

STRUTTURA DEL CORSO

- Organizzazione del genoma: concetto di cromosoma e di plasmide
- Meccanismi di evoluzione: trasferimento genico orizzontale e perdita di materiale genetico
- Origine ed evoluzione del genoma di batteri patogeni modello: *Shigella*, *Yersinia pestis*, *Vibrio cholerae*

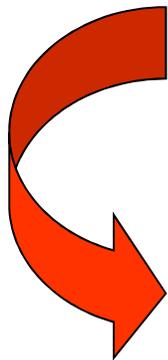
- Genoma minimo e genoma sintetico: verso il batterio artificiale
- Biologia sintetica: concetti di base e potenziali applicazioni
- Differenziamento nei batteri: modelli e regolazione
- Piccoli RNA e regolazione genica nei batteri
- L'immunità nei batteri, le sequenze CRISPR e la resistenza ai batteriofagi

- Microbioma intestinale: analisi genomica e metagenomica
- Microbioma polmonare: analisi metagenomica ed implicazioni nella fibrosi cistica
- Microbioma alimentare e salute dell'uomo

- Nuove strategie antibatteriche: farmaci mirati e terapia fagica
- I nuovi vaccini: dall'analisi genomica agli studi *in vivo*

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 costituita 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

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

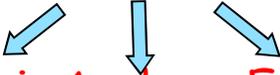
I batteri sono aploidi e si riproducono asexualmente .
Come si definisce quindi una specie?

Concetto di specie nei procarioti:

- due ceppi appartengono alla stessa specie quando la percentuale 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 divergenza)

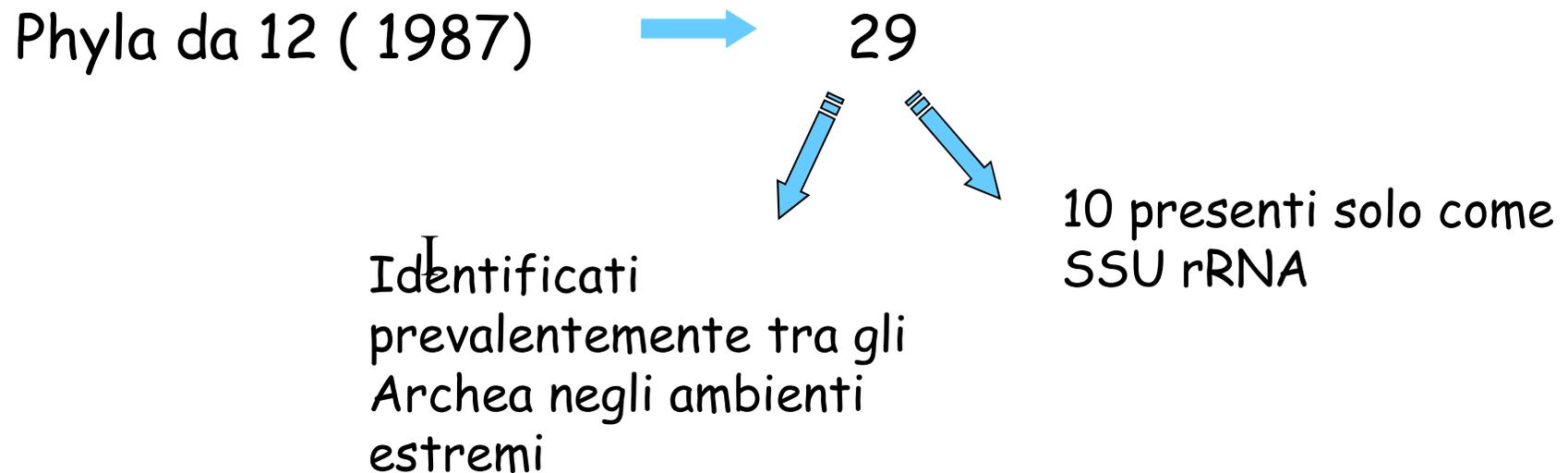
Specie ---Generi---Famiglie---Ordini---Classi---Phyla---- Dominio

Batteri Archea Eucarioti



La gran parte del mondo microbico rimane **NON COLTIVABILE**

Tramite l'estrazione e l'analisi del DNA dall'ambiente ci si è resi conto che questo mondo è molto più ampio di quanto si potesse pensare:



Quante specie ci sono? 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

Soltanto una piccola frazione della comunità microbica coltivabile
0.5-1 %

Possono esserci

- in un campione di suolo fino a 8×10^3 specie diverse
- in un campione d'acqua fino a 8×10^2 specie diverse

Tenendo conto dei parametri di associazione tra comunità diverse del suolo si calcola vi possano essere 10^9 specie di procarioti nel suolo

La differenza nel numero di specie tra suolo ed acqua può essere spiegata:

- alta diversità delle risorse del suolo
- isolamento della specie che ne riduce la competizione

LIMITI dell'identificazione di microrganismi non COLTIVABILI

le conoscenze rimangono limitate alla sequenza ma sulla fisiologia , biochimica regolazione ?????

POSSIAMO grazie alla sequenza possiamo sapere

- quali geni siano dominanti in un certo ambiente
- quali specie siano predominanti
- progettare degli strumenti idonei allo studio dei microrganismi non coltivabili

Gli orologi evolutivi

Alcuni **geni** ed alcune **proteine** possono servire come **orologi evolutivi** ovvero come misura dei cambiamenti che sono avvenuti nel corso dell'evoluzione

Le differenze di sequenza nucleotidica o aminoacidica di macromolecole simili da un punto di vista funzionale (quindi omologhe) ci possono indicare la distanza evolutiva

.

Criteri per la scelta di una molecola ideale come orologio evolutivo

deve essere distribuita universalmente nel gruppo in esame

- deve essere funzionalmente omologa
- deve contenere regioni di sequenza conservate
- non deve aver subito troppi cambiamenti di sequenza

I migliori candidati:

RNA ribosomiale (componenti chiave del processo di traduzione)

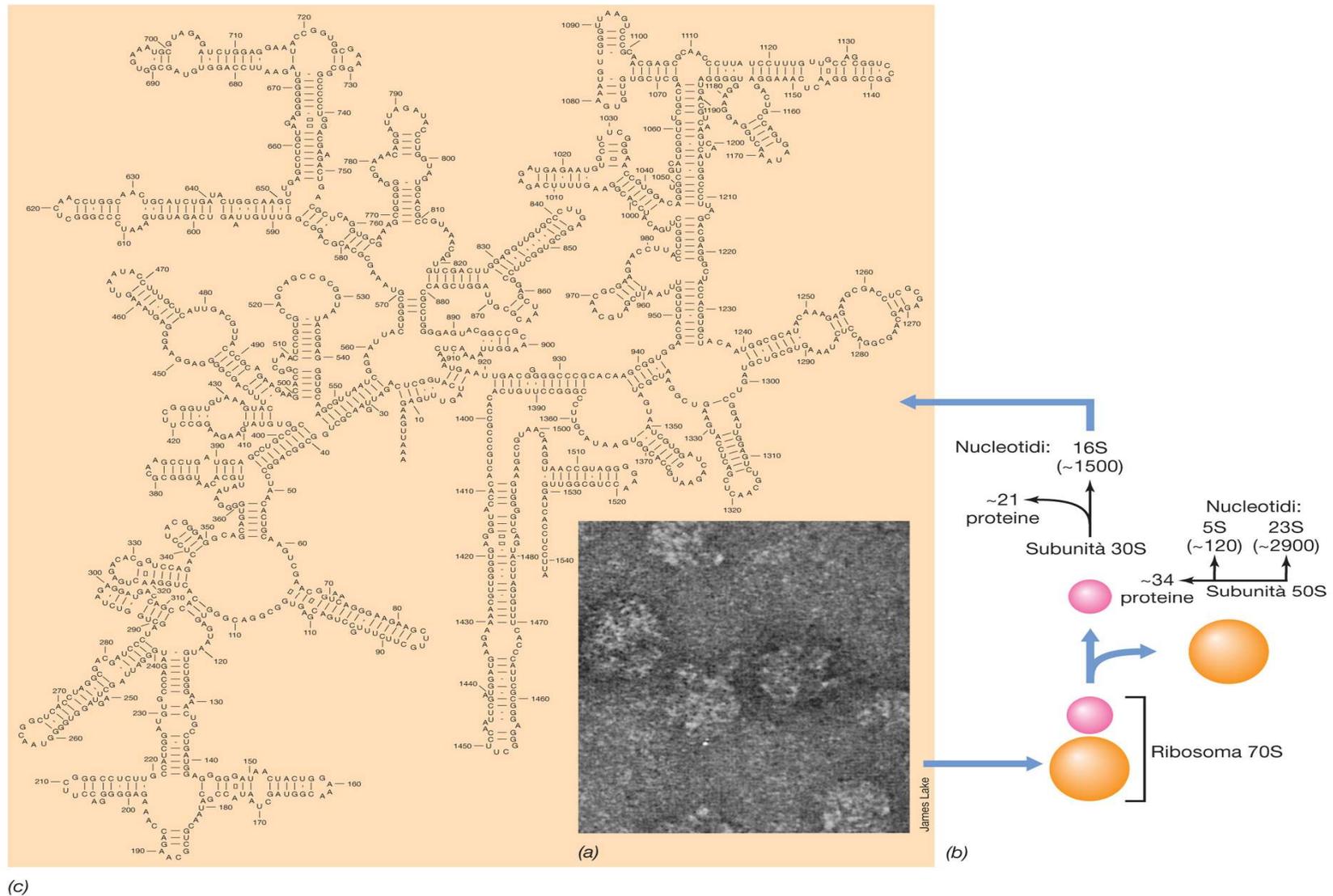
•

Le proteine con attività ATPasi (i componenti enzimatici che idrolizzano ATP)

•

RecA (la proteina chiave della ricombinazione)

Ribosomal Database Project (RDP) contiene oltre 100.000 sequenze di rRNA <http://rdp.cme.msu.edu/>



(c)

RNA ribosomiale come orologio evolutivo

- eccellente orologio evolutivo
- molecole relativamente grandi
- dotate della medesima funzione
- presenti in tutti gli organismi
- contengono regioni nucleotidiche ben conservate

Esistono 3 molecole di RNA ribosomiale

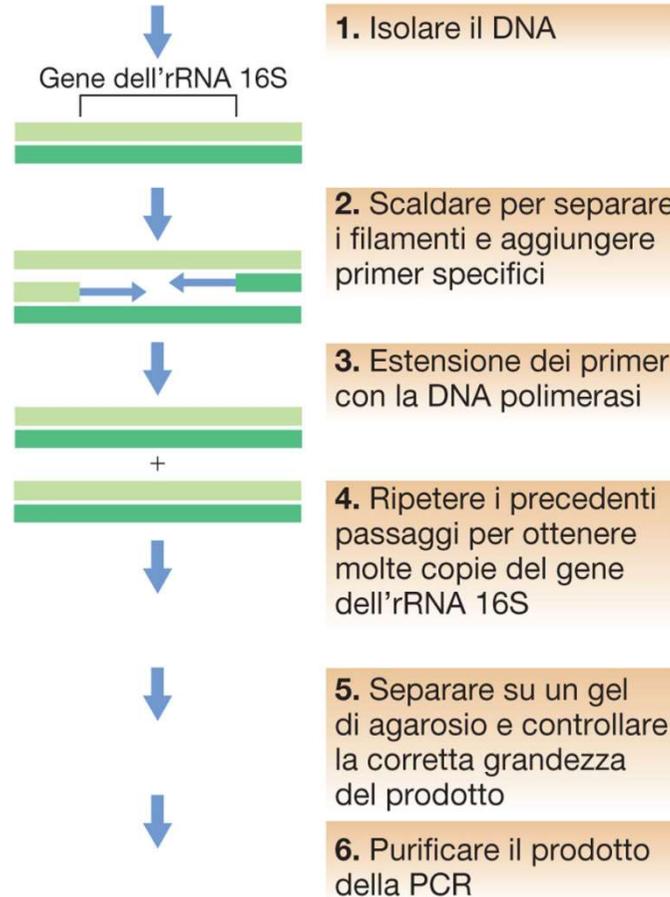
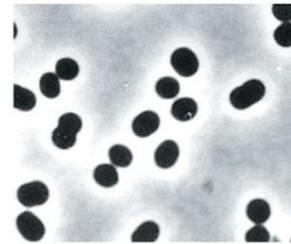
5S 16S e 23S

RNA 16 S è stato molto utilizzato negli studi di filogenesi nei procarioti
RNA 18S negli Eucarioti

SSU : dato che entrambe le molecole fanno parte della subunità piccola del ribosoma (30S o40S) per SSU si intende sequenziamento della Small Subunit

Quali sono le principali tecniche molecolari per l'identificazione dei microrganismi?

Sequenziamento dell'RNA ribosomiale subunità 16 S (SSU RNA)



Albero filogenetico basato sulle distanze evolutive tra sequenze di RNA ribosomiale SSU

Organismo	Sequenza	Analisi
A	C G U A G A C C U G A C	Per A → B, esistono tre differenze su un totale di dodici; quindi $\frac{3}{12} = 0,25$
B	C C U A G A G C U G G C	
C	C C A A G A C G U G G C	
D	G C U A G A U G U G C C	

(a) Allineamento e analisi della sequenza

Due algoritmi ampiamente utilizzati sono

- distanza
- parsimonia

	Distanza evolutiva	Distanza evolutiva corretta
E_D	A → B 0,25	0,30
E_D	A → C 0,33	0,44
E_D	A → D 0,42	0,61
E_D	B → C 0,25	0,30
E_D	B → D 0,33	0,44
E_D	C → D 0,33	0,44

(b) Calcolo della distanza evolutiva

Distanza evolutiva (E_D)

Si calcola la distanza evolutiva sulla base delle differenze presenti per ogni singolo nucleotide una volta che le sequenze sono state allineate

- la E_D rappresenta quindi la percentuale di nucleotidi non omologhi tra gli RNA di una qualsiasi coppia
Si può quindi costruire una matrice di distanza che mostra le ED tra tutte le coppie di nucleotidi presenti
- la E_D viene poi modificata in base ad un fattore di correzione che tiene conto dei cambiamenti multipli che possono essere avvenuti ad ogni sito. Si applica una correzione statistica che tiene conto delle retromutazioni che riconducono al genotipo originale sia di altre mutazioni che possono essere avvenute nello stesso sito.

PARSIMONIA :

altro algoritmo importante per l'analisi filogenetica

Gli alberi evolutivi sono generati sulla base dell'assunzione che

la **DIVERGENZA** tra due linee, a partire da un antenato comune, si sia prodotta in seguito alla **QUANTITA' MINIMA** di modificazioni necessarie.

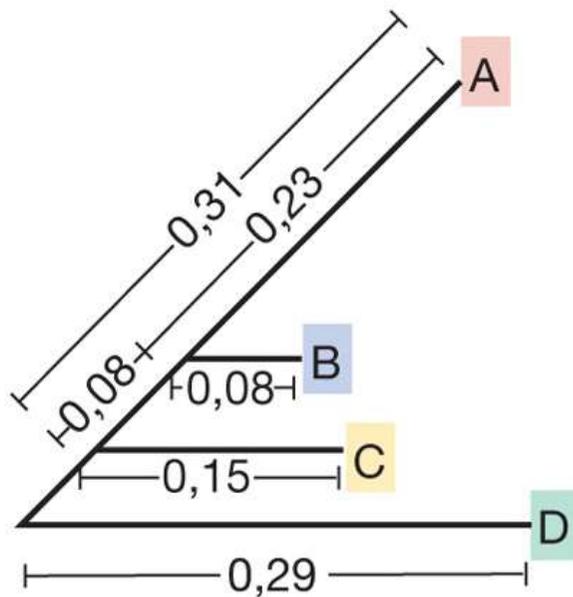
Come il metodo della E_D anche questo richiede la somma del numero di differenze di sequenza in un particolare insieme di dati .

L'ordine di ramificazione degli alberi ottenuti secondo i due parametri può talvolta differire.

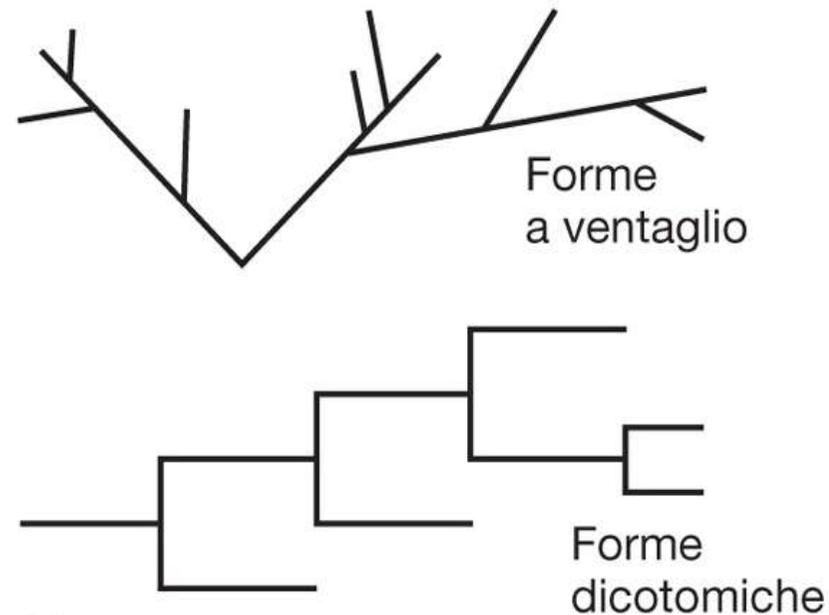
Alberi filogenetici

Vengono generati sulla base della distanza evolutiva

- forme a ventaglio o forme dicotomiche



(c) Albero filogenetico

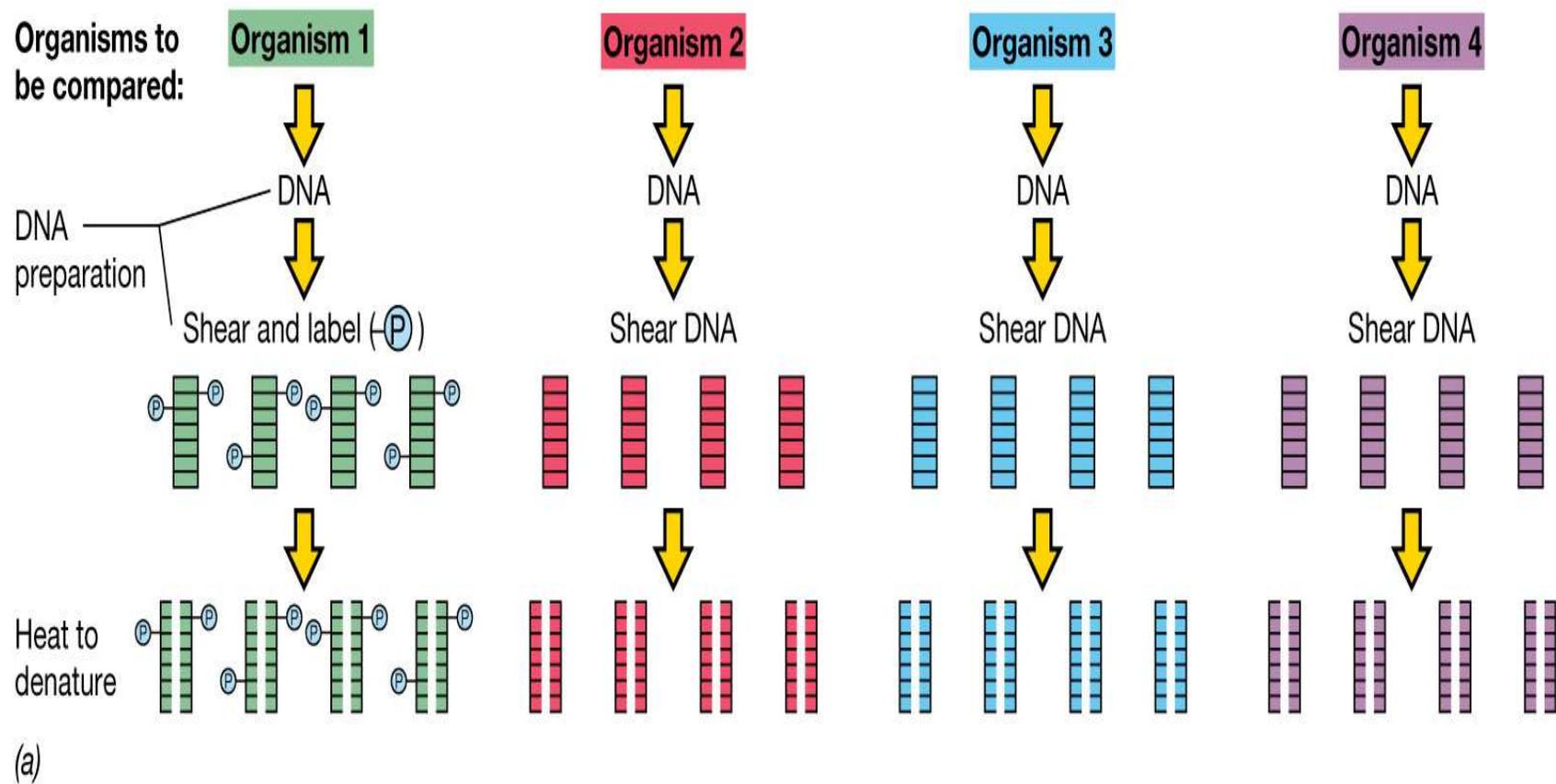


(d) Topologie dell'albero

In entrambi i formati la lunghezza totale dei rami che separano una coppia di organismi è proporzionale alla distanza evolutiva esistente tra essi

Tecniche di ibridazione DNA-DNA per classificare le specie

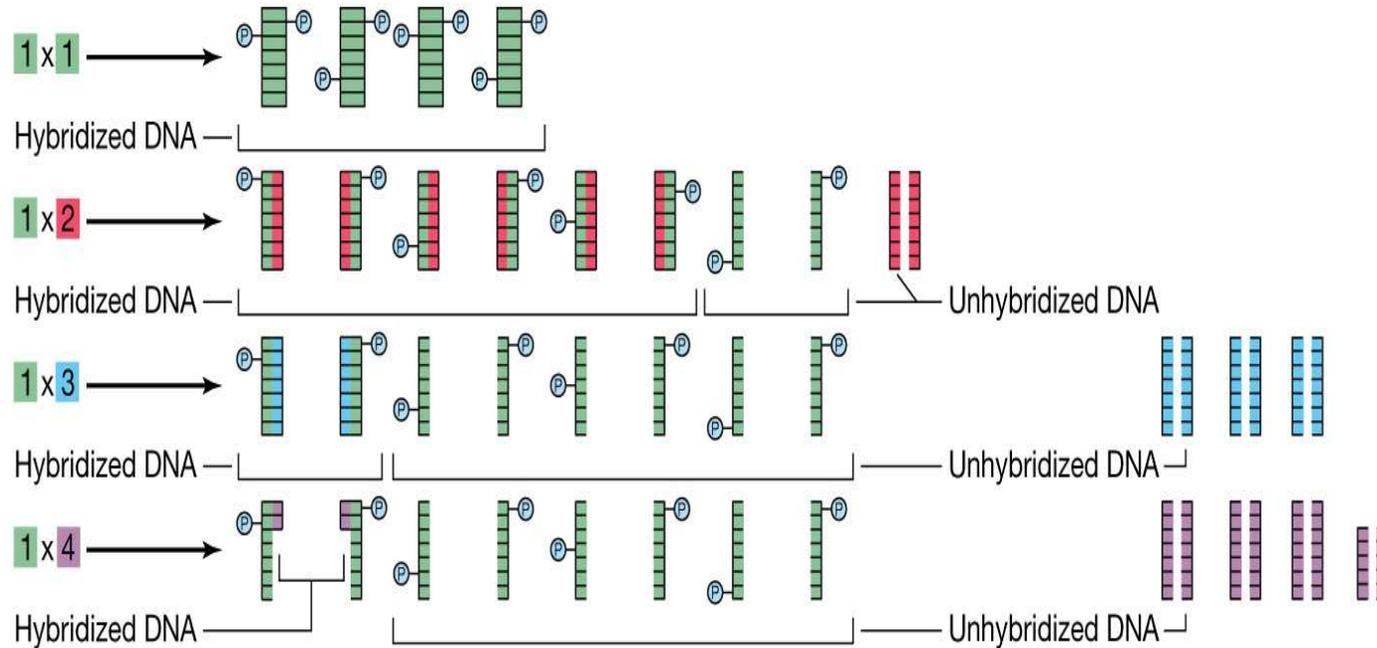
L'ibridazione genomica misura il grado di omologia di sequenza ed è particolarmente utile per differenziare microrganismi correlati



La percentuale di DNA ibridizzato fornisce una stima dell'omologia tra le due specie

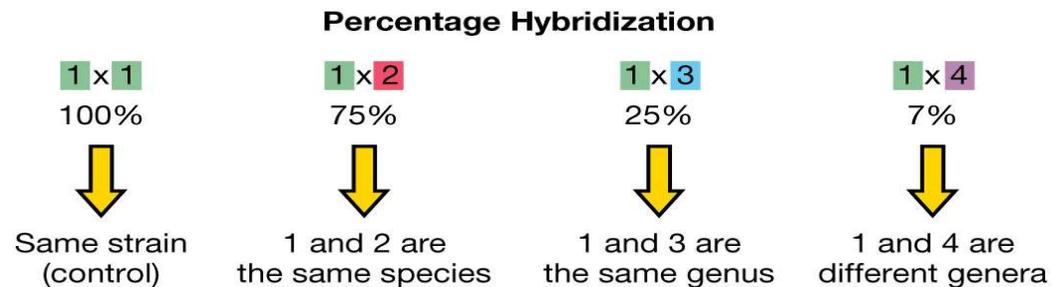
Hybridization

experiment: Mix DNA from two organisms—unlabeled DNA is added in excess:



(b)

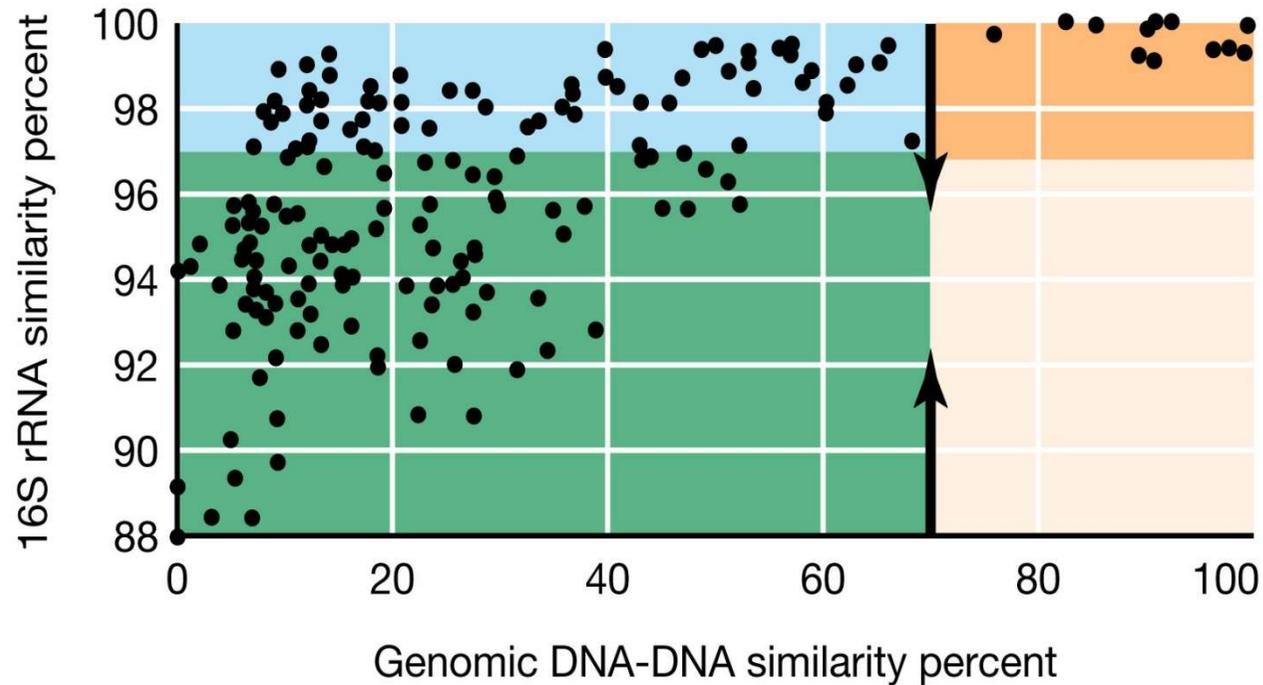
Results and interpretation:



(c)

Relazione tra diversi microrganismi ottenute utilizzando la tecnica del sequenziamento del SSU RNA e della ibridazione DNA-DNA

Punti nella box ARANCIONE= elevata omologia stessa specie
Punti nella box VERDE = bassa omologia specie diverse
Punti nella Box AZZURRA = risultati contrastanti specie diverse



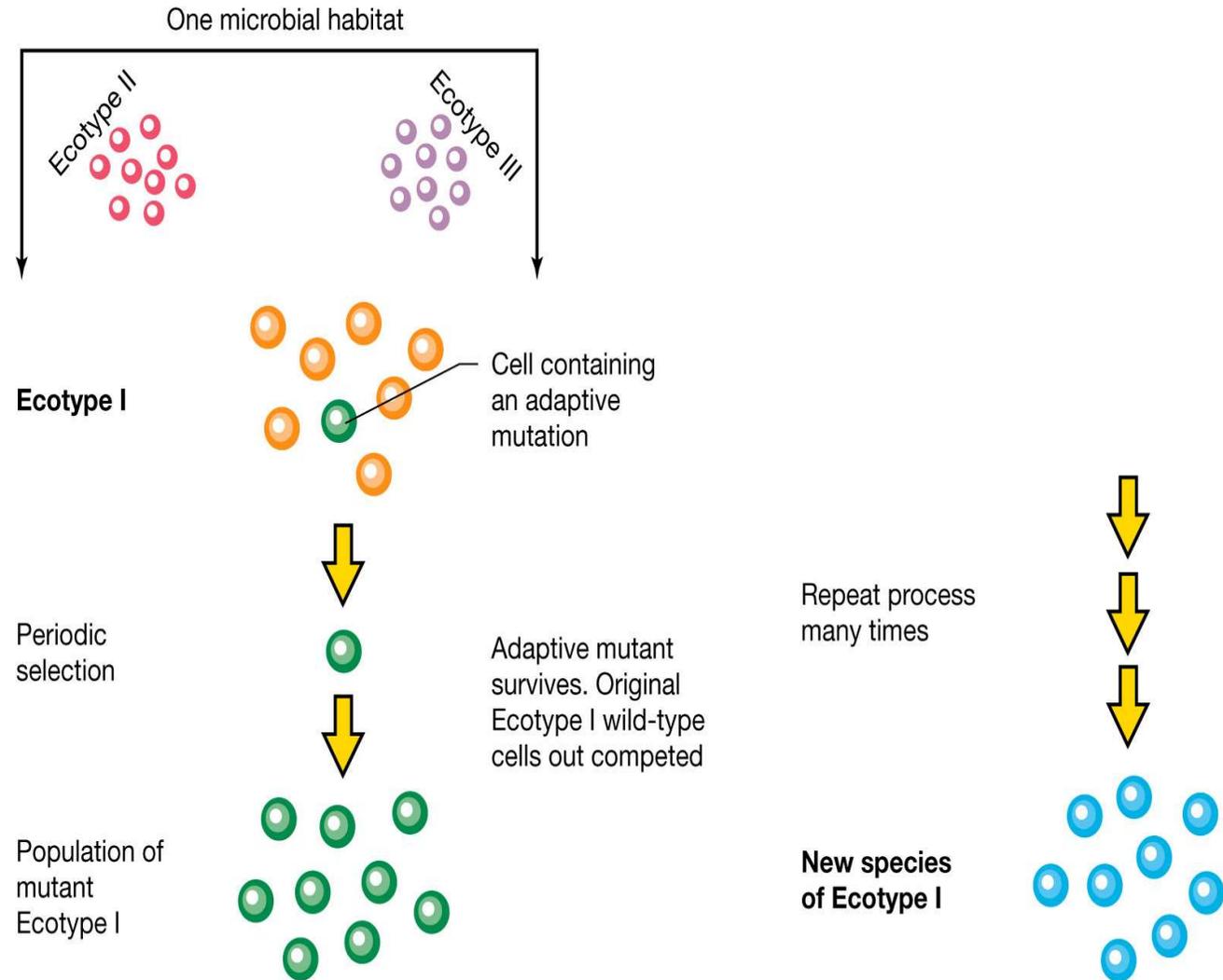
La tecnica di ibridazione DNA-DNA rimane una tecnica molto sensibile e riesce a rivelare differenze tra due specie anche quando non siano viste con la sequenza di RNA SSU ribosomale e con analisi fenotipiche

Possibile meccanismo di speciazione nei batteri

Diversi ecotipi possono coesistere nella stessa nicchia

Se interviene una modificazione che conferisce un vantaggio selettivo questa popolazione prenderà il sopravvento sulla parentale

Se il processo si ripete più volte si originerà una popolazione geneticamente distinta



Organizzazione del genoma nei procarioti

- *Escherichia coli* : dotato di un solo cromosoma circolare
- *Bacillus subtilis* : dotato di un solo cromosoma circolare

I due microrganismi maggiormente studiati da un punto di vista genetico e biochimico sembrano avere la stessa organizzazione del materiale genetico

Da qui la generalizzazione o " dogma" che il genoma batterico sia costituito da un singolo cromosoma circolare.

L'idea che i batteri potessero avere più di un cromosoma inizia a sorgere a meta degli anni '80

I batteri possono avere più di un cromosoma?

Rhodobacter sphaeroides, un microrganismo

- appartenente al gruppo alfa dei Proteobatteri rossi e non sulfurei
- utilizzato come modello negli studi sulla struttura, regolazione e biofisica del processo di fotosintesi batterica
- ha un genoma costituito da 2 cromosomi CI (2.9 Mb) e CII (0.9 Mb) oltre a 5 plasmidi compresi tra 100 e 42 kb)

Perché CII è un cromosoma e non un grande plasmide ?

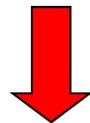
PLASMIDE : metamorfosi di un nome

Il termine PLASMIDE fu coniato da Joshua Lederberg nel 1952 per identificare qualunque forma di elemento genetico extracromosomico stabilmente mantenuto

Non vi erano distinzioni tra elementi di origine procariotica o eucariotica

- mitocondri , cloroplasti particelle virali o fagiche

1970-1980 introduzione delle tecniche di clonaggio



Il termine PLASMIDE descrive piccoli elementi di DNA circolare

Questa definizione risulta ben presto limitata in quanto :

- alcuni plasmidi non sono piccoli (2-25 kb) ma grandi (25-200 kb)
- alcuni plasmidi sono lineari

Alcuni plasmidi sono più grandi di genomi  MEGAPLASMIDI

Sinorhizobium meliloti ha due megapalsmidi di 1.35 (pSymBA) e 1.68 Mb(pSymB)

Treponema pallidum una spirocheta agente eziologico della sifilide ha un cromosoma di 1.1 Mb; *Mycoplasma genitalium* 0.58 Mb

La definizione STRUTTURALE (i.e. dimensioni) per differenziare plasmide e cromosoma non può essere valida

RICERCA di una definizione FUNZIONALE:

PLASMIDI codificano funzioni accessorie per la vita della cellula (antibiotico resistenza, utilizzazione di substrati, detossificazione metalli pesanti)

pSymB codifica l'unico tRNA che riconosce la tripletta CCG in *Rhizobium*
: è quindi un cromosoma?

Definizione funzionale di CROMOSOMA:

- elemento cellulare che porta l'informazione ereditaria in forma di geni
- cromosoma è sempre indispensabile a differenza del plasmide che indispensabile solo in talune condizioni

Prima di distruggere il dogma dell'"unicità" del cromosoma batterico si è cercato di rimuovere i cromosomi aggiuntivi :

se si rimuove "curing"
se non si rimuove

→ MEGAPLASMIDE
→ CROMOSOMA



- : - contiene funzioni essenziali
- non ci sono tecniche efficaci di curing

GENI ESSENZIALI geni per gli RNA ribosomiali 1 su CI
2 su CII

CAPIRE QUALI SIANO LE FUNZIONI CODIFICATE DA UN
GRANDE ELEMENTO GENETICO

Definizione funzionale di CROMOSOMA si basa sulla presenza di uno o più geni essenziali in ogni condizione di crescita (???) - può il laboratorio riprodurre le condizioni di vita naturali?
ma che cosa vuol dire GENE ESSENZIALE ?

È un concetto generale o variabile da specie a specie?

Per esempio quello che è essenziale per *E.coli* non lo è per *Mycoplasma*
capacità di sintetizzare nucleotidi essenziale per *E.coli*

GENE ESSENZIALE è un gene che non può essere distrutto perché porterebbe a morte la cellula

geni RNA polimerasi o geni per le proteine ribosomiali

I geni per la biosintesi degli AA sono essenziali?

E se un gene codifica per una funzione la cui mancanza rende la cellula molto malata è un gene essenziale (i.e. *recA* o *mutL*)?

SOLTANTO un a dettagliata conoscenza del "lifestyle" di un batterio può permetterci di definire quali siano i suoi geni essenziali

MULTIPLE CHROMOSOME -LIKE REPLICONS

MCLR

diffuso tra molti batteri specialmente Proteobatteri gruppo alfa
identificato tramite PFGE nella gran parte dei casi
solo con il sequenziamento in altri,

La presenza di geni per RNA 16S o di operoni *rrn* definisce

MCLR  CROMOSOMA

	N. Cromosomi	Dimensioni in Mb		
<i>Agrobacterium tumefaciens</i>	2	3	2	
<i>Brucella</i>	2	2.1	1.2	
<i>Rhodobacter sphaeroides</i>	2	3.1	0.9	
<i>Burkholderia cepacia</i>	3	3.7	3.2	1.1
<i>Vibrio cholerae</i>	2 (1)	3	1	
<i>Vibrio parahaemolyticus</i>	2	3.2	1.9	

ORIGINE DEI CROMOSOMI MULTIPLI?



- 1 - Frazioni di un unico grande cromosoma
- 2 - Sono derivati da plasmidi per l'aggiunta di geni essenziali
- 3 - Catturati durante eventi di scambio genetico

1. *Brucella melitensis* possiede 2 cromosomi circolari
CI di 2.1Mb e CII di 1.1 Mb.

Entrambi hanno origini di replicazione di tipo cromosomico ricche in AT
3 operoni RNA ribosomiali 2 su CI e 1 su CII

Ipotesi : i due cromosomi potrebbero essersi originati da un singolo
cromosoma ancestrale tramite ricombinazione degli operoni rrn

Esperimento in sostegno di quest'ipotesi: Suddividendo il cromosoma di
B.subtilis in due cromosomi la cellula è vitale

2. Integrazione di geni cromosomici in un plasmide

L'analisi comparata del genoma di

S.melitoti (1 cromosoma circolare e due megaplasmidi)

e di *A.tumefaciens* (1 cromosoma circolare, 1 lineare e due megaplasmidi)

ha rivelato che le due specie sono strettamente correlate e si sono evolute recentemente da un progenitore comune.

- I cromosomi circolari mostrano omologia nell'ordine dei geni

- Il cromosoma lineare di *A.tumefaciens* ha un'origine di replicazione e le proteine di segregazione simili a quelle dei due megaplasmidi.

Ipotesi : il cromosoma lineare (protocromosoma) si è evoluto da un plasmide ancestrale e si è poi diffuso in ceppi diversi.

Ipotesi 1 (o del frazionamento)  diffusione verticale

Ipotesi 2 (origine plasmidica)  diffusione orizzontale.

Ipotesi 3. Cattura di un cromosoma da un'altra specie. Differenze sostanziali nel livello di contenuto in CG tra cromosomi non è ancora stata evidenziata

Synorizobium melitoti
Agrobacterium tuefaciens

PUÒ ESSERCI UN VANTAGGIO SELETTIVO NELL' AVERE PIÙ CROMOSOMI?

Ipotesi 1: può servire per sfruttare più rapidamente nicchie ecologiche diverse?

Microrganismi correlati come *Rhodobacter spheroides* o *R. capsulatum* che occupano nicchie simili hanno un diverso numero di cromosomi

Ipotesi 2: può la ricombinazione intermolecolare essere più rapida ed efficiente di quella intramolecolare?

Non vi sono dati di cinetica di ricombinazione

Ipotesi 3 : può favorire gli scambi interspecie?

Se rimangono come megaplasmidi trasmissibili favoriscono la diversità genetica

C'è una suddivisione conservativa tra i geni nei diversi cromosomi?

- In *Vibrio cholerae* c'è un grande sbilanciamento nella distribuzione dei geni : sul grande cromosoma sono presenti tutti i geni importanti sia nei processi vitali della cellula che i geni coinvolti nella patogenicità

- In *Rhodobacter spheroides* 1/5 dei geni appare duplicato : i geni per alcuni pathway biosintetici come triptofano sono suddivisi tra i due genomi

CII appare derivato per frammentazione da un grande cromosoma

In *Ralstonia solanacearum* sul secondo cromosoma la distribuzione dei geni non è casuale

sono localizzati geni coinvolti nell'adattamento ad ambienti diversi, un sistema di secrezione di tipo III, sintesi di esopolisaccaridi

Cromosomi multipli possono conferire un vantaggio selettivo ma quale sia questo vantaggio è ancora da chiarire!

SECONDO CROMOSOMA o CROMIDE?

Introducing the bacterial 'chromid': not a chromosome, not a plasmid

Peter W. Harrison, Ryan P.J. Lower, Nayoung K.D. Kim and J. Peter W. Young

Department of Biology, University of York, PO Box 373, York, YO10 5YW, UK

La discussione è ancora aperta : in questo lavoro (2010) si critica l'attribuzione del nome "secondo cromosoma" a dei repliconi che portano geni essenziali per la crescita ma contengono origini di replicazione simili a quelle dei plasmidi .

Infatti nessun secondo cromosoma ha un sistema di replicazione di tipo cromosomico ma usano una replicazione simile a quella dei grandi plasmidi.

Da qui la proposta di definire questi elementi cromidi.

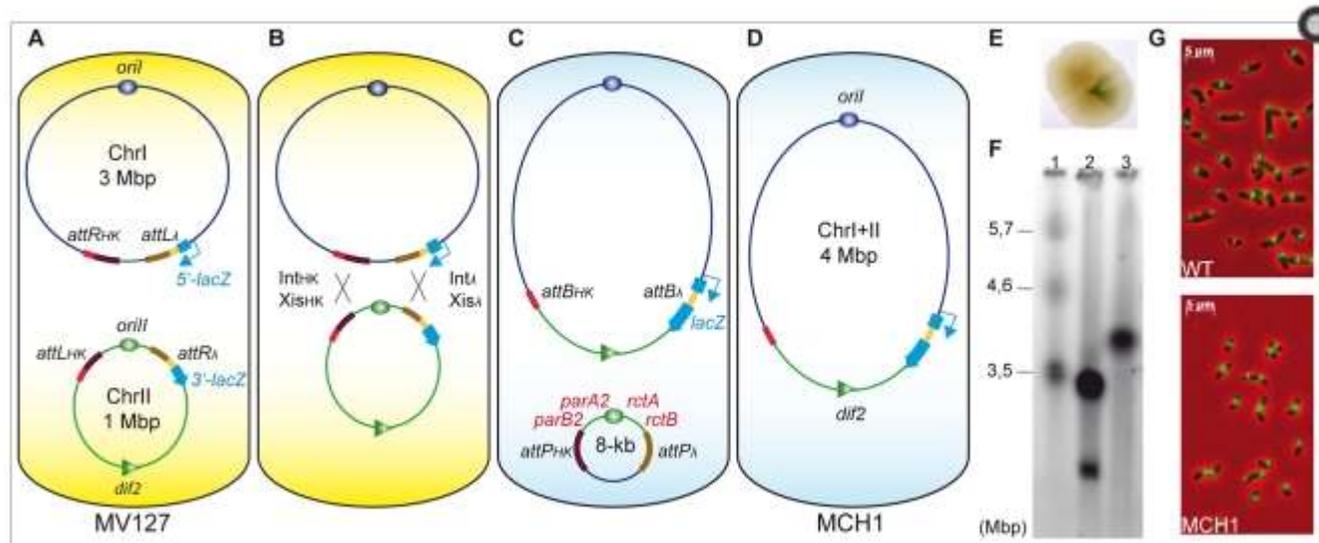
Un'analisi effettuata nel 2010 su circa 900 genomi batterici ne ha identificati ben 80 con un secondo cromosoma caratterizzato da un contenuto in G+C simile a quello del cromosoma.

La gran parte dei batteri dotati di più cromosomi appartengono ai Proteobatteri (Rhizobium, Burkholderia, Rhodobacter, Vibrio) ma anche alle Spirochete ed ai Cianobatteri.

L'uso dei codoni nei secondi cromosomi (cromidi) è molto più simile a quello dei cromosomi che a quello dei plasmidi a dimostrazione che hanno trascorso già un lungo periodo di coabitazione con il cromosoma del batterio che li ospita .

Non vi sono geni "cromidi" specifici comuni a tutti questo tipo di repliconi ma contengono sia geni del Core genome o geni più specifici di adattamento etc.

Un esempio *Vibrio cholerae*



A mono-chromosomal *V. cholerae* model, MCH1.

A. *V. cholerae* MV127 strain with *attR/L* sites from λ and HK022 phages inserted at precise loci. Recombination sites [*attR_{HK}* and *attL_λ*] replaced *dif1* on chrI and [*attL_{HK}* and *attR_λ*] flanked [*parAB2-oriII-rctAB*] on chrII. B. Recombination [*attR_λ* × *attL_λ*] and [*attR_{HK}* × *attL_{HK}*] mediated by the expression of *Int_λ* + *Xis_λ* and *Int_{HK}* + *Xis_{HK}*. C. Recombination events [*attR_λ* × *attL_λ*] regenerate *lacZ*, allowing for phenotypic detection of rearranged chromosomes. D. Without selection, the 8 kb excised molecule (carrying a kanamycin resistance gene) was lost. E. Blue sector appearing within single conjugant on X-Gal supplemented LB-agar plates indicates recombination events between [*attR_λ* × *attL_λ*]. F. Ethidium bromide stained PFGE of genomic DNA: Lane 1, *S. pombe* marker (BioRad) ; Lane 2, WT (N16961) ; Lane 3, MCH1. G. Microscopic observation of WT (top panel) versus MCH1 (bottom panel). Nucleoids of exponentially growing cells stained with DAPI (green) merged with phase-contrast images (red).

Coordinamento nella replicazione : Il caso di *Vibrio cholerae*

One of the important differences between plasmids and chromosomes is that the latter replicate during a defined period of the cell cycle, ensuring a single round of replication per cell.

Vibrio cholerae carries two circular chromosomes, Chr1 and Chr2, which are replicated in a well-orchestrated manner with the cell cycle and coordinated in such a way that replication termination occurs at the same time.

The initiation of Chr2 replication is triggered by the replication of a 150-bp locus positioned on Chr1, called *crtS*. This *crtS* replication-mediated Chr2 replication initiation mechanism explains how the two chromosomes communicate to coordinate their replication.

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