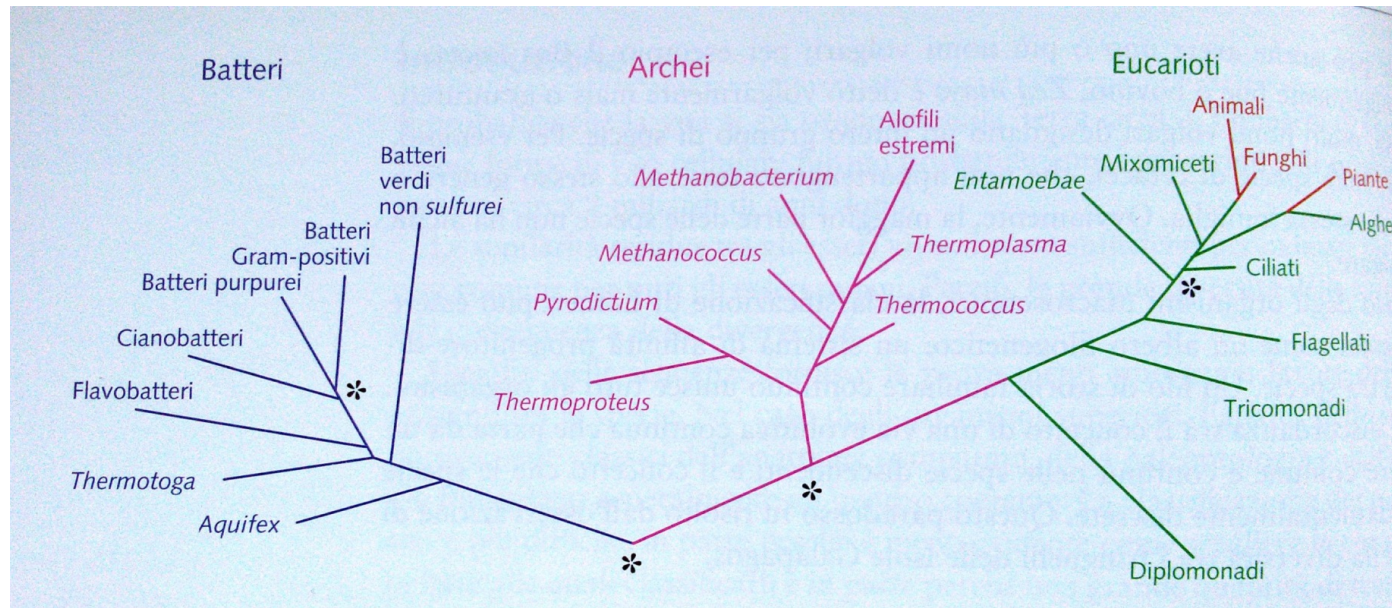


Lezione 2: Genomi: struttura, contenuto, organizzazione

Dimensioni e organizzazione dei genomi

Origine della vita sulla terra: 3,5 miliardi di anni fa...eucarioti 2 miliardi di anni dopo

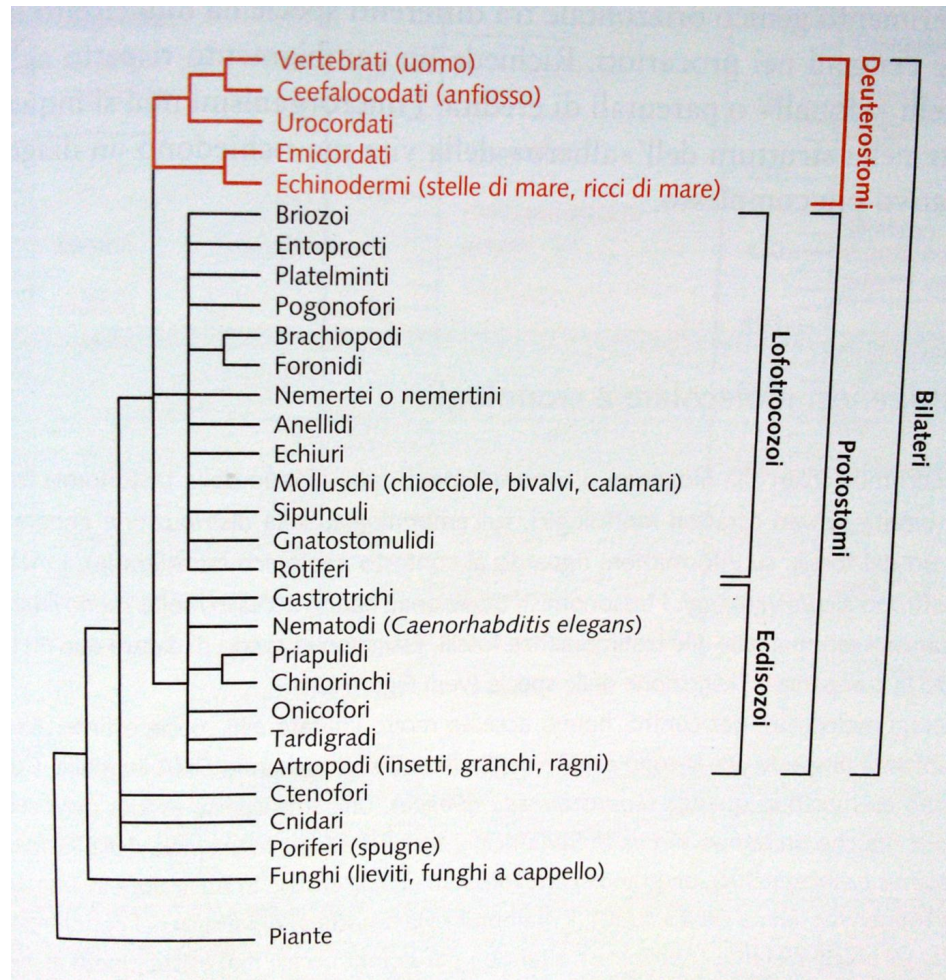
La genomica comparata ci permette di confrontare genomi di specie molto lontane filogeneticamente e di capire **che cosa è comune** e che cosa è **esclusivo** di alcuni rami



Principali divisioni dell'albero della vita. Albero ottenuto con sequenze di rRNA. Rosso + verde: eucarioti, * punti di divergenza cruciali

Dimensioni e organizzazione dei genomi

Protostomi e **deuterostomi**: divergenza circa 670 milioni di anni



Albero filogenetico dei metazoi

Procarioti ed Eucarioti

Esseri viventi: più generale classificazione delle cellule:

Procarioti

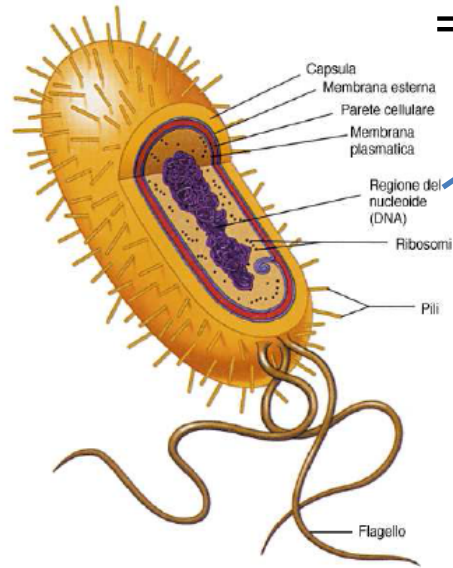
Eucarioti

Differenze tra cellule procariotiche e cellule eucariotiche		
Caratteristica	Caratteristica tipica di una:	
	Cellula procariotica	Cellula eucariotica
Dimensione	10 μm	\approx 0,1 mm
Subcellulare	Assenza di nucleo	Nucleo
Stato del principale componente del materiale genetico	Molecola circolare a cui sono attaccate permanentemente poche proteine	Complessato con istoni per formare cromosomi
Differenze interne	Assenza di struttura subcellulare organizzata	Nuclei, mitocondri, cloroplasti, citoscheletro, reticolo endoplasmatico, apparato di Golgi
Divisione cellulare	Scissione	Mitosi (o meiosi)

Forma e organizzazione del materiale genetico

cellula procariote.

DNA E. coli: 1,6 mm > cellula 2 μm x 0,8 μm (1 μm = 0,001 mm!)

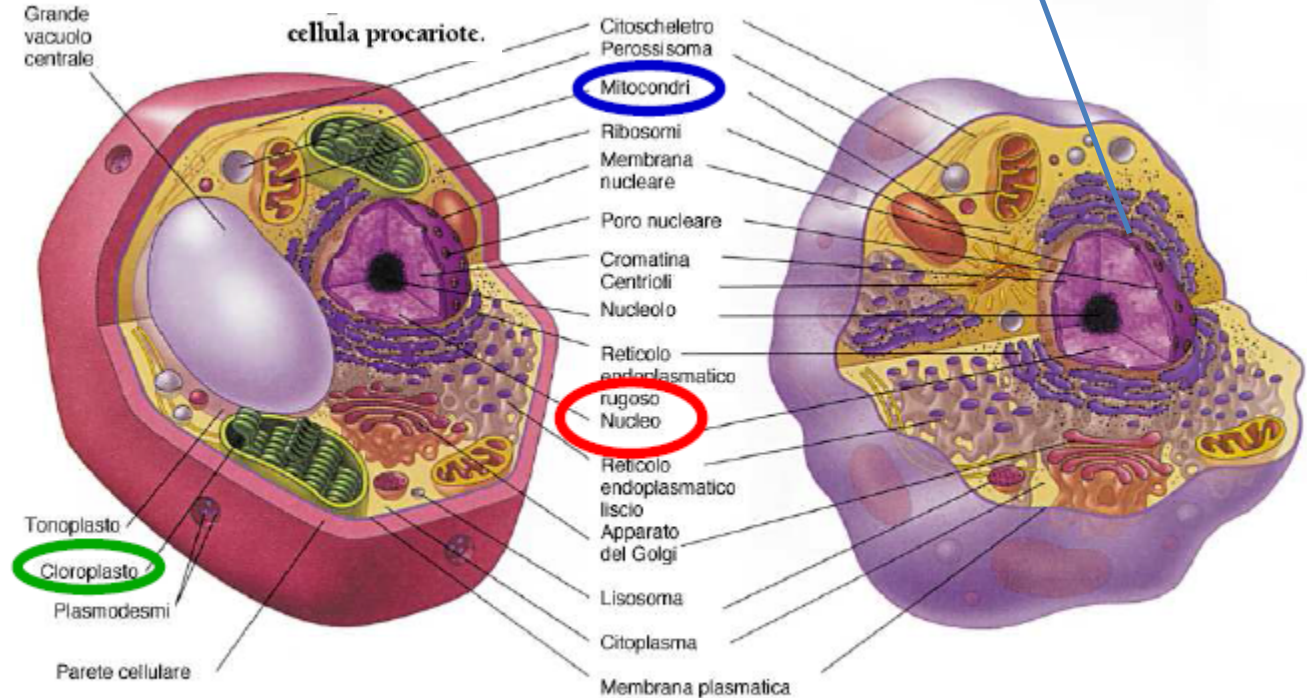


DNA umano: 2 m > diametro del nucleo 6 μm

a) Cellula vegetale

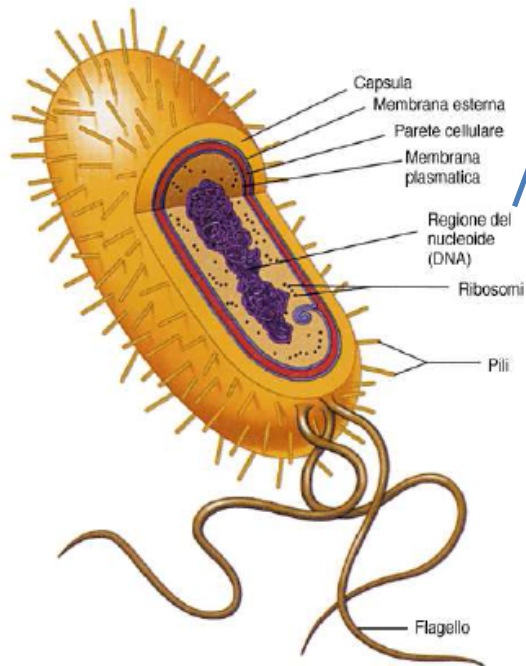
b) Cellula animale

cellula procariote.



Forma e organizzazione del materiale genetico

cellula procariote.



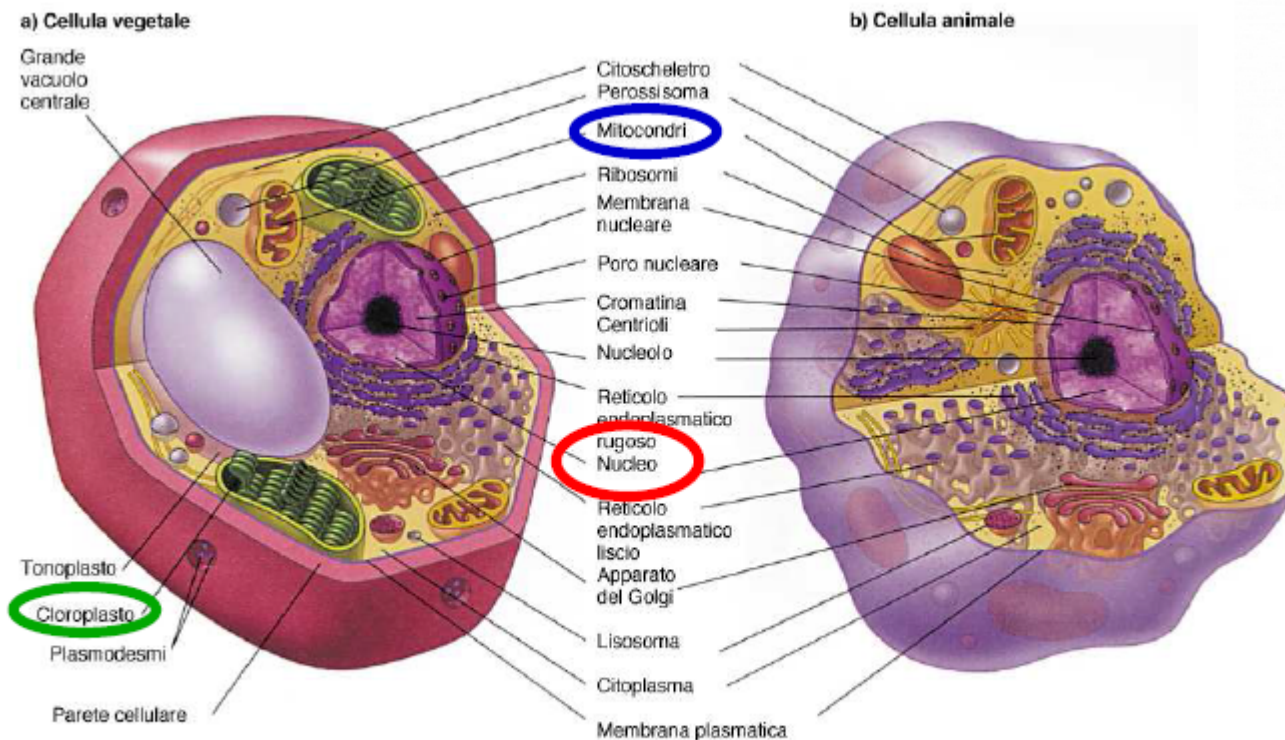
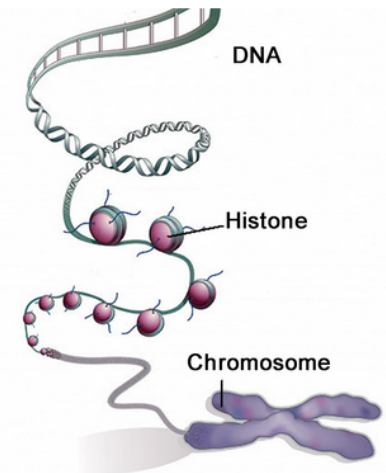
Nucleoide: DNA circolare + proteine attaccate alla membrana (alcuni batteri hanno più di una molecola circolare, altri lineare)

Plasmidi (piccoli DNA circolari non complessati con proteine e slegati dalla membrana)

Forma e organizzazione del materiale genetico

DNA quasi tutto nel nucleo (sito della duplicazione del DNA e della sintesi dell'RNA per la trascrizione)

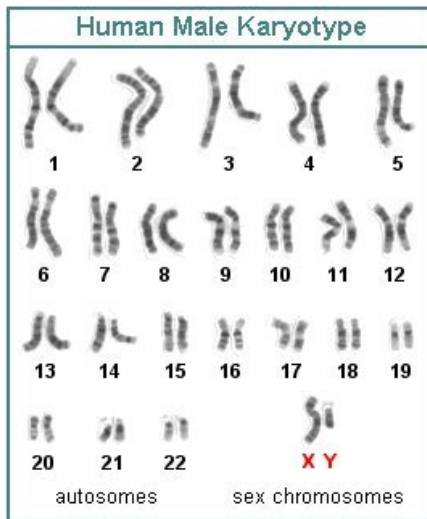
DNA + istoni > cromosomi



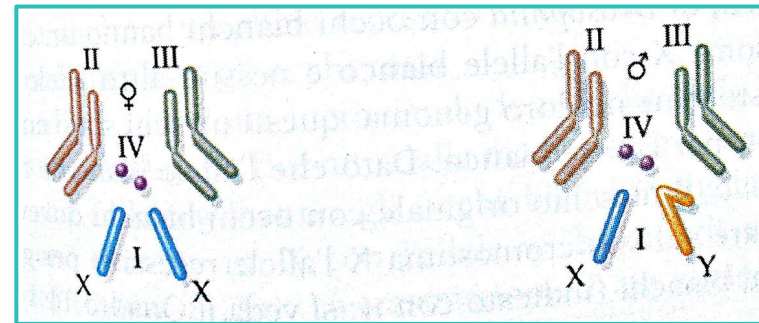
Forma e organizzazione del materiale genetico

Diverse specie > diversi numeri di cromosomi > diverse lunghezze

Uomo: 46 cromosomi di cui
22 coppie di autosomi
+ XX F o XY M



Drosophila melanogaster:



Il sesso nella *Drosophila* è determinato cellula per cellula dal rapporto tra i cromosomi X e gli autosomi. Nella femmina il rapporto è uguale a 1, mentre nel maschio è uguale a 0,5.

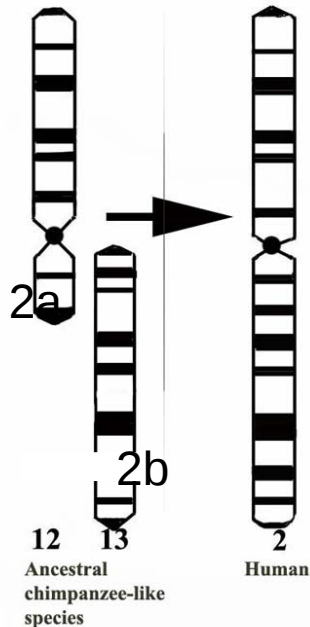
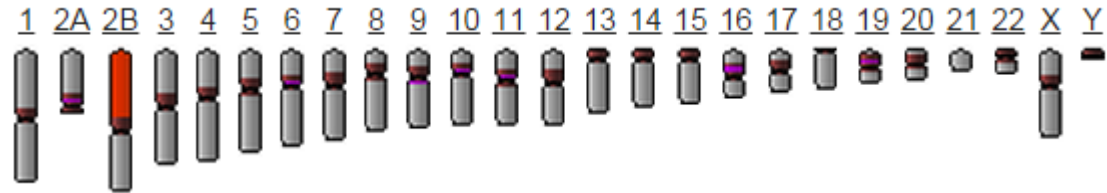
Forma e organizzazione del materiale genetico

Diverse specie > diverso numero di cromosomi > diverse lunghezze

Tabella 11.1 Numeri cromosomici in diversi organismi ^a

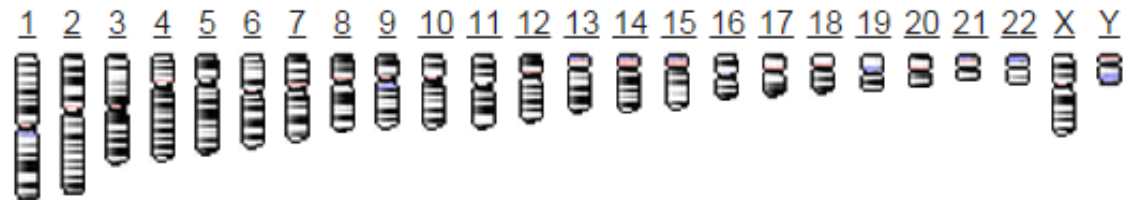
Organismo	Numero totale di cromosomi
Uomo	46
Scimpanzé	48
Cane	78
Gatto	72
Topo	40
Cavallo	64
Pollo	78
Rospo	36
Pesce rosso	94
Stella di mare	36
Moscerino della frutta (<i>Drosophila melanogaster</i>)	8
Zanzara	6
Formica australiana (<i>Myrecia pilosula</i>)	♂1, ♀2
Nematode	♂11, ♀12
<i>Neurospora</i> (aploide)	7
Muschio (aploide)	23
Equiseto	216
Sequoia gigante	22
Tabacco	48
Cotone	52
Patata	48
Pomodoro	24
Grano tenero	42
Lievito (<i>Saccharomyces cerevisiae</i>) (aploide)	16

^a Tutti i numeri cromosomici sono riferiti a cellule diploidi, tranne quando indicato.



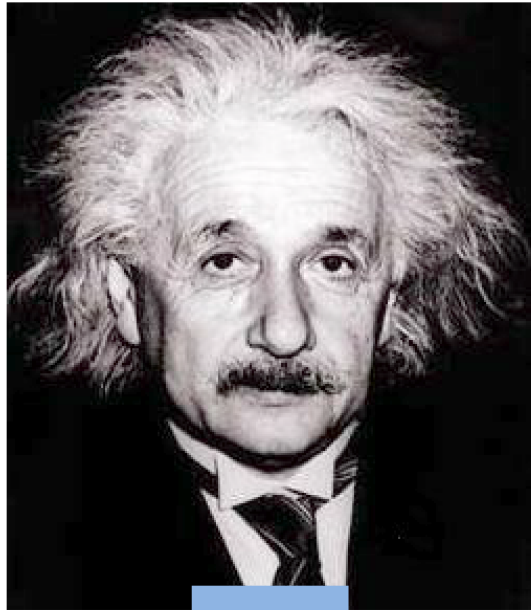
Homo sapiens (human)

Human genome projects have generated an unprecedented amount of knowledge about human genetics and health.



Paradosso del valore **K**: la complessità non correla con il **numero di cromosomi**.

Homo sapiens



46

Lysandra atlantica



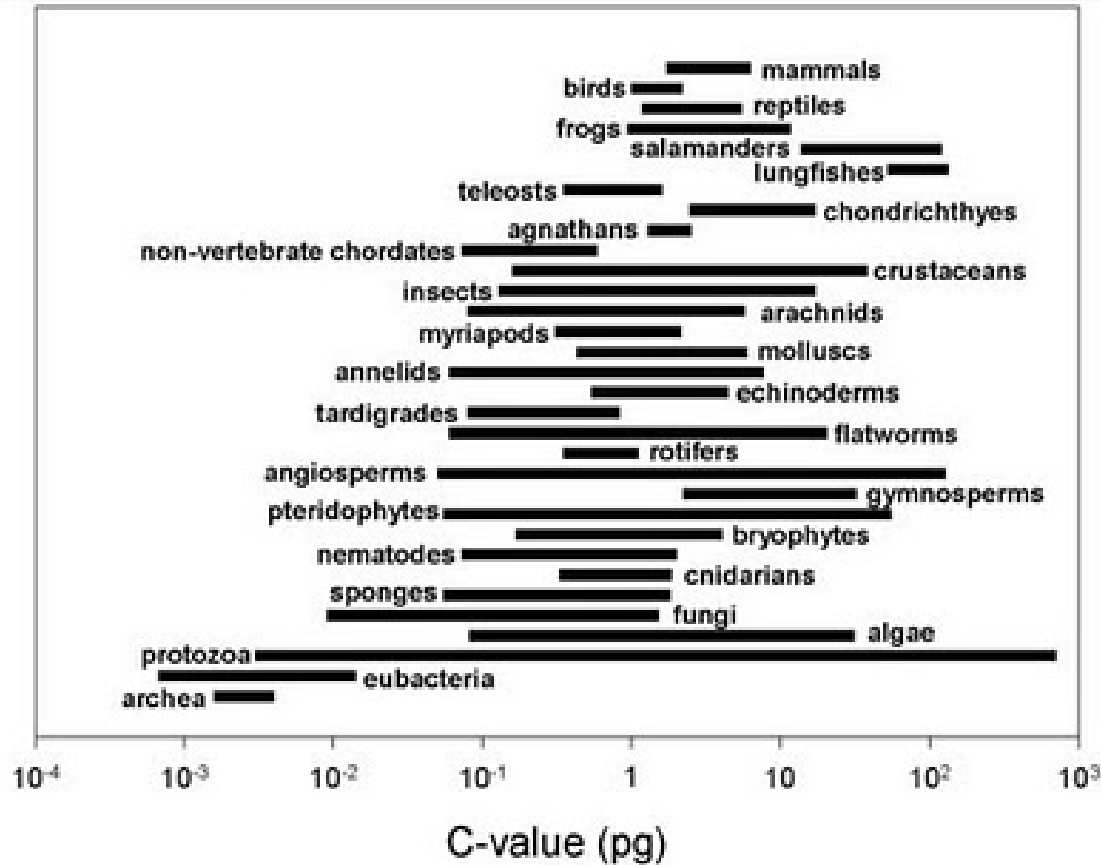
250

Ophioglossum reticulatum



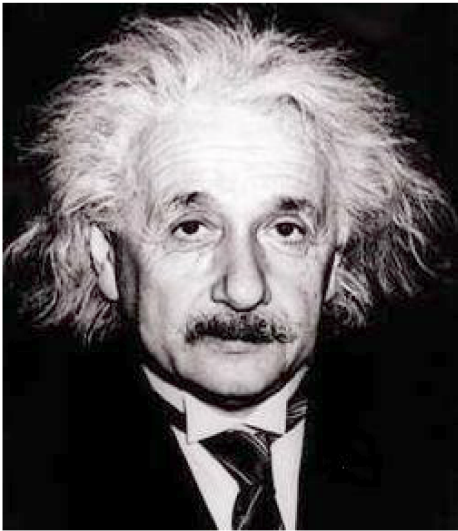
~1260

Il valore C (quantità di DNA contenuto nel nucleo di una cellula aploide di un organismo)



Cellula aploide? Per esempio un gamete

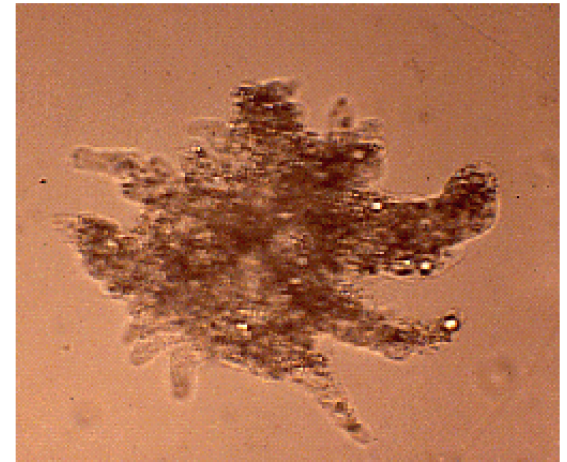
Paradosso del valore **C**: la complessità non correla con la **grandezza del genoma**.



3.4×10^9 bp
Homo sapiens

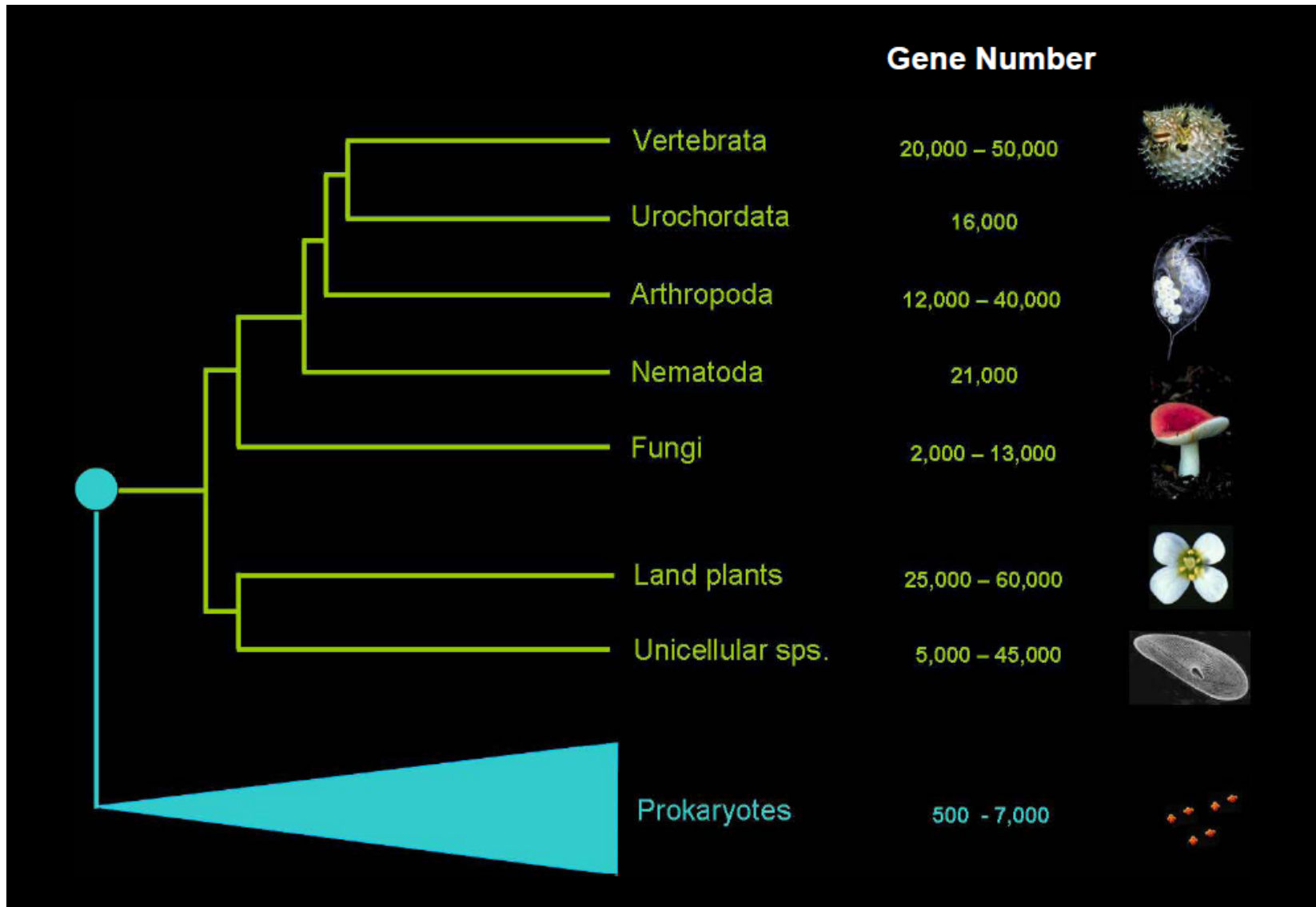


1.5×10^{10} bp
Allium cepa

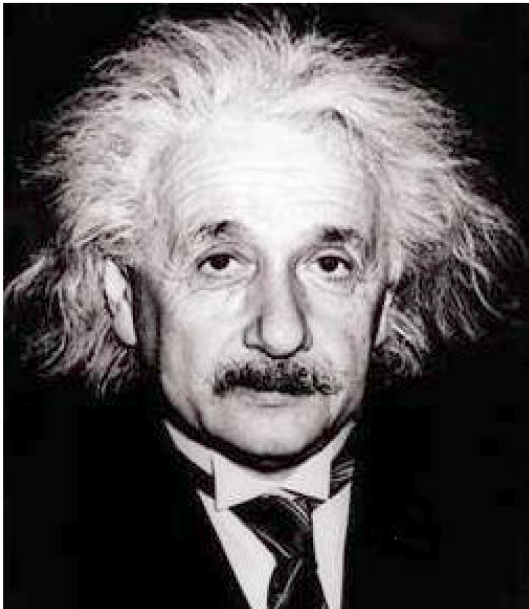


6.8×10^{11} bp
Amoeba dubia

Paradosso del valore **N**: Il **numero di geni** e la complessità degli organismi non sono correlati



Paradosso del valore **N**: Il **numero di geni** e la complessità degli organismi non sono correlati



~21000 geni

~25000 geni

~60000 geni

Organelli

Mitocondri (sia vegetali che animali) e cloroplasti (solo vegetali)

DNA circolare non associato ad istoni come nei procarioti

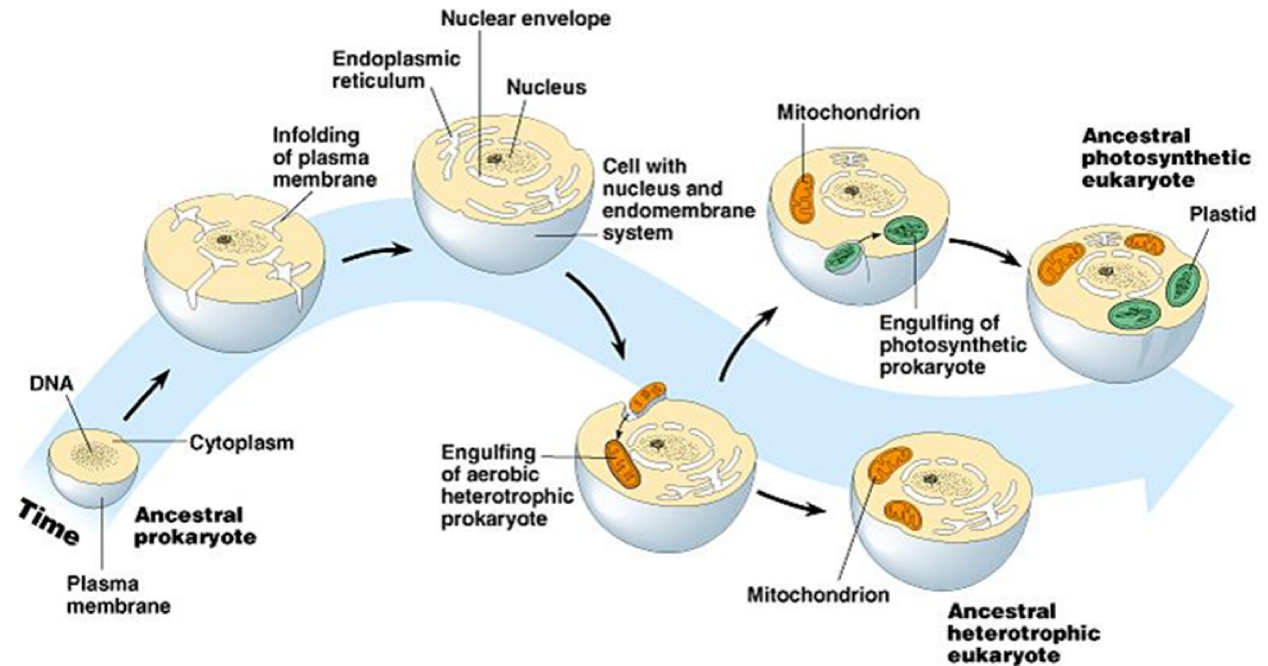
Sintesi proteica: codice genetico specifico (vedi lezione precedente)

Trasduzione dell'energia:

mt > fosforilazione ossidativa (ATP)

chl > fotosintesi

Organelli



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

trasferimento di geni al genoma nucleare

Esempi:

Lievito: 850 prodotti di geni nucleari sono importati come proteine nel mt

Invertebrati: circa 2000

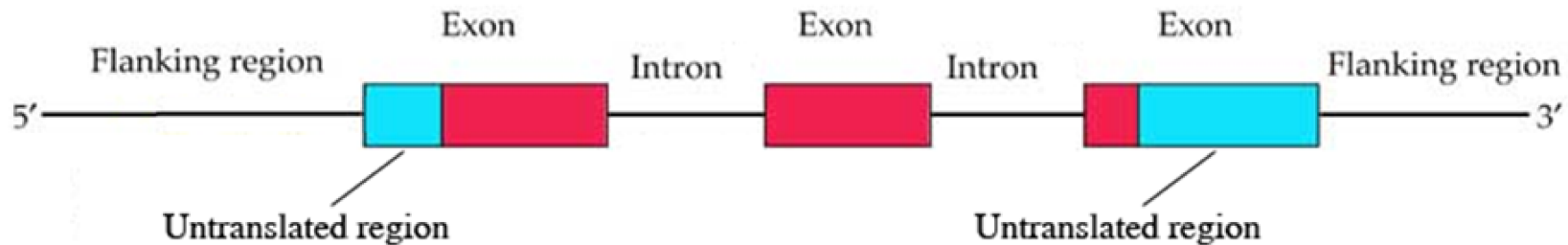
Vertebrati: circa 4000

Non tutti, ma molti di questi geni hanno avuto origine da trasferimenti di geni mt → nucleo

Geni: il concetto tradizionale

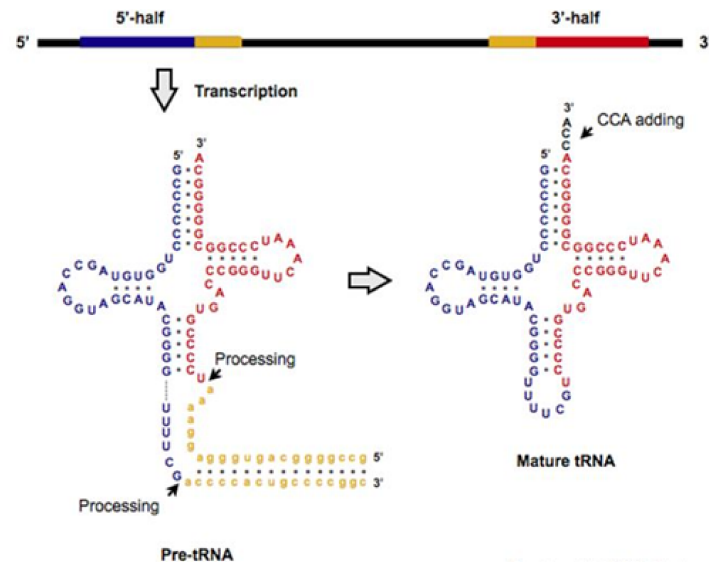
Regioni che vengono espresse come proteine (trascritte > tradotte): contengono un *open reading frame* (ORF), o fase di lettura a perta.

Un ORF comincia con un codone di inizio ATG e termina con un codone di STOP. Un'ORF è una regione che potenzialmente codifica per proteine.



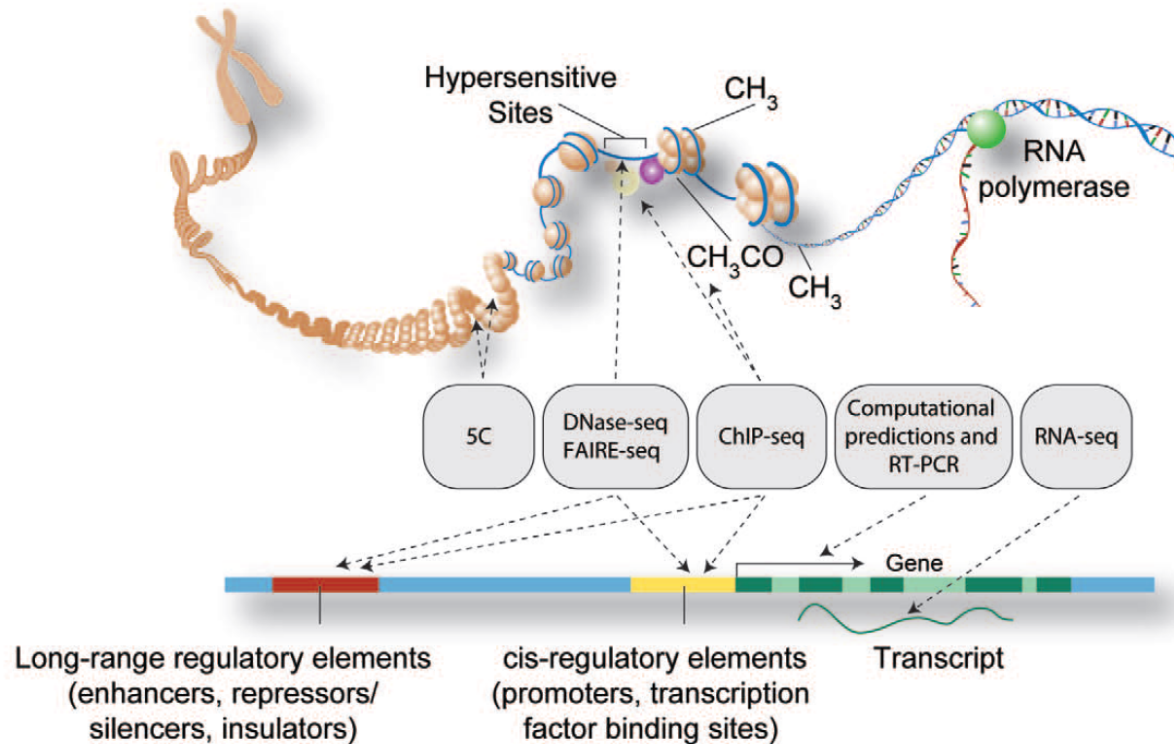
Geni: il concetto esteso a «regione funzionale»

Regioni che vengono espresse come RNA strutturale: presentano solitamente regioni di auto-complementarietà locale corrispondenti ad anse a forcina. Ne sono un esempio gli RNA transfer con la loro struttura a trifoglio.



Geni: il concetto esteso a «regione funzionale»

Regioni bersaglio di interazioni regolatorie.



Overview of the Encyclopedia of DNA Elements (ENCODE) project. Genomic elements that are the targets of the ENCODE project (dashed arrows) and some of the methods (gray boxes) used to quantify them in more than 150 human cell lines.

<https://www.encodeproject.org/>



Geni codificanti...quanti?

Ritornando ai paradossi: quanta parte del DNA di un genoma è codificante?

Specie	Dimensione del genoma (Mb)	Codificante (%)	Numero approssimato di geni	Densità genica stimata (kb/gene)
<i>Escherichia coli</i>	4,64	88	4 300	0,95
<i>Saccharomyces cerevisiae</i> (lievito di birra)	12,5	70	6 000	2,1
<i>Tetraodon nigroviridis</i> , un tetraodontide (o pesce palla)	365	15	30 000	10
<i>Arabidopsis thaliana</i> (arabetta comune)	115	29	25 000	4,5
<i>Homo sapiens</i> (uomo)	3289	1,3	30 000	27

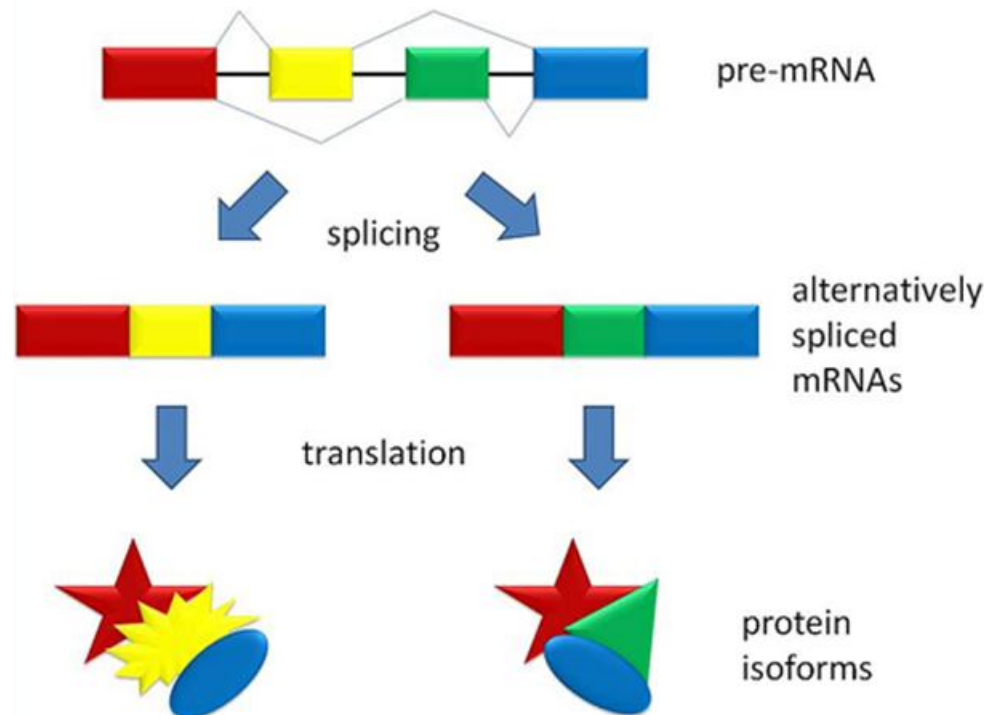
20-30.000?

Megabase: 1.000.000 bases

Geni codificanti...quanti?

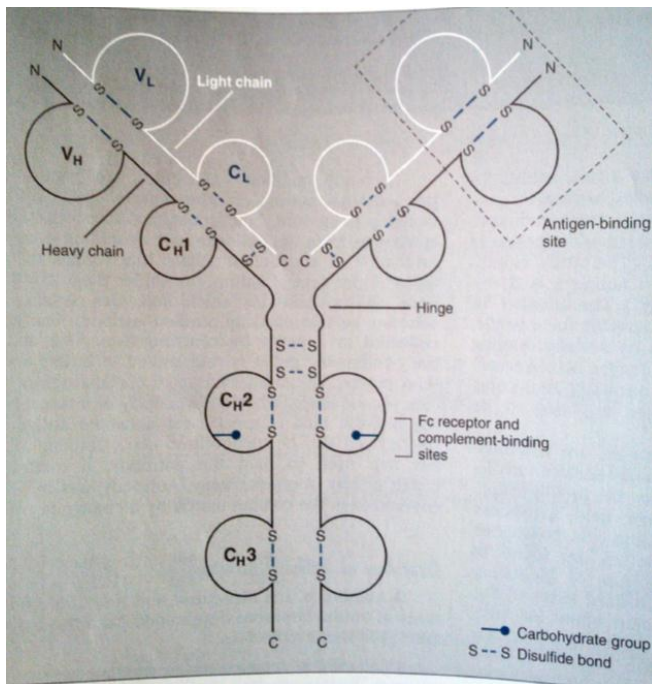
La stima del numero dei geni è resa ancora più complessa dall'esistenza dello splicing alternativo

Splicing alternativo: sintesi di differenti proteine a partire dallo stesso trascritto di RNA attraverso l'unione di differenti serie di esoni



La dimensione/complessità del genoma è molto diversa dalla complessità del **proteoma**

Esempio estremo: **sistema immunitario dei mammiferi**.
Miliardi di anticorpi distinti si originano da meno di 100 esoni.

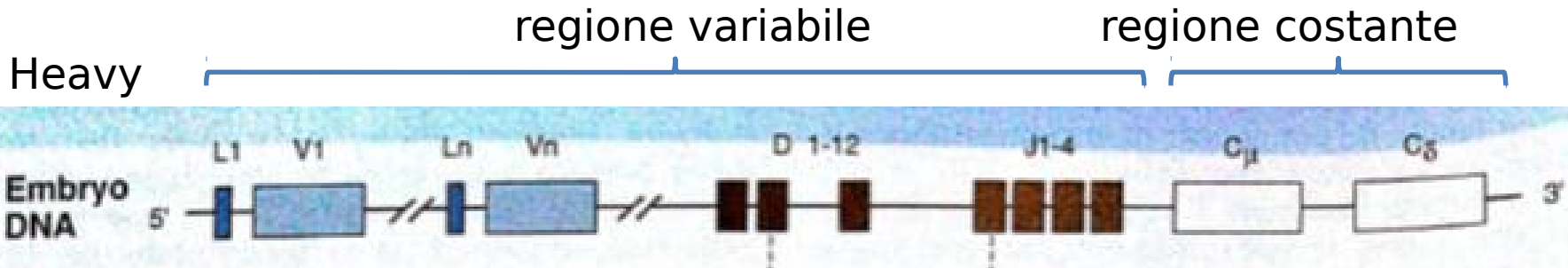


In nero: catene pesanti (heavy).
Definiscono l'isotipo (IgA, IgD, IgE, IgG, IgM nell'uomo e nel topo). Sono composte da una parte costante (C_H) e da una variabile (V_H). Diversi isotipi Heavy sono coinvolti in diverse funzioni.

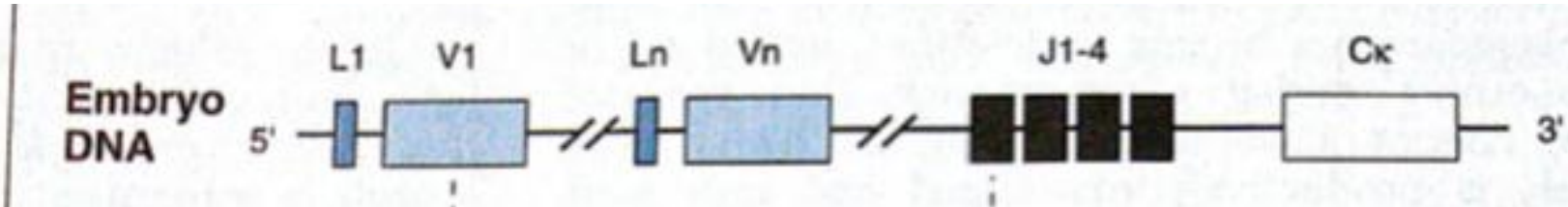
In bianco: catene leggere (light).
Due isotipi: k e λ. Sono composte da una parte costante (C_L) e da una variabile (V_L)

Isotipi degli anticorpi umani

Antibody	Subtypes	H Chain (Designation)	
IgA	IgA1	$\alpha 1$	• Immunità a livello delle mucose
	IgA2	$\alpha 2$	
IgD	None	δ	• ?
IgE	None	ϵ	• coinvolte in numerose allergie
IgG	IgG1	$\gamma 1$	• anticorpi maggiormente impiegati durante la risposta immunitaria secondaria (prodotte tardivamente e in maniera massiva dai linfociti B differenziatisi in plasmacellule).
	IgG2	$\gamma 2$	
	IgG3	$\gamma 3$	
	IgG4	$\gamma 4$	
IgM	None	μ	• prodotte nelle fasi più precoci dell'infiammazione, efficacia minore



Light



Esone L: 60-90 bp, codifica per il segnale di inizio della traduzione e per alcuni residui idrofobici segnale (o leader) > guidano al reticolo endoplasmatico e poi vengono rimossi

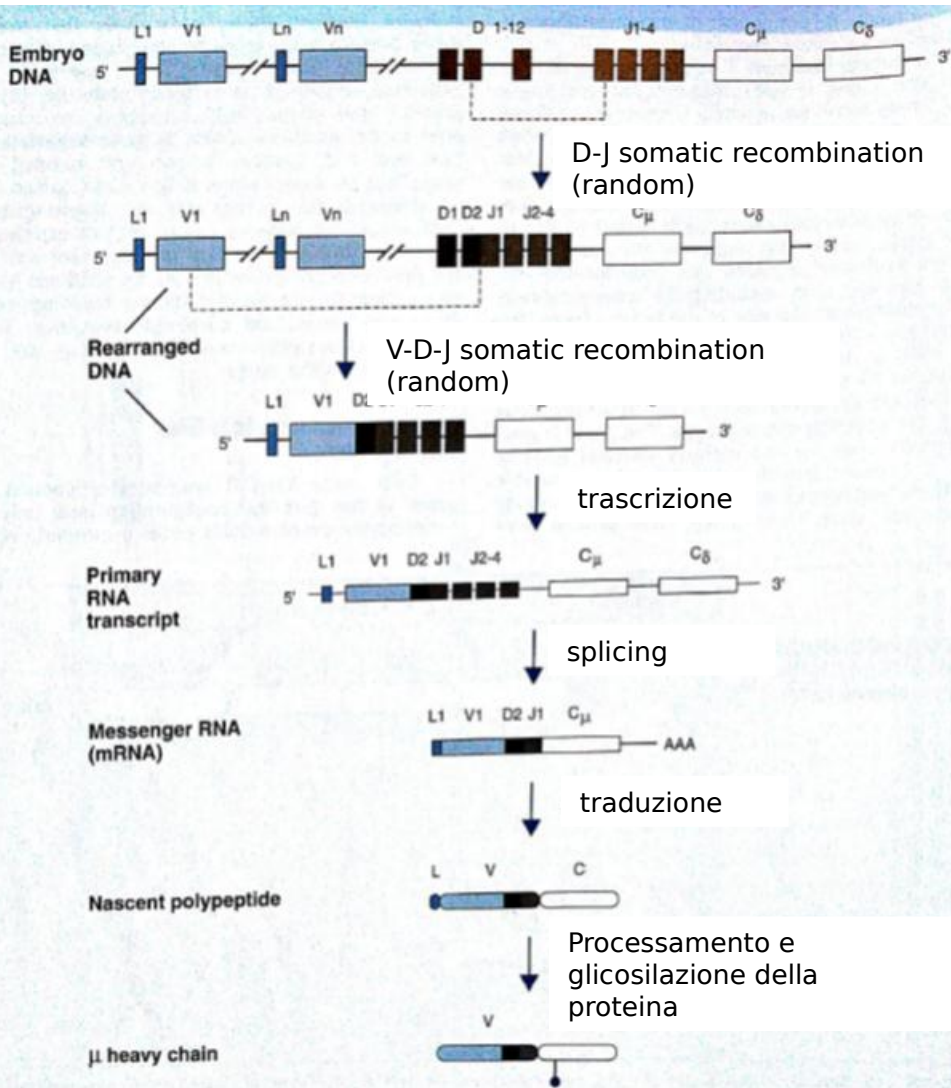
Esone V: possono arrivare a 1000 diversi in alcune specie, nell'uomo 100-200

Esone J: 30-50 bp joining segments (sia nelle catene pesanti che leggere)

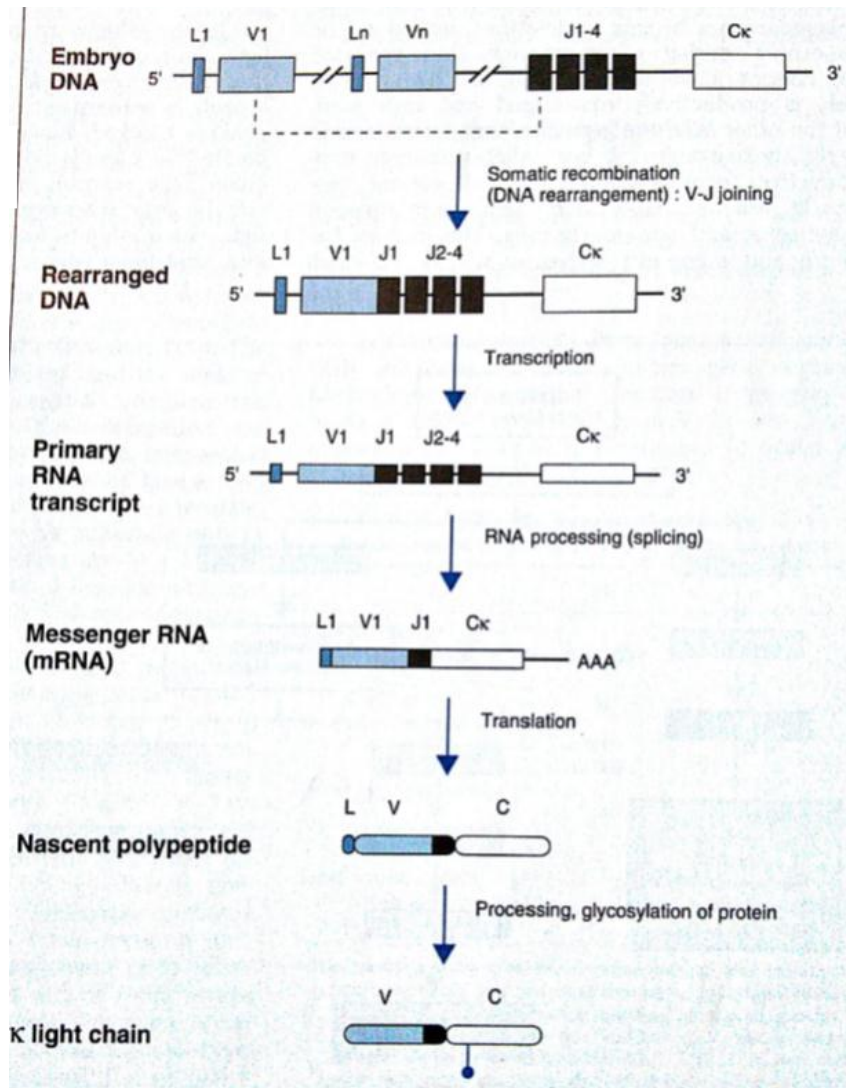
Esone D: diversity segments (solo catene pesanti)

Esone C: meno variabilità, ma comunque alcune copie in tandem, almeno nelle catene pesanti

Riarrangiamento genico, trascrizione e sintesi della catena pesante μ del topo (IgM)



Riarrangiamento genico, trascrizione e sintesi della catena leggera Igk del topo



(Abbas et al. "Cellular and molecular Immunology" ,Fig. 4-6 and 4-7)

Complessità genomica crescente

- Virus
- Batteri
- Eucarioti: uomo come esempio

Viral Genomes

- RNA viruses
 - Proteins translated from RNA
 - RNA replicates
- DNA viruses
 - Transcription to mRNA,, translation to protein
 - DNA replicates
- Retroviruses
 - Proteins can be translated from RNA
 - Replication cycle includes DNA intermediate

RNA Viruses

- Three classifications
 - Double-stranded RNA viruses
 - Positive-sense single stranded RNA viruses
 - SARS, polio, Hepatitis C Hepatitis A, West Nile
 - Negative-sense single-stranded RNA viruses
 - Ebola, measles, mumps, Hepatitis D, Influenza
- RNA replication more error prone than DNA
 - RNA viruses tend to mutate, evolve quickly

Scheda 2.4: tipi di genoma virale (Lesk)

Dopo essere stato assemblato all'interno del virione, un genoma virale può essere costituito da:

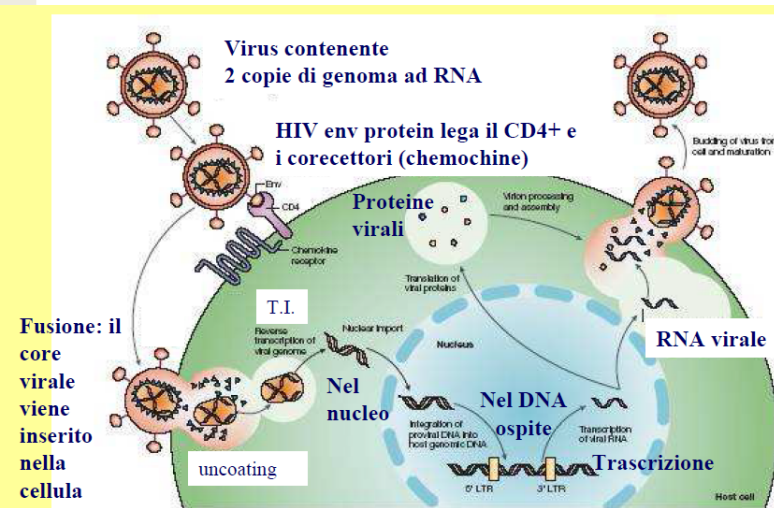
Acido nucleico

Esempi

- DNA a singolo filamento batteriofagi ϕ X-174 e M13
- DNA a doppio filamento adenovirus, virus del vaiolo, virus di Epstein-Barr, batteriofago λ
- RNA a singolo filamento batteriofagi MS2, Q β , virus del mosaico del tabacco, HIV-1
- RNA a doppio filamento virus della bluetongue (malattia della lingua blu)

I genomi virali a DNA a singolo filamento vengono generalmente convertiti in DNA a doppio filamento dall'ospite. La replicazione dei virus a RNA è incline a mutazione perché non si applicano i meccanismi di correzione di errore attivi nella replicazione del DNA dell'ospite. Ciò aiuta i virus a sfuggire al sistema immunitario degli ospiti e facilita il loro salto tra specie ospiti. [Le malattie virali emergenti, comprendenti la sindrome da immunodeficienza acquisita (AIDS, acquired immunodeficiency syndrome) e l'influenza aviaria, ma non solo queste, si basano tipicamente su virus con genomi a RNA.]

Un genoma virale costituito da RNA a singolo filamento può essere:



Evoluzione interna all'ospite

Alto tasso di mutazione

RNA $\xrightarrow{\text{TI}}$ DNA $\xrightarrow{\text{RNAPol II}}$ RNA
 0.2 errori/generazione Altri errori

Veloce generazione virale ~ 2.5 giorni

~ $10^{10} - 10^{12}$ virioni / giorno

Alto tasso di ricombinazione

DNA viruses

- Most are double-stranded but some bacterial (and other) viruses are single-stranded
- Must replicate DNA and transcribe RNA
- Examples
 - Smallpox, (VAIOLO)
 - Human herpesvirus, Chickenpox (VARICELLA)
 - Human papillomavirus

Retroviruses

- RNA genome
 - Reverse transcriptase used to make DNA intermediate
 - DNA transcribed to make new viral genome
 - And to be translated into viral proteins
- Enveloped viruses
 - Have lipid-based envelopes like cell membranes
- HIV, Hepatitis B, some cancer-related viruses

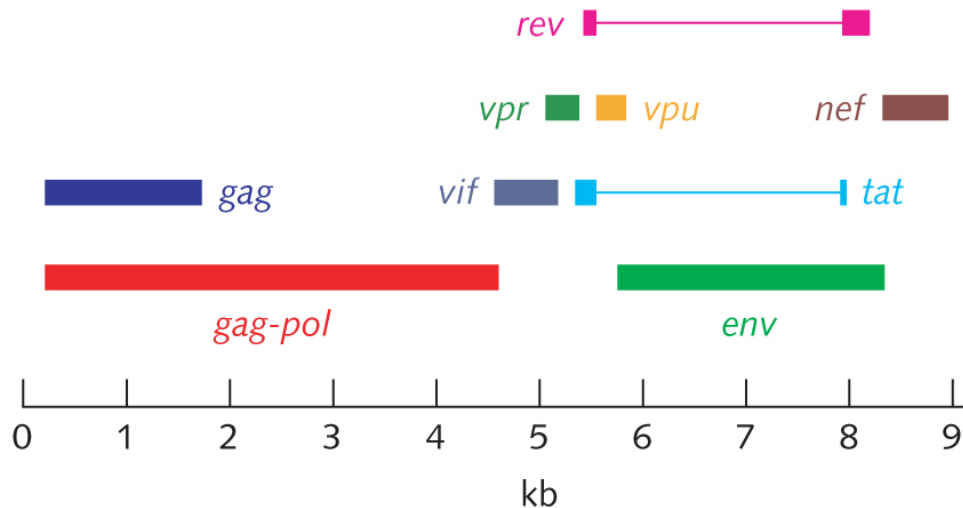
Genomi virali: l'esempio di un retrovirus

Proteine specifiche per attaccare ed invadere le cellule ospiti

Acido nucleico che entra nella cellula ospite

1. Produzione di più copie del genoma virale
2. Sintesi di proteine virali (enzimi attivi nella cellula ospite + proteine per assemblare i futuri virioni)
3. Assemblaggio dei virioni e abbandono dell'ospite

Esempio: il genoma di HIV-1, lunghezza circa 9,8 kb



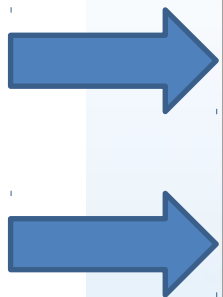
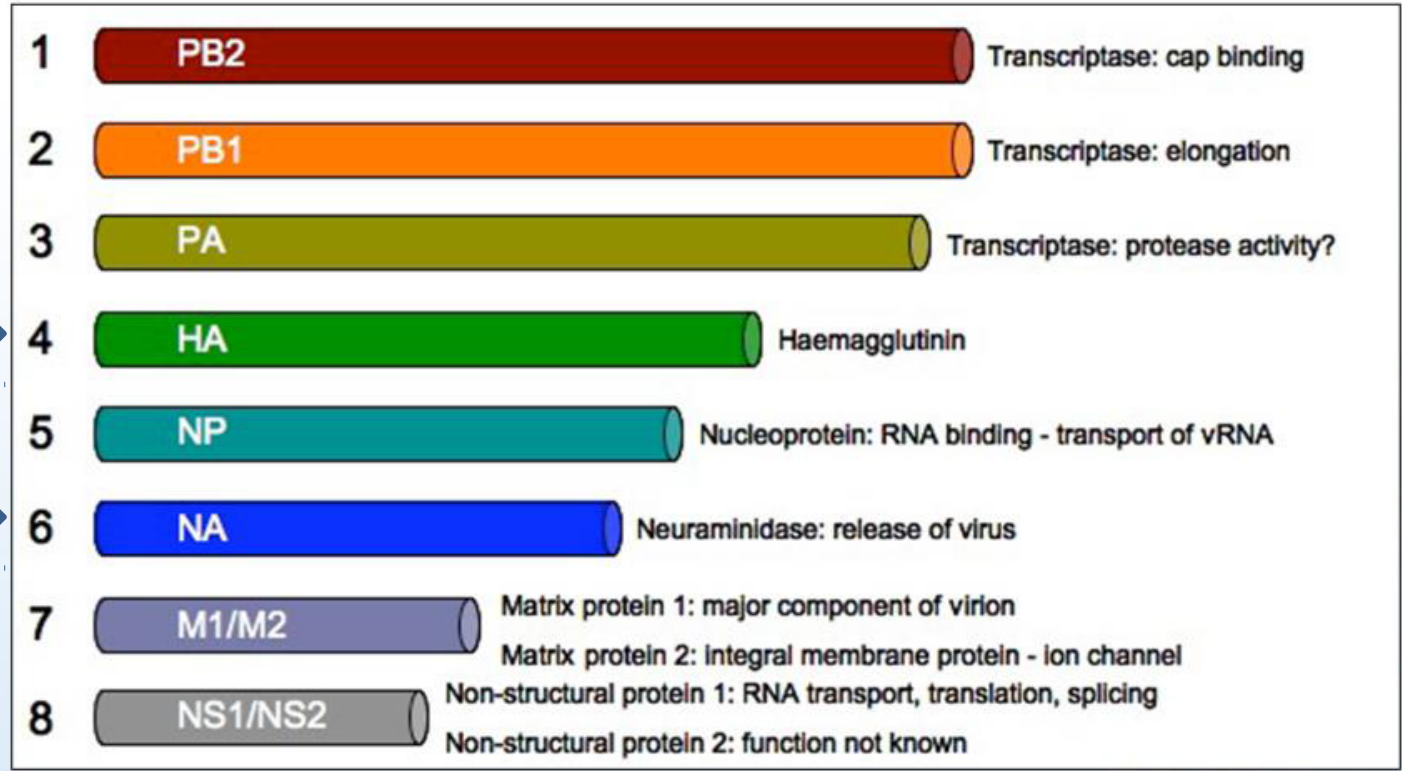
Gene	Proteine	Funzione
<i>gag</i>	p24, p6, p7, p17	proteine strutturali del capsid e della matrice
<i>pol</i>	trascrittasi inversa, integrasi, proteasi	integrazione nel genoma dell'ospite e scissione delle poliproteine codificate dal virus
<i>env</i>	precursori di gp120, gp41	proteine dell'involucro del virus, attive nell'attacco e nella fusione alle cellule ospiti
<i>tat</i>	Tat	facilita la trascrizione dell'RNA virale
<i>rev</i>	Rev	facilita l'esportazione citoplasmatica dei trascritti
<i>nef</i>	Nef	interferisce con la funzione immunitaria dell'ospite
<i>vif</i>	Vif	interferisce con le difese dell'ospite
<i>vpr</i>	Vpr	necessaria per l'importazione nucleare dell'acido nucleico virale
<i>vpu</i>	Vpu	promuove l'assemblaggio e il rilascio della progenie virale; inoltre stimola la degradazione delle proteine CD4 dell'ospite, disabilitando il sistema immunitario dell'ospite

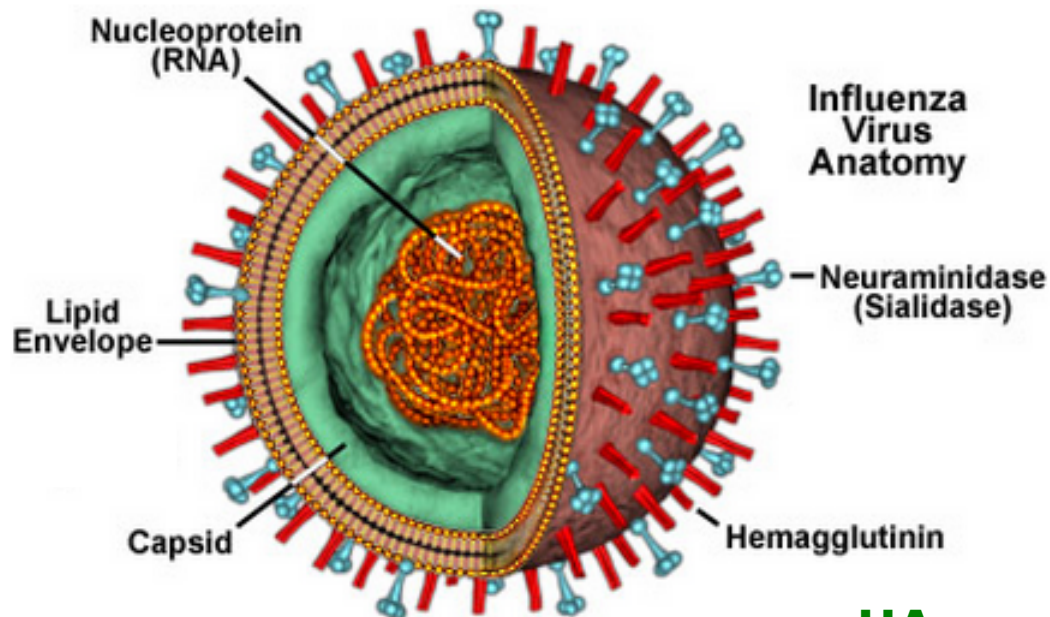
Gag, Pol, Env

- Standard retroviral genome
 - Three genes
- Gag - Group Specific Antigen
 - Proteins for viral interior
- Pol – Polymerase
 - Reverse transcriptase
- Env - Envelope
 - Proteins for lipid bilayer

Genomi virali: l'esempio di un genoma ad RNA Negative strand antisense RNA

Influenza Genome





NA

Enzima che taglia l'acido sialico
Promuove l'uscita del virus dall'ospite (9 tipi: N1-N9)
Buon target per il sistema immunitario e per farmaci antivirali

HA

Glicoproteina di superficie
Lega i recettori delle cellule ospiti per promuovere l'ingresso del virus (18 tipi: H1-H18)

Influenza: una minaccia del passato e del presente

Tre tipi di virus influenzali: A, B e C. I virus umani A e B causano epidemie stagionali quasi ogni inverno

Il tipo A è il più pericoloso

Pandemia 1918-1919 (spagnola) e di nuovo focolai nel 2009 (febbre suina):

H1N1

500 milioni di persone infettate in tutto il mondo

50-100 di morti, 3-5% della popolazione mondiale

Attuale influenza aviaria:

H5N1

<http://wwwnc.cdc.gov/eid/article/12/1/pdfs/05-0979.pdf>

Influenza Virus Resource

- NCBI resource
 - <http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>
 - Large collection of viral sequences
- Can do phylogenetic analysis to see how viruses from certain regions are related

The comparative genomics of viral emergence

Edward C. Holmes^{a,b,1}

^aCenter for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, PA 16802; and ^bFogarty International Center, National Institutes of Health, Bethesda, MD 20892

RNA viruses are the main agents of emerging and re-emerging diseases. It is therefore important to reveal the evolutionary processes that underpin their ability to jump species boundaries and establish themselves in new hosts. Here, I discuss how comparative genomics can contribute to this endeavor. Arguably the most important evolutionary process in RNA virus evolution, abundant mutation, may even open up avenues for their control through “lethal mutagenesis.” Despite this remarkable mutational power, adaptation to diverse host species remains a major adaptive challenge, such that the most common outcome of host jumps are short-term “spillover” infections. A powerful case study of the utility of genomic approaches to studies of viral evolution and emergence is provided by influenza virus and brought into sharp focus by the ongoing epidemic of swine-origin H1N1 influenza A virus (A/H1N1pdm). Research here reveals a marked lack of surveillance of influenza viruses in pigs, coupled with the possibility of cryptic transmission before the first reported human cases, such that the exact genesis of A/H1N1pdm (where, when, how) is uncertain.

http://www.pnas.org/content/107/suppl_1/1742.full.pdf+html

Genomi nei procarioti

Singola molecola circolare a doppio filamento
superavvolto (nucleoide)

0,6-10 milioni di basi

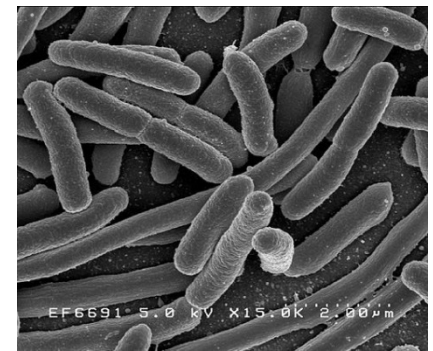
Molti contengono plasmidi

Eccezioni: DNA lineare (*Borrelia burgodferi*, Lyme disease) + plastidi

E. coli: 4.6 Mb; circa 4300 geni su entrambi i filamenti

87.8% di DNA codificante proteine

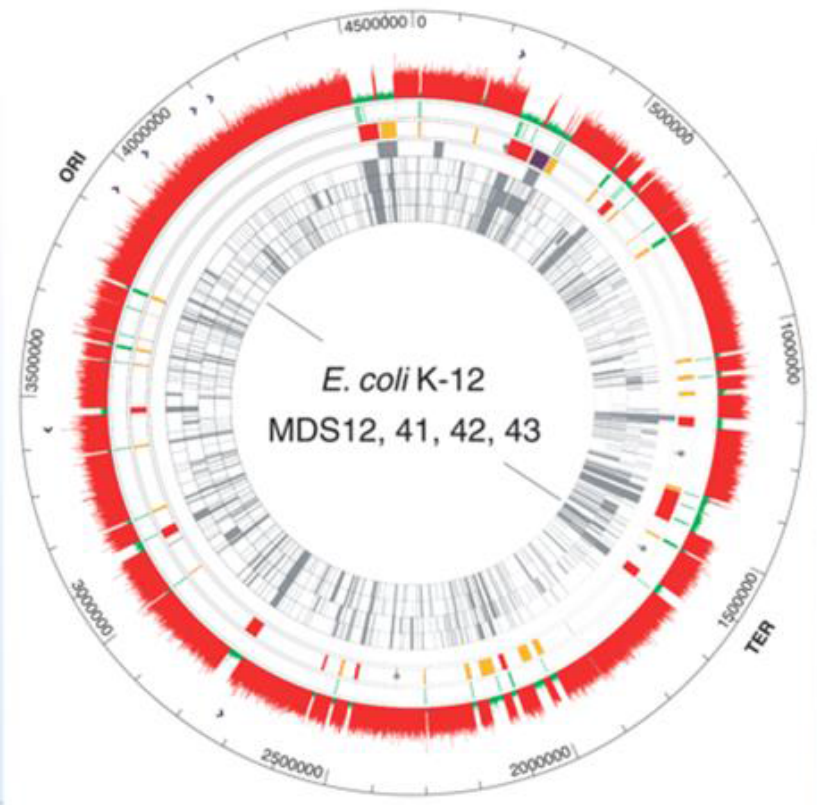
0.8% RNA strutturali



Specie	Seq codificanti	Densità genica media
<i>E. coli</i>	> 90%	1 gene / kb
<i>Tetraodon nigroviridis</i>	15%	1 gene / 10 kb
Uomo	5%	1 gene / 30 kb

E. coli Genome

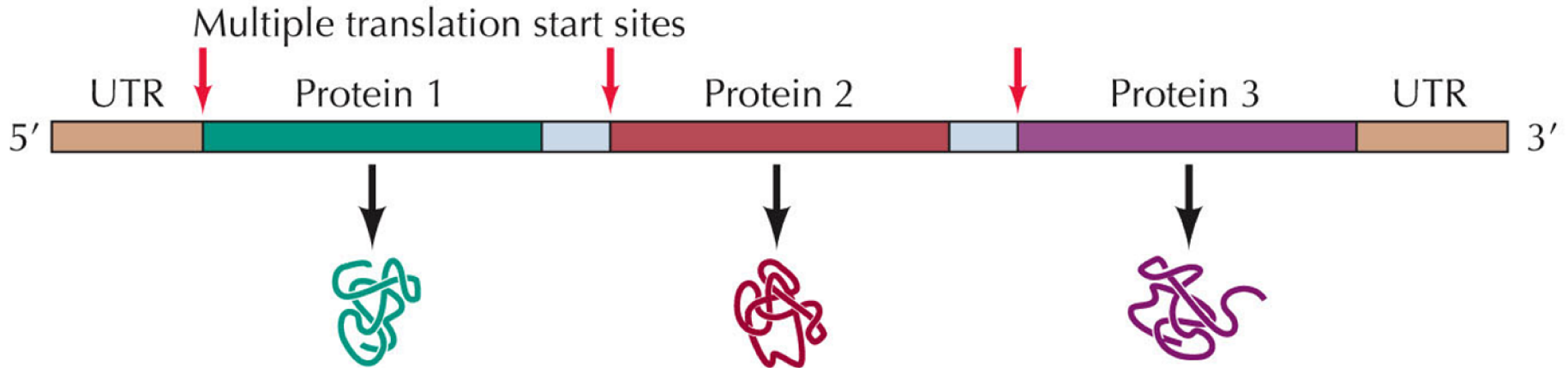
- Note position 0 at top of circle
 - Base pairs numbered in clockwise fashion
 - Position 0 represents origin of replication
- + strand genes run clockwise
- - (complementary) strand genes run counterclockwise



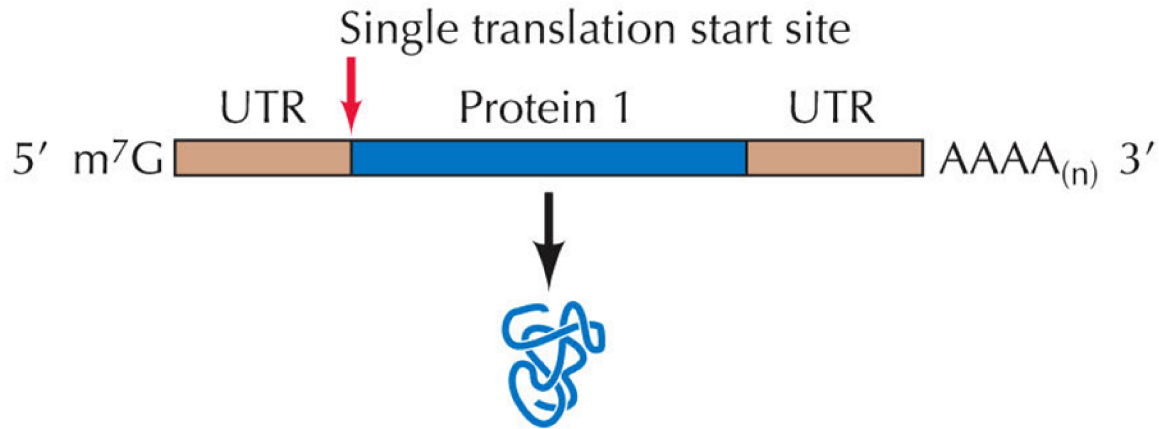
Polycistronic mRNA

- Eukaryotic transcripts usually have only one coding region
 - One mRNA makes one protein
 - Actually many molecules of same single protein
- Prokaryotic transcripts often have more than one coding region
 - One mRNA make multiple proteins
 - Actually many molecules of each protein

Prokaryotic mRNA

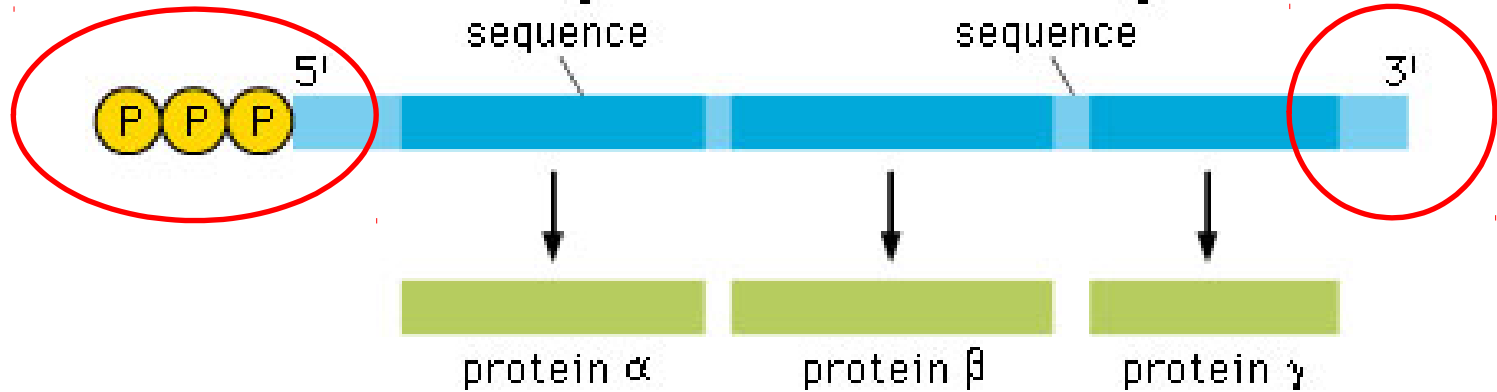


Eukaryotic mRNA

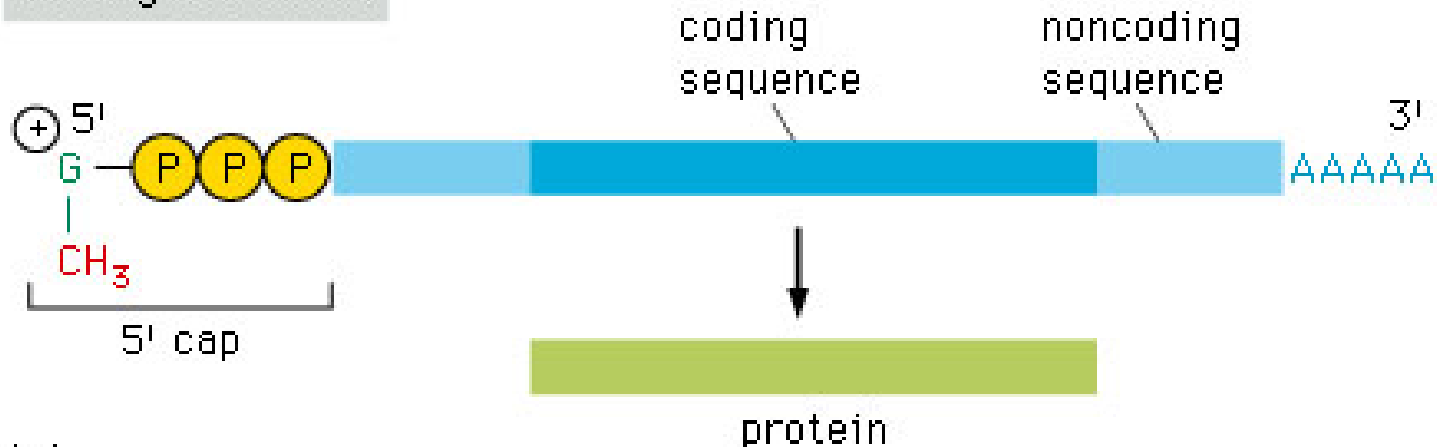


Notare la posizione degli UTR

procaryotic mRNA

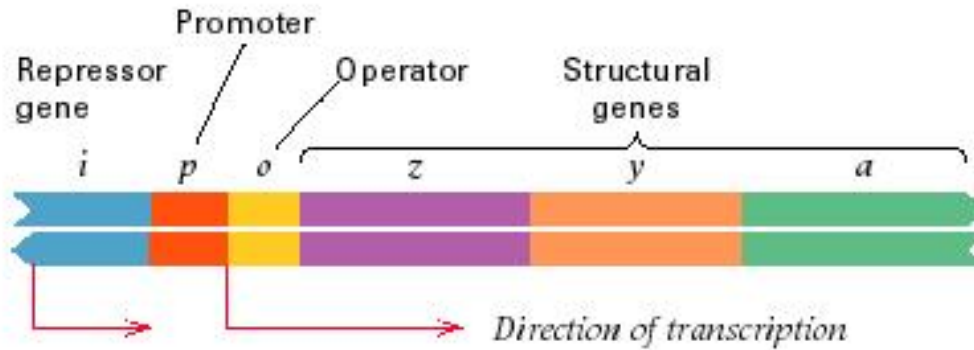


eucaryotic mRNA



(A)

L'operone «Lactose»



Più geni nello stesso trascritto

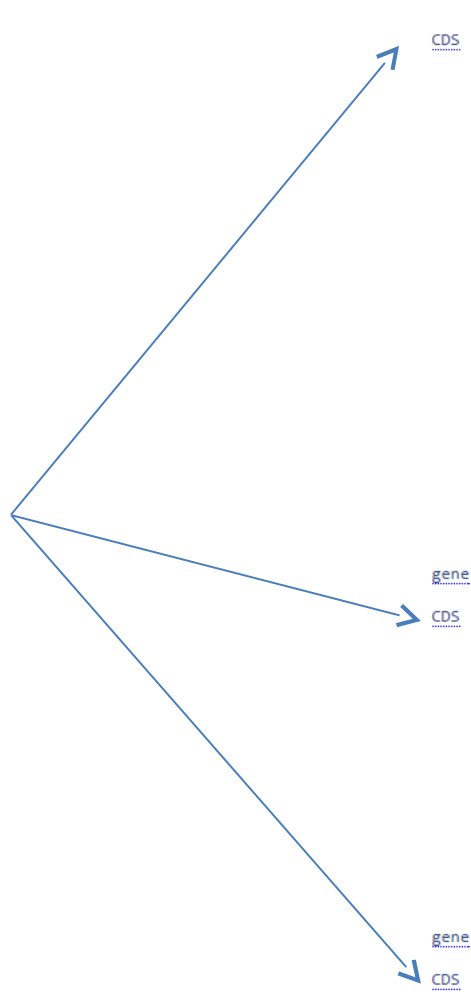
Regioni di controllo della trascrizione nella parte 5'

```

mRNA /note="lacI (repressor) mRNA; preferred in vivo 3' end
51..1230
gene /gene="lacI"
79..1161
CDS /gene="lacI"
79..1161
/note="lac repressor protein (gtg start codon)"
/codon_start=1
/transl_table=11
/protein_id="AAA24052.1"
/db_xref="GI:146576"
/translation="MKPVTLYDVAEYAGVSYQTVSRVWNQASHVSAKTRKVEAAMAE
LNYIPNRVAQQLAGKQSLIGVATSSALHAPSQIVAAIKSRADQLGASVVVSMVERS
GVEACKAAVHNLQAQRVSLIINYPLDDQDAIAVEAACTNVPALFLDVSQDTPINSII
FSHEDGTRLGVEHLVALGHQOIALLAGPLSSVSARLRAGWHKYLTRNMQPFAEREG
DWSAMSGFQQTINQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYYDDT
EDSSCYIPPSSTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNT
QTASPRALADSLMQLARQVSRLESGQ"

```

Multipli CDS (coding sequences) su un unico mRNA



```

/gene="lacZ"
1284..4358
/gene="lacZ"
/note="beta-d-galactosidase"
/codon_start=1
/transl_table=11
/protein_id="AAA24053.1"
/db_xref="GI:146577"
/translation="MTMITDSLAVVLQRRDWNENPGVTLNRLAAHPPFASWRNSEEAR
TDRPSQQLRSLNGEWRFAWFPAPAEVPESELCELDPEADTVVPSNQMHGYDAPITY
NVTYPIITVNPFPVTENPTGCYSLTFNVDSEWLQEGQTRIIDFGVNSAFHLWCNGRW
GYGQDSRLPSEFDLSAFLRAGENRLAVMVLWRVSDGSYLEDDQMWRMSGIFRDVSLHK
PTTQISDFHVATRFNDPFSRAVLEAEVQMGELRDYLRVTVSLWQGETQVASTGAPFG
GEIIDERGGYADRVTLRNLVFNPKLWSAEIPNLRYAVVLEHTADGTLIEAEACDVGRF
EVRIEENGLLLNGKPLLRIRGVNRHEHPLHGQVMDQTMVQDILLMKQNNFNAVRCSS
YPIHPLWYTLCDRYGLYVDEANIEETHGMVPMNRLTDDPRWLPAMSERVTRMVRDRN
HPSVILWLSGNEGSHGANHDALYRNKISVDPSPRVQYEGGGADTTATDIIICPMYARVD
EDQFPFAPVKWSIKKWSLPGETRPLICEYAHAMGNSLGGFAKYWQAFRQYPRLQGG
FVWDWVQSLIKYDENGPNWASAYGGDGTDPNDRQFCMNGLVFADRTPHPALTEAKHQ
QQFFQFRLSGQTEVTSYELFRHSDNELLHMVVALDGGKPLASGEVPLDVAPQKQLIE
LPELPQPEAGQLWLVTRVVPNATAWSEAGHISAWQWRLEAENSVTLPAASHAIPH
LTTSEMDFCIELGKRWQFNRSQGFLSQMMIGDKKQLLTPLRDQFTRAPLDNDIGVSE
ATRIDPNAWVERWKAAGHYQAEALLQCTADTLADAVLITTAHANQHOGKTLFISRKT
YRIDSGGQMAITVDVEVASDTPHARIGLNCQLAQVAERVMWLGLPQENYPRDLTAA
CFDRWDLPLSDMYTPYVFPSENGLRGCTRELNYGPHQNRGDFQFNIISRYSQQLMIETS
HRHLHAEEGTWNLDGFMHIGGDDSNPSVSAEFQLSAGRYHYQLVWCQK"
4410..5663
/gene="lacY"
4410..5663
/gene="lacY"
/note="lactose permease"
/codon_start=1
/transl_table=11
/protein_id="AAA24054.1"
/db_xref="GI:146578"
/translation="MYLKNITNFWMFGLFFFFYFFIMGAYFPFFIWLHDINHISKSD
TGIIFAAISLFSLLFQPLFGLLSDKLGRLKYLWIIITGMLVMFAPFFIFIFGPLLQYN
ILVGSIVGGIYLGFCFNAGAPAVEAFIEKVSRRSNFEFGRARMFQVCGWALCASIVGI
MFTINNQFVWLVGSGCALILAVLLFFAKTDAPSSATVANAVGANHSFSLKLALLELFR
PKLWFLSLYIVGSCTYDVFQDQFANFFTSFFATGEQGRVFGVYVTTMGEELNASIM
FFAPLIINRIGGKNALLAGTIMSVRIIGSSFATSALVILKTLHMFVFPFLLVGCF
KYITSQFEVRSATIYLVCFCKFKQLAMIFMSVLAGNMYESIGFGAYLVGLVALGF
TLISVFTLSGPGPLSLLRRQVNEVA"
5727..6338
/gene="lacA"
5727..6338
/gene="lacA"
/note="thiogalactoside acetyltransferase (ttg start codon)"
/codon_start=1
/transl_table=11
/protein_id="AAA24055.1"
/db_xref="GI:551814"
/translation="MMMPMTERIRAGKLFDMCEGLEPKRLRGKTLHYEFNHSHPSEV
EKRESLIKEMFATVGENAWVEPPVYFSYGSNIHIGRNFYANFLTIVDYDVTIIGDNL
LIAPNVTLSVTGHPVHLELRKNGEMYSPFITIGMNVWIGSHVINPGVTIGDNSVIGA
GSIVTKDIPPVVAAGVPCRVIREINDRDKHYFFDYKVESSV"

```

About

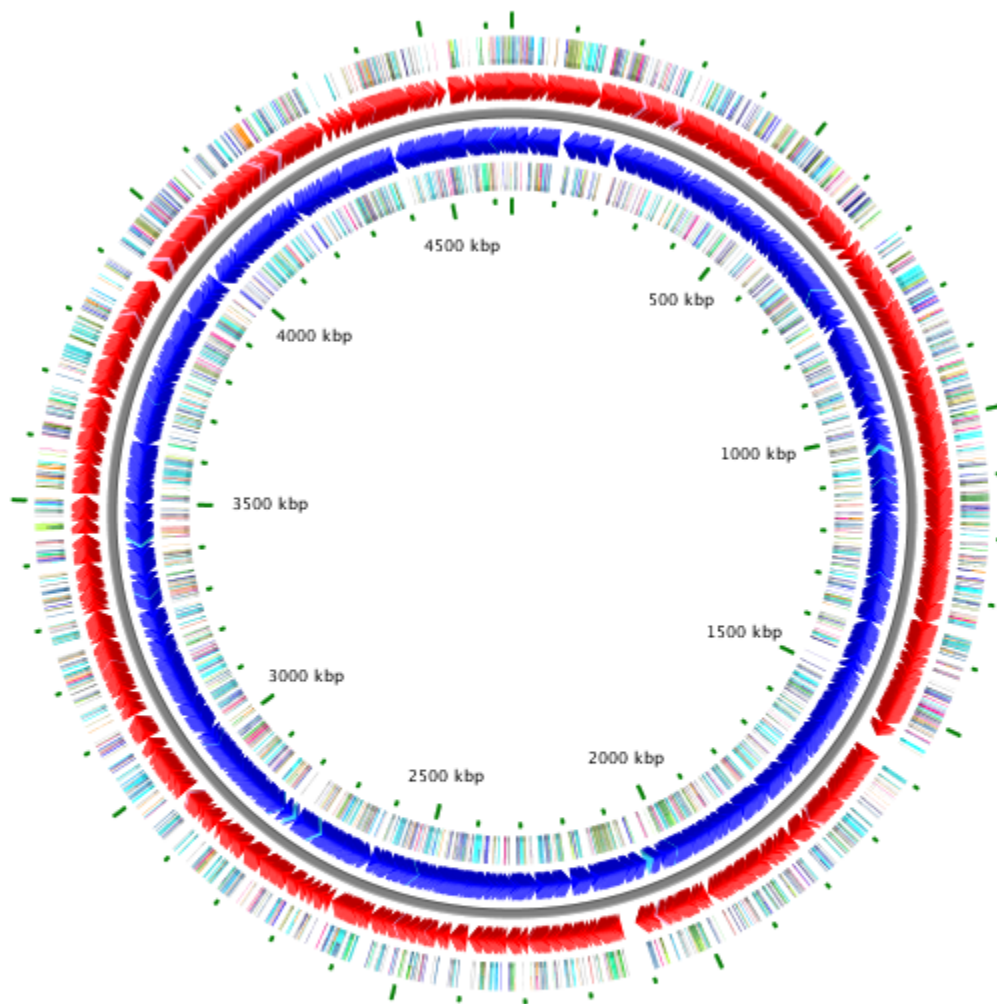
BacMap is an interactive atlas of fully sequenced bacterial genomes. It contains fully labeled, zoomable and searchable chromosome maps for 1,790 sequenced bacterial (archaeobacterial and eubacterial) species. This covers essentially all completely sequenced genomes in GenBank. For every genome and its genetic elements, BacMap used various bioinformatics programs and web-mining tools to generate detailed graphic and textual reports such as genome biography, genome maps, gene ontology and related metabolic pathways. All bacterial genome maps are supplemented with separate prophage genome maps as well as separate tRNA and rRNA maps. Each bacterial chromosome entry in BacMap contains graphs and tables on a variety of gene and protein statistics. BacMap allows the user to query, sort or select genes and genomes by taxonomy and/or phenotype classification and provides genome specific BLAST.

BacMap is developed by the Wishart group at the University of Alberta.

Database Statistics

Number of Bacteria	1,671
Number of Archaeobacteria	119
Number of Chromosomes	1,872
Number of Plasmids	1,117
Number of Bases	6,094,854,497
Number of Genes	4,838,288

BacMap



PROTEINS

- Forward strand
- Reverse strand

FUNCTIONAL RNA

- Forward strand
- Reverse strand

INFORMATION STORAGE AND PROCESSING

- Translation, ribosomal structure and biogenesis
- RNA processing and modification
- Transcription
- Replication, recombination and repair
- Chromatin structure and dynamics

CELLULAR PROCESSES AND SIGNALING

- Cell cycle control, cell division, chromosome partitioning
- Nuclear structure
- Defense mechanisms
- Signal transduction mechanisms
- Cell wall/membrane/envelope biogenesis
- Cell motility
- Cytoskeleton
- Extracellular structures
- Intracellular trafficking, secretion, and vesicular transport
- Posttranslational modification, protein turnover, chaperones

METABOLISM

- Energy production and conversion
- Carbohydrate transport and metabolism
- Amino acid transport and metabolism
- Nucleotide transport and metabolism
- Coenzyme transport and metabolism
- Lipid transport and metabolism
- Inorganic ion transport and metabolism
- Secondary metabolites biosynthesis, transport and catabolism

POORLY CHARACTERIZED

- General function prediction only
- Function unknown
- No COG

Accession: NC_000913

Topology: circular; Length: 4,639,675 bp; Genes: 4,255

[Expand -](#)
[Expand +](#)
[Full view](#)
[Rotate -](#)
[Rotate +](#)
[?](#)

Click tick marks to expand the view.

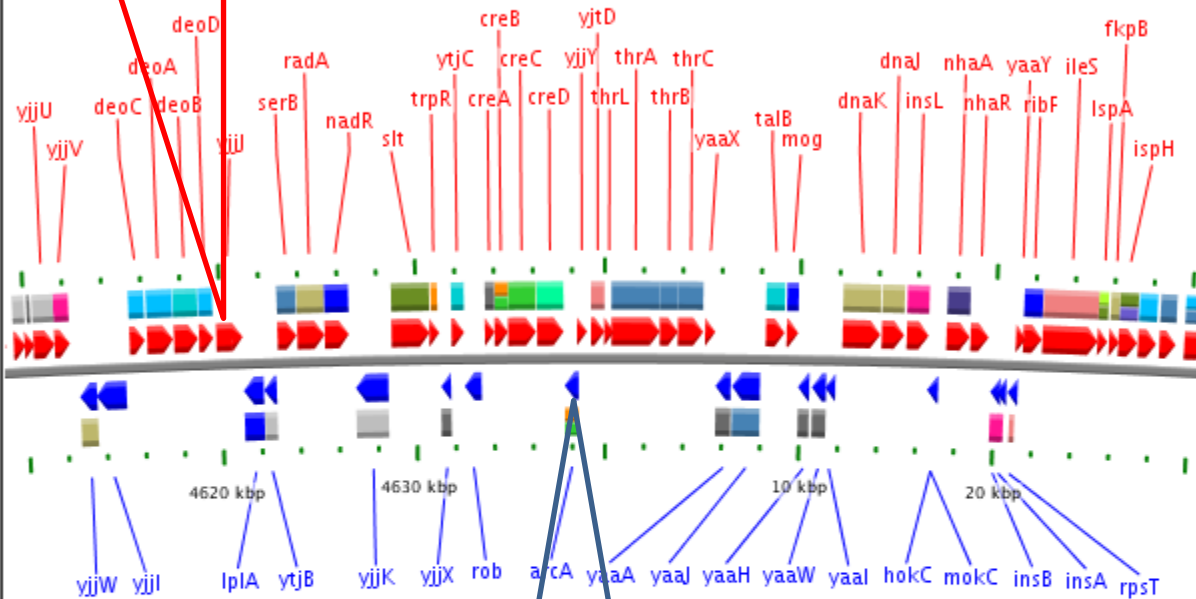
Valid XHTML 1.0; Valid CSS.

Displayed PNG file size: 190 kb.

Centered on base 1; Zoom = 1.

BacMap

Regioni codificanti
proteine sul filamento
forward



Regioni codificanti
proteine sul filamento
reverse

Accession: NC_000913

Topology: circular; Length: 4,639,675 bp; Genes: 4,255

Expand - Expand + Full view Rotate - Rotate + ?

Click tick marks to expand the view.

Valid XHTML 1.0; Valid CSS.

COG: Cluster of orthologous groups
(www.ncbi.nlm.nih.gov/COG/)

UCSC Microbial Genome Browser

=====
Help
=====

=====
FAQ
=====

=====
Tutorial
=====

=====
Credits
=====

=====
Lowe Lab
=====

<http://microbes.ucsc.edu/>

UCSC Genome Bioinformatics