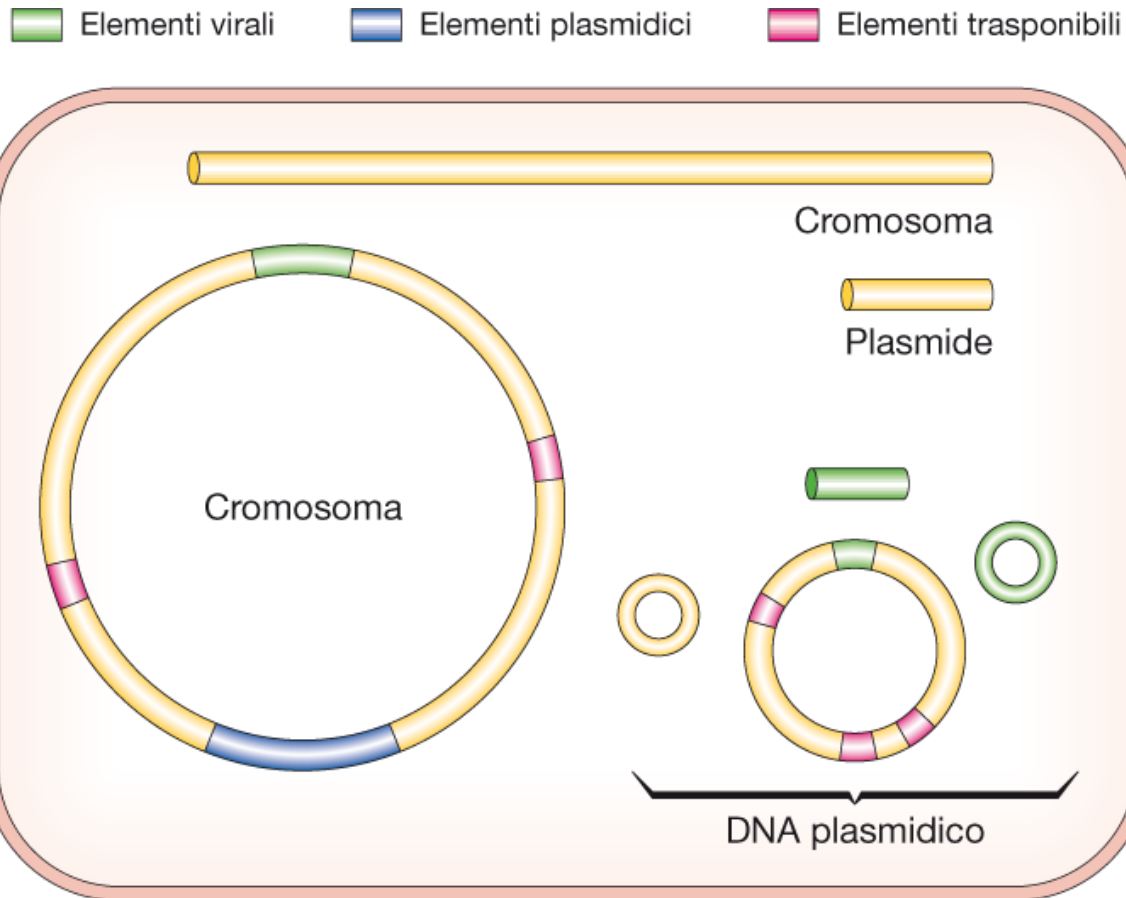


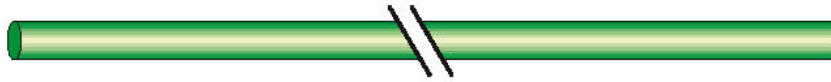
**Tabella 8.2 Dimensioni, numero e struttura dei cromosomi e dei plasmidi di alcune specie microbiche.**

ORGANISMO	CROMOSOMI			PLASMIDI		
	DIMENSIONE (Mpb)	NUMERO	STRUTTURA	DIMENSIONE (kpb)	NUMERO	STRUTTURA
<i>Agrobacterium tumefaciens</i>	2,1 + 3,0	2	Lineare + circolare	450 + 200	2	Circolare
<i>Bacillus subtilis</i>	4,2	1	Circolare			
<i>Bacillus thuringiensis</i>	5,7	1	Circolare	> 50 ognuno	6	
<i>Borrellia</i>	0,91	1	Lineare	5-200	Molteplici	Circolare + lineare
<i>Bradyrhizobium japonicum</i>	8,7	1	Circolare			
<i>Brucella melitensis</i>	2,1 + 1,2	2	Circolare			
<i>Brucella suis</i> biovars 1, 2, 4	1,0 + 2,0	2	Circolare			
<i>Brucella suis</i> biovar 3	3,1	1	Circolare			
<i>Buchnera</i> sp. ceppo APS	0,640	1	Circolare	< 7,8 ognuno	2	Circolare
<i>Deinococcus radiodurans</i>	2,6 + 0,4	2	Circolare	177 + 45	2	
<i>Escherichia coli</i> K-12	4,6	1	Circolare			
<i>Leptospira interrogans</i>	4,7 + 0,35	2	Circolare			
<i>Paracoccus denitrificans</i>	2,0 + 1,1 + 0,64	3	Circolare			
<i>Rhizobacterium meliloti</i>	3,4 + 1,7	2	Circolare	1400	1	Circolare
<i>Rhodobacter sphaeroides</i>	3,0 + 0,3	2	Circolare			
<i>Ureaplasma urealyticum</i>	1,75	1	Circolare			
<i>Vibrio cholerae</i>	2,9 + 1,1	2	Circolare			
<i>Vibrio parahaemolyticus</i>	3,2 + 1,9	2	Circolare			
<i>Xylella fastidiosa</i>	2,7	1	Circolare	5,1 + 1,3	2	Circolare

# Organizzazione del genoma nei procarioti da un cromosoma a più cromosomi



Cromosoma lineare di 910 Kb

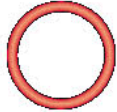


Plasmidi circolari

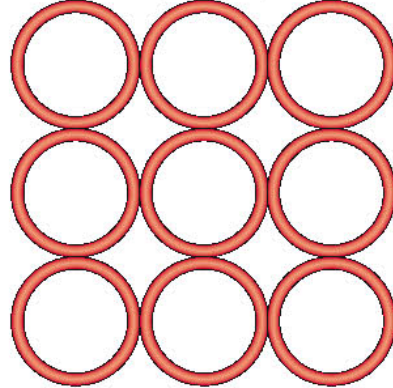
cp9



cp26



cp32 (1-9)



Plasmidi lineari

lp5



lp17



lp21



lp21



lp28-1



lp28-2



lp28-3



lp28-4



lp36



lp39



lp54



lp56



## Il genoma di *Borrellia burgdorferi*

Alcune specie di *Borrellia* (una spirocheta) sono responsabili della malattia di Lyme trasmessa all'uomo tramite le zecche.

Il genoma è costituito da:

- un cromosoma circolare di 1 Mb
- 21 (...!!) plasmidi alcuni circolari, altri lineari

- Alcuni plasmidi sia lineari che circolari (in rosso) sono importanti per la virulenza.

- Crescendo i batteri a T dell'insetto 23°C o a 37°C nell'uomo, si è potuto capire quali fossero i geni essenziali per la colonizzazione dei 2 habitat.

## SEQUENZE GENOMICHE ED EVOLUZIONE ; quale contributo?

La disponibilità di intere sequenze genomiche da molti microrganismi costituisce un'opportunità unica per ottenere una descrizione più obiettiva delle relazioni evolutive tra i diversi microrganismi

La scoperta che i processi di

- trasferimento genico orizzontale
- perdita di geni

svolgono un ruolo cruciale nell'evoluzione dei microrganismi ha profondamente influito sugli studi dei genomi batterici.

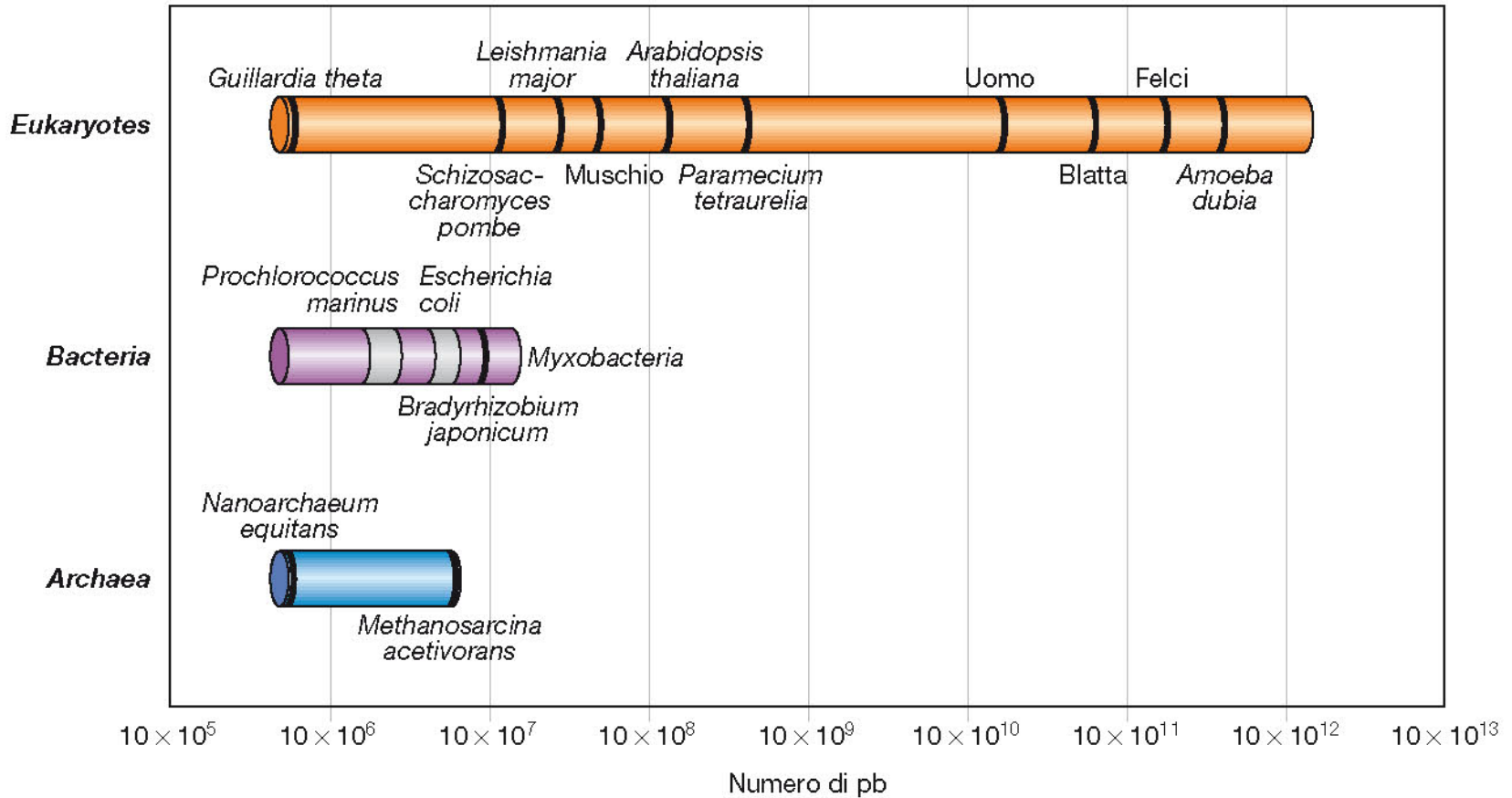
In particolare è cambiato il concetto che mutazioni all'interno dei singoli geni fossero la fonte principale delle variazioni fenotipiche in funzione della selezione naturale.

Il sequenziamento dei genomi prosegue con un ritmo sempre più accelerato

Dai primi dati risulta evidente che bisogna sequenziare più di un microrganismo all'interno della stessa specie per capirne pienamente la funzionalità

Le sequenze genomiche effettuate finora sono prevalentemente quelle di microrganismi di importanza clinica con un genoma di circa 3 Mb visione altamente parziale del mondo microbico

# Quanti geni nei procarioti??

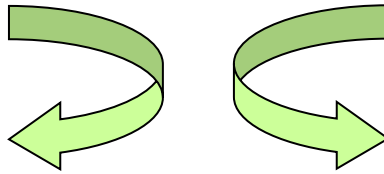


Dall'analisi dei genomi sequenziati il numero dei geni nei Procarioti (Batteri e Archea) varia da 500 a 10.000 mentre negli Eucarioti da 2.000 a 35.000. queste stime non includono i geni che sono trascritti in sRNAs. A livello di dimensioni del genoma vediamo però che i batteri arrivano ad un massimo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

# 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

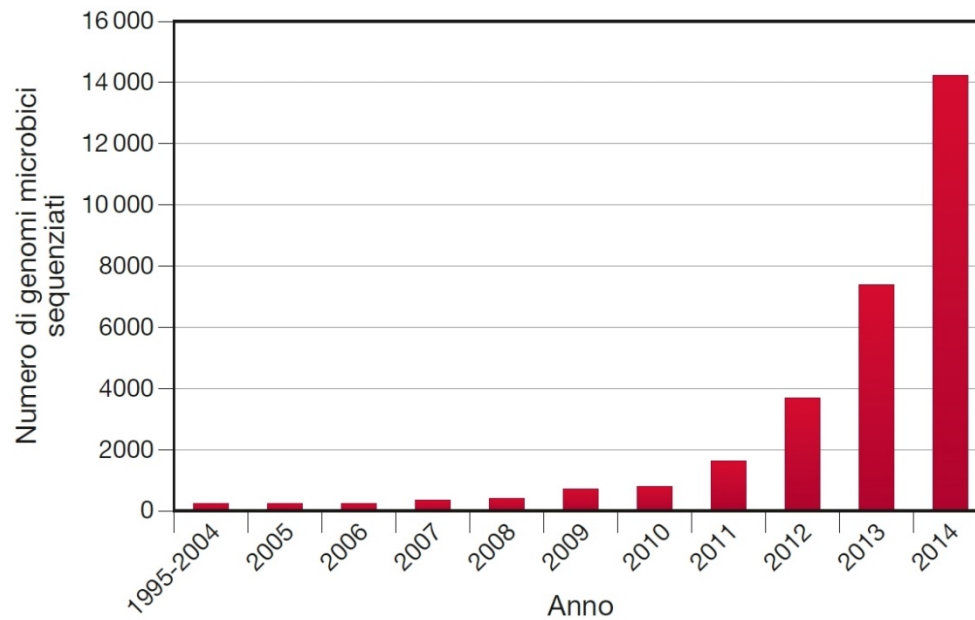
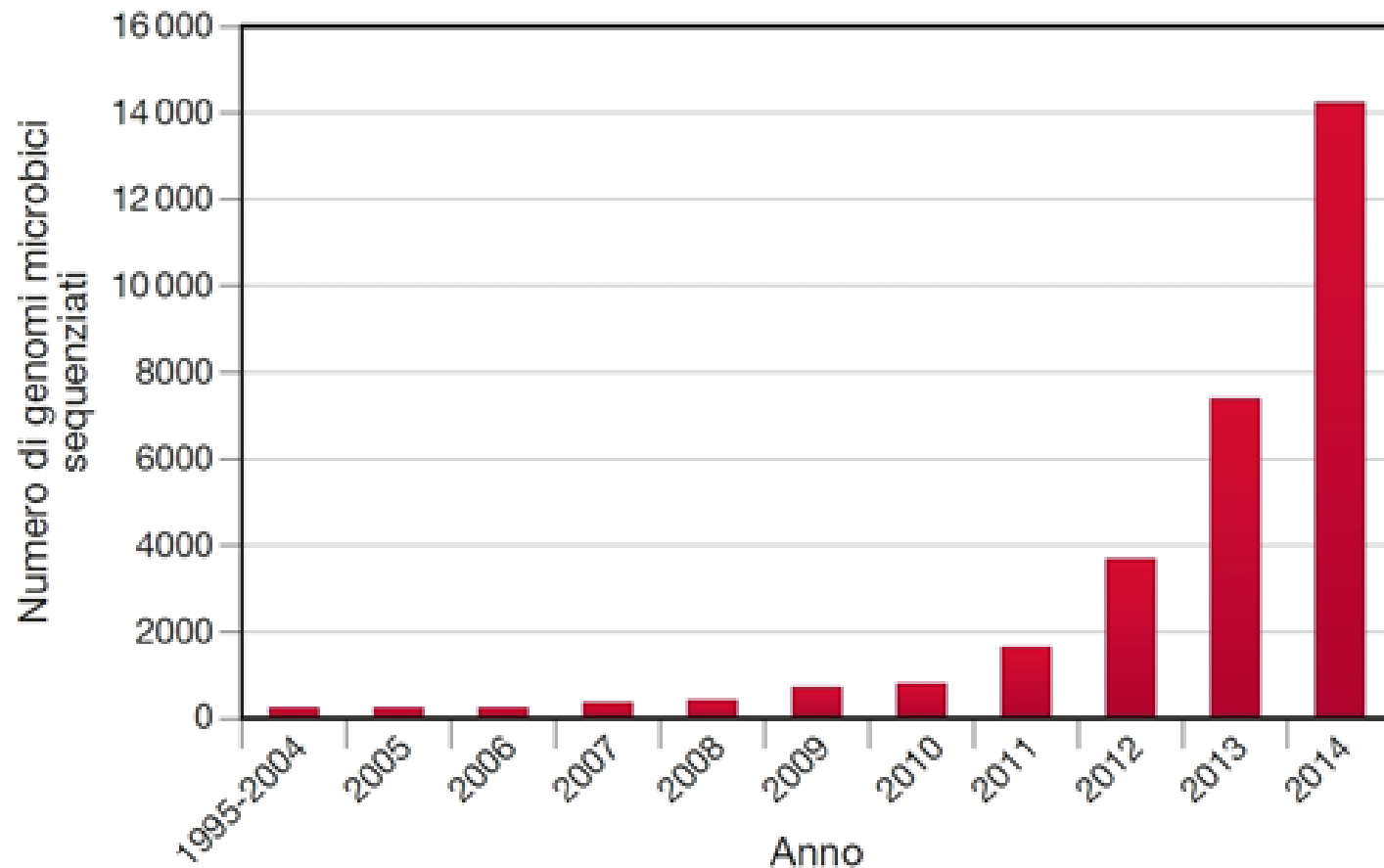


Figura 15.2 INCREMENTO DEL NUMERO DI SEQUENZE DI GENOMI MICROBICI DEPOSITATE PRESSO LA BANCA DATI GENBANK DAL 1995 AL 2014.



# Incremento del numero di sequenze di genomi microbici depositate in GeneBank



**Organismo****Dimensioni  
(in paia di basi)****ORF<sup>b</sup>****Commenti****Genomi di Batteri sequenziati****Bacteria**

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

# Genomi degli Archea sequenziati

## Archaea

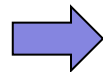
<i>Nanoarchaeum equitans</i>	490 885	552	Il più piccolo genoma conosciuto (▶ vol. 2A, cap. 17.11)
<i>Thermoplasma acidophilum</i>	1 564 905	1509	Termofilico, acidofilo (▶ vol. 2A, cap. 17.5)
<i>Methanocaldococcus jannaschii</i>	1 664 976	1738	Metanogeno (▶ vol. 2A, cap. 17.4)
<i>Aeropyrum pernix</i>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

L'analisi dei genomi procarioti ha evidenziato **una forte relazione tra le dimensioni di un genoma e il suo contenuto in ORF ( Open Reading Frame)**

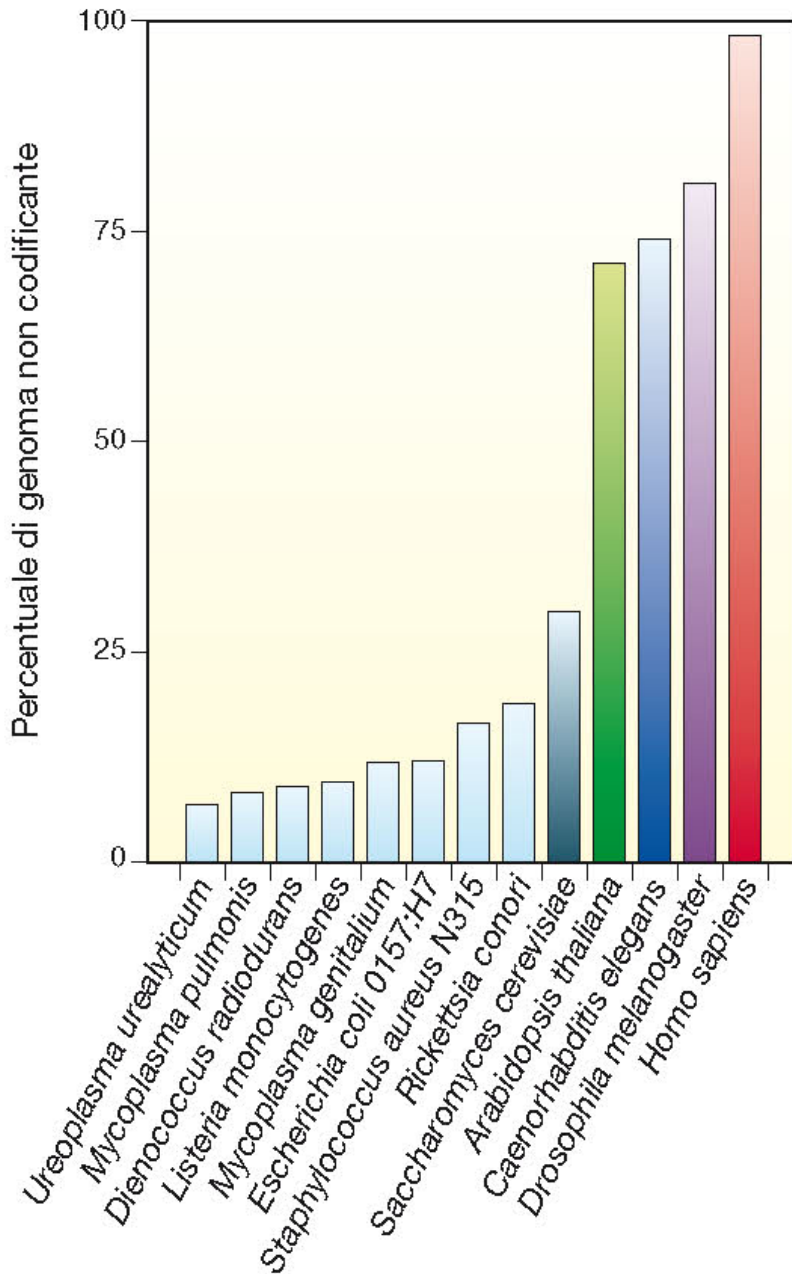
Indipendente dal microrganismo per ogni megabase di DNA) procariotico (Mb=1000 kb) si possono contare in media 1000 ORF.

A differenza degli eucarioti dove il DNA non codificante può costituire una larga frazione del genoma ( specialmente negli eucarioti superiori) nei procarioti all'aumento delle dimensioni del genoma corrisponde un aumento del numero di geni.

Curiosità

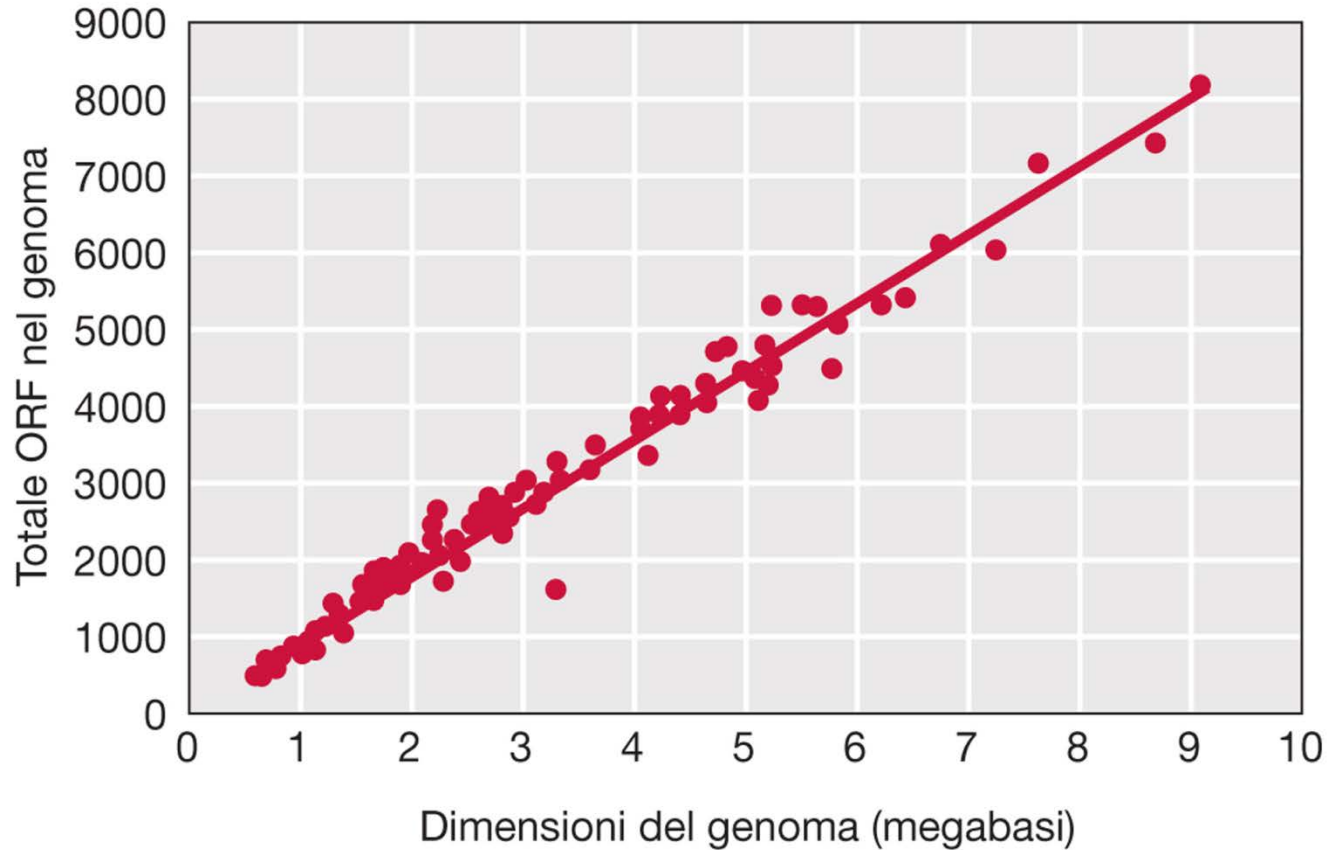


Il genoma di **Mycoplasma genitalium** è più piccolo di quello del virus Chlorella e del batteriofago G



Vi è quindi un enorme disparità nelle capacità codificanti dei genomi dei procarioti ( l'intero genoma) rispetto a quello del genoma umano ( 2%)

## Correlazione tra dimensioni del genoma e numero di ORF nei procarioti



Dati derivanti dall'analisi di 115 genomi procariotici ( Batteri ed Archea)

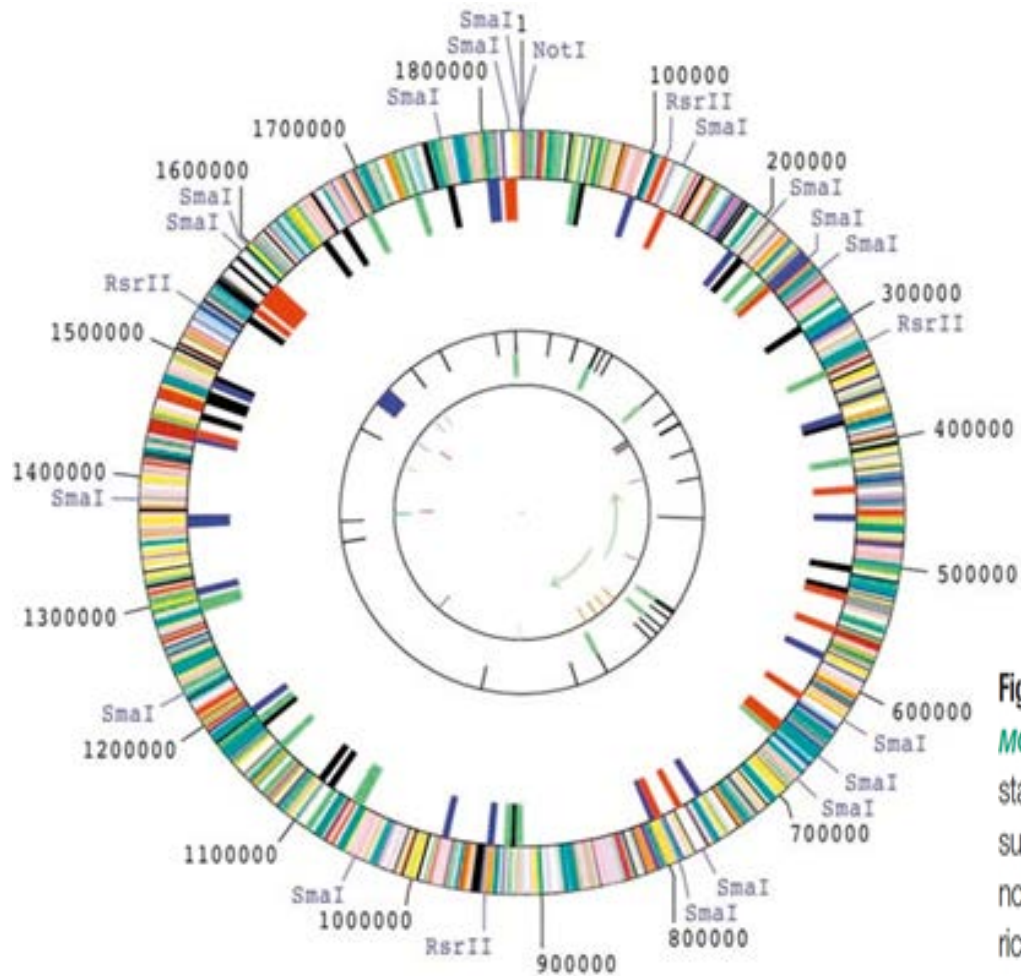
# Annotazioni dei genomi

= conversione dei dati di sequenza in una lista di geni presenti nel genoma.

- la maggior parte dei geni dei microrganismi codificano proteine
- la stragrande maggioranza del genoma è costituito da sequenze codificanti
- ricerca di ORF funzionali

I geni che codificano per gli rRNA o per i tRNA non sono riconosciuti dai programmi per la ricerca delle ORF. Vengono generalmente localizzati sulle sequenze dei genomi considerando che le sequenze di questi RNA sono molto conservate tra le diverse specie.

## Il genoma di *Haemophilus influenzae*



**Figura 15.1 MAPPA DELLA SEQUENZA COMPLETA DEL GENOMA DI *HAEMOPHILUS INFLUENZAE* CEPPLO RD.** La figura, apparsa sulla copertina della rivista *Science* nel luglio del 1995, mostra la localizzazione delle 1737 ORF predette sul genoma circolare composto da 1 830 137 coppie di basi. Sul perimetro esterno sono riportate le coordinate (in pdb) della sequenza e la posizione dei siti di riconoscimento di tre enzimi di restrizione (*NotI*, *RsrII* e *SmaI*). Il cerchio esterno indica le regioni codificanti per ORF, colorate in base alla loro funzione. Il secondo cerchio indica con colori diversi le regioni con alto e basso contenuto in G+C e le regioni ricche in A+T. I segmenti blu del cerchio successivo indicano la posizione dei cloni sequenziati. Il quarto cerchio indica la posizione dei sei operoni ribosomali (*in verde*), dei tRNA (*in nero*) e di un profago (*in blu*). Il quinto cerchio indica la posizione delle ripetizioni in tandem. L'origine di replicazione è evidenziata da due frecce *in verde* divergenti da un punto vicino alla base 603 000. Due presunti siti di terminazione della replicazione sono indicati dall'altra parte del cerchio *in rosso*.



# Come si rappresenta il genoma di un microrganismo ?

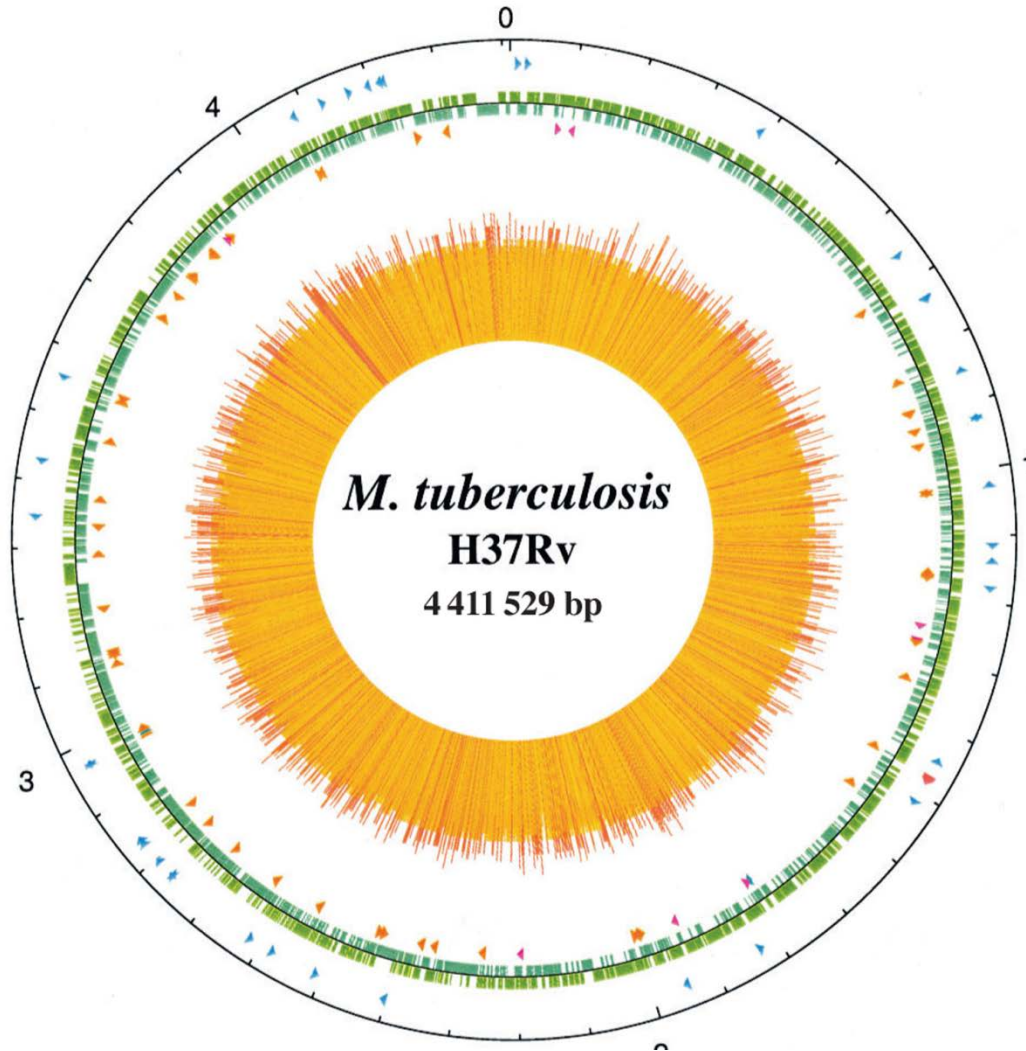
La mappa di *Mycobacterium tuberculosis*, l'agente eziologico della tubercolosi, ottenuta con clonaggio shotgun

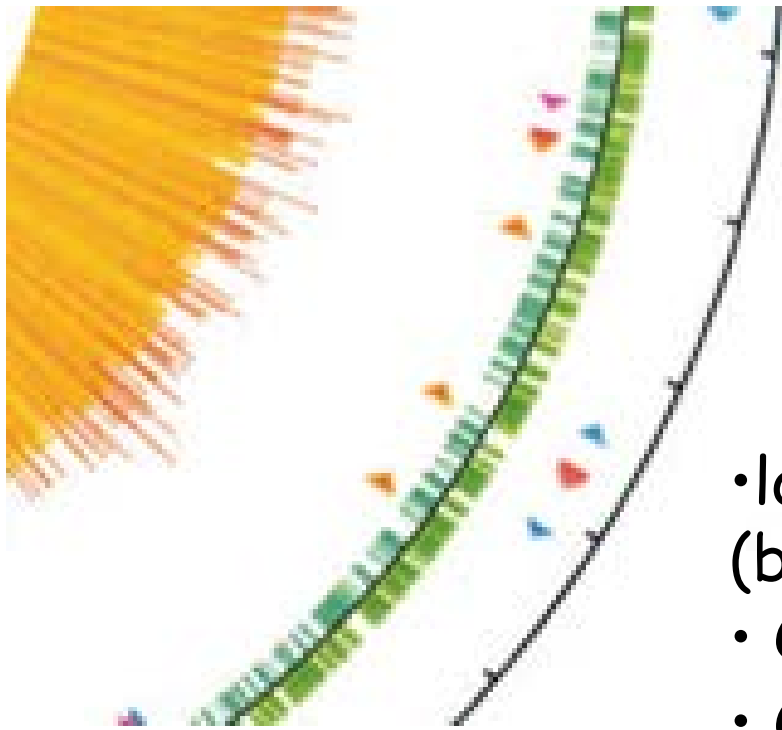
Gli anelli dall'esterno verso l'interno

- Dimensioni del genoma a partire dall'origine di replicazione

contenuto in GC con raggi gialli (minore 65%)

Raggi arancioni (maggiore 65%)



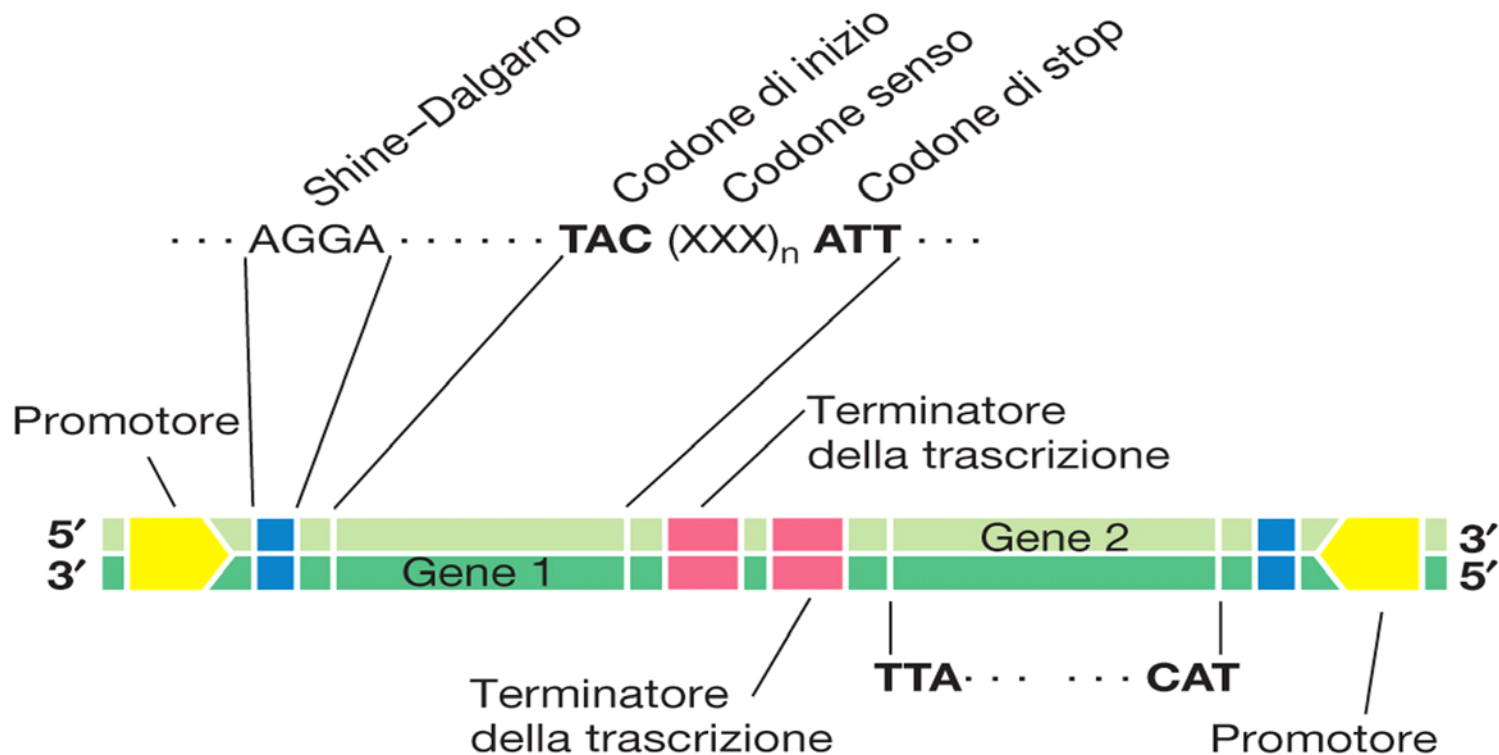


- localizzazione dei geni per i tRNA (blu) e per gli rRNA (arancio)
- ORF trascritte in senso orario
- ORF trascritte in senso antiorario
- elementi di DNA ripetuto (IS arancio)

## Identificazione di una ORF

Una ORF funzionale codifica una proteina nella cellula

- ricerca di una sequenza Shine Dalgarno di legame dei ribosomi
- ricerca del codone di inizio ATG (TAC)
- spaziatura di almeno 100 codoni in fase codificanti AA
- codone di stop UAA (ATT)



## Identificazione di una ORF: preferenza dei codoni

I singoli organismi mostrano una preferenza nell'utilizzo dei codoni sinonimi.

La preferenza dei codoni può suggerire se una ORF è funzionale o no.

Se l'uso dei codoni in una data ORF è molto differente da quello generalmente utilizzato da un determinato microrganismo, l'ORF potrebbe rivelarsi funzionale (o no) ma deriverebbe da un evento di trasferimento genico orizzontale.

## ORF Funzionale ?

Una ORF viene considerata funzionale

- anche se la sua sequenza è simile ad altre ORF presenti in genomi di altri microrganismi
- se alcune porzioni della ORF presentano sequenze conosciute in quanto codificano domini proteici funzionali

Proteine che svolgono la stessa funzione vengono considerate proteine omologhe e quindi correlate in senso evolutivo e possono condividere caratteristiche di sequenza e di struttura.

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

Le 470 ORF di *M. genitalium* sono presenti anche in *M. pneumoniae*

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.

Tra i più piccoli genoma finora identificati vi è quello di un Archea *Nanoarchaeum equitans* il cui genoma è di 90 kb più piccolo di quello di *M.genitalium*.

*N.equitans* pur essendo privo di geni per la sintesi di proteine coinvolte nei processi di catabolismo ed anabolismo ha un genoma interamente codificante ed un numero di ORF superiore a *M.genitalium*

### Genomi molto grandi

*Bradyrhizobium japonicum* ( responsabile della fissazione dell'azoto) contiene 2800 ORF in più rispetto al genoma di *Saccharomyces cerevisiae*.

*Streptomyces coelicolor* ha un genoma di 8 Mb, 7846 ORF circa 1800 in più rispetto a *S. cerevisiae*.

## 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 circa 3000 dei 4.300 potenziali geni

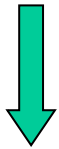
Un ORF non identificata contiene tutte le caratteristiche di un gene ipotetico quando possiede un sito di legame dei ribosomi un codone di inizio ed un codone di stop.

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



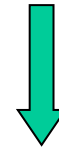
Una volta nota la sequenza di un genoma la funzione dei geni e delle ORF viene inizialmente assegnata ricercando in database i geni omologhi.

I geni omologhi possono essere suddivisi  
in geni **ORTOLOGHI** e geni **PARALOGHI**



derivano da un progenitore comune e codificano proteine con la stessa funzione in specie diverse

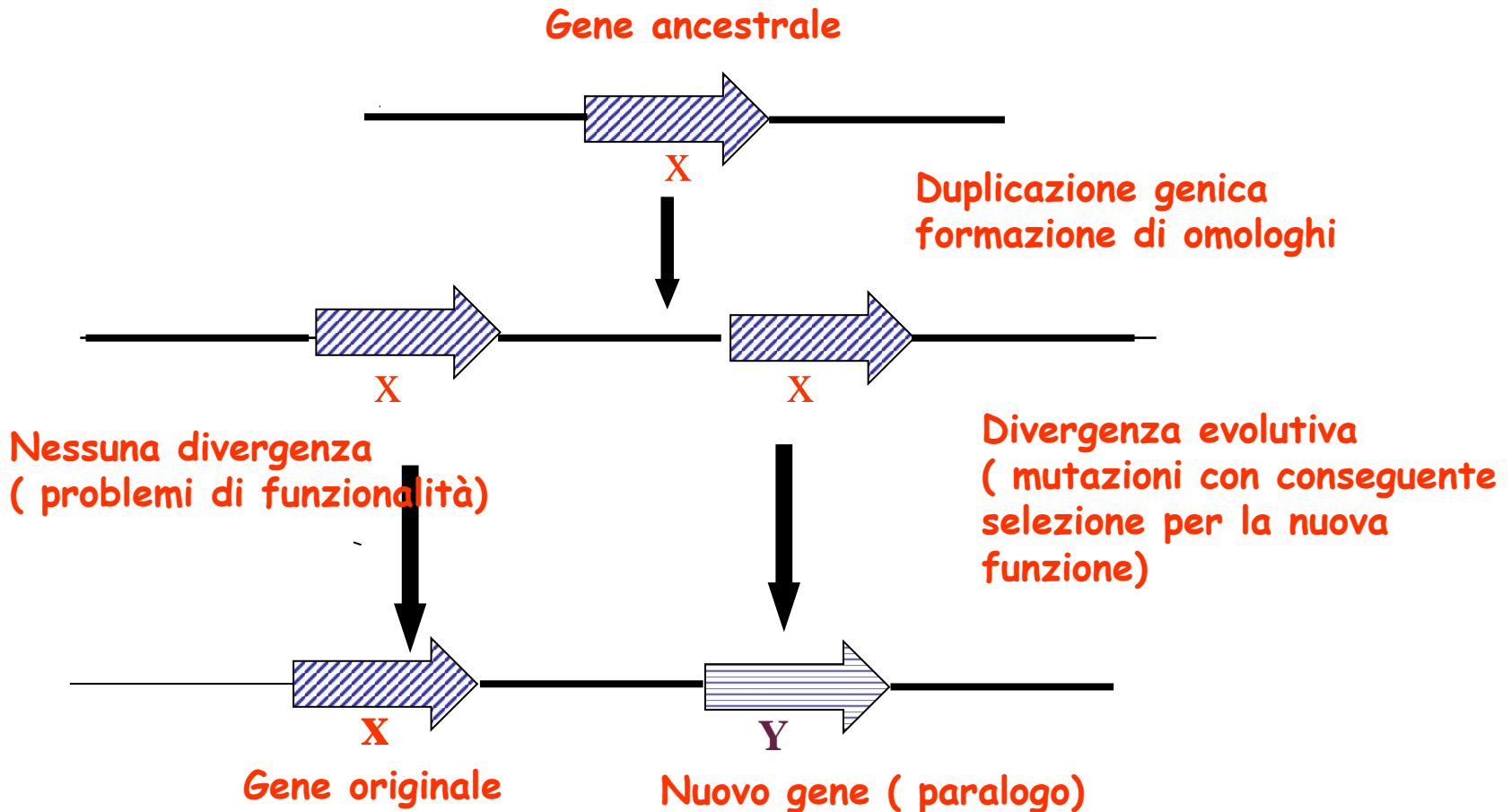
La storia di questi geni riflette la storia della specie



si sono evoluti per duplicazione genica e possono codificare proteine con funzione simile ma non identica

L' omologia è il risultato di duplicazione piuttosto che di speciazione

# Duplicazione genica e divergenza : formazione dei paraloghi



Progenitore

X

A

B

Duplicazione genica

Speciazione

Speciazione

Speciazione

A<sub>1</sub>

B<sub>1</sub>

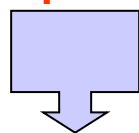
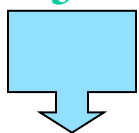
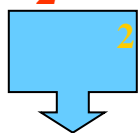
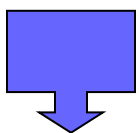
A<sub>2</sub>

C

B<sub>3</sub>

A<sub>4</sub>

B<sub>4</sub>



Specie 1

Specie 2

Specie 3

Specie4

A<sub>1</sub> A<sub>2</sub> A<sub>4</sub>  
A B C

ortologi  
paraloghi

# Identificazione dei geni ortologhi

Nel confronto tra due genomi i geni ortologhi dovrebbero avere la più elevata omologia di sequenza ed essersi diversificati recentemente rispetto ai paraloghi che si sono originati per duplicazione prima della speciazione.

Difficile definire un gene ortologo solo sulla base della sequenza.

Si può quindi cercare :

ordine genico osservare se i geni fiancheggiati siano ortologhi

Identificazione degli ortologhi è ostacolata da

- divergenza genica
- duplicazione genica accompagnata da perdita di un gene (A1 e B3 nell'esempio precedente sono paraloghi non ortologhi)
- struttura a moduli di molte proteine con struttura a domini multipli proteine non ortologhe possono avere domini ortologhi

# Quale il ruolo della duplicazione genica ?

La duplicazione genica ha permesso:

- una rapida diversificazione di reazioni enzimatiche
- un aumento delle dimensioni del genoma con potenzialità per lo sviluppo di nuove proprietà enzimatiche
- in seguito alla duplicazione uno dei due geni può diventare **DISPENSABILE** e quindi può subire una serie di mutazioni o riarrangiamenti

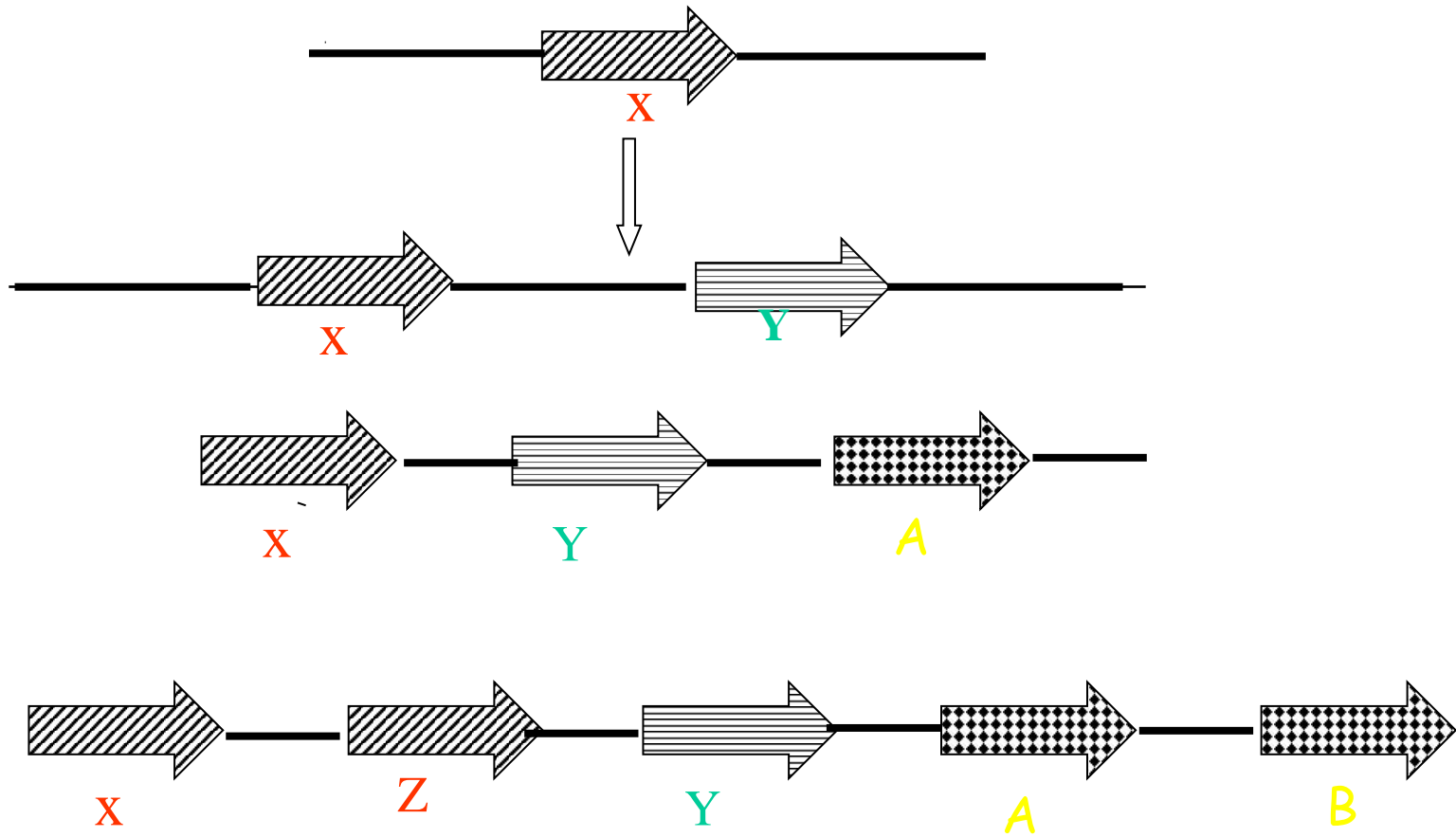
In genere i geni paraloghi svolgono funzioni diverse ,seppur simili, all'interno dello stesso microrganismo .

La duplicazione genica può generare però copie di un gene che mantiene la stessa funzione permettendo la produzione di grande quantità di RNA o proteine ( rRNA)

Due paraloghi possono poi dar origine a loro volta a duplicazione formando una famiglia di geni paraloghi

# Evoluzione di una famiglia di geni paraloghi

Gene ancestrale

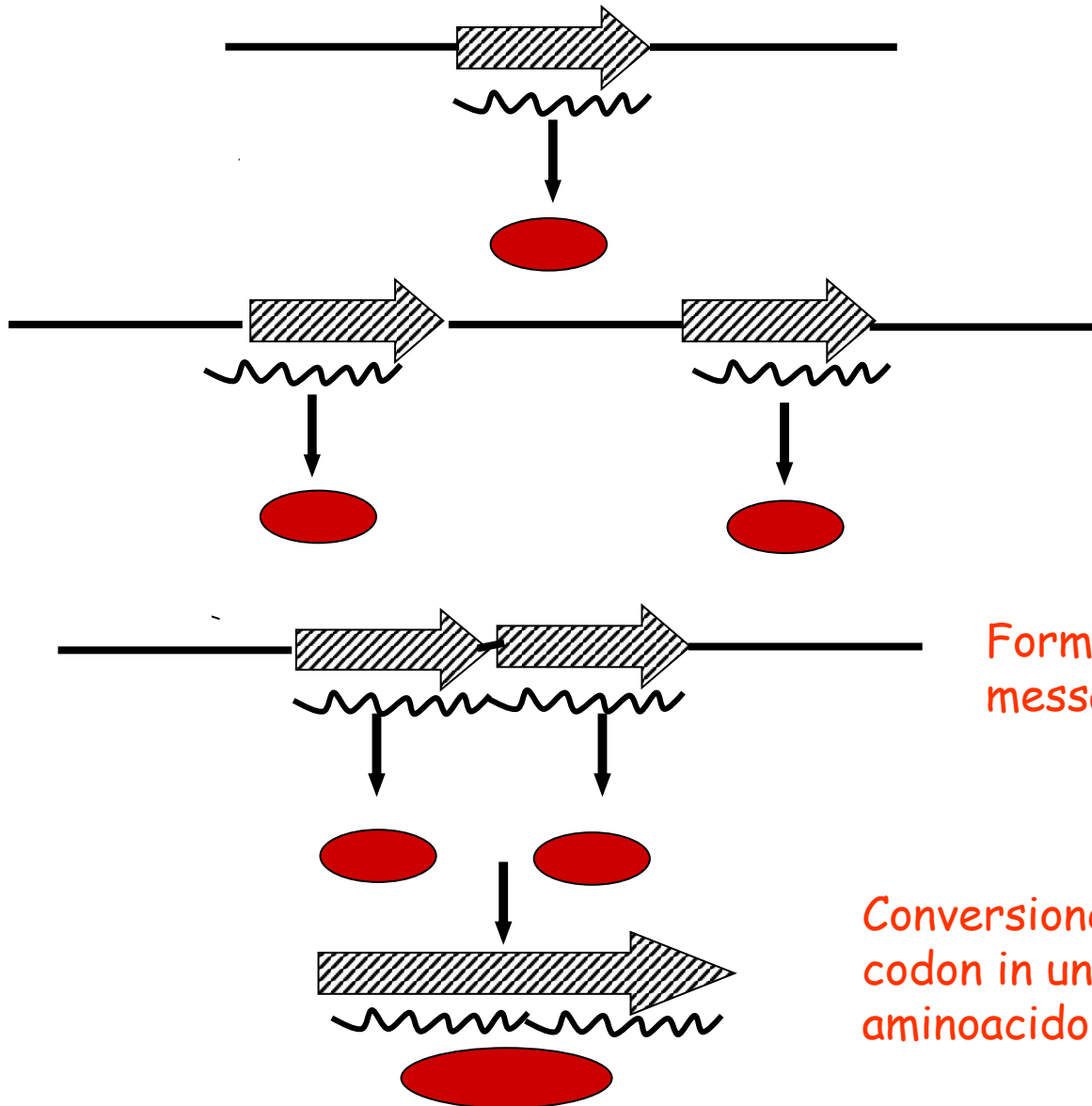


Famiglia di geni paraloghi

## Tipi di duplicazione del DNA

- intero gene
  - parte di un gene
  - segmenti di DNA comprendenti uno o più geni
  - interi operoni
  - parte di un cromosoma
- 
- Elongazione di un gene:allungamento di un gene è una tappa importante nel processo evolutivo.
- 
- Il meccansimo prevede una duplicazione dei geni posti in arrangiamento a tandem.
  - Per eliminazione della sequenza tra i due geni si ha dapprima la formazione di un mRNA policistronico
  - formazione di un gene costituito da due porzioni di paraloghi
  - Molte delle proteine attuali contengono moduli ripetuti o domini funzionali che si possono essere formati per duplicazione genica.

# Formazine di proteine di fusione



Formazione di un  
messaggero policistronico

Conversione di uno stop  
codon in un codone per un  
aminoacido



# Destino delle sequenze duplicate

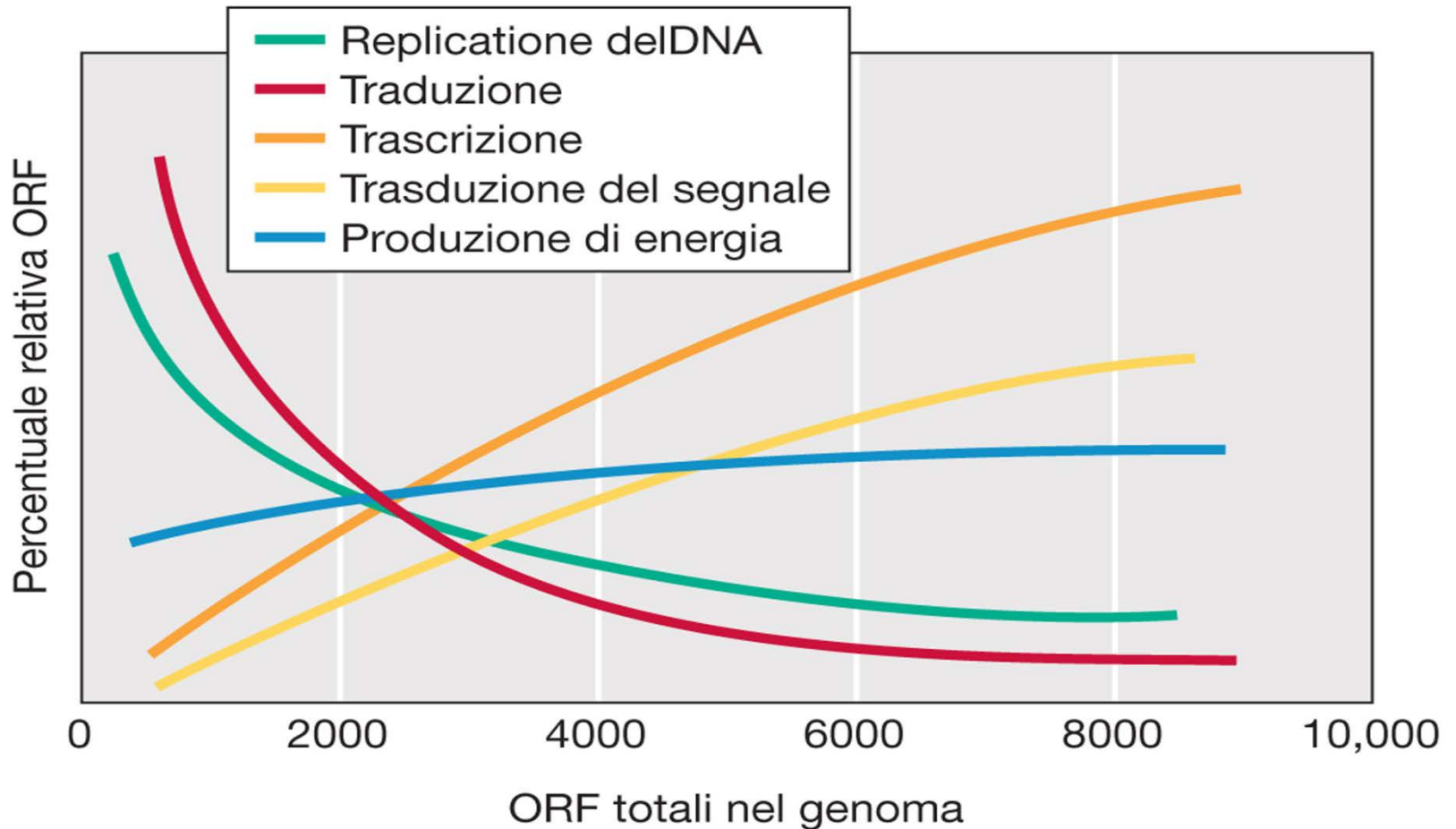
1. Una delle due copie può accumulare mutazioni e diventare un nuovo gene
2. una copia può essere inattivata da mutazioni
3. se non divergono possono indurre la maggiore quantità di prodotto genico (RNA e/o proteina)
4. si può avere allungamento genico ( duplicazione in tandem)
5. si può avere formazione di geni chimerici con fusione di una copia in un gene adiacente

## Funzione genica nei genomi batterici : rapporto funzione dimensioni

<b>Categorie funzionali</b>	<b>Percentuale di geni sul cromosoma nella categoria di riferimento</b>		
	<b><i>Escherichia coli</i> (4,64 Mbp)<sup>a</sup></b>	<b><i>Haemophilus influenzae</i> (1,83 Mbp)<sup>a</sup></b>	<b><i>Mycoplasma genitalium</i> (0,58 Mbp)<sup>a</sup></b>
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

<sup>a</sup> 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.



Organismi di dimensioni maggiori mostrano un numero più elevato di geni coinvolti nei processi di trascrizione o di regolazione rispetto ai microrganismi con genomi più piccoli.

Questi meccanismi regolativi permettono alla cellula di rispondere in maniera migliore alla disponibilità di substrati diversi attraverso l'espressione di geni specifici.

Gli organismi di piccole dimensioni fanno a meno di questi processi regolativi e sono in genere parassiti

## *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.

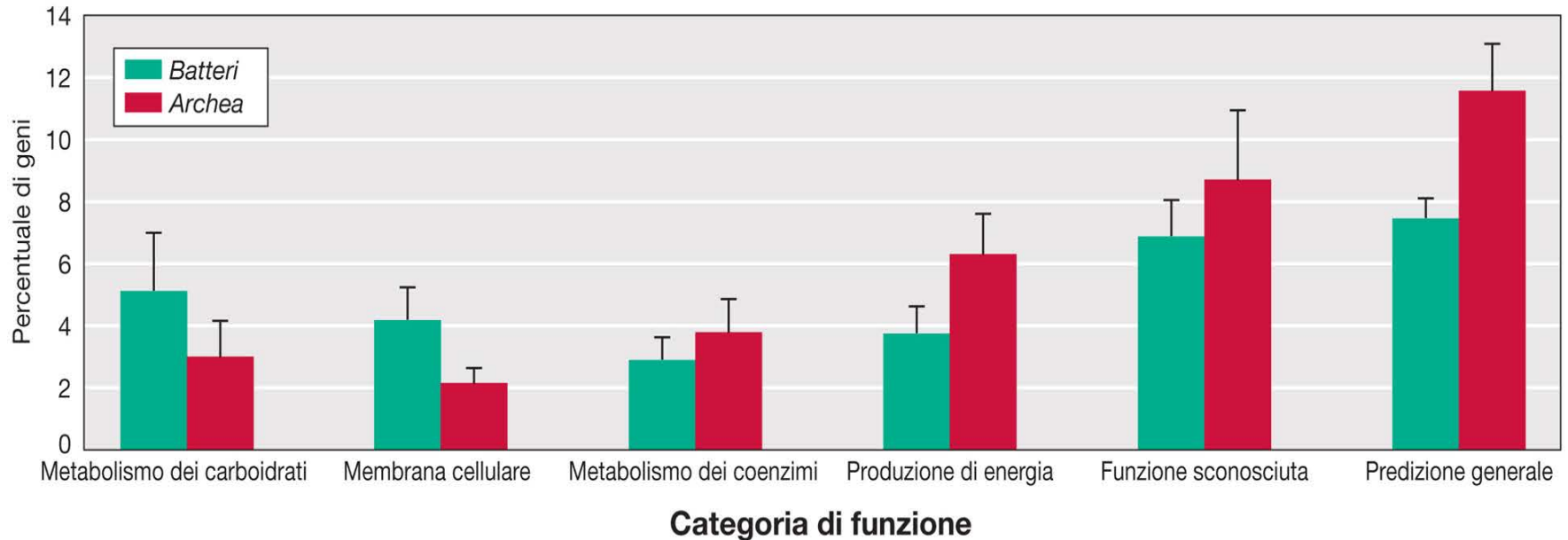
I meccanismi di regolazione permettono alla cellula di rispondere in modo migliore alla disponibilità di substrati differenti attraverso l'attivazione di geni specifici.

**Uno degli habitat principali dei microrganismi è il suolo.**

Nella gran parte dei casi i microrganismi con genomi di dimensioni maggiori alle 6 Mb 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 intermittente.

# Variazione delle categorie geniche nei Batteri e negli Archea



Funzione sconosciuta = geni noti che codificano proteine la cui funzione è sconosciuta

Predizione generale = geni che codificano proteine ipotetiche che potrebbero esistere o meno

# 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 carboidrati
- numero ridotto di geni collegati alle funzioni della membrana cellulare come il trasporto e la biosintesi delle membrane

## I genomi dei Batteri sono caratterizzati

- 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



**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
<i>Mycoplasma genitalium</i> (470)	+	-	-	-	-
<i>Buchnera</i> spp. (588)	+	-	+	+	+
<i>Rickettsia prowazekii</i> (834)	-	+	-	-	-
<i>Chlamidia trachomatis</i> (894)	+	-	+	-	-
<i>Treponema pallidum</i> (1014)	+	-	-	-	-
<i>Mycobacterium leprae</i> (1604)	parziale	in decadimento	+	+	+

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

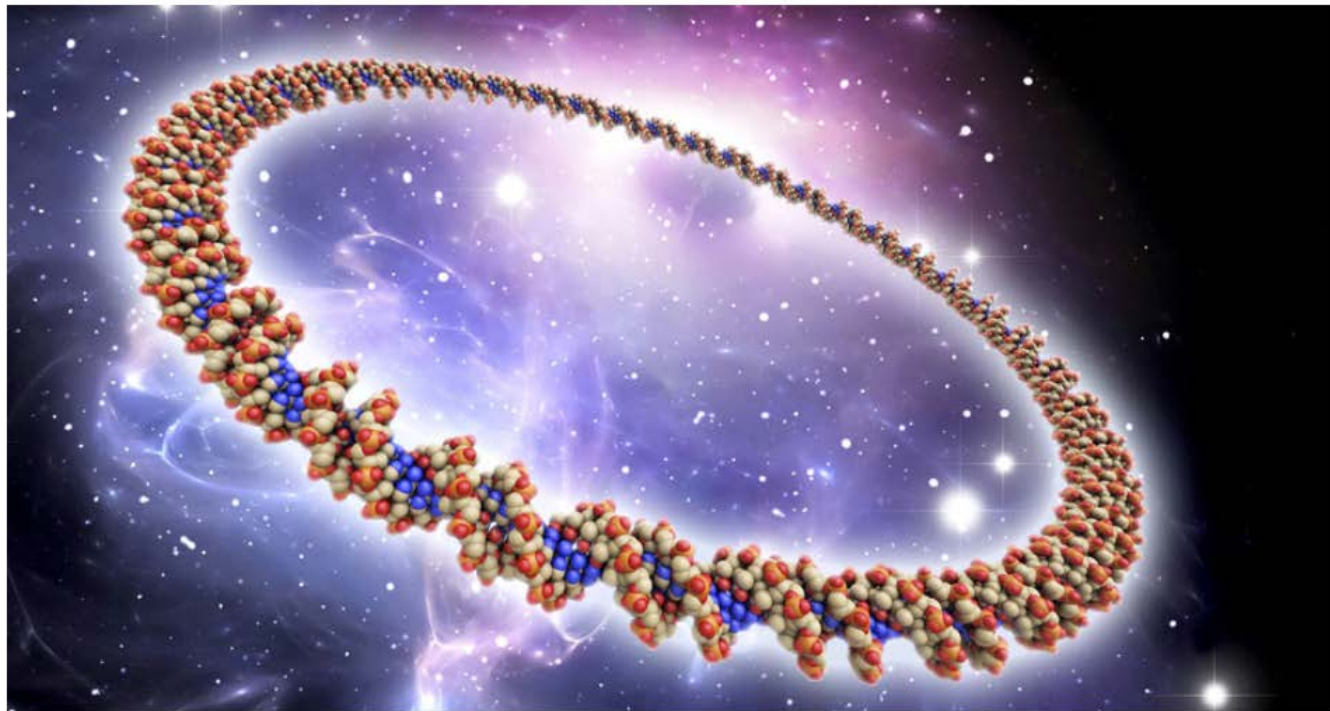
In *Mycobacterium leprae* circa 50 % dei geni sono pseudogeni che mantengono omologia con *Mycobacterium tuberculosis*

GENETICS MICROBES

# Genes: How few needed for life?

People need more than 20,000, but one bacterium was rebuilt — and lived — with a mere 473

BY TINA HESMAN SAEY APR 5, 2016 — 7:15 AM EST



Una domanda che scaturisce dal confronto dei genomi riguarda la dimensione minima sufficiente perchè vi sia VITA.

E' possibile che esistano specie con un genoma più piccolo di 0.58Mb ovvero quello di *M.genitalium* che per ora sembra la forma di batterio con minore geni (che non sia endosimbionte degli insetti)

E' stata fatta quindi in silico un 'analisi genomica comparativa tra i genomi più piccoli per capire quali fossero i geni conservati

Mutagenesi di un gene alla volta per verificare se questo fosse essenziale alla vita della cellula

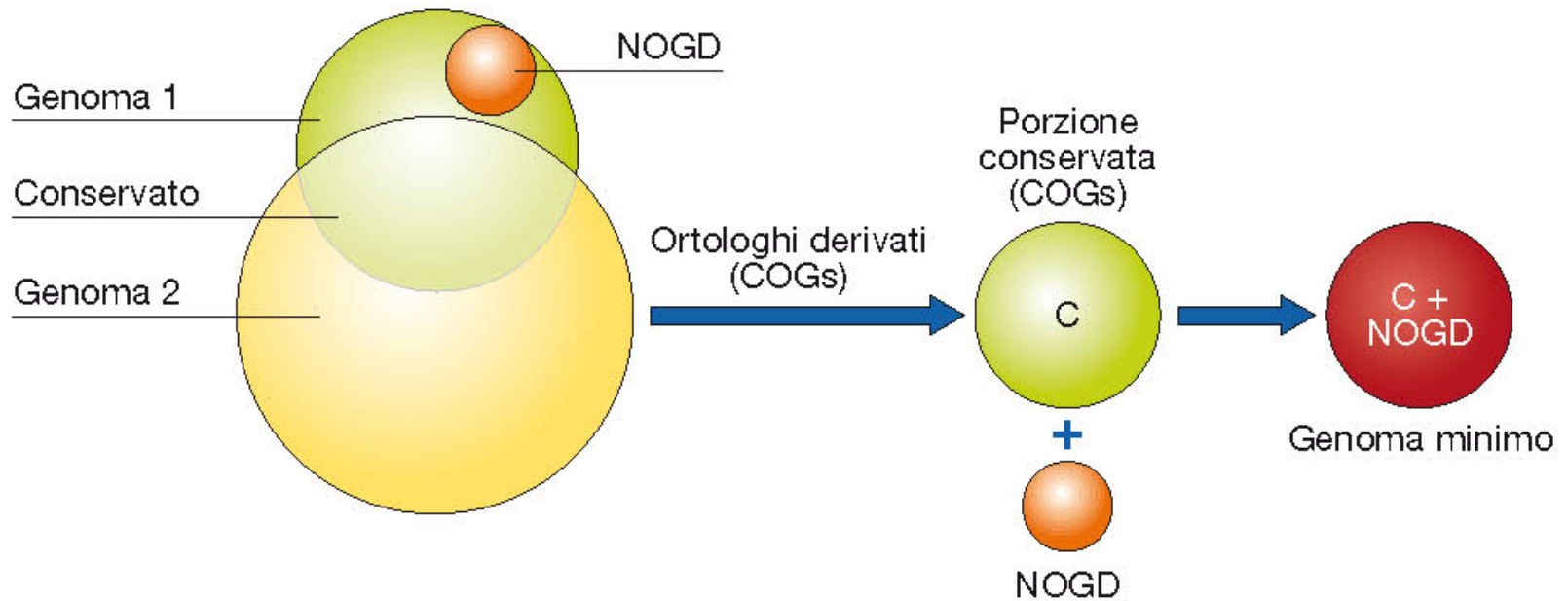
La prima proposta di genoma minimo viene dall'analisi dei 2 genomi sequenziati per primi *Haemophilus influenzae* ( 1815 geni) e *Mycoplasma genitalium* 525 geni) che rivela un gruppo comune di **geni 256** molto più piccolo del numero di geni presenti nel genoma di *Mycoplasma*.

Da qui l'idea di introdurre il metodo di global transposon mutagenesis che ha permesso, effettuando uno studio sulla vitalità delle cellule portatrici di mutazioni in ciascuno dei 525 geni di *Mycoplasma* di stabilire che

150 geni sono NON ESSENZIALI

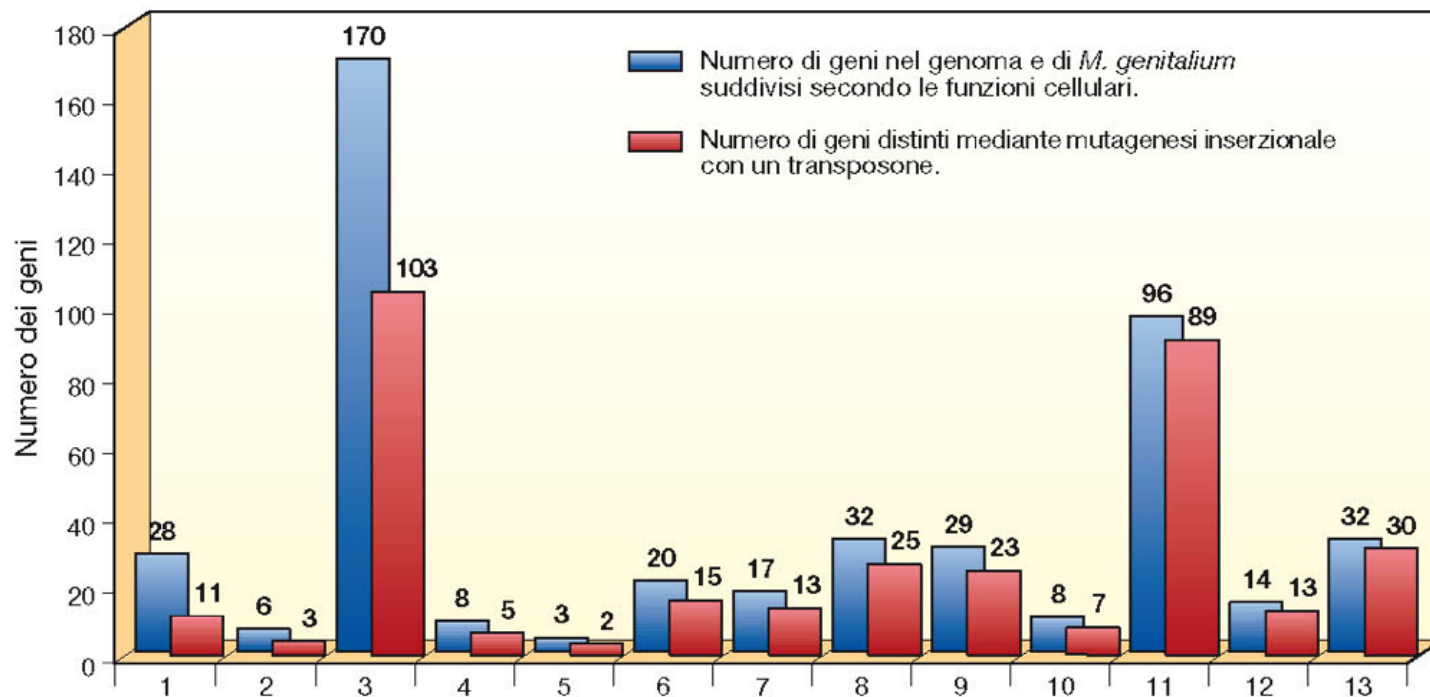
375 geni sono ESSENZIALI

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



L'analisi comparativa delle sequenze dei genomi di due organismi filogeneticamente distanti mostra l'esistenza di geni conservati nelle due specie (geni ortologi o COGs Cluster of Orthologous Genes) e di geni non ortologi specifici dell'uno o dell'altro organismo (NOGD Non Orthologous Gene Displacement). Il genoma minimo è il numero dei geni essenziali per la vita di uno specifico organismo ed è la somma dei geni COGs più i NOGS che saranno più o meno ampi.

Nel caso di *M. genitalium* , l'analisi di 1354 inserzioni casuali di trasposoni nel genoma mostra che dei 480 geni codificanti , 265-350 sono essenziali nelle condizioni di laboratorio, tra questi circa 100 geni codificano per funzioni non note.



#### Funzioni cellulari delle classi dei geni

- |  |   |
|--|---|
| 1. Membrana                              | 8. Trasporto                                    |
| 2. Regolazione                           | 9. Replicazione/ricombinazione/riparazione      |
| 3. Sconosciuto                           | 10. Metabolismo degli acidi grassi e dei lipidi |
| 4. Metabolismo principale                | 11. Traduzione                                  |
| 5. Biosintesi di cofattori               | 12. Processi cellulari                          |
| 6. Metabolismo delle purine e pirimidine | 13. Produzione di energia                       |
| 7. Trascrizione                          |   |

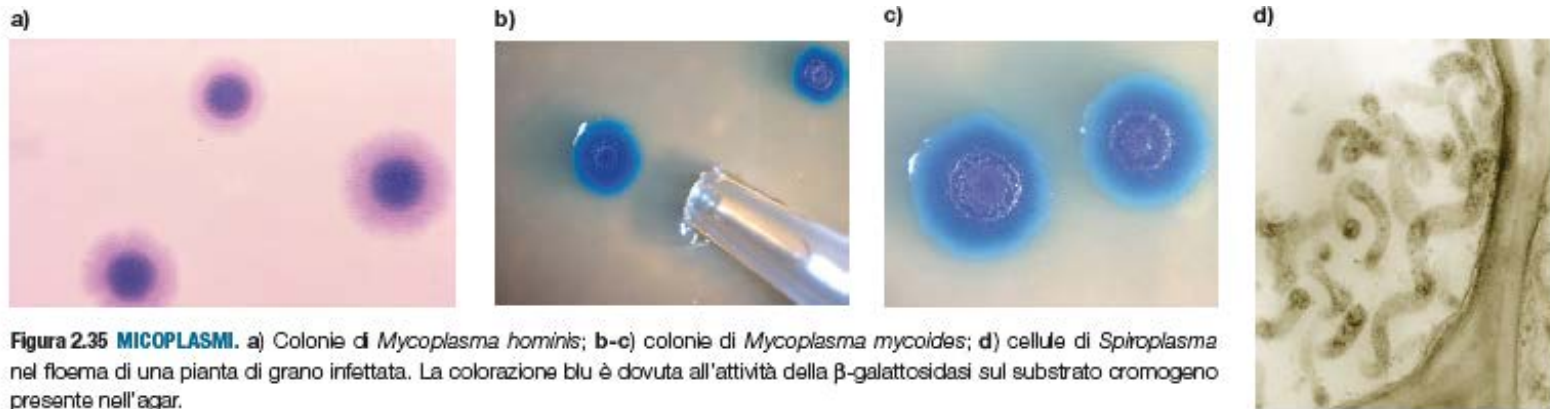


Figura 2.35 MICOPLASMI. a) Colonie di *Mycoplasma hominis*; b-c) colonie di *Mycoplasma mycoides*; d) cellule di *Spiroplasma* nel floema di una pianta di grano infettata. La colorazione blu è dovuta all'attività della  $\beta$ -galattosidasi sul substrato cromogeno presente nell'agar.

I micoplasmi batteri senza parete caratterizzati da un genoma molto piccolo (meno di 1 megabase), molte richieste nutrizionali.

Dal RNA16S appartengono ai Gram+, sono tutti commensali parassiti o patogeni, infezioni difficili da trattare con antibiotici



## Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jaysree Zaveri, Timothy B. Stockwell, Anushka Brownley, David W. Thomas, Mikkel A. Algire, Chuck Merryman, Lei Young, Vladimir N. Noskov, John I. Glass, J. Craig Venter, Clyde A. Hutchison III, Hamilton O. Smith\*

We have synthesized a 582,970–base pair *Mycoplasma genitalium* genome. This synthetic genome, named *M. genitalium* JCVI-1.0, contains all the genes of wild-type *M. genitalium* G37 except MG408, which was disrupted by an antibiotic marker to block pathogenicity and to allow for selection. To identify the genome as synthetic, we inserted “watermarks” at intergenic sites known to tolerate transposon insertions. Overlapping “cassettes” of 5 to 7 kilobases (kb), assembled from chemically synthesized oligonucleotides, were joined by *in vitro* recombination to produce intermediate assemblies of approximately 24 kb, 72 kb (“1/8 genome”), and 144 kb (“1/4 genome”), which were all cloned as bacterial artificial chromosomes in *Escherichia coli*. Most of these intermediate clones were sequenced, and clones of all four 1/4 genomes with the correct sequence were identified. The complete synthetic genome was assembled by transformation-

genome, we needed to reliable methods for the much larger synthetic I.

Strategy for synthesizing 580,076-bp *M. gen* (*Mycoplasma genitalium* genomic sequence; acc was partitioned into 10 roughly 5 to 7 kb in 10 individually synthesized and then joined together cassette boundaries we so that each cassette c complete genes. This deletion or manipulation ual cassettes. Most ca adjacent neighbors by segments overlapped t Cassette 101 overlappi pleting the circle.

Short “watermark” in cassettes 14, 29, 39, 5 inserted or substituted sa or encode information in can be either in noncod

*Science* (2008) :319, 1215

## Global Transposon Mutagenesis and a Minimal *Mycoplasma* Genome

Clyde A. Hutchison III,<sup>1,2\*</sup> Scott N. Peterson,<sup>1\*†</sup> Steven R. Gill,<sup>1</sup> Robin T. Cline,<sup>1</sup> Owen White,<sup>1</sup> Claire M. Fraser,<sup>1</sup> Hamilton O. Smith,<sup>1‡</sup> J. Craig Venter<sup>1‡§</sup>

*Mycoplasma genitalium* with 517 genes has the smallest gene complement of any independently replicating cell so far identified. Global transposon mutagenesis was used to identify nonessential genes in an effort to learn whether the naturally occurring gene complement is a true minimal genome under laboratory growth conditions. The positions of 2209 transposon insertions in the completely sequenced genomes of *M. genitalium* and its close relative *M. pneumoniae* were determined by sequencing across the junction of the transposon and the genomic DNA. These junctions defined 1354 distinct sites of insertion that were not lethal. The analysis suggests that 265 to 350 of the 480 protein-coding genes of *M. genitalium* are essential under laboratory growth conditions, including about 100 genes of unknown function.

*Science* (1999) :286, 2165

## RESEARCH ARTICLE SUMMARY

### SYNTHETIC BIOLOGY

## Design and synthesis of a minimal bacterial genome

Clyde A. Hutchison III,<sup>†</sup> Ray-Yuan Chuang,<sup>‡</sup> Vladimir N. Noskov, Nacrya Assad-García, Thomas J. Deerinck, Mark H. Ellisman, John Gill, Krishna Kannan, Bogamil J. Karas, Li Ma, James F. Peletier, Zhi-Qing Qi, R. Alexander Richter, Elizabeth A. Strychalski, Lijie Sun, Yo Suzuki, Bilyana Tsvetanova, Kim S. Wise, Hamilton O. Smith, John I. Glass, Chuck Merryman, Daniel G. Gibson, J. Craig Venter\*

**INTRODUCTION:** In 1984, the simplest cells capable of autonomous growth, the mycoplasmas, were proposed as models for understanding the basic principles of life. In 1995, we reported the first complete cellular genome sequences (*Haemophilus influenzae*, 1815 genes, and *Mycoplasma genitalium*, 525 genes). Comparison of these sequences revealed a conserved core of about 250 essential genes, much smaller than either genome. In 1999, we introduced the method of global transposon mutagenesis and experimentally demonstrated that *M. genitalium* contains many genes that are nonessential for growth in the laboratory, even though it has the

smallest genome known for an autonomously replicating cell found in nature. This implied that it should be possible to produce a minimal cell that is simpler than any natural one. Whole genomes can now be built from chemically synthesized oligonucleotides and brought to life by installation into a receptive cellular environment. We have applied whole-genome design and synthesis to the problem of minimizing a cellular genome.

**RATIONALE:** Since the first genome sequences, there has been much work in many bacterial models to identify nonessential genes and

define core sets of conserved genetic functions, using the methods of comparative genomics. Often, more than one gene product can perform a particular essential function. In such cases, neither gene will be essential, and neither will necessarily be conserved. Consequently, these approaches cannot, by themselves, identify a set of genes that is sufficient to constitute a viable genome. We set out to define a minimal cellular genome experimentally by designing and building one, then testing it for viability. Our goal is a cell so simple that we can determine the molecular and biological function of every gene.

**RESULTS:** Whole-genome design and synthesis were used to minimize the 1079-kilobase pair (kbp) synthetic genome of *M. mycoides* JCVI-syn1.0. An initial design, based on collective knowledge of molecular biology in combination

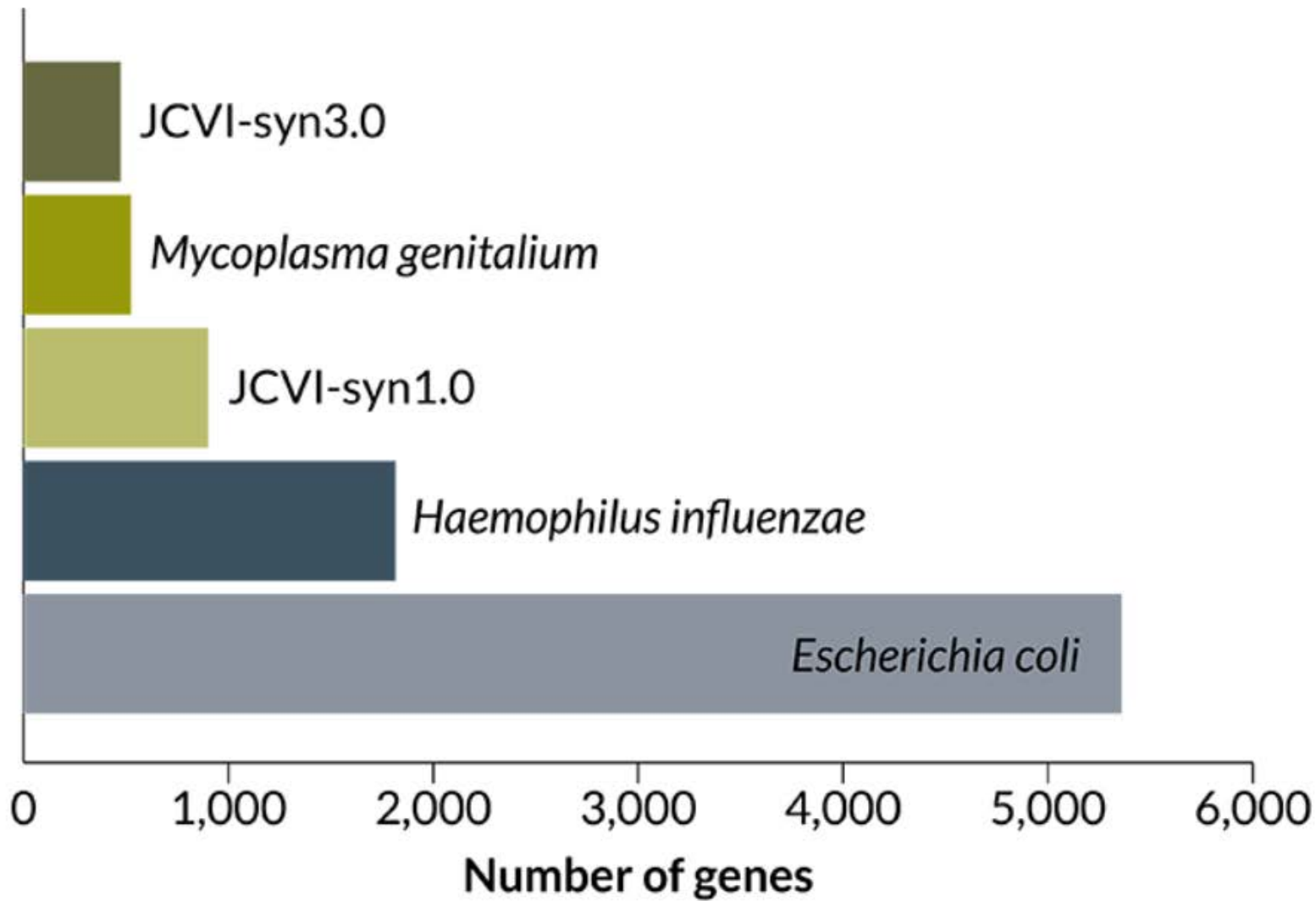
#### ON OUR WEB SITE

Read the full article at <http://dx.doi.org/10.1126/science.aad6253>

with limited transposon mutagenesis data, failed to produce a viable cell. Improved transposon mutagenesis methods revealed a class of quasi-essential genes that are needed for robust growth, explaining the failure of our initial design. Three more cycles of design, synthesis, and testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531 kbp, 473 genes). Its genome is smaller than that of any autonomously replicating cell found in nature. JCVI-syn3.0 has a doubling time of

Corrected 28 June 2016; see full text.

*Science* (2016) :351, aad6253



Il primo genoma sintetizzato artificialmente riguarda quello di *M. mycoides*, un *Mycoplasma* in grado di crescere più rapidamente.

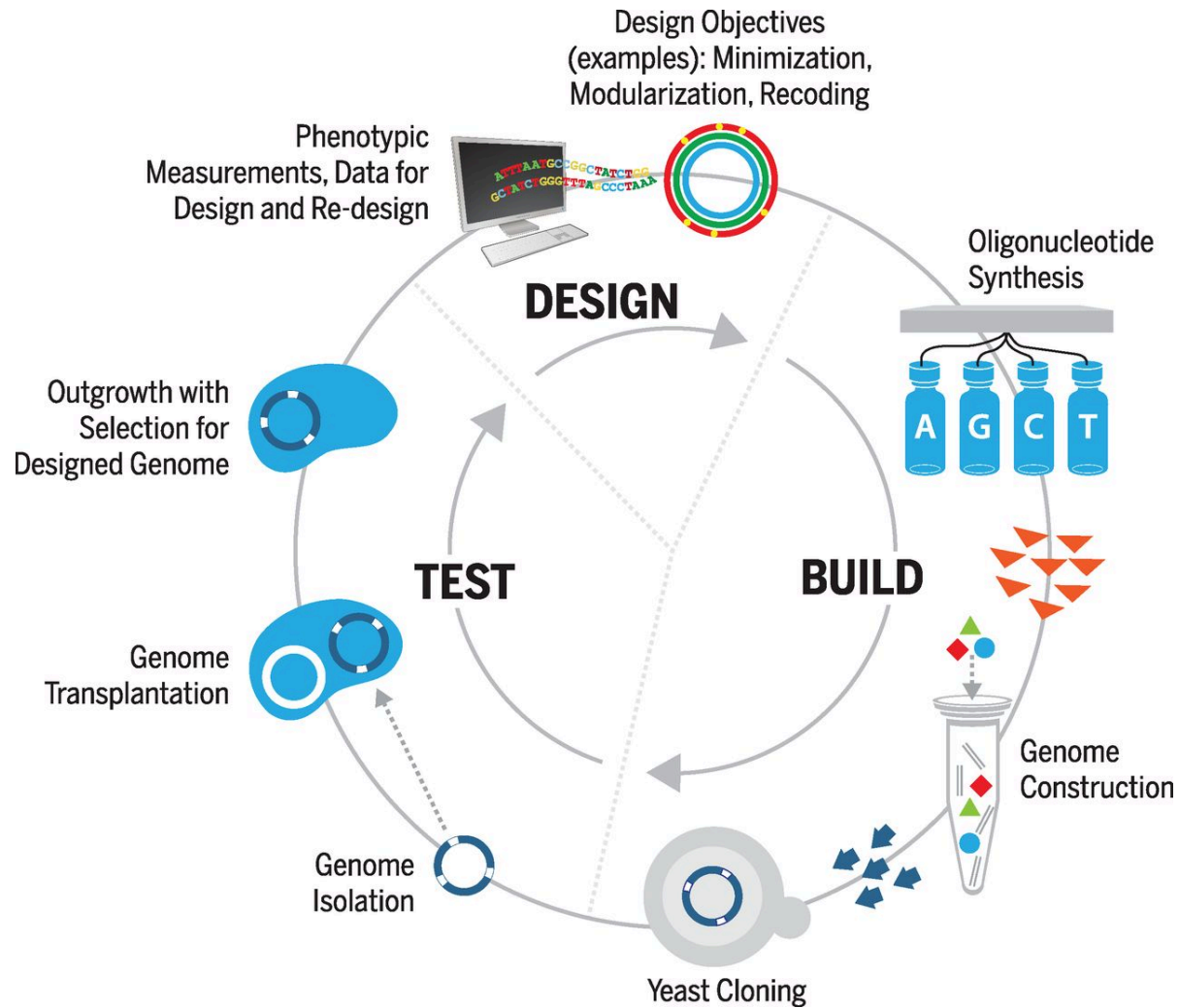
Questo genoma definito syn 1.0 è stato trapiantato all'interno di una specie diversa *M. capricolum* e corrisponde esattamente al genoma di *M. mycoides* con aggiunta di alcuni marcatori e di sequenze del vettore.

A partire da questo genoma è stato creato syn 3.0 costituito da 531 kb che codifica 438 proteine e 35 sRNA.

Il suo genoma è più piccolo di *M. genitalium* ma si duplica 5 volte più velocemente.

Il primo ipotetico genoma minimo è stato costruito sulla base di dati derivanti dalla mutagenesi e delezioni e da informazioni di letteratura sui pathway più importanti

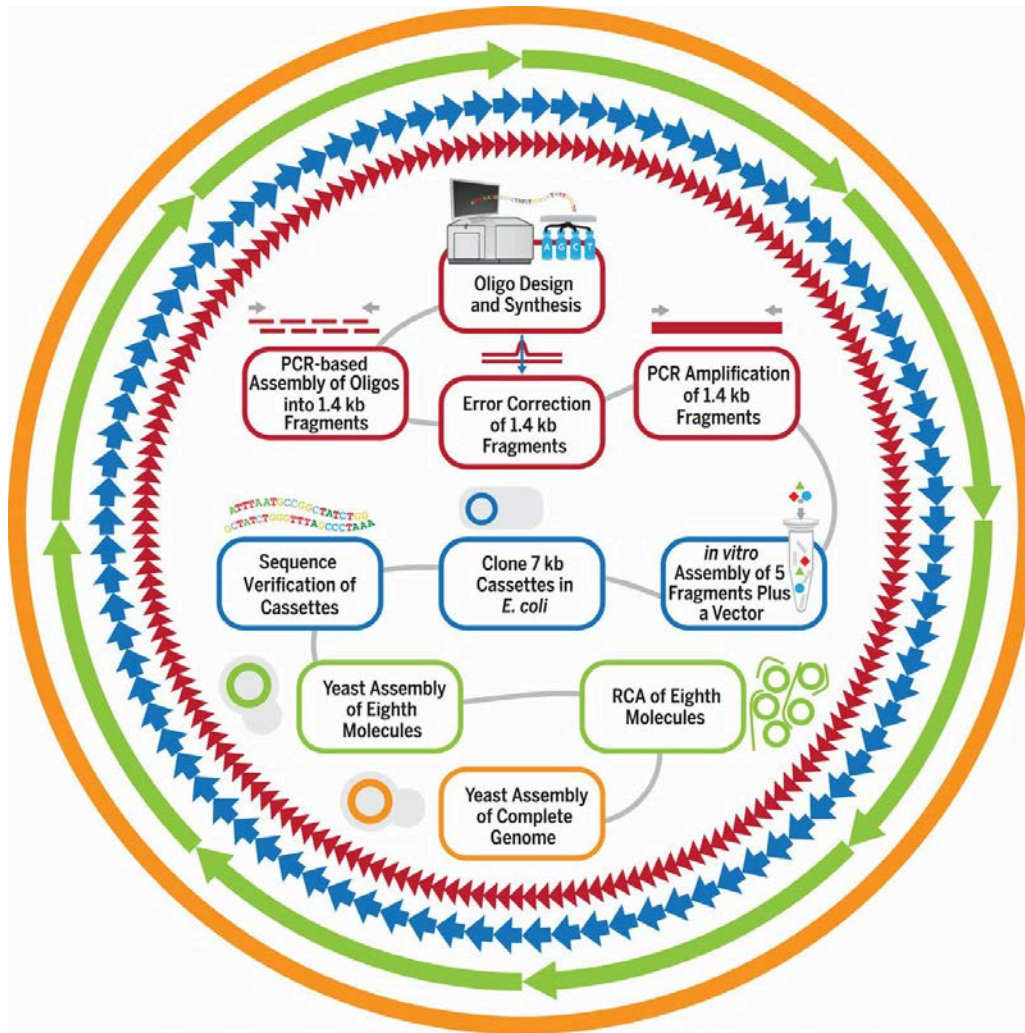
**Fig. 1 The JCVI DBT cycle for bacterial genomes.**



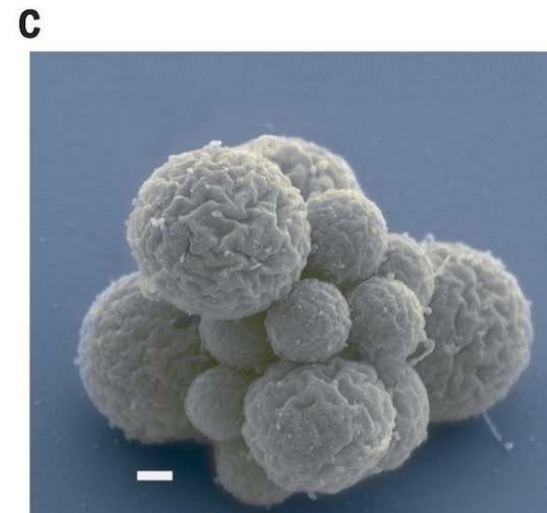
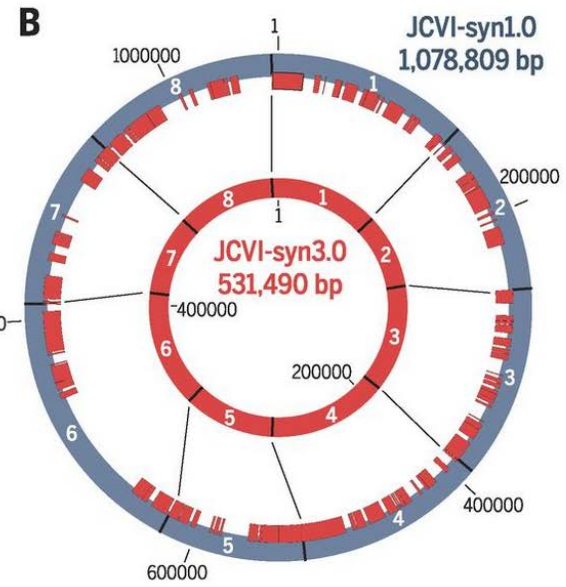
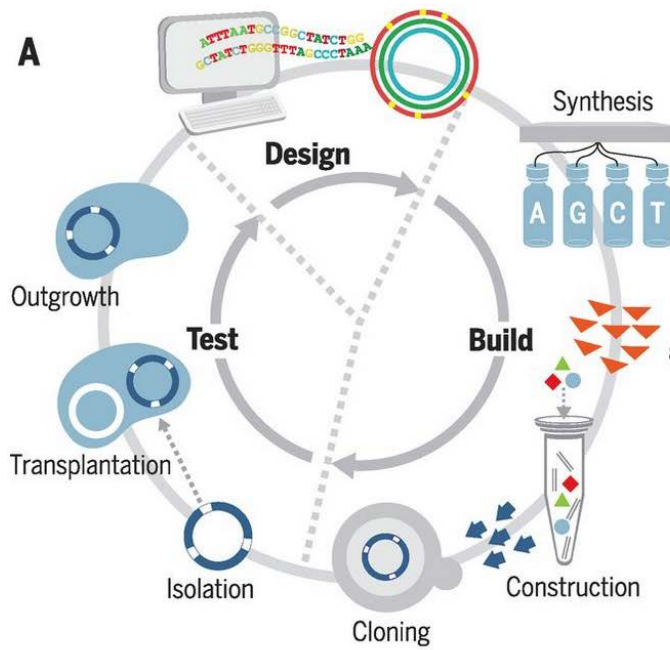
Clyde A. Hutchison III et al. *Science* 2016;351:aad6253



## Strategy for whole-genome synthesis.



Overlapping oligonucleotides (oligos) were designed, chemically synthesized, and assembled into 1.4-kbp fragments (red). After error correction and PCR amplification, five fragments were assembled into 7-kbp cassettes (blue). Cassettes were sequence-verified and then assembled in yeast to generate one-eighth molecules (green). The eight molecules were amplified by RCA and then assembled in yeast to generate the complete genome (orange).

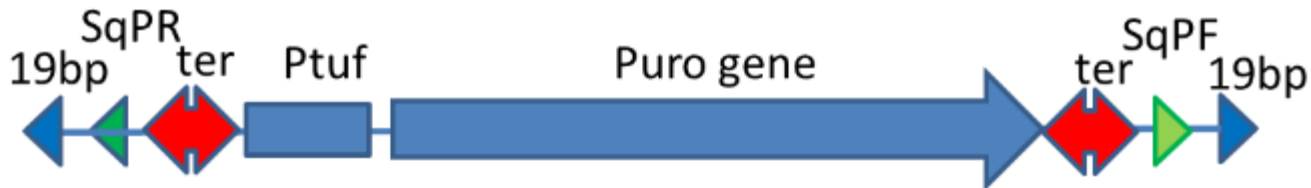


## Procedura

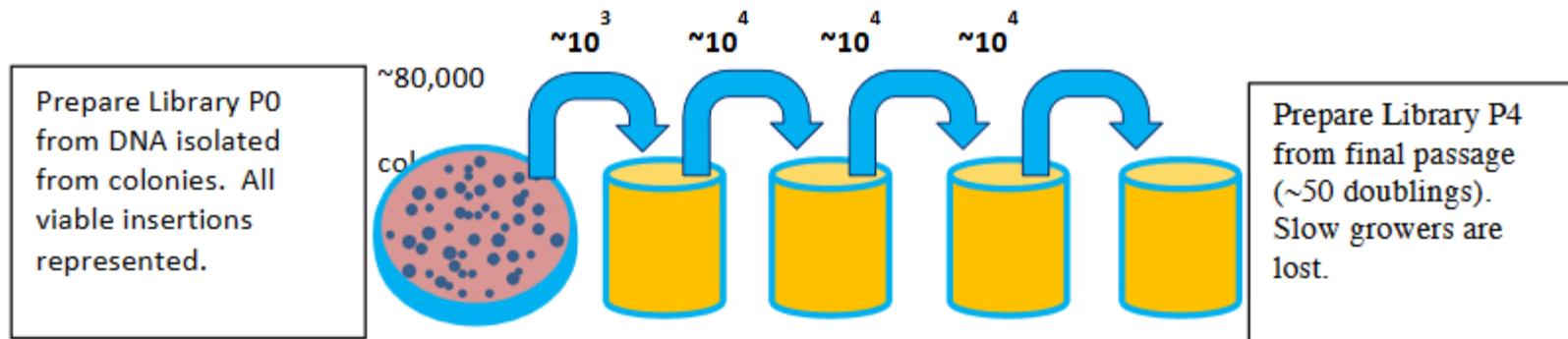
Partendo dal genoma syn 1 , effettuata mutagenesi con miniTn5 purR. Dal Pull di colonie contenenti Tn inserito in siti diversi (PO) si sono fatte una serie di passaggi seriali ( circa 40 passaggi )per controselezionare i ceppi che crescevano lentamente (P4) .

- I geni che non venivano mai colpiti da mutagenesi ESSENZIALI
- Quelli colpiti da mutagenesi e presenti sia nel PO che nel P4 definiti NON ESSENZIALI
- I geni colpiti nel PO ma non nel P4 definiti QUASI ESSENZIALI : se la delezione di questi geni generava crescita ridotta definiti i (I genes impairment growth)

**Step 1.** Construct Tn5 transposon containing 19 bp mosaic ends, sequencing primer sites, terminator sequences, and a selectable marker (puromycin-resistance gene). Bind Tn5 transposase (Epicentre) to the 19bp termini to form the active transposome.



**Step 2.** Introduce transposome into *Mycoplasma mycoides* JCVI-syn1.0 R-M (minus) strain by polyethylene glycol (PEG) transformation method. Collect puromycin-resistant colonies and serially propagate to eliminate slow growers.

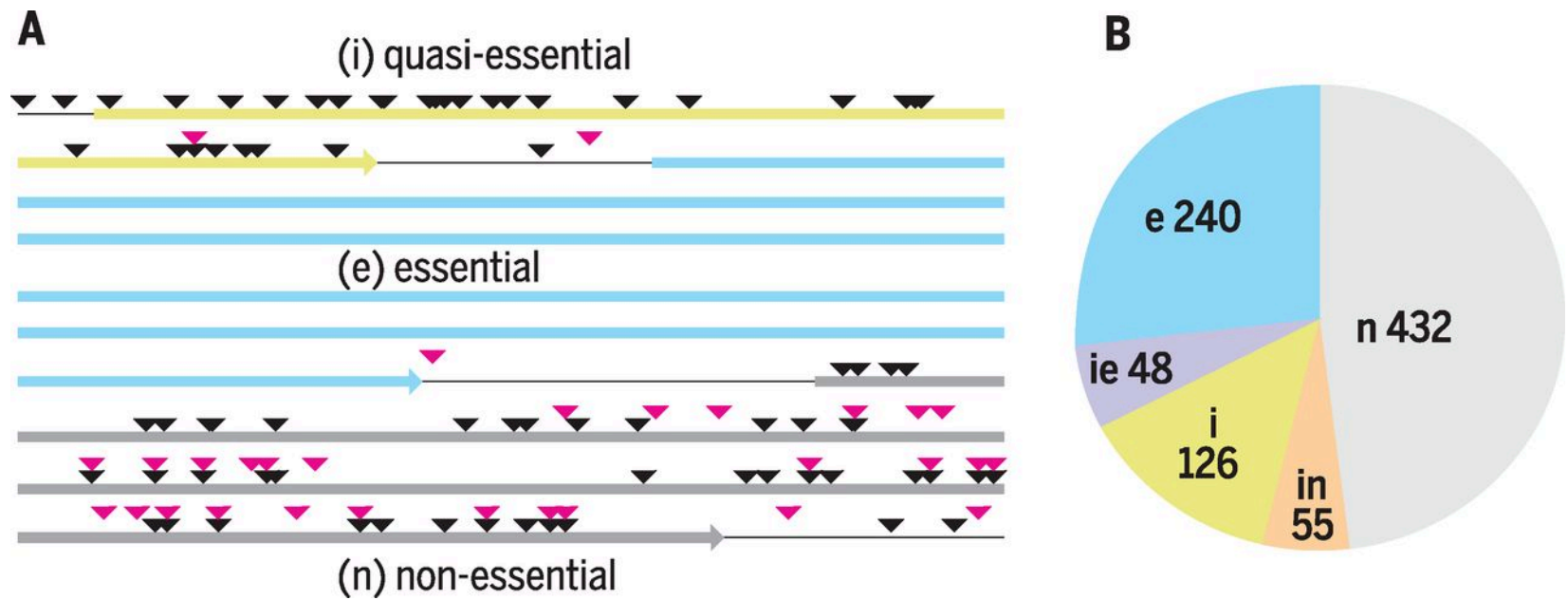


## Procedura

Partendo dal genoma syn 1 , effettuata mutagenesi con miniTn5 purR.  
 Dal Pull di colonie contenenti Tn inserito in siti diversi (P0) si sono fatte una serie di passaggi seriali ( circa 40 passaggi )per controselezionare i ceppi che crescevano lentamente (P4) .

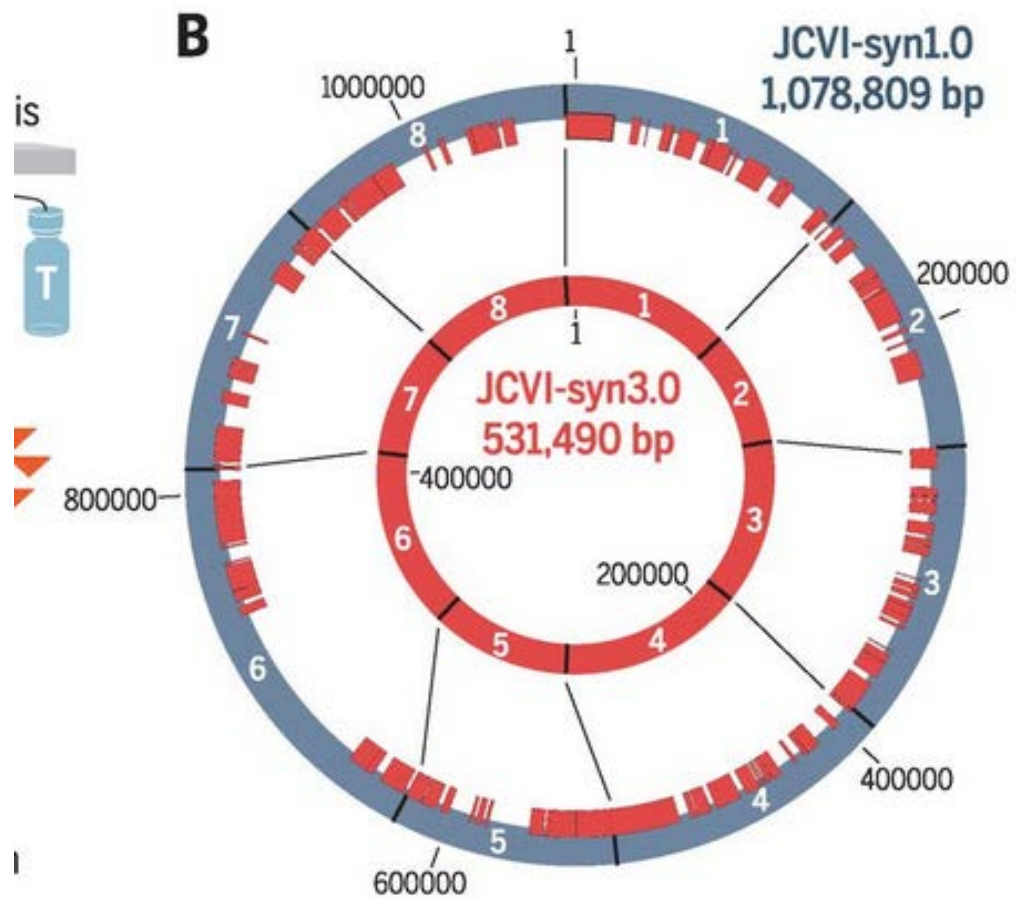


**Fig. 3 Classification of gene essentiality by transposon mutagenesis.**



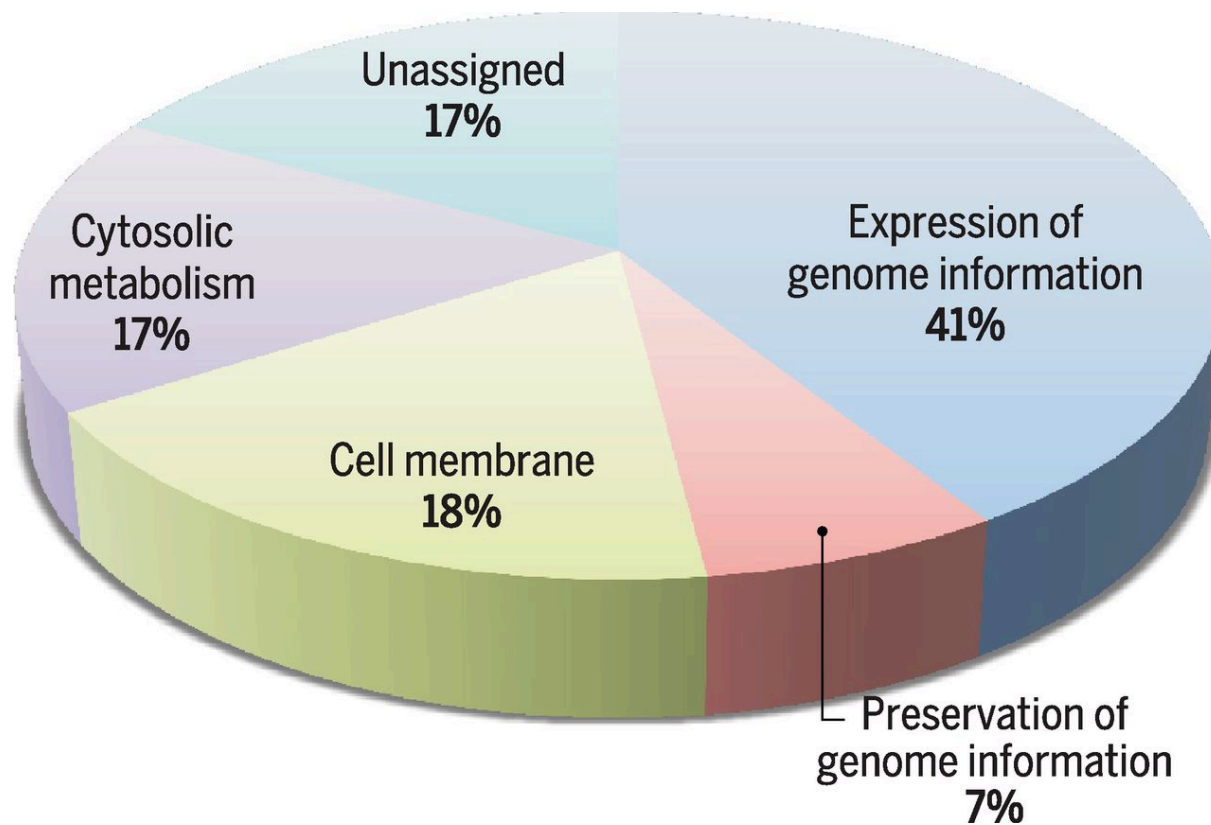
Clyde A. Hutchison III et al. Science 2016;351:aad6253



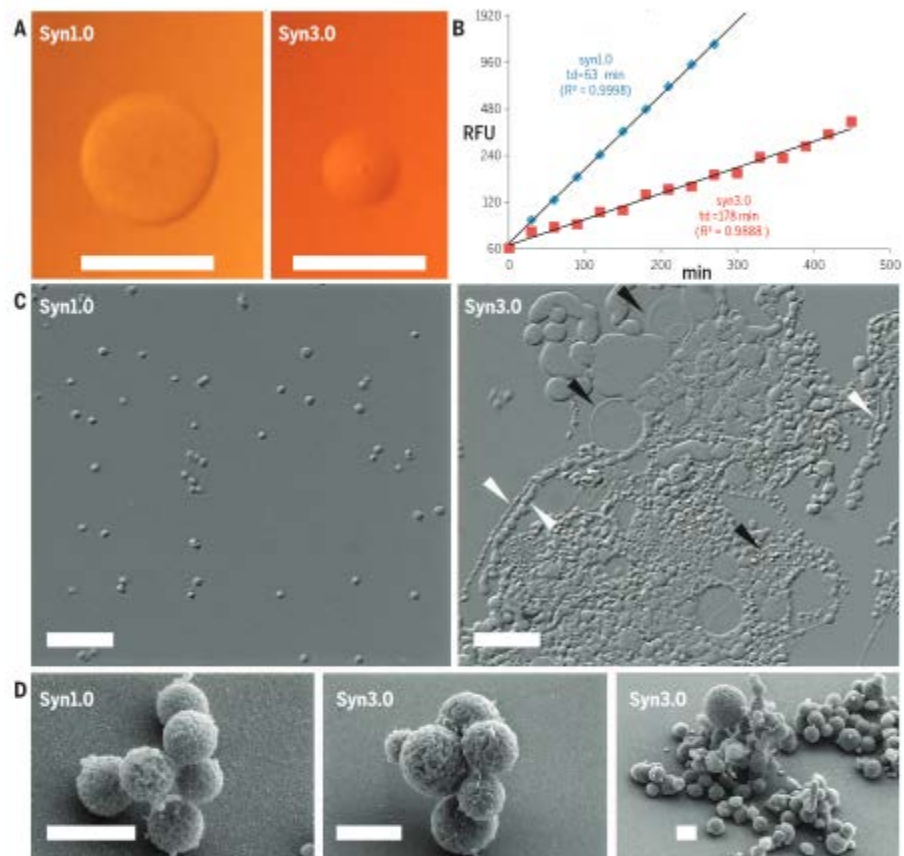




**Fig. 6 Partition of genes into four major functional groups.**



Partition of genes into four major functional groups. Syn3.0 has 473 genes. Of these, 79 have no assigned functional category (Table 1). The remainder can be assigned to four major functional groups: (i) expression of genome information (195 genes); (ii) preservation of genome information (34 genes); (iii) cell membrane structure and function (84 genes); and (iv) cytosolic metabolism (81 genes). The percentage of genes in each group is indicated.



**Fig. 7. Comparison of *syn1.0* and *syn3.0* growth features.** (A) Cells derived from 0.2  $\mu\text{m}$ -filtered liquid cultures were diluted and plated on agar medium to compare colony size and morphology after 96 hours (scale bars, 1.0 mm). (B) Growth rates in liquid static culture were determined using a fluorescent measure (relative fluorescent units, RFU) of double-stranded DNA accumulation over time (minutes) to calculate doubling times (td). Coefficients

of determination ( $R^2$ ) are shown. (C) Native cell morphology in liquid culture was imaged in wet mount preparations by means of differential interference contrast microscopy (scale bars, 10  $\mu\text{m}$ ). Arrowheads indicate assorted forms of segmented filaments (white) or large vesicles (black). (D) Scanning electron microscopy of *syn1.0* and *syn3.0* (scale bars, 1  $\mu\text{m}$ ). The picture on the right shows a variety of the structures observed in *syn3.0* cultures.

## Genoma di E.coli : esempio di grande variabilità

La sequenza dell'intero genoma di E.coli ha rivelato una variabilità intraspecie estremamente elevata

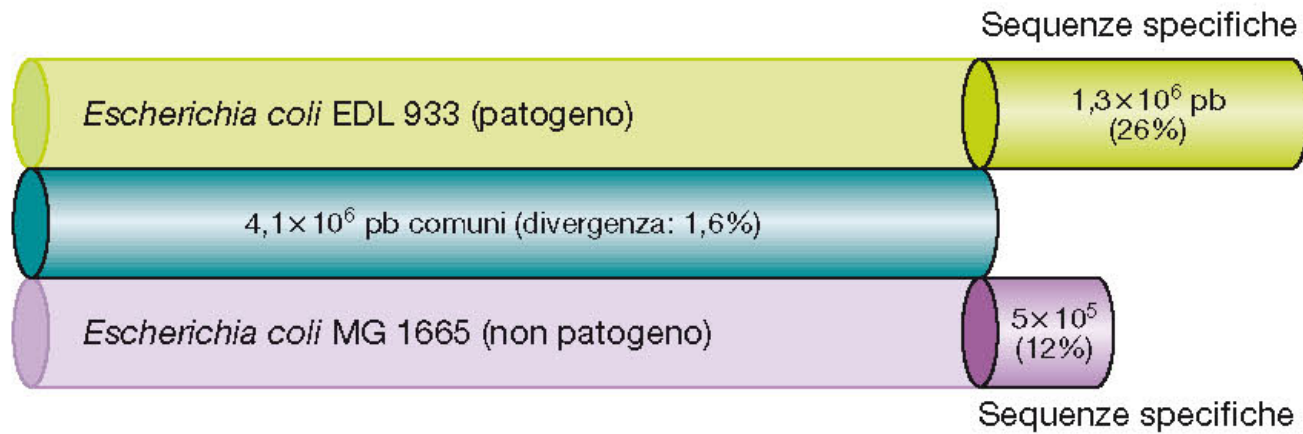
Sono disponibili 4 sequenze genomiche di E.coli

Analisi genomica comparativa ha rivelato che *E.coli* 0157 ha un genoma di 1 Mb più grande di quella di *E.coli* K12 e circa 25% dei geni non sono conservati nel genoma di *E.coli* K12.

Molti dei geni presenti in 0157 si pensa siano stati acquisiti tramite eventi di trasferimento orizzontale e tramite elementi genetici mobili quali fagi, profagi e sequenze IS

Soltanto 3.000 geni sono in comune tra i 4 genomi di E.coli mentre erano 4.000 tra E.coli K12 e 0157

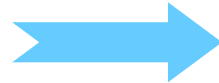
I 3000 geni comuni presentano SINTENIA suggerendo una base di trasmissione verticale



La diversità all'interno di una specie è caratteristica solo di E.coli?

Streptococcus pneumoniae

Burkholderia cepacia



Grande Variabilità

Mycobacterium tuberculosis soltanto 1-2% di variabilità

solamente microrganismi patogeni con habitat ristretto