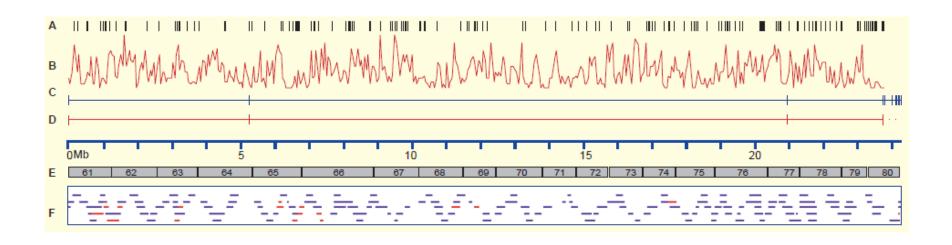
Proyecto de Secuenciación de Drosophila

Genomic resources for Drosophila.

Туре	Description	Resolution	Contribution Location of whole-genome scaffolds to chromosomes; confirmation of accuracy of assembly	
BAC-based STS content map	STS content map constructed by screening ~23× genome coverage of BAC clones; a tiling path of BACs spanning each chromosome arm was selected	50 kb		
Polytene map	Tiling-path BACs hybridized to polytene chromosomes	30 kb	Location of STSs and BACs to chromosomes; validation of BAC map	
BAC end-sequence	\sim 500 bp of sequence from each end of a BAC clone	Two reads per \sim 130 kb	Long-range association of sequence contigs	
Finished clone-based sequence	BAC, P1, and cosmid clones completely sequenced to high accuracy	~29 Mb of total sequence	Assessment of accuracy of Celera sequence and assembly	
Draft sequence from mapped BACs	≥1.5× shotgun sequence coverage of 825 clones from the tiling path of BAC and P1 clones	384 reads distributed across \sim 160 kb	Location of sequence contigs to a small genomic region; templates for gap closure	



Ensamblaje del genoma de Drosophila



- A) Elementos transponibles
- B) Densidad génica
- C) Andamio del ensamblaje por contigs
- D) Andamio del ensamblaje por shotgun
- E) Division de los cromosomas politenos
- F) Ensamblaje basado en clonas pequeñas



Identificación de Genes en un genoma

Consideraciones:

Genoma bacteriano, de archea o en eucarionte

Genoma bacteriano, o de archea.

GEN=Cadena ininterrumpida de DNA ente un codon de inicio usualmente ATG, (en alguna pequeña proporción de genes GTG, TTG, o CTG) y un codon de parada codon TAA, TGA, o TAG.

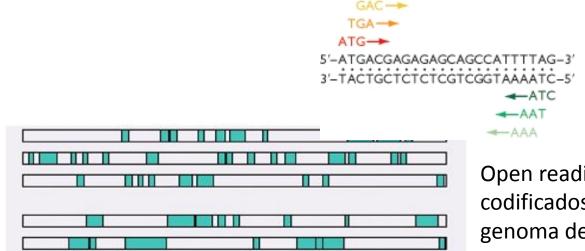
No hay limites estrictos en el tamaño de un gen. Eg *rpmJ* solo tiene 111 bp de largo, mientras que *PksK* en *B. subtilis* tiene 13,343 bp de largo.

mRNAs menores a 30 codones se traducen pobremente, por lo que usualmente los genes que Codifican para proteínas en procariontes tienen por lo menos 100 bp de largo. Por lo tanto En un proyecto de secuenciación ORF menores de 100 bp no son tomados en cuenta.



Análisis bioinformatico seDefine a un Gen como el ORF más largo de una región determinada de DNA.

Traducción de la secuencia de DNA en los 6 marcos de lectura se puede hacer online usando un servidor como ExPASy (http://www.expasy.org/tools/dna.html)



Open reading frames mayores de 100 bp codificados en un fragmento de 10-kb en el genoma de *Escherichia coli* K12.

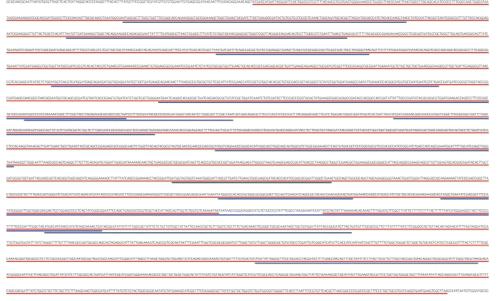


Problemas asociados con la predicción de genes:

- •En raras ocasiones entre dos ORF que se sobrelapan, el mas pequeño es el gen real
- •La existencia de un ORF en la dirección contraria a una secuencia que codifica para una proteína es mayor que en una secuencia al azar debido a las propiedades de la región.

 Región del operon de la lactosa.

ORF mayores de 50 codones se marcan en azul, los genes reales lacZ y lacY se marcan en rojo





Debido a estas complicaciones Siempre es recomendable tener otras lineas de evidencia de que un ORF codifica Realmente para una proteína. e.g.

El ORF en cuestion codifica para una proteína que es similar a otra descrita anteriormente (buscar homologos en la base de datos de proteínas)

EL ORF tiene el contenido tipico de GC, la frecuencia de codones o la composición de oligonucleotidos (con respecto al resto de los genes del genoma).

EL ORF es precedido por un tipico sitio de unión a ribosoma (se busca la secuencia Shine-Dalgarno rio arriba del ORF).

EL ORF es precedido por un tipico promotor (se buscan rio arriba secuencias promotoras similares a la secuencia consenso si se conocen para el organismo.)



- 1.-Identificamos ORF
- 2.-Traducimos la secuencia y se buscan secuencias homologas
- 3.-Comparamos la secuencia traducida con la base de datos del Clusters of Orthologous Groups of proteins (COG)



Identificación de Genes en un genoma eucarionte unicelular

En Eucariontes unicelulares sus genomas son extremadamente diversos en tamaño, La porción del genoma ocupada por genes y la frecuencia de intrones.

En *S. cerevisiae* 67% de su genoma contiene genes que codifican para proteínas, y solo 223 genes (menos de 4% del genoma) tiene intrones.

En Encephalitozoon cuniculi solo 12 genes contiene intrones.

Plasmodium falciparum el 43% de sus genes se encuentran en el cromosoma 2 y contiene uno o dos más intrones.

Prostistas con genomas más grandes tiene un densidad mayor de intrones e.g . Physarum polycephalum tiene en promedio 3.7 intornes por gen. Además el tamaño promedio de un exon es de 165 ± 85 bp y el de un intron 138 ± 103 bp En este organismo la prediccion de genes se hace más complicada

Identificación de Genes en un genoma eucarionte unicelular

No hay una sola forma eficiente de predecir genes en eucariontes unicelulares Debido a usa diversidad genomica.

e.g. Para levadura se usa el mismo sistema que para procariontes.

Para genomas ricos en intrones, en el modelaje de genes se incluye la información de los sitios de splicing, los cuales se pueden obtener comparando la secuencia genomica Contra una biblioteca de EST del mismo organismo.

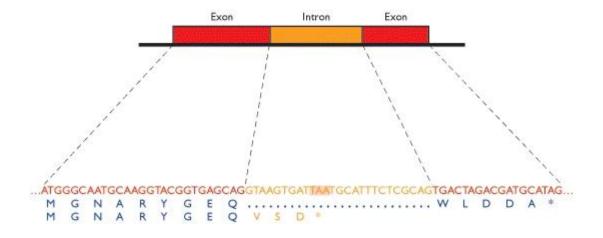


Genoma eucariontes multicelulares

Eucariontes multicelulares la mayoria de los genes estan interrumpidos por intrones

El promedio de un exon es de 150bp, muchos intrones son muy largos lo que produce genes de algunas megabases de largo.

Predicción de genes es muy complicada.





Genoma eucariontes multicelulares

Idealmente la predicción de genes deberia identificar todos los exones e intrones.

Para la identificación de intrones se han aplicado los siguientes criterios:

Codon bias (refiere al hecho de que no todos los codones se usan de forma igual en los genes en un mismo organismo).

Frontera entre exones e intrones se busca la siguiente secuencia:PPPPPPNC↓AG

Buscan regiones regulatorias rio arriba de los genes para determinar donde comienza un gen adémas que como los intrones tiene caracteristicas particulares que los distinguen de los exones



Se buscan homologos de la proteína traducida

Genoma eucariontes multicelulares predicción de genes

Incluso los mejores programas solo predicen correctamente al 40% de los genes

Los principales errores se cometen con genes con intrones muy grandes, los cuales Pueden ser predecidos como secuencias intergenicas, lo que resulta en la fisión de genes.

Con genes que tiene secuencias intergenicas muy cortas que pueden ser predecidas Como intrones, lo que resulta en la fusion de genes.



Programas de predicción de genes

Program(s)	Author(s), WWW site	Program versions available for:
GeneMark, GenMark.hmm	Mark Borodovsky, http://opal.biology.gatech.edu/ GeneMark	Human, mouse, rat, chicken, C. elegans, Drosophila, rice, Arabidopsis, yeast, many bacteria and archaea
Glimmer, GlimmerM	Steven Salzberg, <u>http://www.tigr.org/softlab</u>	Many bacteria and archaea, Plasmodium, Aspergillus, rice, Arabidopsis
Grail, GrailEXP	Edward Uberbacher, http://compbio.ornl.gov	Human, mouse, <i>Drosophila, Arabidopsis, E. coli</i>
GenScan	Christopher Burge, http://genes.mit.edu/ GENSCAN.html	Human, <i>Arabidopsis</i> , maize
GeneBuilder	Igor Rogozin, Luciano Milanesi <u>http://</u> www.itba.mi.cnr.it/webgene/	Human, mouse, rat, fugu, Drosophila, C. elegans, Arabidopsis, Aspergillus
Genie	David Kulp, David Haussler, http://www.cse.ucsc.edu/ ~dkulp/cgi-bin/genie	Human
GeneID	Roderick Guigo, http://www1.imim.es/software/geneid	Human, <i>Drosophila</i>
GeneFinder, Fgenes, Fgenesh	Victor Solovyev, http://genomic.sanger.ac.uk/gf/gf.shtml	Human, Drosophila, C. elegans, Arabidopsis, yeast
HMMgene	Anders Krogh, Anders Pedersen, Søren Brunak, http://www.cbs.dtu.dk/services	Human (vertebrates), C. elegans, Arabidopsis
GeneFinder, MZEF	Michael Zhang, http://www.cshl.org/genefinder	Human, mouse, <i>Arabidopsis</i> , <i>S. pombe</i>
GeneParser	Erik Snyder http://mcdb.colorado.edu/~eesnyder/ GeneParser.html	Available only for download

Análisis de la función de genes por análisis bioinformatico

Genes Homologos son aquellos que comparten un ancestro común, revelado por su similitud a nivel de secuencia.

Dos genes homologos NO tienen exactamente la misma secuencia nucleotidica, debido a que ha sufrido al azar diferentes tipos de mutaciones, pero como los cambios se han producido en la misma secuencia inicial, conservan gran parte de la secuencia inicial

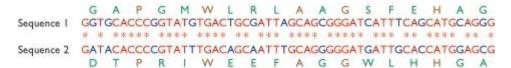
equence I GGTGAGGGTATCATCCCATCTGACTACACCTCATCGGGAGACGGAGCAGT
equence 2 GGTCAGGATATGATTCCATCACACTACACCTTATCCCGAGTCGGAGCAGT

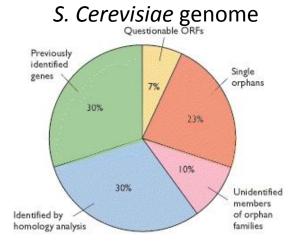


Busqueda de homologos

Esté analisis se basa en el hecho de que un gen que es similar a otro a nivel de secuencia, se deduce que tiene un relación evolutiva y por lo tanto se puede inferir que la función de los dos genes es similar.

La homologia o su falta es más facilmente observable a nivel proteíco





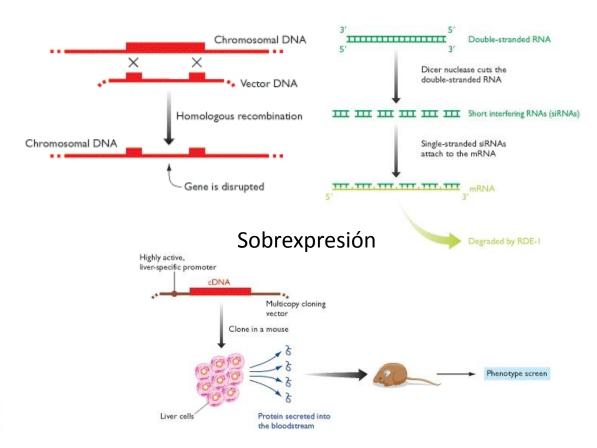
GUTE: Excelencia y Calidad Académica.

Asesores con Posgrado y vocación docente. 56-59-67-97 ext. 304



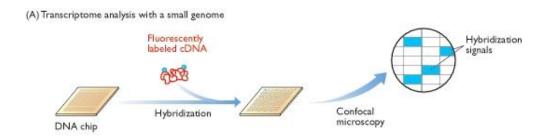
Análisis de la función de genes por análisis experimental

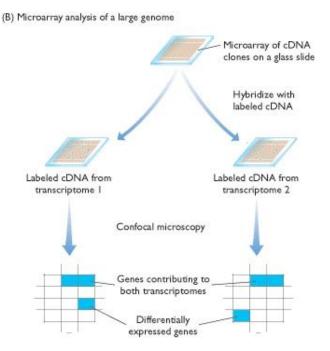
Inactivación





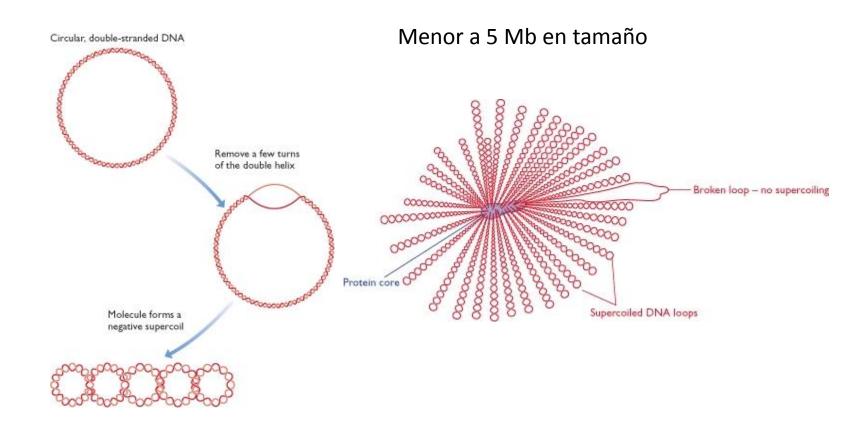
Análisis de la función de genes por análisis experimental





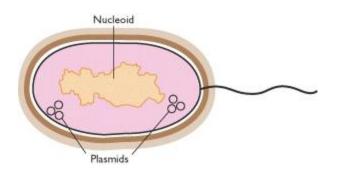


Estructura del genoma procarionte





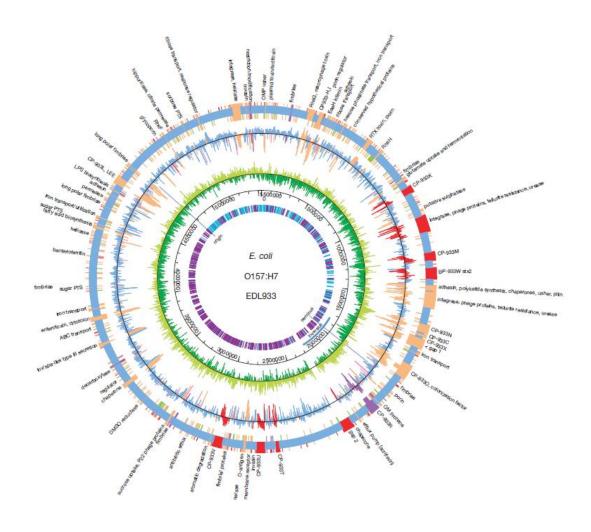
Estructura del genoma procarionte



Type of plasmid	Gene functions	
Resistance	Antibiotic resistance	:
Fertility	Conjugation and DNA transfer between bacteria	
Killer	Synthesis of toxins that kill other bacteria	
Degradative	Enzymes for metabolism of unusual molecules	
Virulence	Pathogenicity	

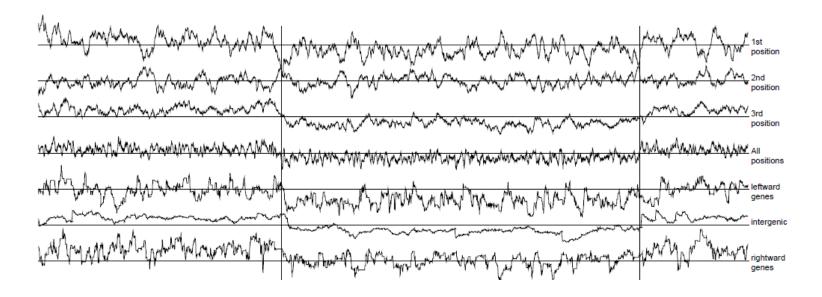


Genoma de *E.coli*





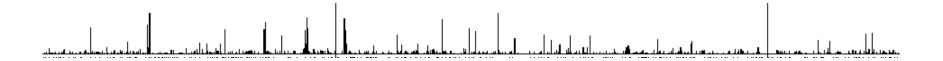
Composición de bases del genoma de E.coli



Grafica del sesgo de GC para cada una de las 3 posiciones presentada de forma separada



Genoma de E.coli



Histograma que muestra la similitud de proteínas de fagos con proteínas que se encuentranen el genoma de *E.coli*



Genoma de E.coli

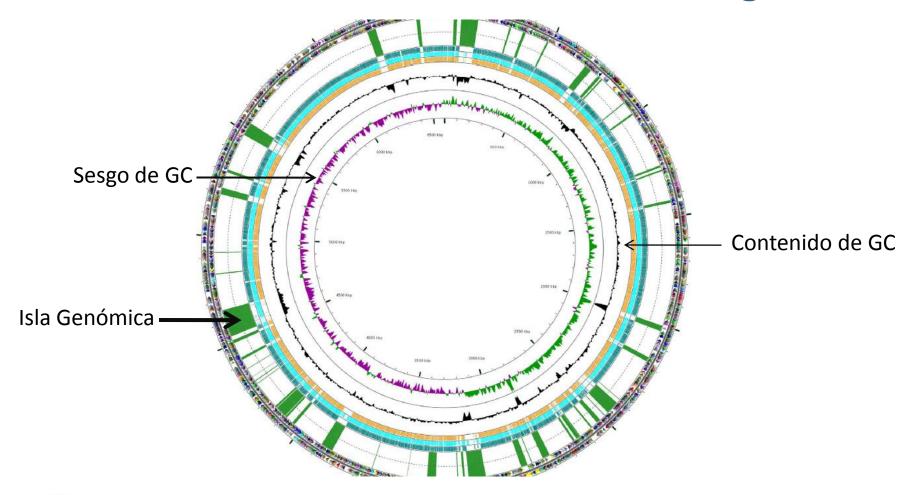
Distribution of E. coli proteins among 22 functional groups (simplified schema).

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

^{*}Total of these rounded values is 99.97%.



Genoma de Pseudomonas aeruginosa



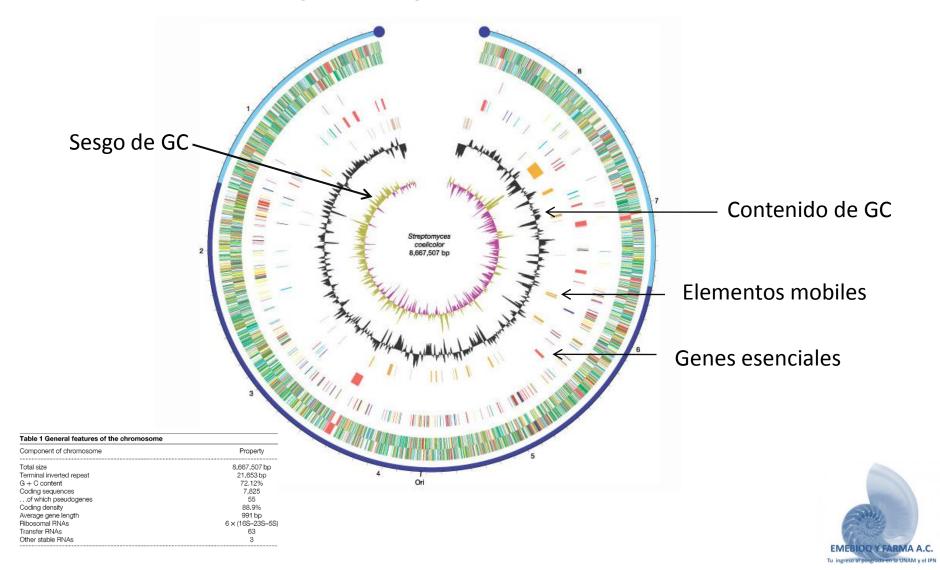


Caracteristicas genomicas de P.aeruginosa

	PA7	PAO1*	PA14*	LESB58*
Conomo Sizo (hp)				
Genome Size (bp)	6,588,339	6,264,404	6, 537, 648	6,601,757
G+C content	66.5	66.6	66.3	66.3
protein coding genes	6286	5566	5892	5925
% coding	89	89	89	88
structural RNAs [†]	75	77	72	81
pseudogenes	8	5	none	34
Assigned function [‡]				
Translation, ribosomal structure and biogenesis	206	205	205	199
Transcription	530	516	537	501
DNA replication, recombination and repair	235	160	185	145
Cell division and chromosome partitioning	37	34	35	34
Posttranslational modification, protein turnover, chaperones	215	200	210	201
Cell envelope biogenesis, outer membrane	260	265	266	261
Cell motility and secretion	152	150	154	149
Inorganic ion transport and metabolism	355	376	377	313
Signal transduction mechanisms	346	337	345	337
Energy production and conversion	336	329	340	330
Carbohydrate transport and metabolism	250	252	249	196
Amino acid transport and metabolism	571	587	590	490
Nucleotide transport and metabolism	105	108	110	104
Coenzyme metabolism	192	191	192	210
Lipid metabolism	245	244	248	234
Secondary metabolites biosynthesis, transport and catabolism	198	205	212	171
General function prediction only	759	756	771	603
Function unknown	503	476	493	500



Caracteristicas genomicas de Streptomyces coelicolor



Caracteristicas genomicas de Streptomyces coelicolor

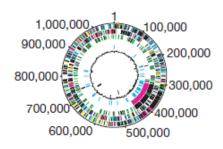
Majority description*	Sco	Mtu	Cdi	Bsu	Mio	Eco
ECF sigma factor (13)	41 (0.52)	10 (0.25)	7 (0.29)	7 (0.17)	16 (0.23)	1 (0.02
Sigma factor (54)	14 (0.17)	3 (0.07)	2 (0.08)	8 (0.19)	3 (0.04)	4 (0.09
Two-component sensor kinase (1.3)	27 (0.34)	8 (0.20)	5 (0.20)	12 (0.29)	41 (0.60)	20 (0.46)
Two-component sensor kinase (1.15)	6 (0.07)	0	0	0	0	0
Two-component regulator (1.6)	50 (0.63)	2 (0.05)	5 (0.20)	9 (0.21)	5 (0.07)	7 (0.16)
Two-component regulator (1.5)	24 (0.30)	11 (0.28)	6 (0.24)	13 (0.31)	21 (0.31)	14 (0.32)
Serine/threonine protein kinase (6.1)	44 (0.56)	13 (0.33)	5 (0.20)	8 (0.19)	14 (0.20)	8 (0.18)
Regulator (Lacl) (2.4)	33 (0.42)	1 (0.02)	2 (0.08)	12 (0.29)	15 (0.22)	13 (0.30)
Regulator (ROK) (36)	23 (0.29)	3 (0.07)	3 (0.12)	3 (0.07)	6 (0.08)	7 (0.16)
Regulator (TetR) (112)	18 (0.22)	1 (0.02)	0	0	0	0
Regulator (KorSA/GntR) (2.9)	10 (0.12)	0	0	0	0	0
Regulator (WhiB-like)	8 (0.10)	4 (0.10)	3 (0.12)	0	0	0
DNA-binding (86)	25 (0.31)	0	0	0	0	0
ABC transport (ATP-binding) (2.1.3)	27 (0.34)	4 (0.10)	8 (0.33)	6 (0.14)	3 (0.04)	3 (0.06)
Transport (permease) (2.3)	36 (0.45)	4 (0.10)	1 (0.04)	8 (0.19)	25 (0.37)	3 (0.06)
Transport (sugar) (2.2)	36 (0.45)	4 (0.10)	1 (0.04)	8 (0.19)	26 (0.38)	3 (0.06)
Transport (substrate-binding) (1.8)	35 (0.44)	4 (0.10)	1 (0.04)	6 (0.14)	24 (0.35)	3 (0.06
Integral membrane (59)	14 (0.17)	14 (0.35)	3 (0.12)	2 (0.04)	0	0
Membrane ATPase (42)	13 (0.16)	12 (0.30)	6 (0.24)	4 (0.09)	3 (0.04)	1 (0.02)
Secreted hydrolase (7)	100 (1.27)	19 (0.48)	8 (0.33)	21 (0.51)	8 (0.11)	9 (0.20)
Secreted hydrolase (7.3)	17 (0.21)	0	0	0	1 (0.01)	0
Secreted chitinase (7.8)	5 (0.06)	Ö	0	0	0	0
Secreted protease (191)	10 (0.12)	3 (0.07)	0	0	0	0
Secreted protesse (7.6)	8 (0.10)	0	0	0	0	0
Secreted cellulase (7.4)	7 (0.08)	1 (0.02)	0	1 (0.02)	0	0
Secreted hypothetical (17)	25 (0.31)	11 (0.28)	8 (0.33)	13 (0.31)	8 (0.11)	9 (0.20)
Hypothetical (63)	25 (0.31)	0	0	6 (0.14)	0	0
Conservon (Ovn1-4; 178, 177, 214, 180)	13 (0.16)	1 (0.02)	0	0	0	0
Hypothetical (204)	13 (0.16)	0	0	0	0	0
Hypothetical (19)	12 (0.15)	0	Ö	0	0	0

Numbers indicate absolute number of proteins from each genome in each family and the percentage of the total proteins in that genome in parentheses. Family number is indicated in parentheses in the monthly description column. The hierarchical numbering system reflects use of higher BlastP thresholds to be break large complex families into discrete subfamilies. Complete data are available from http://www.sanger.ac.uk/Projects/s_Coelicolor/. Soc. Scoelicolory, Mtu. M. tuberculosis; Col. C. diphtheriae; Bsu. B. subtlis; Mto. M. tot; Eco. E. coli.

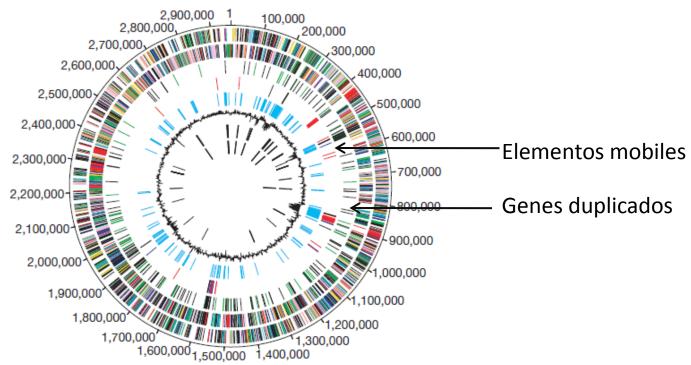
[&]quot;Groupings are based on sequence similarity, so individual families do not necessarily include all representatives of each type of protein in each genome [see Methods].



Caracteristicas genomicas de Vibrio cholerae



	Chromosome 1	Chromosome 2
Size (bp)	2,961,151	1,072,914
Total number of sequences	36,797	14,367
G+C percentage	47.7	46.9
Total number of ORFs	2,770	1,115
ORF size (bp)	952	918
Percentage coding	88.6	86.3
Number of rRNA operons (16S-23S-5S)	8	0
Number of tRNA	94	4
Number similar to known proteins	1,614 (58%)	465 (42%)
Number similar to proteins of unknown function*	163 (6%)	66 (6%)
Number of conserved hypothetical proteins†	478 (17%)	165 (15%)
Number of hypothetical proteins‡	515 (19%)	419 (38%)
Number of Rho-independent terminators	599	193





Caracteristicas genomicas de *Deinococcus* radiodurans

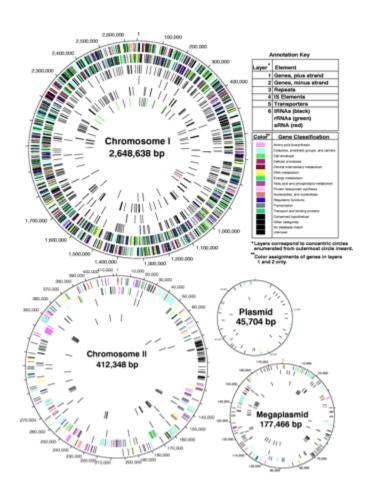
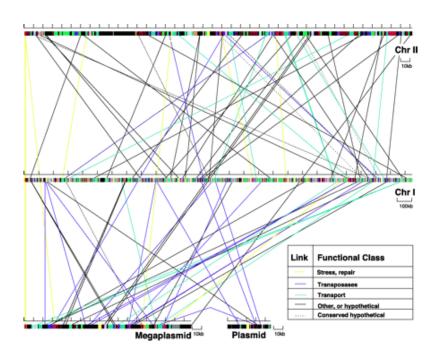


Table 1. General features of the D. radiodurans genome.					
Molecule	Length	Average ORF length (bp)	Protein coding regions	GC content	Repeatcontent
Chromosome I		913	90.8%	67.0%	1.8%
Chromosome II	412,348	1,044	93.5%	66.7%	1.4%
Megaplasmid	177,466	1,100	90.4%	63.2%	9.2%
Plasmid	45,704	928	80.9%	56.1%	13.0%
All	3,284,156	937	90.9%	66.6%	3.8%



Caracteristicas genomicas de Deinococcus radiodurans



Repeat elements of the	Repeat elements of the <i>D. radiodurans</i> genome.			
Repeat	Length	Copies		
SRE	160	84		
SMR1	139	41		
SMR2	114	92		
SMR4	147	7		
SMR5	215	38		
SMR7	140	18		
SMR8	131	24		
SMR9	105	6		
SMR10	60	6		
Total		316		



Caracteristicas genomicas de Borrelia burgdorferi

910,725 bp (28.6% G+C)

Coding sequences (93%) RNAs (0.7%) Intergenic sequence (6.3%)

853 coding sequences

500 (59%) with identified database match 104 (12%) match hypothetical proteins 249 (29%) with no database match

Plasmids

Chromosome

cp9	9,386 bp (23.6% GC
cp26	26,497 bp (26.3% GC
lp17	16,828 bp (23.1% GC
lp25	24,182 bp (23.3% GC
lp28-1	26,926 bp (32.3% GC
lp28-2	29,771 bp (31.5% GC
lp28-3	28,605 bp (25.1% GC
lp28-4	27,329 bp (24.4% GC
lp36	36,834 bp (26.8% GC
lp38	38,853 bp (26.1% GC
lp54	53,590 bp (28.1% GC

Coding sequences (71%) Intergenic sequence (29%)

430 coding sequences

70 (16%) with identified database match 110 (26%) match hypothetical proteins 250 (58%) with no database match

Ribosomal RNA Chromosome coordinates 16S 444581-446118 23S 438590-441508 5S 438446-438557

 23\$
 438590-441508

 5\$
 438446-438557

 23\$
 435334-438267

 5\$
 435201-435312

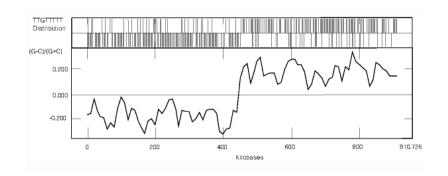
Stable RNA

tmRNA 46973-47335 mpB 750816-751175

Transfer RNA

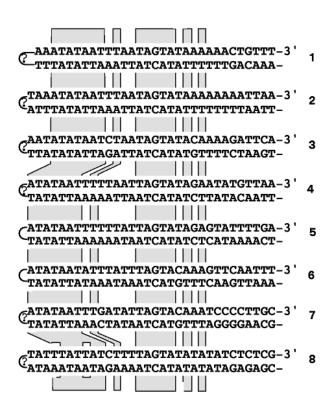
34 species (8 clusters, 14 single genes)

Distribución del octamero TTGTTTTT en el cromosoma de *B. burgdorferi*





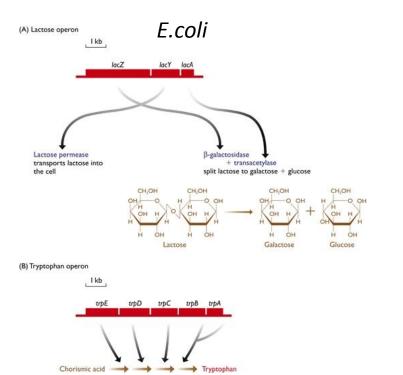
Caracteristicas genomicas de *Borrelia* burgdorferi



- 1- Termino del lado izquierdo del Cromosoma Sh-2-82
- 2- Termino del lado izquierdo del Cromosoma B31
- 3 -B.afzelii Termino del lado derecho del Cromosoma R-IP3; 4, B.
- 4- Termino del lado derecho del Cromosoma B31
- 5- Termino del lado izquierdo del plasmido Ip17
- 6- Termino del lado derecho del plasmido Ip17
- 7- *B. hermisii* Terminos del lado derecho de los plasmidos bp7E y pb21E
- 8-Termino del lado derecho del plasmido Ip28-1



Genoma procarionte

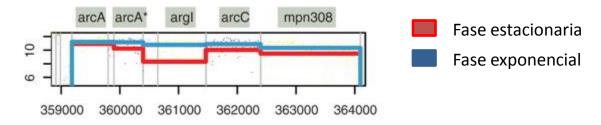


Aquifex aeolicus



gatC, glutamyl-tRNA aminotransferasa subunidad C recA proteina de recombinacion RecA pilU proteina de la movilidad por twitching cmk citidilato kinasa, requiere para la sintesis de nucleotidos de citidiana pgsA fosfatidilglicerofosfato sintasa recJ Endonucleasa de cadena

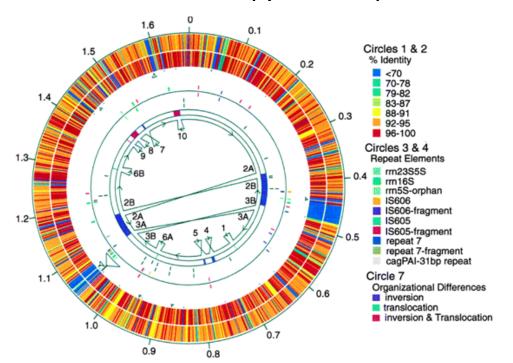
Mycoplasma pneumoniae





"Concepto Especie" en procariontes

Helicobacter pylori 2 cepas secuenciadas



Tamaño genomas 1.67 Mb y1 .64 Mb Numero de genes 1552 y 1495 genes

1406 genes estaban presentes en ambas cepas

Entre el 6–7% del genes de cada cepa es unico a ella



"Concepto Especie" en procariontes

Escherichia coli 2 cepas secuenciadas

Cepa K-12 O157:H7

Tamaño 4.64 Mb 5.53 Mb

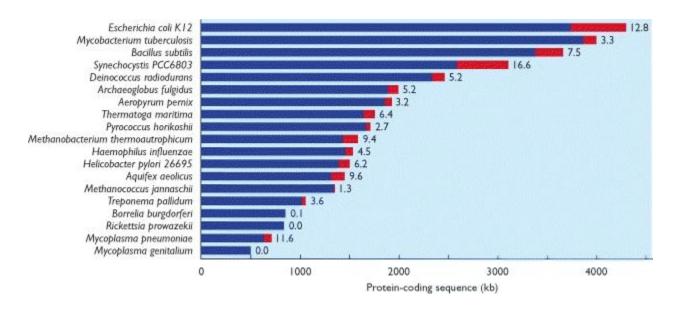
genoma

Genes 1387 528

especificos



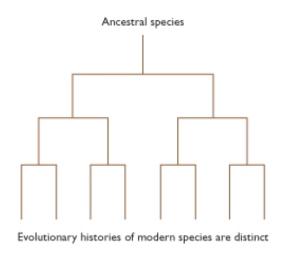
"Concepto Especie" en procariontes

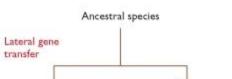


En azul el DNA unico a una especie en rojo el DNA acquirido por transferencia horizontal

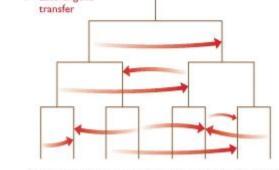


"Concepto Especie" en procariontes



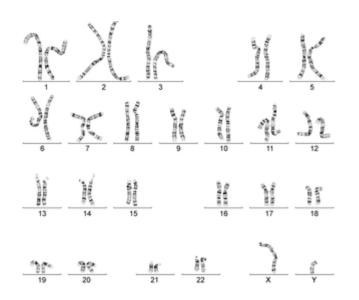


(B) Lateral gene transfer occurs between species



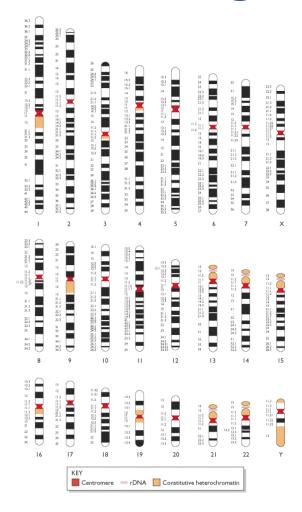




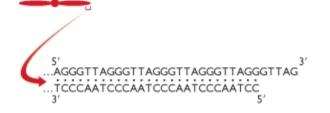


Species	Genome size (Mb)	
Fungi		
Saccharomyces cerevisiae	12.1	
Aspergillus nidulans	25.4	
Protozoa		
Tetrahymena pyriformis	190	
Invertebrates		
Caenorhabditis elegans	97	
Drosophila melanogaster	180	
Bombyx mori (silkworm)	490	
Strongylocentrotus purpuratus (sea urchin)	845	
Locusta migratoria (locust)	5000	
Vertebrates		
Takifugu rubripes (pufferfish)	400	
Homo sapiens	3200	
Mus musculus (mouse)	3300	
Plants		
Arabidopsis thaliana (vetch)	125	
Oryza sativa (rice)	430	
Zea mays (maize)	2500	
Pisum sativum (pea)	4800	
Triticum aestivum (wheat)	16 000	
Fritillaria assyriaca (fritillary)	120 000	



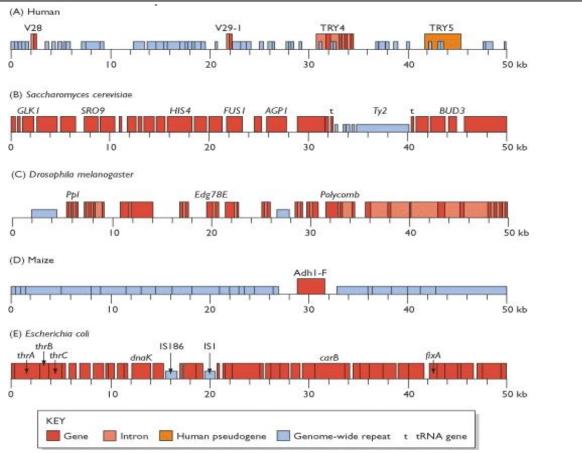






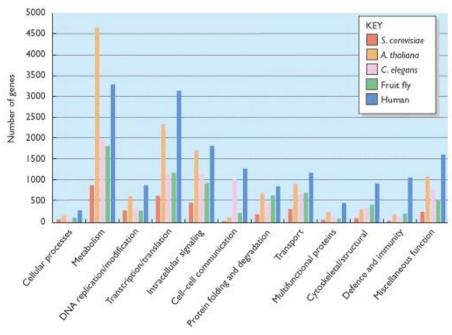


Feature	Yeast	Fruit fly	Human
Gene density (average number per Mb)	479	76	11
Introns per gene (average)	0.04	3	9
Amount of the genome that is taken up by genome-wide repeats	3.4%	12%	44%





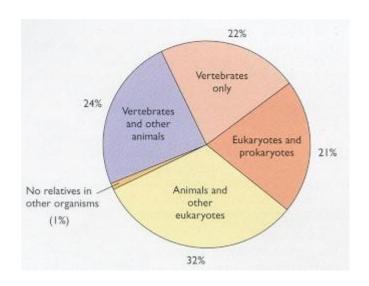
Genes presentes en el genoma eucarionte



		Number of genes in the genome containing the domain				
Domain	Function	Human	Fruit fly	Caenorhabditis	Arabidopsis	Yeast
Zinc finger, C ₂ H ₂ type	DNA binding	564	234	68	21	34
Zinc finger, GATA type	DNA binding	11	5	8	26	9
Homeobox	Gene regulation during development	160	100	82	66	6
Death	Programmed cell death	16	5	7	0	0
Connexin	Electrical coupling between cells	14	0	0	0	0
Ephrin	Nerve cell growth	7	2	4	0	0



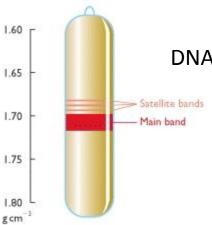
Genoma Humano ¿Que tiene de humano?





DNA repetitivo

DNA repetido en tandem tambien se conoce como DNA satelite.



DNA satelite puede contener grandes series de repetidos en tandem hasta de cientos de Kb de largo.

Un genoma puede tener diferentes tipos de DNA satelite, En donde su unidad repetitiva puede variar entre 5 y 200 bp

La mayoria se encuentra en los centromeros



Minisatelites

Conglomerados de DNA de hasta 20 Kb de largo en donde la unidad repetitiva tiene hasta 25 bp

El DNA telomerico es un ej. de DNA minisatelite ya que se tienen cientos de copias de la secuencia 5'-TTAGGG-3'



Microsatelites

Conglomerados de DNA de hasta 150 bp en donde la unidad repetitiva tiene 13 bp o menos.

Tipico microsatelite consiste de una unidad de 1,2,3 o 4 bp repetida entre 10 y 20 veces. Como los microsatelites que se encuentran en el locus del receptor β de las celulas T que tiene las siguientes secuencia:

5-'CACACACACAC-3'

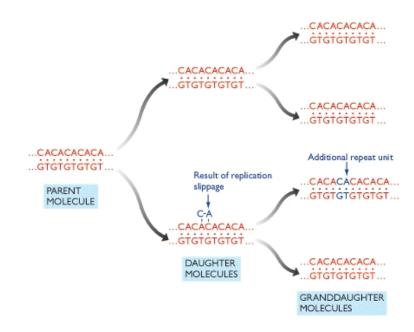
Llegan a constituir hasta 0.8% del genoma humano i.e. 8MB





DNA repetitivo ¿como se produce?

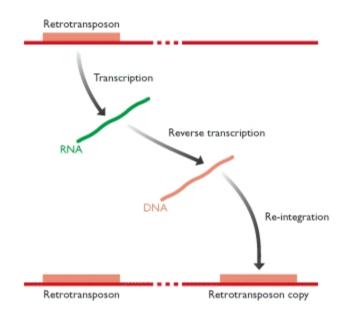
Por expansion de una secuencia progenitora, ya sea por error en la replicación (slippage) o mediante recombianción y dan como resultado repetidos adyacentes.





Repeticiones esparcidas

Una secuencia que se repite en muchas posiciones dispersas dentro de un genoma.





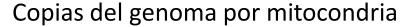
Tamaño de los genomas mitocondriales

Species	Type of organism	Genome size (kb)
Mitochondrial genomes		
Plasmodium falciparum	Protozoan (malaria parasite)	6
Chlamydomonas reinhardtii	Green alga	16
Mus musculus	Vertebrate (mouse)	16
Homo sapiens	Vertebrate (human)	17
Metridium senile	Invertebrate (sea anemone)	17
Drosophila melanogaster	Invertebrate (fruit fly)	19
Chondrus crispus	Red alga	26
Aspergillus nidulans	Ascomycete fungus	33
Reclinomonas americana	Protozoa	69
Saccharomyces cerevisiae	Yeast	75
Suillus grisellus	Basidiomycete fungus	121
Brassica oleracea	Flowering plant (cabbage)	160
Arabidopsis thaliana	Flowering plant (vetch)	367
Zea mays	Flowering plant (maize)	570
Cucumis melo	Flowering plant (melon)	2500



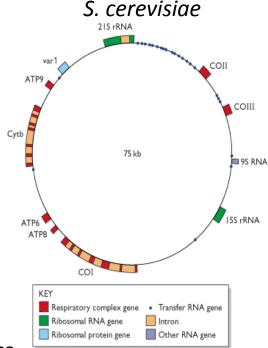
Genomas mitocondriales

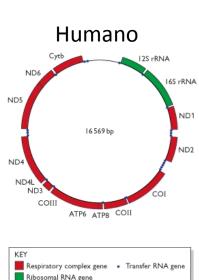
Generalmente son circulares aunuqe algunos eucariontes microscopicos tienen genomas lineales e.g. Paramecium, Chlamydomonas





S. cerevisiae: 100







GUTE: Excelencia y Calidad Académica.

Asesores con Posgrado y vocación docente. 56-59-67-97 ext. 304

Genomas mitocondriales

Feature	Plasmodium falciparum	Chlamydomonas reinhardtii	Homo sapiens	Saccharomyces cerevisiae	Arabidopsis thaliana	Reclinomonas americana
<u>Total number of</u> genes	5	12	37	35	52	92
Types of genes	***************************************				***************************************	
<u>Protein</u> -coding genes	5	7	13	8	27	62
Respiratory complex	0	7	13	7	17	24
Ribosomal proteins	0	0	0	1	7	27
Transport proteins	0	0	0	0	3	6
RNA polymerase	0	0	0	0	0	4
Translation factor	0	0	0	0	0	1
Non-coding RNA genes	0	5	24	26	25	30
Ribosomal RNA genes	0	2	2	2	3	3
Transfer RNA genes	0	3	22	24	22	26
Other RNA genes	0	0	0	1	0	1
Number of introns	0	1	0	8	23	1
Genome size (kb)	6	16	17	75	367	69

Genomas mitocondriales

Tamaño de los genomas mitocondriales NO se relaciona con la complejidad del organismo que los contiene.

Animales multicelulares tienen genomas pequeños con organización genética compacta (los genes tienen poco espacio entre ellos)

Mientras que los eucariontes inferiores (levaduras) y algunas plantas con flores Tiene genomas más grandes y menos compactos.

Todos los genomas contienen genes para rRNAs y algunos componentes de la cadena respiratoria



Genomas en los cloroplastos

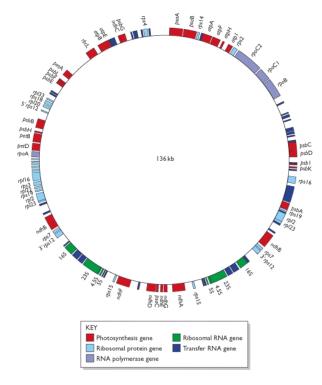
Tamaño MAS homogeneo que el de los genomas mitocondriales

Pisum sativum	Flowering plant (pea)	120
Marchantia polymorpha	Liverwort	121
Oryza sativa	Flowering plant (rice)	136
Nicotiana tabacum	Flowering plant (tobacco)	156
Chlamydomonas reinhardt	ii Green alga	195

Es un genoma circular, aunque puede haber DNA circular que contiene Solo algunos de los elementos del genoma



Genomas en los cloroplastos



Los genes que contienen codifican para rRNAs, tRNAs, proteínas ribosomales y proteínas que participan en la fotosintesis.

