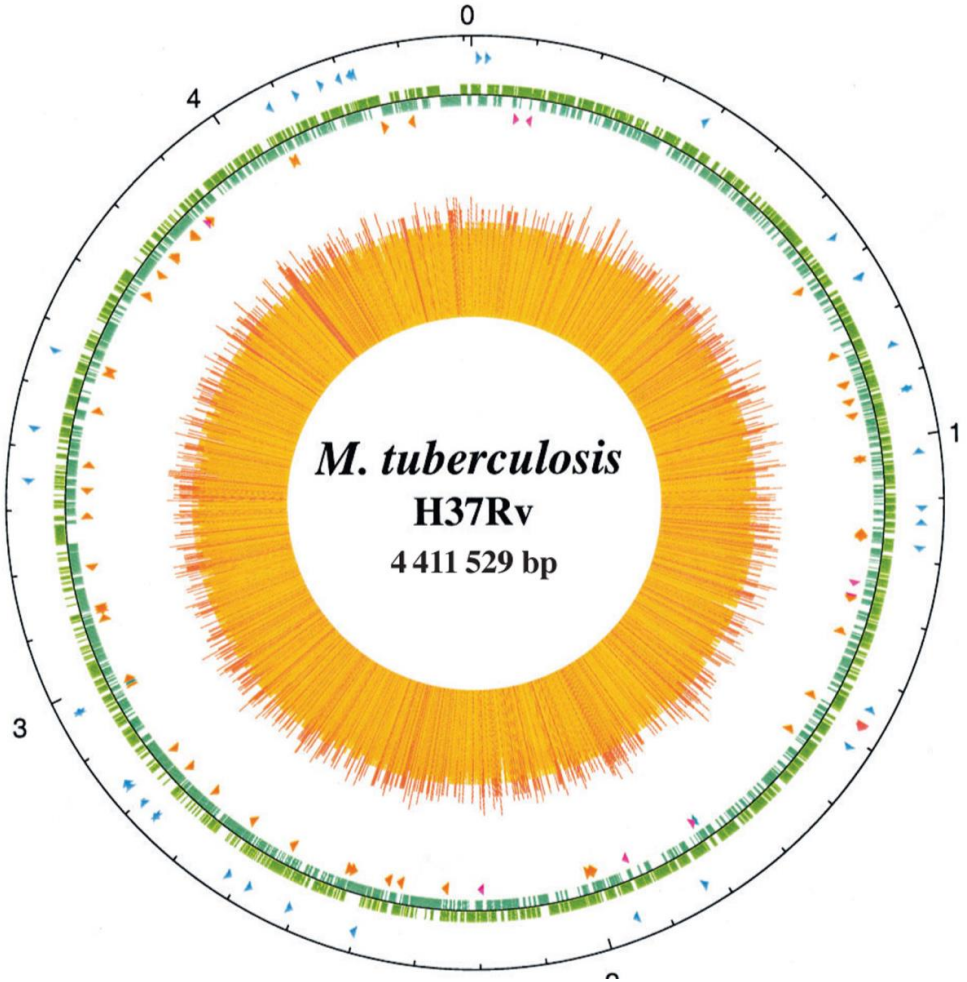
- **Batteri a vita libera (generalisti):**

genomi grandi (media 2-4 Mb), riarrangiamenti genomici frequenti, alto grado di variabilità nel contenuto genico sia a livello inter- che intra-specie, maggior contenuto di repetizioni.

- **I genomi archebatterici** hanno dimensioni poco variabili e generalmente comprese tra 1.5 e 2.5 Mbp.

Come si rappresenta il genoma di un microrganismo ?

- **La mappa di *Mycobacterium tuberculosis*, l' agente eziologico della tubercolosi, ottenuta con clonaggio shotgun**



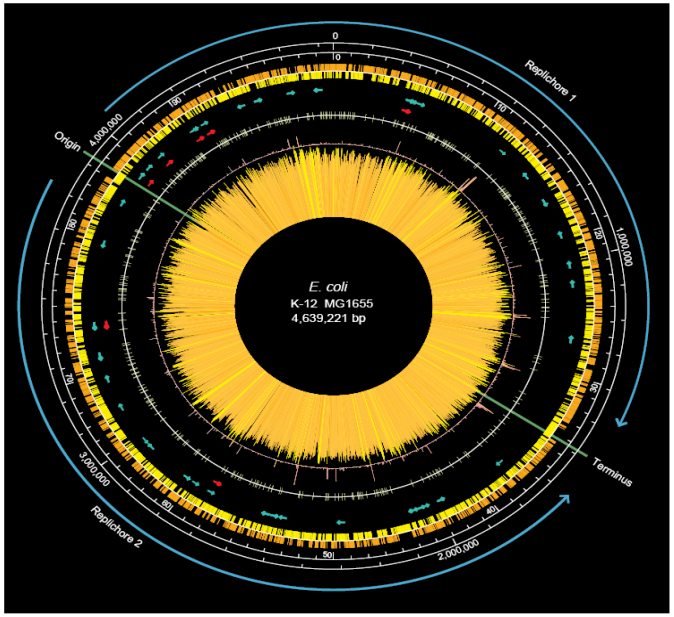
Gli anelli dall'esterno verso l'interno

- ✓ Dimensioni del genoma a **partire dall'origine di replicazione**
- ✓ contenuto in GC con raggi gialli (minore 65%)
- ✓ Raggi arancioni (maggiore 65%)

The Complete Genome Sequence of *Escherichia coli* K-12

Frederick R. Blattner,* Guy Plunkett III,* Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, Ying Shao

www.sciencemag.org • SCIENCE • VOL. 277 • 5 SEPTEMBER 1997



Tra il 1995 ed il 1997 è stata determinata la sequenza dell'intero genoma dei seguenti microrganismi:

Haemophilus influenzae
Micoplasma genitalium
Saccharomyces cerevisiae
Methanococcus jannaschii

- Genoma di *Escherichia coli* contains 4·639·221 bp for a total of 4288 genes (87,8% of the genome)

- 1897 genes coding for known proteins

-397 unidentified open-reading frames

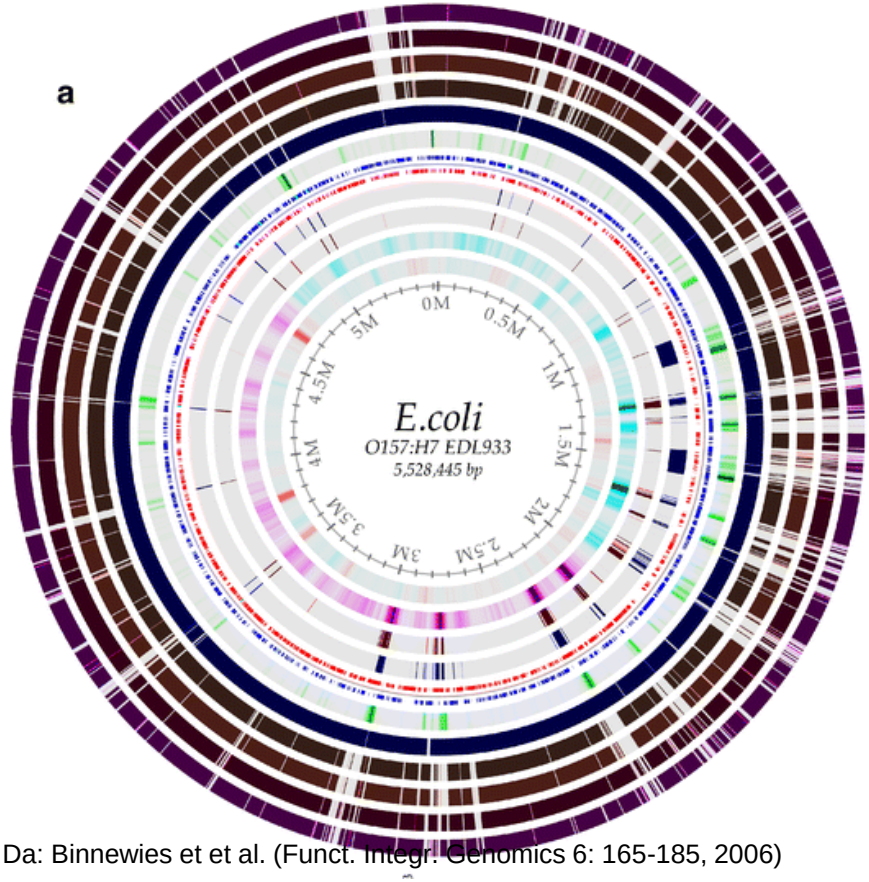
Table 4. Distribution of *E. coli* proteins among 22 functional groups (simplified schema).

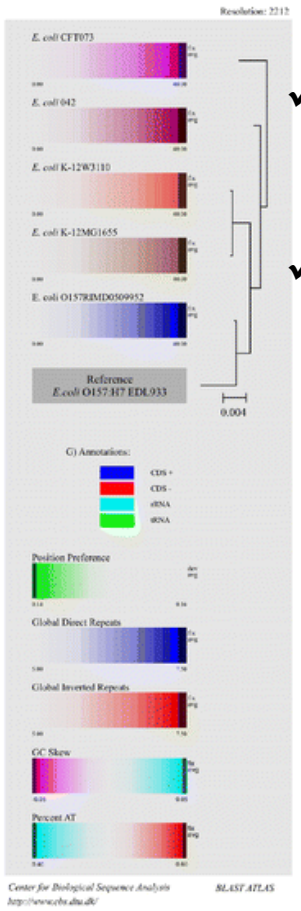
Functional class	Number	Percent of total
Regulatory function	45	1.05
Putative regulatory proteins	133	3.10
Cell structure	182	4.24
Putative membrane proteins	13	0.30
Putative structural proteins	42	0.98
Phage, transposons, plasmids	87	2.03
Transport and binding proteins	281	6.55
Putative transport proteins	146	3.40
Energy metabolism	243	5.67
DNA replication, recombination, modification, and repair	115	2.68
Transcription, RNA synthesis, metabolism, and modification	55	1.28
Translation, posttranslational protein modification	182	4.24
Cell processes (including adaptation, protection)	188	4.38
Biosynthesis of cofactors, prosthetic groups, and carriers	103	2.40
Putative chaperones	9	0.21
Nucleotide biosynthesis and metabolism	58	1.35
Amino acid biosynthesis and metabolism	131	3.06
Fatty acid and phospholipid metabolism	48	1.12
Carbon compound catabolism	130	3.03
Central intermediary metabolism	188	4.38
Putative enzymes	251	5.85
Other known genes (gene product or phenotype known)	26	0.61
Hypothetical, unclassified, unknown	1632	38.06
Total	4288	100.00*

*Total of these rounded values is 99.97%.

I Genomi dei Procarioti: Dimensione

- Si osserva una **marcata eterogeneità** delle dimensioni anche **nell'ambito di una stessa specie.**
- Ad esempio, 20 genomi di *E.coli* sequenziati hanno dimensioni comprese tra 4,6 e 5,5 Mb.





Da: Binnewies et et al. (Funct. Integr. Genomics 6: 165-185, 2006)

- ✓ Confronto tra *E.coli* O157:H7 e i genomi di altri 4 ceppi di *E.coli*.
- ✓ Il ceppo patogeno O157 ha 1,34 Mb specifiche per fattori di virulenza, assenti nel ceppo di laboratorio K12, che a sua volta contiene 0,53 Mb assenti nel ceppo O157.

Bacterial Genome structures

- **Genome size**
- **Within species variation**

- **Insertion Sequences (IS)**

- One of the main reason of within-species genome poly-morphism
- IS are DNA sequences inserted in the genome, present among certain individuals in a population
- Typically IS can be
  - lysogenic phages or
  - sequences acquired by **horizontal transfer**

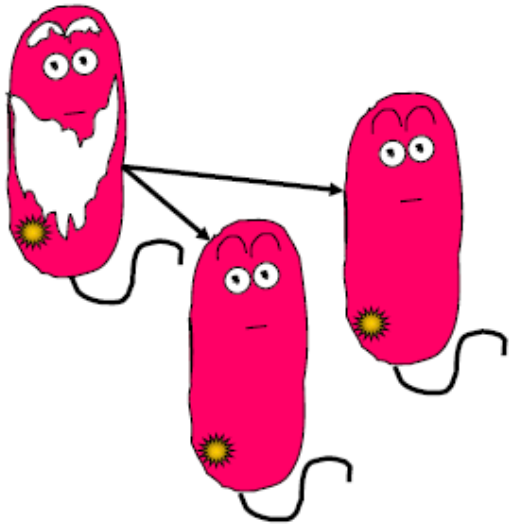
Il trasferimento genico orizzontale

Il passaggio di geni tra microrganismi diversi, è detto

TRASFERIMENTO GENICO ORIZZONTALE

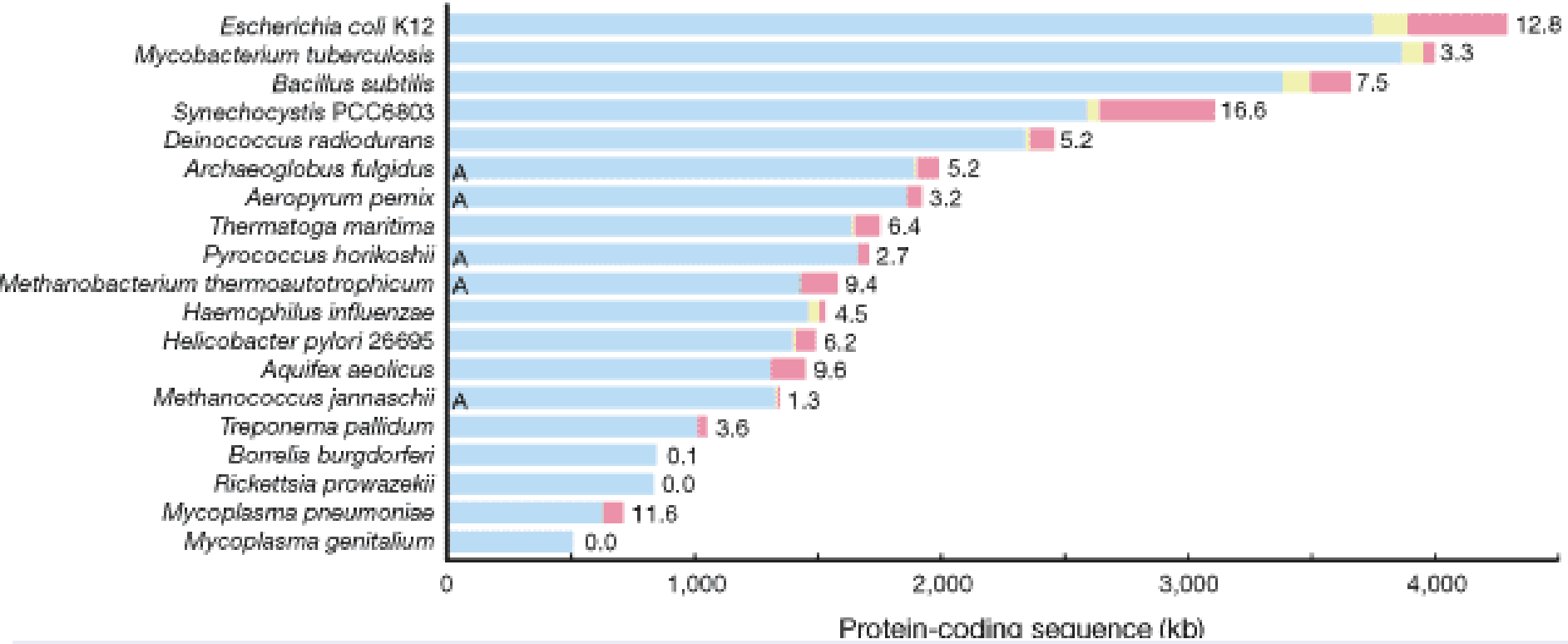


In contrapposizione al
TRASFERIMENTO VERTICALE



(passaggio di un gene da una cellula alla propria progenie)

Distribution of horizontally acquired (foreign) DNA in sequenced bacterial genomes.



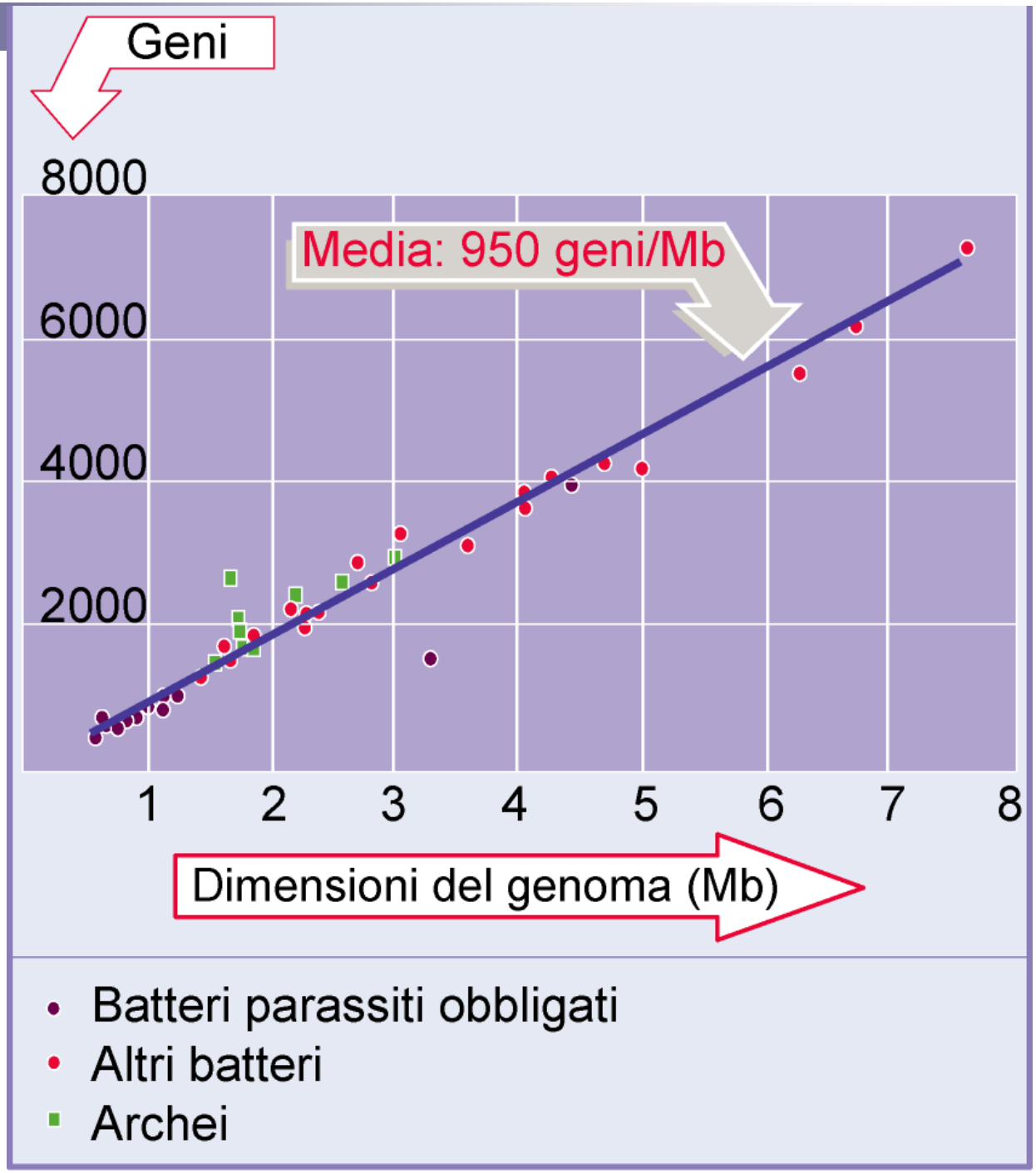
- Lengths of bars denote the amount of protein-coding DNA.
- For each bar, **the native DNA is blue;**
- foreign DNA identifiable as mobile elements, including
  - **transposons and bacteriophages**, is yellow and other foreign DNA is red.
  - The percentage of foreign DNA is noted to the right of each bar.

'A' denotes an Archaeal genome.

Insertion Sequences and pathogenicity

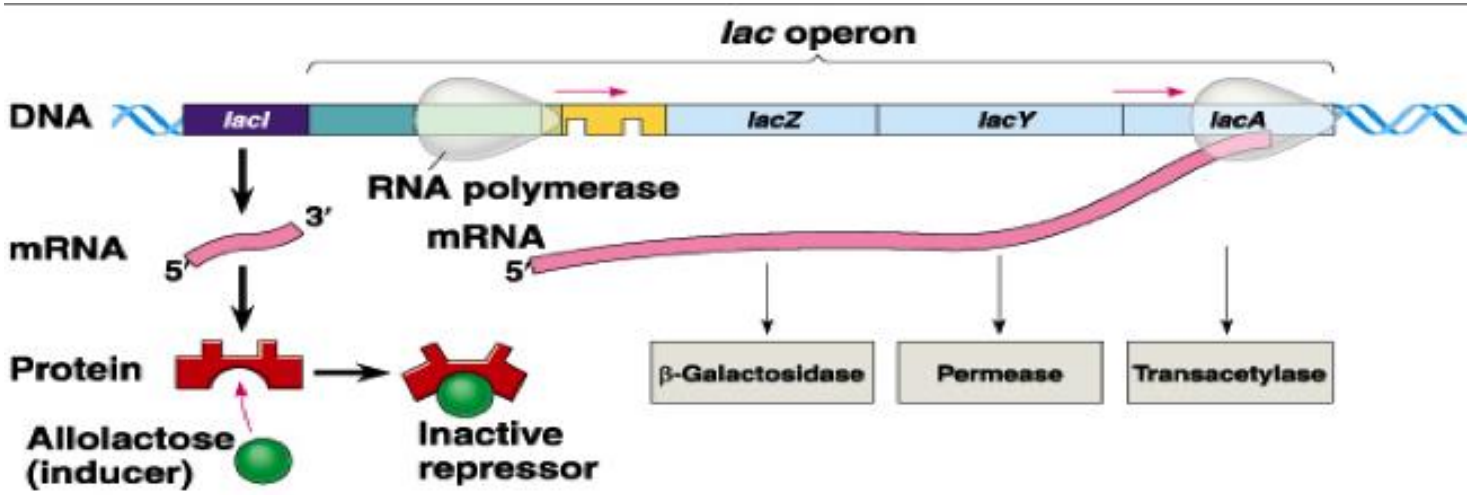
- **Many bacteria are pathogens because of acquired IS, e.g. from phages** (prophages).
- Examples:
  - *Y. pestis* acquired the toxicity protein from a phage
  - *E. coli O157:H7* is a pathogenic strain of *E. coli*, only because of added IS.
  - *P. aeruginosa PA01* contains inserted sequences of bacteriocins, designed by phages to kill bacteria

Le dimensioni dei genomi batterici sono correlate al numero di geni



I Genomi dei Procarioti: Contenuto genico

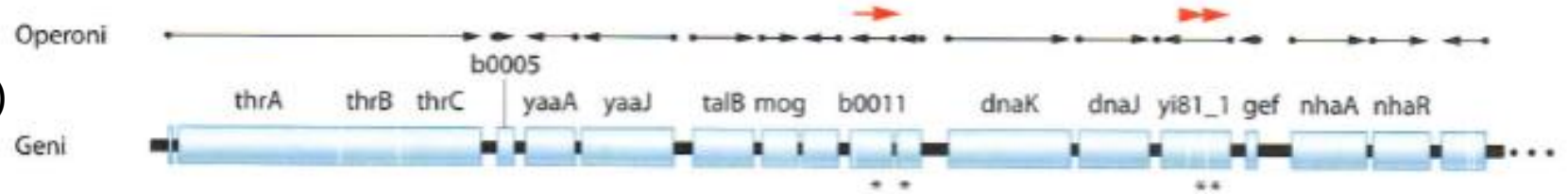
- **densità genica molto elevata.**
- In media, i geni codificanti proteine occupano l'85% del genoma.
- **Non interrotti da introni**
- Organizzati in **unità trascrizionali policistroniche** (operoni).



- Il numero di geni, così come le dimensioni del genoma, riflette lo stile di vita del batterio.
- I parassiti specializzati (es. *Mycoplasma*) hanno circa 500-600 geni,
- mentre i batteri generalisti hanno un numero di geni molto maggiore, tipicamente compreso tra 4,000 e 5,000.
- Gli Archea hanno un numero di geni compreso tra 1,700 e 2,900.

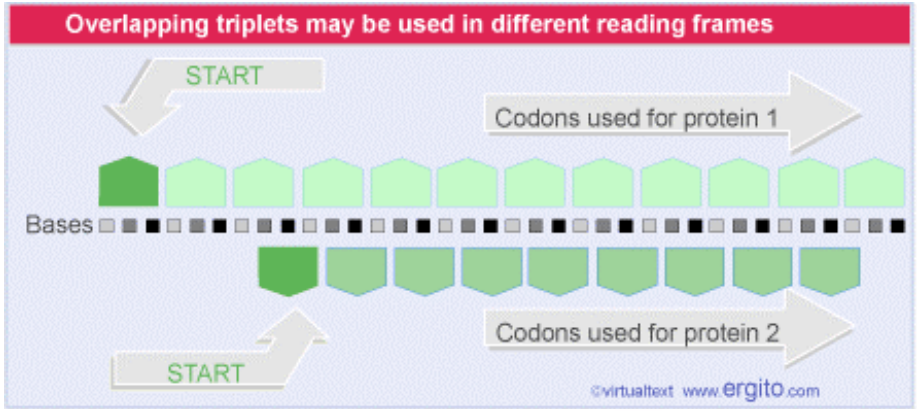
Genomi degli Eubatteri

a)



a) Porzione del cromosoma di E. coli dove sono mostrati geni e operoni.
Un puntino indica il promotore per ciascun gene o operone.
Le frecce indicano la direzione della trascrizione. I geni sovrapposti sono identificati da asterischi.

Geni sovrapposti



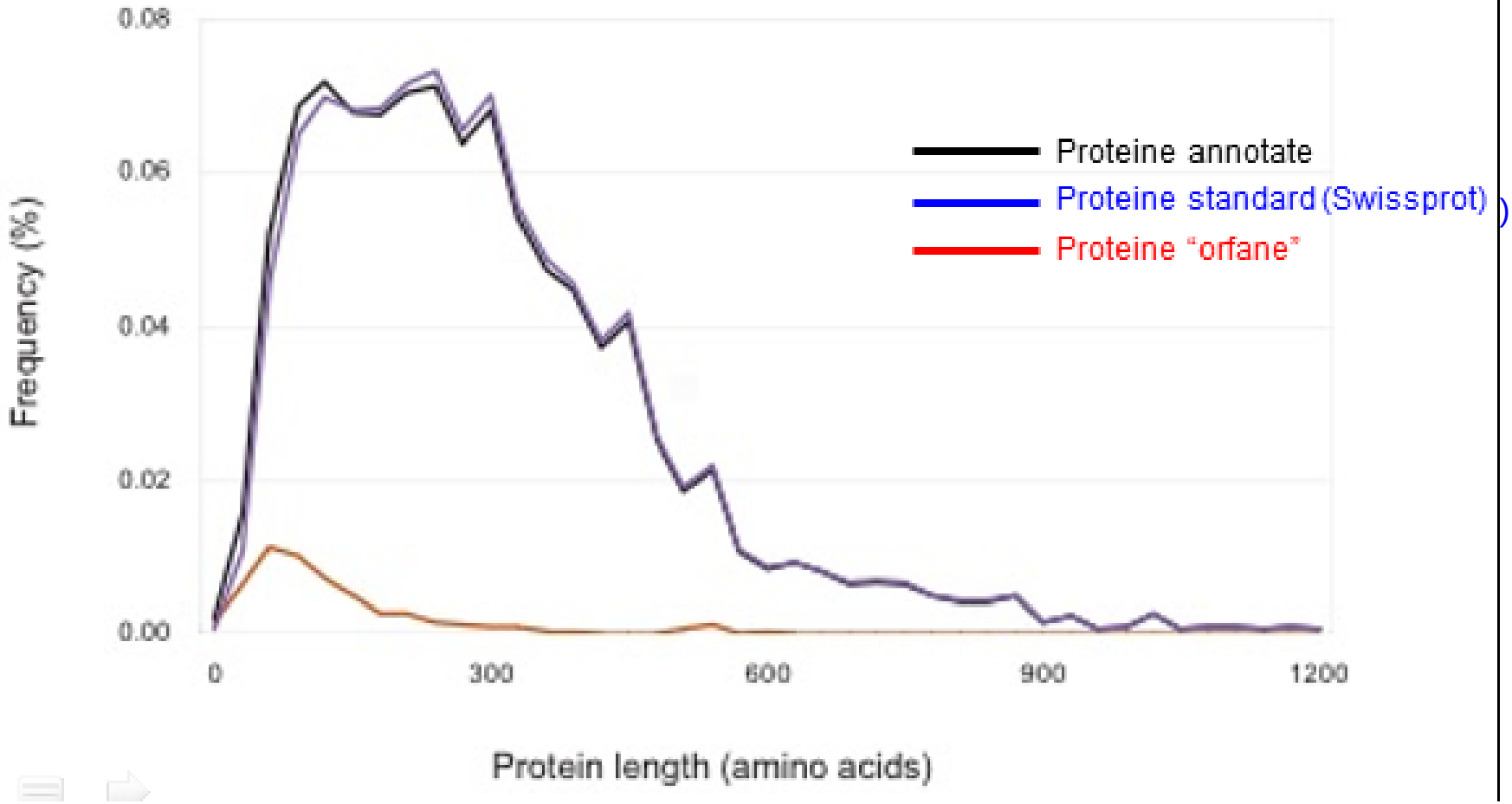
b)



- **Operone del genoma di *A. aeolicus* contiene geni per la sintesi proteica (gatC), la ricombinazione del DNA (*recA e recI*), per una proteina della motilità (pili), per la biosintesi dei nucleotidi (cmk), e per la biosintesi dei lipidi (pgsA1).**
- **Questa organizzazione mette in discussione l'idea convenzionale che i geni di un operane codifichino per prodotti che controllano un percorso metabolico comune.**

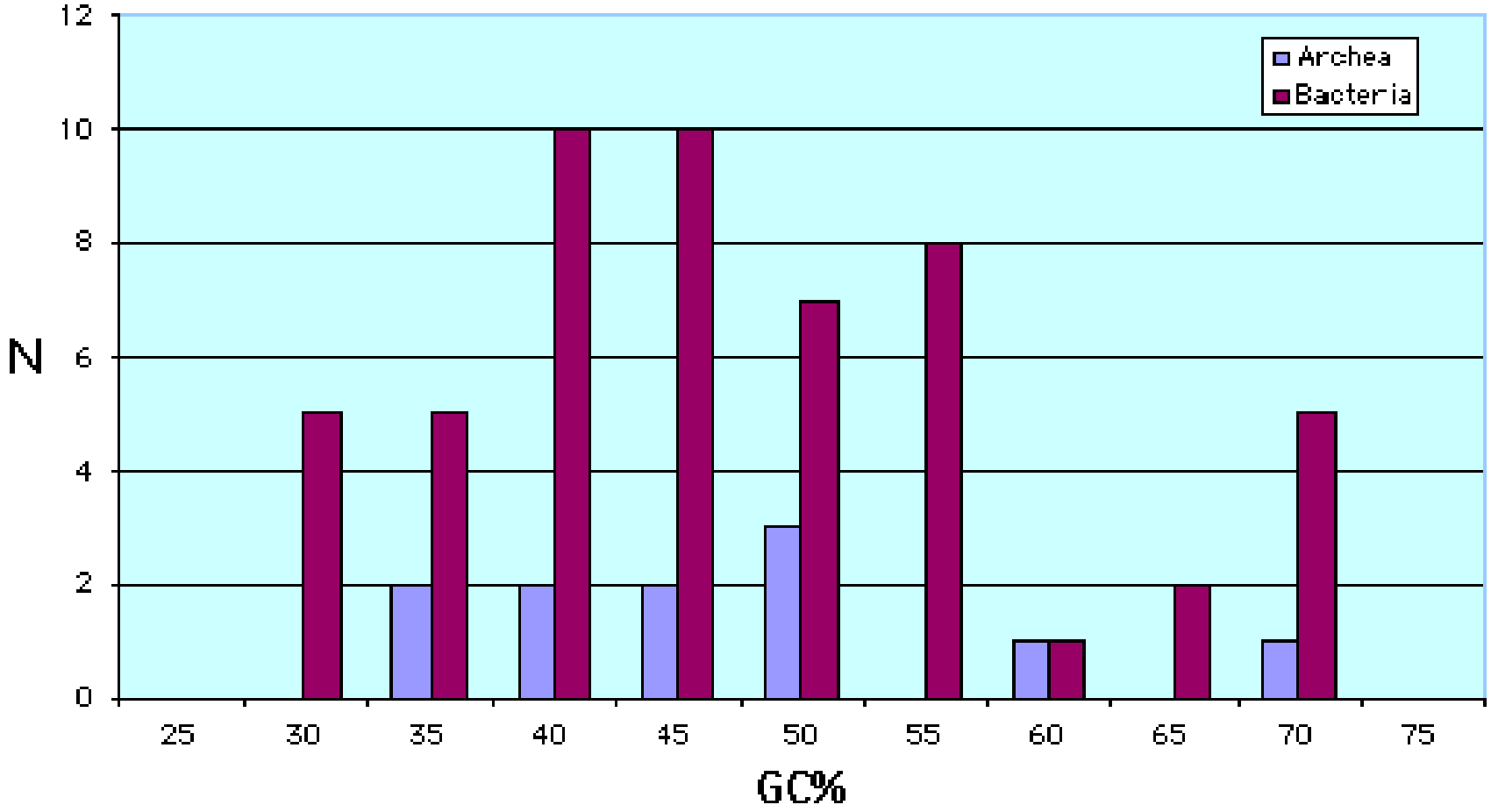
I Genomi dei Procarioti: Dimensioni dei geni

- La lunghezza media di una proteina in *E. coli* K12, normalmente compresa tra 100 e 400 aa, è pari a circa 300 aa.



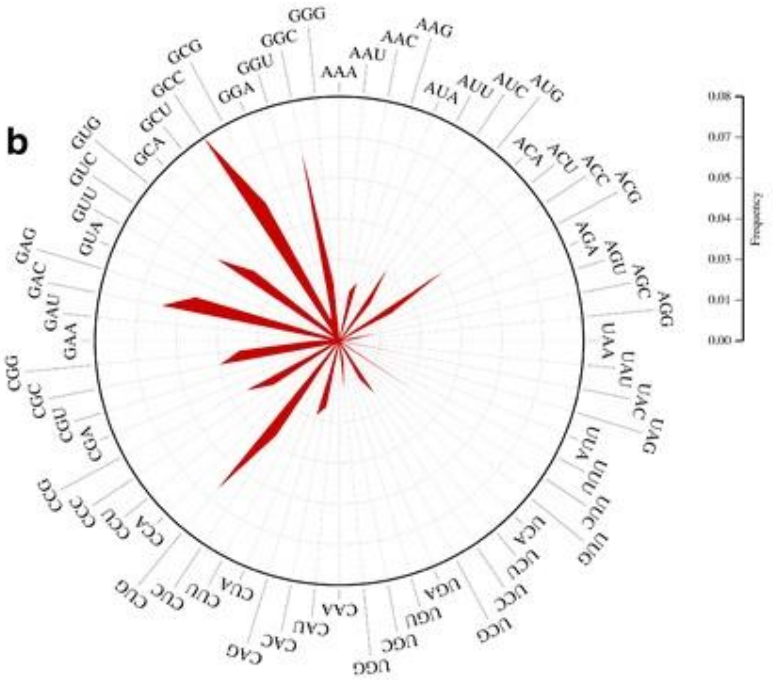
I Genomi dei Procarioti: Composizione in basi

- La composizione in basi
  - **molto variabile:** compresa tra il 25% e il 75%.
  - **non è omogenea** lungo il genoma (formazione di compartimenti, LGT, inserimento di elementi mobili).



I Genomi dei Procarioti: Uso del Codice Genetico

- **I Codoni sinonimi non sono usati in modo omogeneo.**
  - E' possibile misurare quanto l'utilizzo dei codoni (Codon Usage) si discosta da un utilizzo omogeneo.



• **Costrizioni composizionali** (es genoma di *Campylobacter* ha GC di 30,4% e tende a preferire codoni con A/T finale)

• **Selezione a livello di traduzione-**

In *E. coli* a altri batteri i codoni adoperati da geni altamente espressi, come quelli codificanti per proteine ribosomiali o proteine coinvolte in importanti vie biosintetiche, sono correlati ai codoni ottimali per la traduzione, predetti sulla base della abbondanza relativa di tRNA isoaccettori)

ORF non identificate

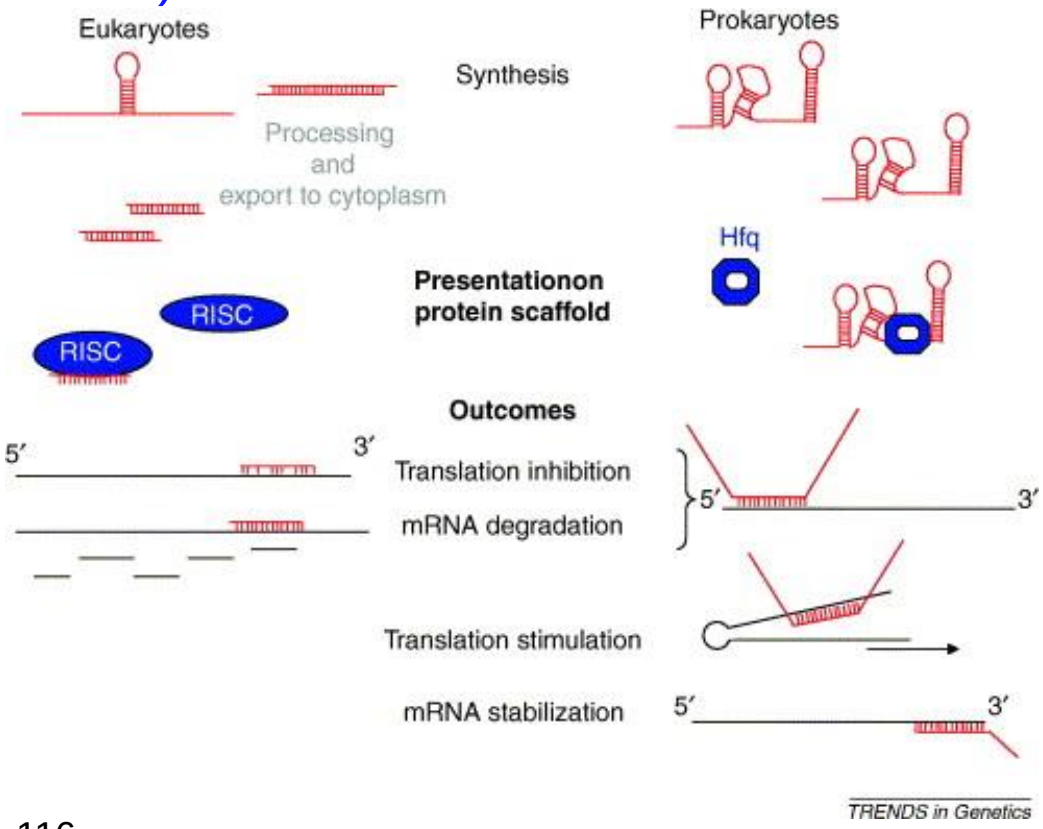
- Nella maggior parte dei casi il numero dei geni identificabili corrisponde al massimo a circa 70% delle ORF individuate.
- In E.coli sono state assegnate 2700 dei circa 4.300 geni

- Un ORF non identificata contiene tutte le caratteristiche di un gene ipotetico quando
  - possiede un sito di legame dei ribosomi
  - un codone di inizio
  - un codone di stop.

- L'**identificazione** avviene tramite **l'analisi con una sequenza aminoacidica sufficientemente omologa a quella di una proteina nota**.

I Genomi dei Procarioti: sncRNA

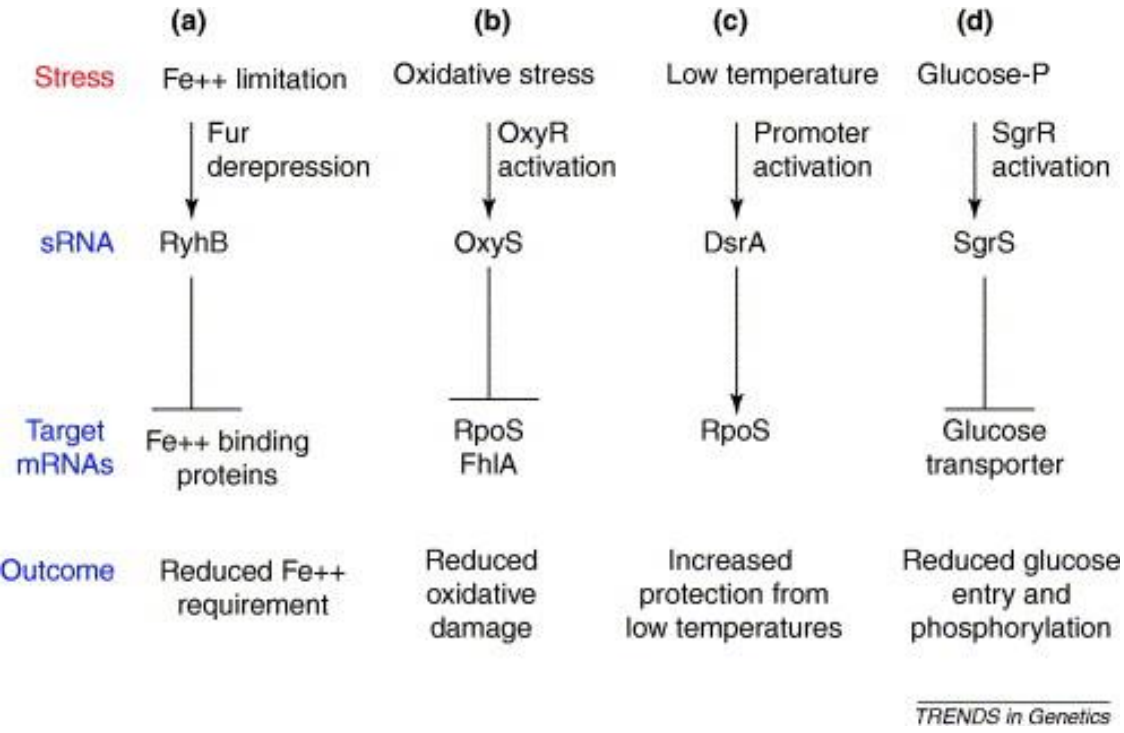
- Anche nei procarioti (come negli eucarioti) sono stati scoperti **numerosi piccoli RNA non codificanti proteine che svolgono importanti funzioni regolatorie** (small noncoding RNAs, sncRNAs) appaiandosi per complementarietà a diversi mRNA bersaglio.
- L'interazione tra sncRNA e mRNA è favorita da una chaperonina (**Hfq** in *E. coli*).

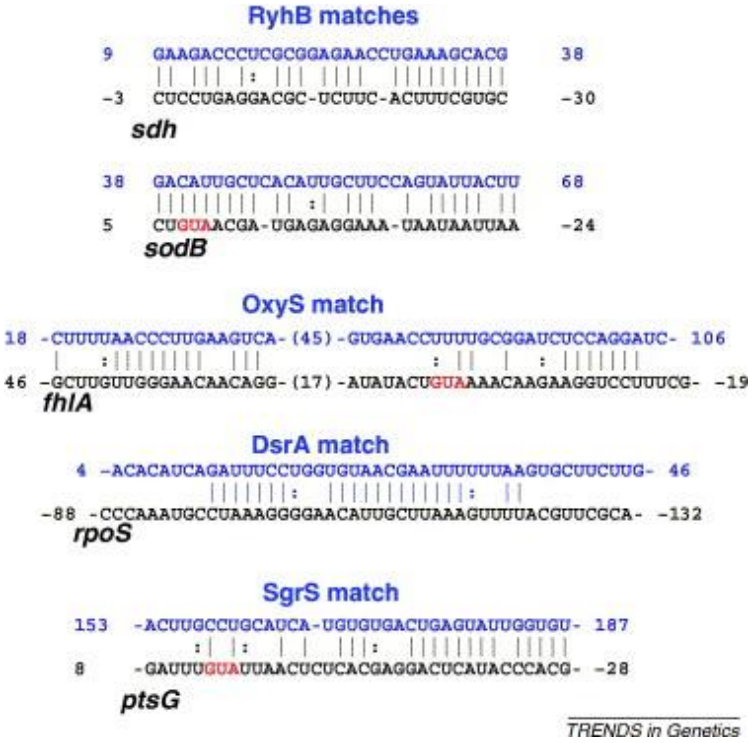


- **Gli sncRNA** vengono prodotti a partire da promotori canonici e normalmente mostrano una elevata propensione a formare strutture secondarie complesse.
- **Senza subire alcun processamento essi si legano alla chaperonina**,
- quindi si appaiano per **complementarietà all'mRNA bersaglio** (anch'esso legato alla chaperonina).

I Genomi dei Procarioti: sncRNA

- **L'appaiamento con l'mRNA target** coinvolge **generalmente l'estremità 5' del trascritto in prossimità del codone di inizio**.
- L'effetto di questo appaiamento è:
  - inibizione della traduzione;
  - degradazione dell'mRNA;
  - attivazione della traduzione;
  - stabilizzazione dell'mRNA.





Da: Gottesman, Trends in Genetics 21: 399-404,

I Genomi dei Procarioti: Regioni non-geniche

- **Le regioni non-geniche (in media circa il 15%) contengono**
  - **geni per RNA non codificanti proteine** (molti dei quali tuttora ignoti)
  - **elementi ripetuti**, costituiti prevalentemente da elementi mobili (**MGE**, mobile genetic element) in grado di spostarsi sia all'interno di uno stesso genoma che tra genomi differenti.
- Gli **MGE** comprendono plasmidi, fagi, trasposoni, elementi IS, ecc.



IS Insertion Sequence
(replic o conserv)



Trasposone composito
(conserv)



Tn3
(replic)

Trasposizione replicativa:
nuova copia dell'elemento donatore in una differente posizione
Trasposizione conservativa:
escissione e reintegrazione in un nuovo sito

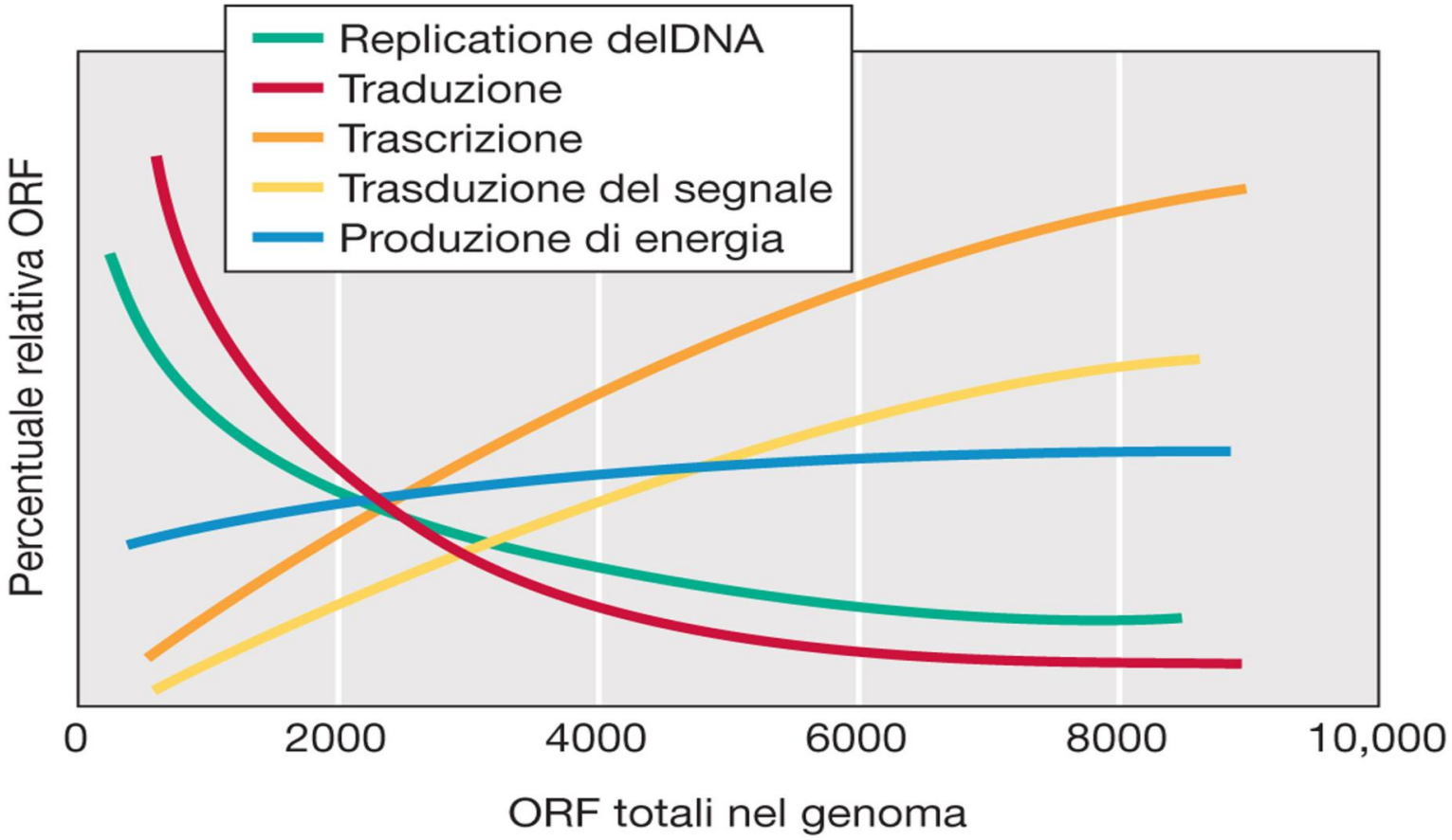
Funzione genica nei genomi batterici : rapporto funzione dimensioni

Categorie funzionali	Percentuale di geni sul cromosoma nella categoria di riferimento		
	Escherichia coli (4,64 Mbp)[a]	*Haemophilus influenzae* (1,83 Mbp)[a]	*Mycoplasma genitalium* (0,58 Mbp)[a]
Metabolismo	21,0	19,0	14,6
Strutturali	5,5	4,7	3,6
Trasporto	10,0	7,0	7,3
Regolazione	8,5	6,6	6,0
Traduzione	4,5	8,0	21,6
Trascrizione	1,3	1,5	2,6
Replicazione	2,7	4,9	6,8
Altri noti	8,5	5,2	5,8
Sconosciuti	38,1	43,0	32,0

[a] Dimensioni dei cromosomi. Ognuno dei microrganismi elencati contiene un solo cromosoma circolare.

Geni coinvolti nella sintesi proteica sono essenziali

- più i genomi sono piccoli maggiore è la componente percentuale dei geni dedicati ai processi di traduzione.



Microrganismi con genomi di grandi dimensioni

- Gli organismi con genomi di grandi dimensioni possiedono la capacità di codificare per molti geni coinvolti sia nel metabolismo che nei processi regolativi.

- I meccanismi di regolazione permettono alla cellula di rispondere in modo migliore alla disponibilità di substrati differenti attraverso l'attivazione di geni specifici.

- Uno degli habitat principali dei microrganismi è il suolo.

- Tutti i microrganismi con genomi di dimensioni maggiori alle 6 Mbp sono microrganismi del suolo.

- Il suolo è un habitat nel quale le fonti di carbonio e di energia sono spesso scarse o disponibili in forme diverse e spesso fruibili in maniera inter-mittente.

Batteri con genomi piccoli dipendono dall'ospite per numerose funzioni:

Tabella 8.3 La perdita di funzioni metaboliche nei batteri patogeni.

SPECIE (NUMERO DI GENI)	GLICOLISI	CICLO ACIDI TRICARBOSSILICI	BIOSINTESI AMINOACIDI	BIOSINTESI PURINE	BIOSINTESI PIRIMIDINE
Mycoplasma genitalium (470)	+	–	–	–	–
Buchnera spp. (588)	+	–	+	+	+
Rickettsia prowazekii (834)	–	+	–	–	–
Chlamidia trachomatis (894)	+	–	+	–	–
Treponema pallidum (1014)	+	–	–	–	–
Mycobacterium leprae (1604)	parziale	in decadimento	+	+	+

- sono state infatti perse funzioni metaboliche importanti come la glicolisi, biosintesi di aminoacidi e purine e pirimidine che vengono fornite dall'ospite.
- Molti di questi genomi possiedono ancora geni per queste funzioni ma non funzionali ovvero pseudogeni.
- In *Mycobacterium leprae* circa 50 % dei geni sono pseudogeni che mantengono omologia con *Mycobacterium tubercolosis*

Come sono sono catterizzati i genomi dei Batteri?

- un numero elevato di **geni per il metabolismo dei carboidrati**
- un numero significativo di **geni per funzioni correlate alla membrana**
- un elevato numero di geni ancora a funzione sconosciuta
- un elevato numero di proteine ipotetiche

❑ **Rispetto agli Archea**

- un numero minore di geni per il metabolismo dei coenzimi
- un numero minore di geni per la produzione di energia

- Oltre alle dimensioni del genoma anche l'appartenenza al Dominio (Batteri o Archea) sembra influenzare la categorizzazione funzionale dei geni nei procarioti

I genomi degli Archea sono caratterizzati

- **maggiore percentuale di geni** coinvolti nella **produzione di energia**
- maggior numero di coenzimi (specialmente negli Archea metanogenici
- numero elevato di geni a funzione sconosciuta
- numero elevato di geni che codificano proteine ipotetiche
- numero ridotto di geni coinvolti nel metabolismo dei carbo-idrati
- numero ridotto di geni collegati alle funzioni della membrana cellulare come il trasporto e la biosintesi delle membrane

Quanti geni sono necessari per la minima funzionalità di una cellula?

- *Mycoplasma genitalium* e *Mycoplasma pneumoniae* sono microrganismi con un genoma estremamente piccolo.

- Le 470 ORF di *M. genitalium* sono presenti anche in *M. pneumonie*
- Attraverso mutagenesi e analisi comparativa con altri genomi si è visto che sono necessari

- **300 geni** codificanti per stabilire la **minima funzionalità di una cellula.**

- Non si conoscono finora genomi con un numero inferiori di ORF.

- La prima proposta di genoma minimo viene dall'analisi dei 2 genomi sequenziati per primi
  - ***Haemophilus influenzae*** (1815 geni) e *Mycoplasma genitalium* 525 geni) che rivela
  - un gruppo comune di **geni 256** molto più piccolo del numero di geni presenti nel genoma di *Mycoplasma.*

- Da qui l'idea di introdurre il metodo di mutagenesi trasposizionale globale che ha permesso, effettuando uno studio sulla vitalità delle cellule portatrici di mutazioni in ciascuno dei 525 geni di *Mycoplamsa* di stabilire che:
  - 150 geni sono NON ESSENZIALI
  - 375 geni sono essenziali

- Nasce la possibilità di produrre un genoma minimo che fosse più piccolo di quelli esistenti in natura ma che sarebbe stato in ogni caso più grande di questi 256 geni identificati come CORE

Applicazioni della Genomica Microbica

1. Identificazione delle differenze essenziali tra batteri patogeni e non patogeni, e tra fattori di virulenza e fattori di colonizzazione.

2. Sviluppo della "**Reverse Vaccinology**"

Genoma → Identificazione "bioinformatica" di proteine secrete o extracellulari → Test di immunogenicità e identificazione di antigeni protettivi

Esempi: *Neisseria meningidites* (serovar B); *Streptococcus pneumoniae*; *Bacillus anthracis*; etc. (cfr. Mora et al. Curr. Op. Microbiol. 9:532-36)

- L'esteso livello di variazione intra-specifica rende l'analisi di un singolo genoma insufficiente alla realizzazione di un vaccino "universale".
- E' necessario quindi **integrare dati di genomi diversi per determinare la combinazione di antigeni ottimale per la** specifica struttura della popolazione del patogeno considerato

Metagenomica

La **metagenomica** è una nuova disciplina che studia l'insieme dei diversi materiali genetici (detto **metagenoma**) derivanti dalle diverse specie presenti in un determinato ambiente.
Ad esempio il materiale genetico dell'uomo e quello derivante dal suo microbiota (detto microbioma), rappresenta il metagenoma umano.

L'**approccio metagenomico** è applicato con successo per rivelare la diversità genetica e tassonomica delle comunità microbiche.

Non richiede l'isolamento e la messa in coltura di ogni singola specie, e la sua identificazione su basi biochimiche o morfologiche.





Il Progetto “Microbioma”

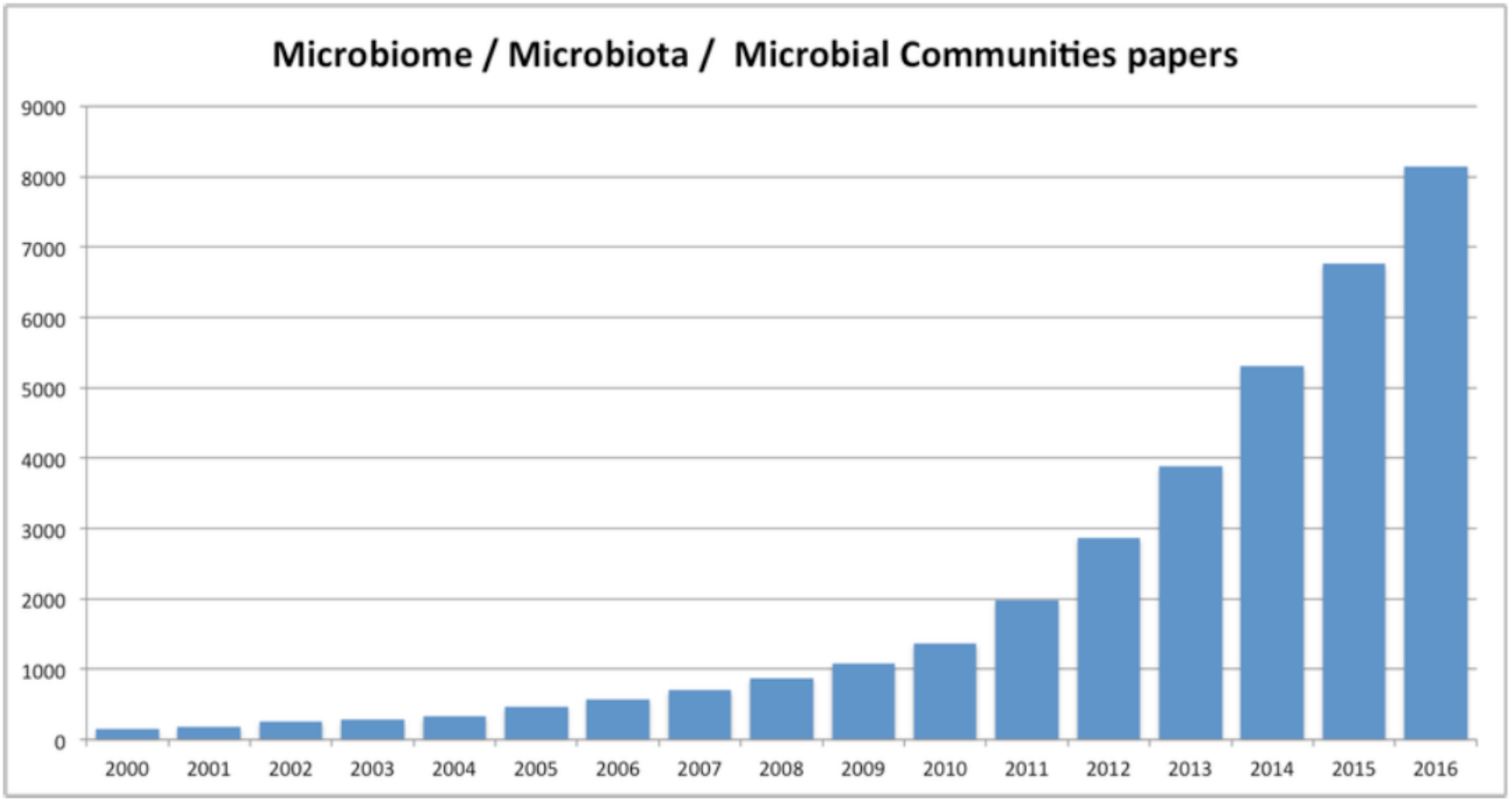
- I microrganismi che vivono nell’uomo (microbiota) superano il numero **di cellule somatiche di un ordine di grandezza (ca. 10^{15}).**

- Suppliscono proprietà fisiologiche e metaboliche che l’organismo non possiede in proprio (es. capacità digestive, detossificazione, sviluppo ed attività del sistema immunitario, etc.).
- Ad esempio, è stato osservato che l’obesità può essere correlata ad una particolare composizione del microbiota.

Per esempio :

- Il programma “Human Microbiome Project” si propone appunto
- di **far luce sulle caratteristiche e proprietà della componente microbica dell’uomo e**
- di come questo **contribuisca**
  - **ai normali processi fisiologici** e
  - nella predisposizione alle malattie.

Microbioma: un universo in via di espansione



Fonte:PubMed

Genomica funzionale di un genoma batterico

ANALISI DEL GENOMA DI *PSEUDOMONAS AERUGINOSA*

Caratteristiche generali

Dimensione del genoma (bp)	6 264 403
Contenuto in G + C	66,6%
Regioni codificanti	89,4%

Sequenze codificanti

Livello di confidenza	ORF	(%)	Definizione
1	372	6,7	Geni di *P. aeruginosa* con funzione dimostrata
2	1059	19,0	Elevata omologia con geni con funzione dimostrata di altri organismi
3	1590	28,5	Geni con funzione putativa basata su ricerca di motivi o omologia limitata
4a	769	13,8	Omologhi a geni riportati a funzione sconosciuta
4b	1780	32,0	Nessuna omologia con nessuna sequenza riportata
Totale	5570	100	

Fonte: Stover, *et al.* 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 406:959–64. Table 1, p. 961.

Strategie per le assegnazioni funzionali di geni sconosciuti

(Pseudomonas Aeruginosa)

Classe funzionale	ORF	
	Numero	%
Adattamento, protezione (per esempio, proteine dello shock da freddo)	60	1,1
Biosintesi e metabolismo degli aminoacidi	150	2,7
Resistenza e sensibilità agli antibiotici	19	0,3
Biosintesi di cofattori, gruppi prostetici e trasportatori	119	2,1
Catabolismo dei composti del carbonio	130	2,3
Divisione cellulare	26	0,5
Parete cellulare	83	1,5
Metabolismo degli intermedi centrali	64	1,1
Proteine "chaperone" e "heat shock"	52	0,9
Chemiotassi	43	0,8
Replicazione, ricombinazione, modificazione e riparazione del DNA	81	1,5
Metabolismo energetico	166	3,0
Metabolismo degli acidi grassi e dei fosfolipidi	56	1,0
Proteine di membrana	7	0,1
Motilità e adesione	65	1,2
Biosintesi e metabolismo dei nucleotidi	60	1,1
Apparato per la secrezione/esporto delle proteine	83	1,5
Enzimi putativi	409	7,3
Simili ai fagi, trasposoni o plasmidi	38	0,7
Fattori secreti (tossine, enzimi, alginati)	58	1,0
Regolatori trascrizionali	403	7,2
Trascrizione, processamento dell'RNA, e degradazione	45	0,8
Traduzione, modificazioni post-trascrizionali, e degradazione	149	2,7
Trasporto di piccole molecole	555	10,0
Sistemi regolatori a due componenti	118	2,1
Ipotetiche	1774	31,8
Ignote (ipotetiche conservate)	757	13,6
Totale	5570	100

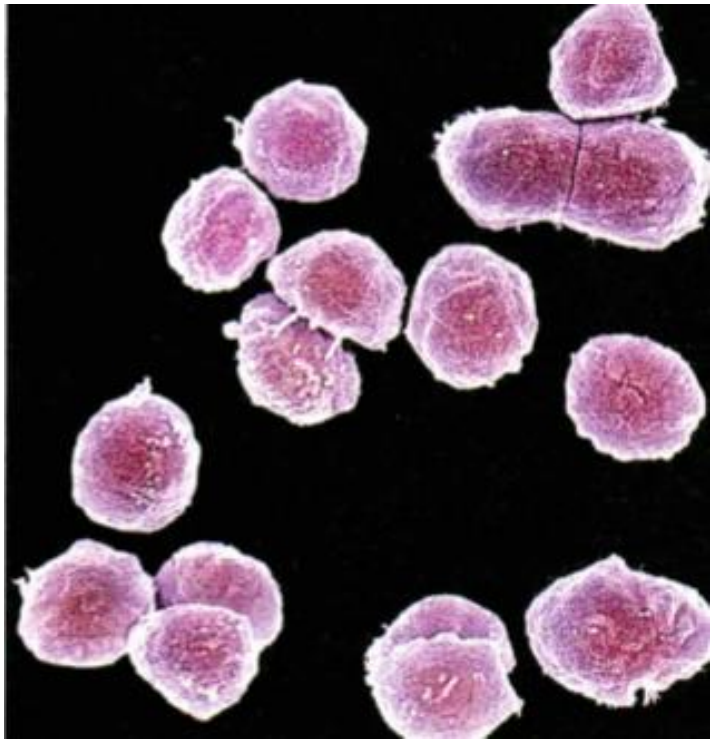
Fonte: Stover, *et al.* 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 406:959–64.

2. Genomi degli Archea

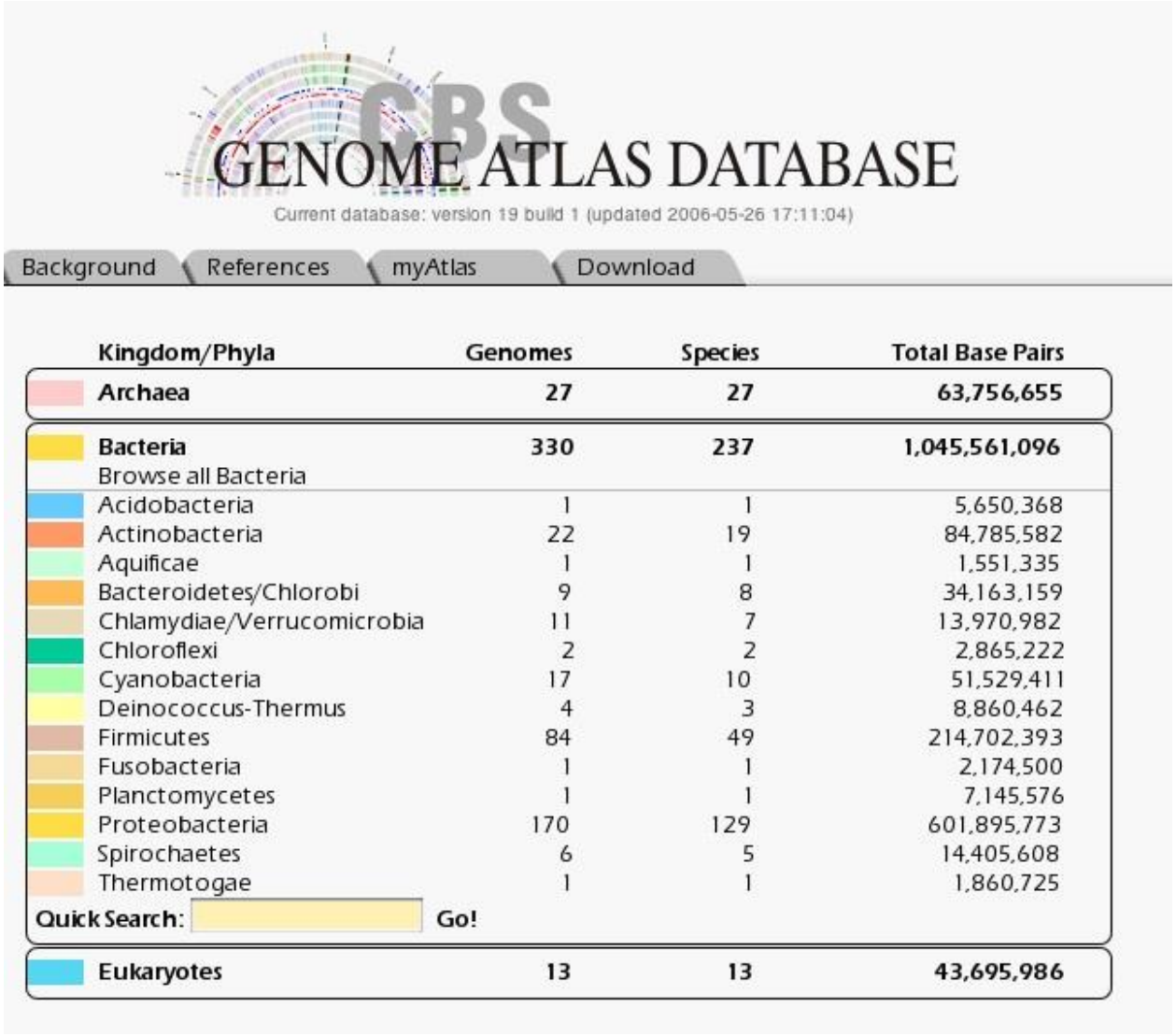
Caratteristiche

- Non hanno nucleo
- vivono in ambienti estremi p.e *Methanococcus j.* isolato a -2600 m/slm , temp di 300° C , 270 atm
- genoma: molecola circolare di 1,66Mb con 1738 geni che codificano proteine,
- maggior parte dei geni (58%) non coincide con alcun altro gene noto
- geni: coinvolti in prod. energia , div. cellulare e metabolismo generale simili a Eubatteri
- geni densamente concentrati, operoni, no introni
- Similitudini con eucatioti : sintesi di RNA, proteine e DNA.
- proteine istoniche , DNA cromosomico organizzato in cromatina.

Microfotografia colorata al ME a scansione che mostra le cellule dell'archeon, *Methanococcus jannasheii,* un procariote con tre cromosomi



Le Banche Dati di Genomi Microbici



Kingdom/Phyla	Genomes	Species	Total Base Pairs
Archaea	**27**	**27**	**63,756,655**
Bacteria	**330**	**237**	**1,045,561,096**
Browse all Bacteria			
Acidobacteria	1	1	5,650,368
Actinobacteria	22	19	84,785,582
Aquificae	1	1	1,551,335
Bacteroidetes/Chlorobi	9	8	34,163,159
Chlamydiae/Verrucomicrobia	11	7	13,970,982
Chloroflexi	2	2	2,865,222
Cyanobacteria	17	10	51,529,411
Deinococcus-Thermus	4	3	8,860,462
Firmicutes	84	49	214,702,393
Fusobacteria	1	1	2,174,500
Planctomycetes	1	1	7,145,576
Proteobacteria	170	129	601,895,773
Spirochaetes	6	5	14,405,608
Thermotogae	1	1	1,860,725
Quick Search: Go!			
Eukaryotes	**13**	**13**	**43,695,986**

http://www.cbs.dtu.dk/services/GenomeAtlas/