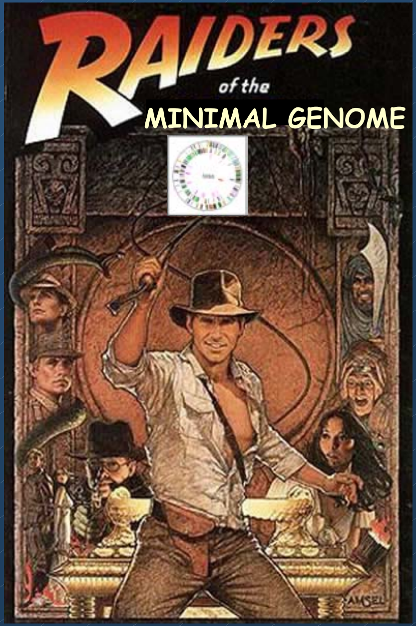
VNIVERSITAT DE VALÈNCIA

Institut Cavanilles de Biodiversitat i Biologia Evolutiva

En busca del
GENOMA MÍNIMO



Dra. Rosario Gil
RVGP, octubre 2004

El concepto de genoma mínimo

Genoma mínimo (Koonin, 2000):

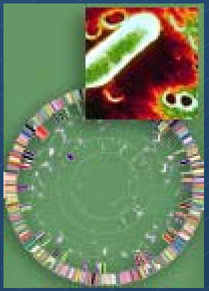
grupo más pequeño de genes que debe ser suficiente para mantener en funcionamiento una forma de vida en las condiciones más favorables imaginables

- en presencia de todos los nutrientes esenciales
- en ausencia de estrés ambiental

Funciones esenciales:

- mantenimiento estructuras celulares
- reproducción
- evolución

El poder de la genómica comparativa



- **Mushegian y Koonin, 1996**
  - ➢ *Haemophilus influenzae* (G-, 1703 genes codificantes de proteínas)
    Mycoplasma genitalium (G+, 470 genes codificantes de proteínas)
  - ➢ **Conjunto mínimo de genes**: 256 genes

Pero...

- **Hutchison *et al.*, 1999**

  Mutagénesis global (Tn) sobre *M. genitalium* y *M. pneumoniae*

  - ➢ Muchos genes presentes en el “conjunto mínimo” **pueden ser interrumpidos**

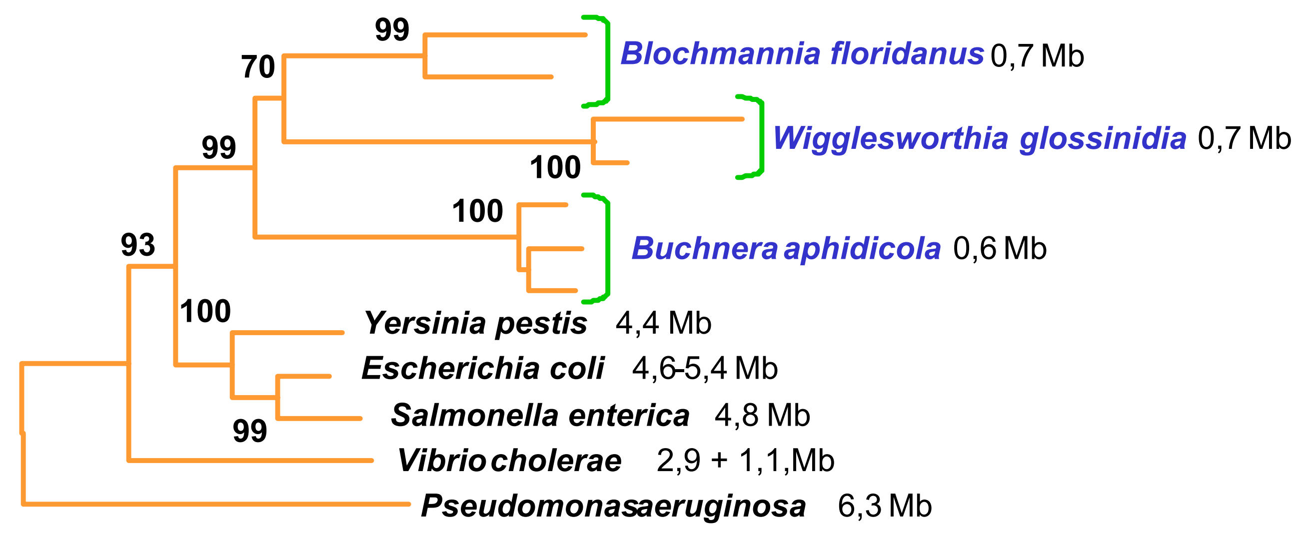
**Si aumentamos el número de genomas utilizados en la comparación
podemos disminuir significativamente
el número de genes considerados como esenciales**

Los reducidos genomas de los endosimbiontes de insectos

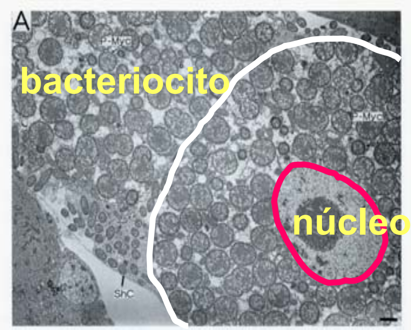

hormiga carpintero

mosca tse-tse


pulgón

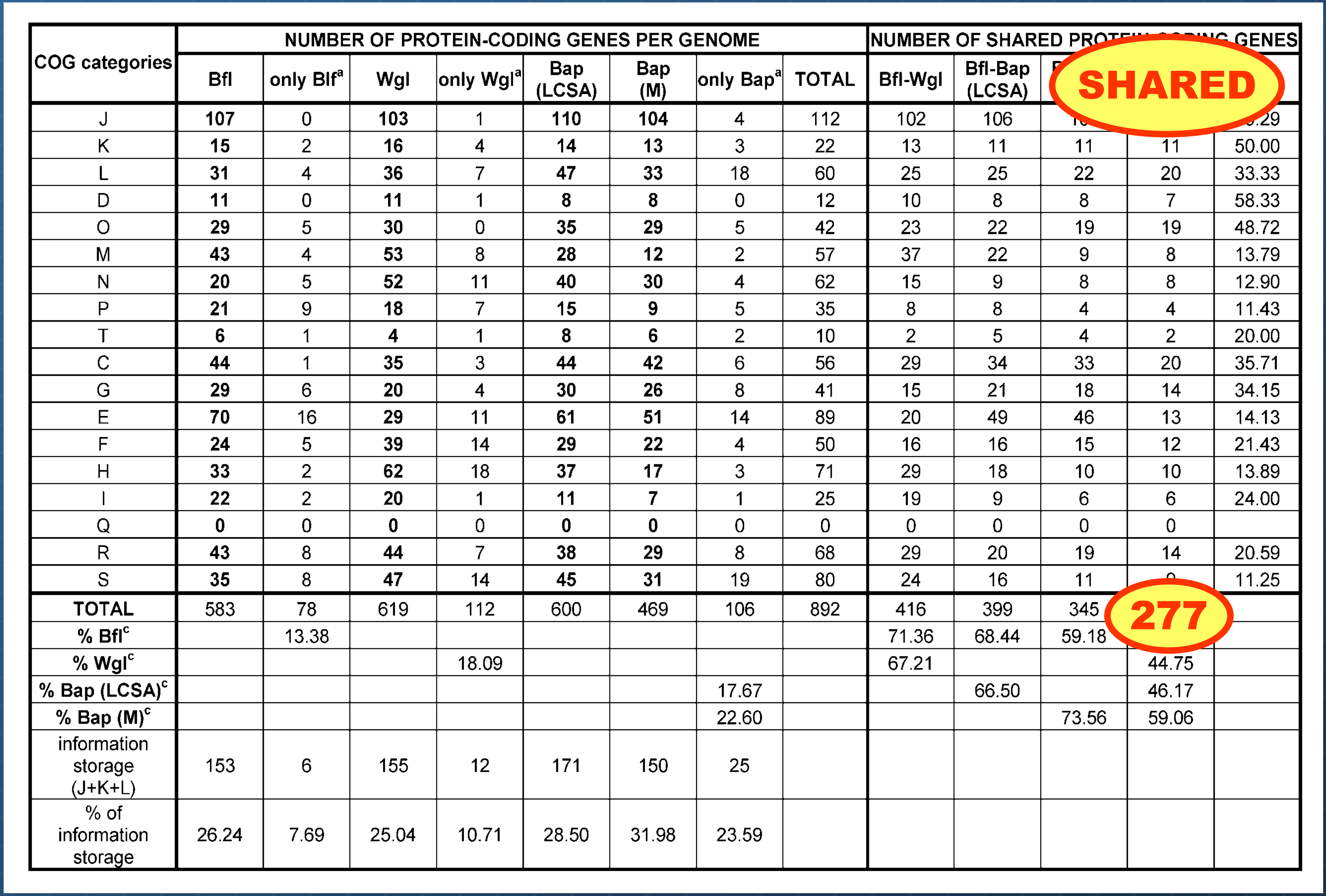


Blochmannia floridanus 0,7 Mb
Wigglesworthia glossinidia 0,7 Mb
Buchnera aphidicola 0,6 Mb
Yersinia pestis 4,4 Mb
Escherichia coli 4,6-5,4 Mb
Salmonella enterica 4,8 Mb
Vibrio cholerae 2,9 + 1,1,Mb
Pseudomonas aeruginosa 6,3 Mb



Análisis funcional comparativo

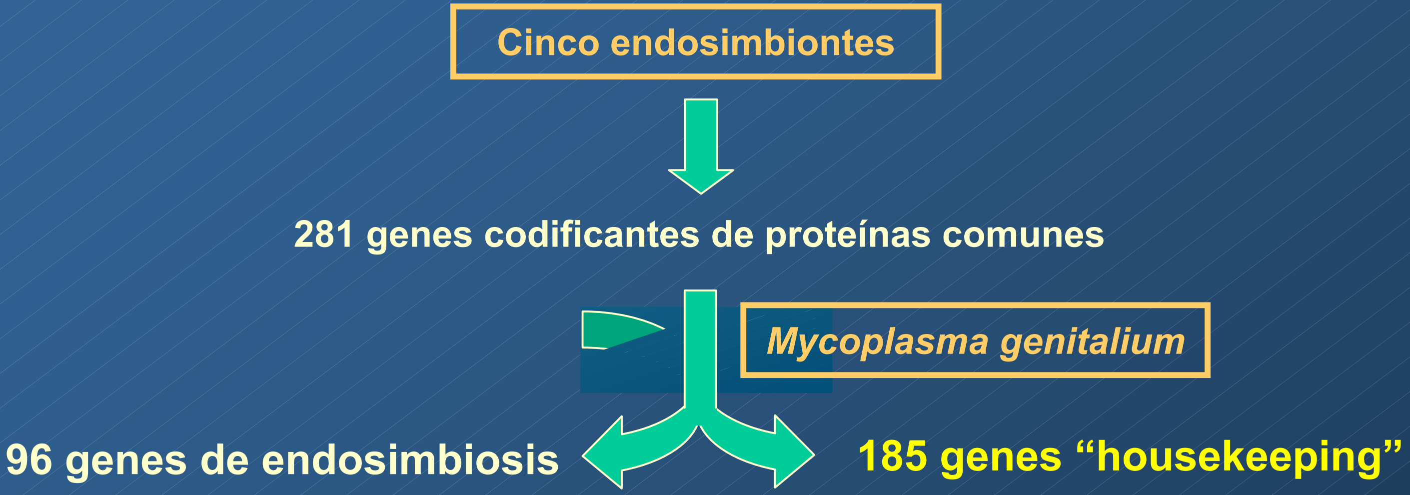
- Basado en las categorías COG



COG categories	NUMBER OF PROTEIN-CODING GENES PER GENOME								NUMBER OF SHARED PROTEIN-CODING GENES				
	Bfl	only Blf[a]	Wgl	only Wgl[a]	Bap (LCSA)	Bap (M)	only Bap[a]	TOTAL	Bfl-Wgl	Bfl-Bap (LCSA)	[illegible]	[illegible]	[illegible]
J	107	0	103	1	110	104	4	112	102	106	[illegible]	[illegible]	[illegible]
K	15	2	16	4	14	13	3	22	13	11	11	11	50.00
L	31	4	36	7	47	33	18	60	25	25	22	20	33.33
D	11	0	11	1	8	8	0	12	10	8	8	7	58.33
O	29	5	30	0	35	29	5	42	23	22	19	19	48.72
M	43	4	53	8	28	12	2	57	37	22	9	8	13.79
N	20	5	52	11	40	30	4	62	15	9	8	8	12.90
P	21	9	18	7	15	9	5	35	8	8	4	4	11.43
T	6	1	4	1	8	6	2	10	2	5	4	2	20.00
C	44	1	35	3	44	42	6	56	29	34	33	20	35.71
G	29	6	20	4	30	26	8	41	15	21	18	14	34.15
E	70	16	29	11	61	51	14	89	20	49	46	13	14.13
F	24	5	39	14	29	22	4	50	16	16	15	12	21.43
H	33	2	62	18	37	17	3	71	29	18	10	10	13.89
I	22	2	20	1	11	7	1	25	19	9	6	6	24.00
Q	0	0	0	0	0	0	0	0	0	0	0	0	
R	43	8	44	7	38	29	8	68	29	20	19	14	20.59
S	35	8	47	14	45	31	19	80	24	16	11	[illegible]	11.25
TOTAL	583	78	619	112	600	469	106	892	416	399	345	277	
% Bfl[c]		13.38							71.36	68.44	59.18	[illegible]	
% Wgl[c]				18.09					67.21			44.75	
% Bap (LCSA)[c]							17.67			66.50		46.17	
% Bap (M)[c]							22.60				73.56	59.06	
information storage (J+K+L)	153	6	155	12	171	150	25						
% of information storage	26.24	7.69	25.04	10.71	28.50	31.98	23.59						

SHARED

Comparación con *M. genitalium*



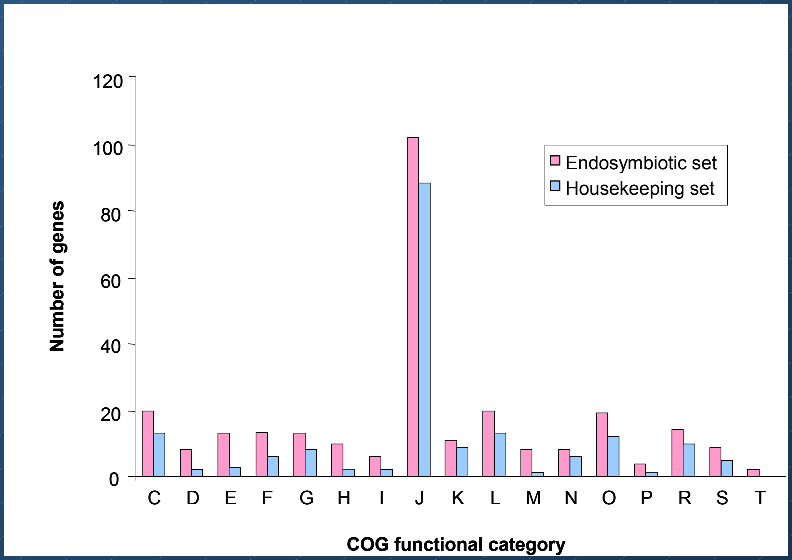
Cinco endosimbiontes

281 genes codificantes de proteínas comunes

Mycoplasma genitalium

96 genes de endosimbiosis

185 genes “housekeeping”



Las aproximaciones experimentales

- Tres formas de identificar genes esenciales bajo unas determinadas condiciones de cultivo:
    - Mutagénesis masiva con transposones
        - genes no esenciales que reducen el crecimiento
        - genes esenciales que toleran inserciones
    - RNA antisentido
        - sólo válido si hay una expresión adecuada del RNA antisentido
    - Inactivación sistemática de cada gen individual
        - genes esenciales redundantes
        - genes no dispensables simultaneamente

¿Y si las juntamos?

1. Nuestra comparación de genomas reducidos

∗ además el genoma de *Phytoplasma asteris*

2. Datos experimentales

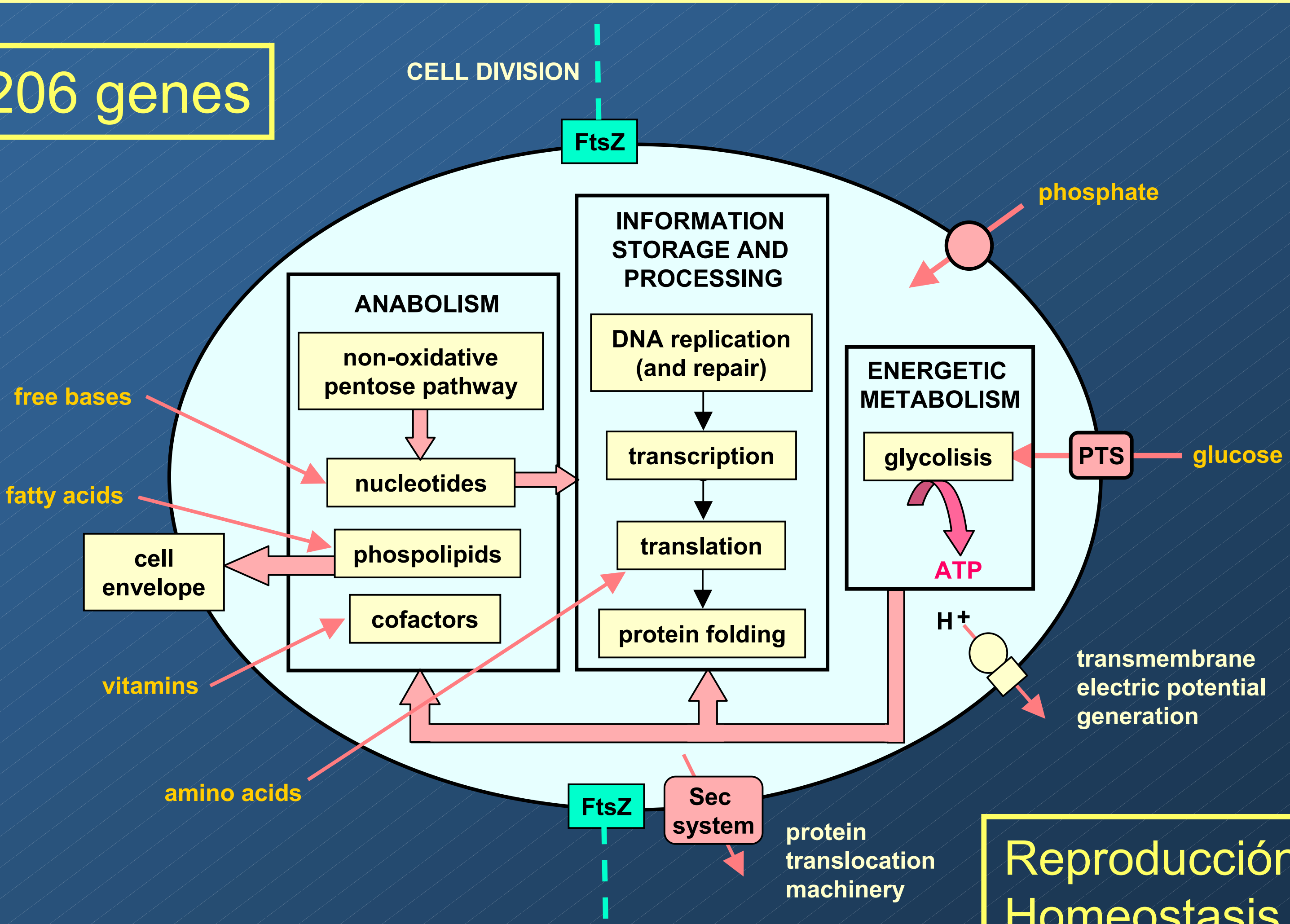
a) *M. genitalium* y *M. pneumoniae* (mutagénesis masiva, Hutchison III *et al*., 1999)
b) *Bacillus subtilis* (inactivación sistemática, Kobayashi *et al*., 2003)
b) *Escherichia coli* (mutagénesis masiva, Gerdes *et al*., 2003)
c) *Staphylococcus aureus* (RNA antisentido, Forsyth *et al*., 2002)

3. Revisar el sistema metabólico

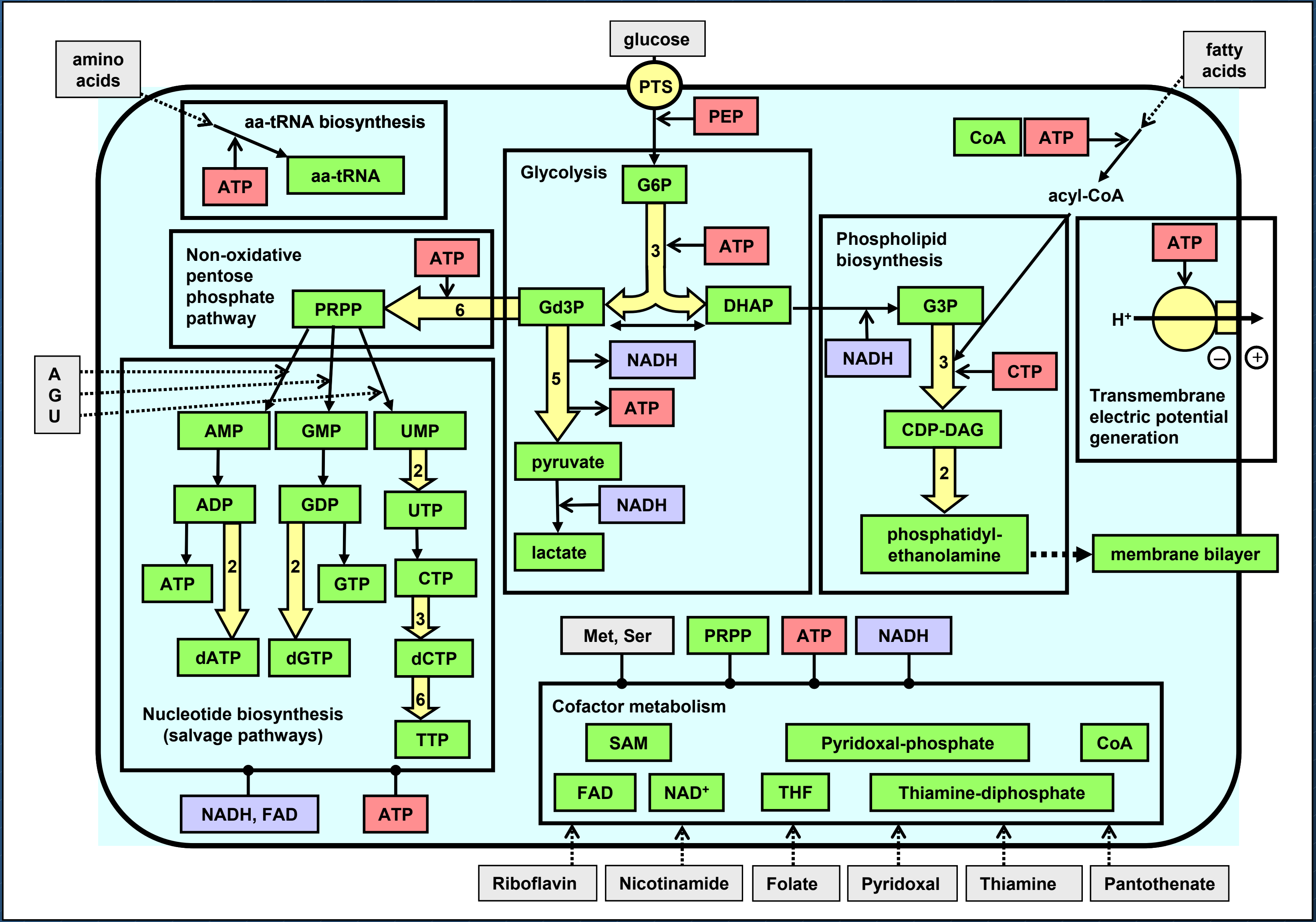
∗ Rutas alternativas
∗ Rutas incompletas
∗ Homeostasis metabólica

“Nuestro” genoma mínimo

206 genes



Reproducción
Homeostasis metabólica
Evolución



glucose

PTS

PEP

amino acids

aa-tRNA biosynthesis

ATP

aa-tRNA

Glycolysis

G6P

3

ATP

Gd3P

DHAP

5

NADH

ATP

pyruvate

NADH

lactate

Non-oxidative pentose phosphate pathway

ATP

PRPP

6

A
G
U

AMP

GMP

UMP

ADP

GDP

2

UTP

ATP

2

2

GTP

CTP

3

dATP

dGTP

dCTP

6

TTP

Nucleotide biosynthesis (salvage pathways)

NADH, FAD

ATP

fatty acids

CoA

ATP

acyl-CoA

Phospholipid biosynthesis

G3P

NADH

3

CTP

CDP-DAG

2

phosphatidyl-ethanolamine

membrane bilayer

ATP

H+

⊖

⊕

Transmembrane electric potential generation

Met, Ser

PRPP

ATP

NADH

Cofactor metabolism

SAM

Pyridoxal-phosphate

CoA

FAD

NAD+

THF

Thiamine-diphosphate

Riboflavin

Nicotinamide

Folate

Pyridoxal

Thiamine

Pantothenate

Microbiology and Molecular Biology Reviews, Sept. 2004, p. 00 Vol. 68, No. 3
1092-2172/04/$08.00+0 DOI: 10.1128/MMBR.68. 000–000.2004

Determination of the Core of a Minimal Bacterial Gene Set†

Rosario Gil,[1,2*] Francisco J. Silva,[1,2] Juli Peretó,[1,3] and Andrés Moya[1,2]

Institut Cavanilles de Biodiversitat i Biologia Evolutiva Universitat de València, Valencia,[1] and
Departament de Genètica[2] and Departament de Bioquímica i Biologia Molecular,[3]
Universitat de València, Burjassot (València), Spain

INTRODUCTION

Complete genome sequences are becoming available for a large number of diverse bacterial species. Comparative genomics shows that most bacterial proteins are highly conserved in evolution, allowing predictions to be made about the functions of most products of uncharacterized genomes based on model organisms, such as *Escherichia coli* and *Bacillus subtilis* (gram-positive and gram-negative bacteria, respectively), for which abundant high-quality genetic and biochemical information has been obtained in the past.

One important question raised by the availability of complete genomic sequences is how many genes are essential for cellular life. Although bacterial genomes differ vastly in their sizes and gene repertoires, no matter how small, they must contain all the information to allow the cell to perform many essential (housekeeping) functions that give the cell the ability to maintain metabolic homeostasis, reproduce, and evolve, the three main properties of living cells (53). Cells usually can import metabolites but not functional proteins; therefore, they have to rely on their own gene products to perform such essential functions.

The determination of the minimal set of protein-coding genes necessary to maintain a living cell is becoming an increasingly appealing issue, considering that such minimal gene set should include "the smallest possible group of genes that would be sufficient to sustain a functioning cellular life form under the most favorable conditions imaginable, that is, in the presence of a full complement of essential nutrients and in the absence of environmental stress" (48). Reconstruction of the minimal gene set can take advantage of the increasing knowledge of completely characterized genomes. In recent years, several research groups have tried to define the essential set of survival protein-encoding genes in bacteria by different experimental and computational methods (reviewed in references 22, 24, 38, and 79). Three different experimental approaches have been used to identify genes that are essential under particular growth conditions: massive transposon mutagenesis strategies (the most widely used approach), the use of antisense RNA to inhibit gene expression (22, 37), and the systematic inactivation of each individual gene present in a genome (31, 45, 64; http://www.genome.wisc.edu/functional/tnmutagenesis.htm). However, all these approaches have limitations. Transposon mutagenesis might overestimate the set

---

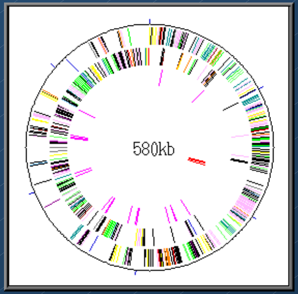
[*] Corresponding author. Mailing address: Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartat Oficial 2085, 46071 València, Spain. Phone: 34 96 354 36 29. Fax: 34 96 354 36 70. E-mail: rosario.gil@uv.es.

† Supplemental material for this article may be found at http://mmbr.asm.org.

Mycoplasma genitalium, “el pequeño de la casa”

- Fraser *et al*., 1995

  secuenciación del genoma completo de *M. genitalium*
  (el más pequeño hasta el momento)



  - 580 Kb
  - sólo 470 genes codificantes de proteínas

- Modelo para conocer el mínimo número de genes y proteínas necesarios para mantener una forma de vida ¿libre?

  Primer paso esencial para “crear” microorganismos para una ámplia variedad de aplicaciones

Este trabajo ha sido posible gracias a ...

VNIVERSITAT DE VALÈNCIA
Institut Cavanilles de Biodiversitat i Biologia Evolutiva

GRUPO DE GENÉTICA EVOLUTIVA

http://www.uv.es/cavanilles/genevol

Rosario Gil
Francisco J. Silva
Juli Peretó
François Delmotte
Fernando González-Candelas
Amparo Latorre
Andrés Moya