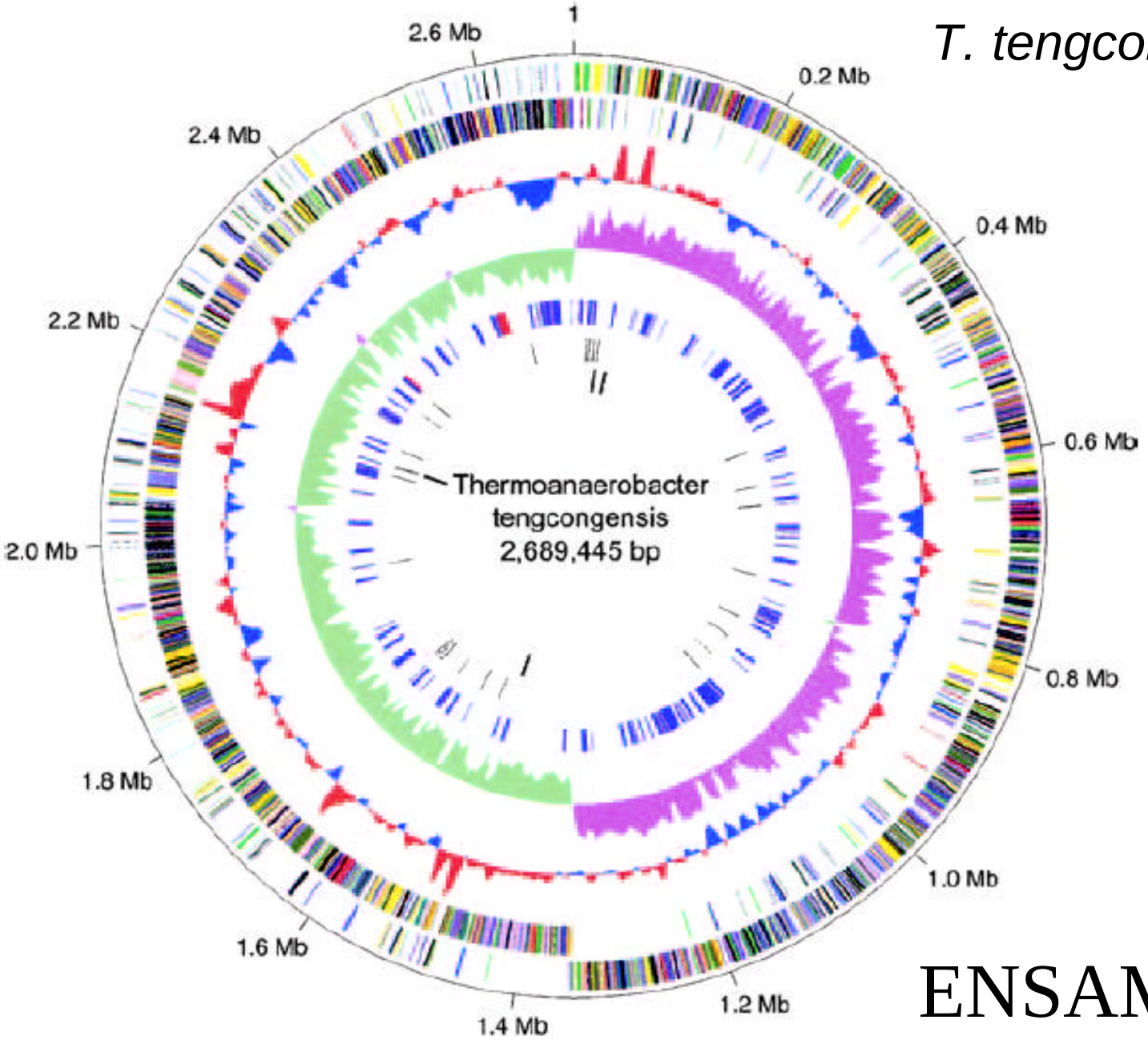
T. tengcongensis



ENSAMBLE

Aislar el material genético de un organismo

1500-3500 Kb

Trozar el DNA (Shot-gun)

1-5Kb

Cloning

genoteca

1.000-100.000 clones

Secuenciación automática

5000 corridas

Lectura de secuencia nucleotídica

1000 pb corridas

5000 kb

ensamblaje

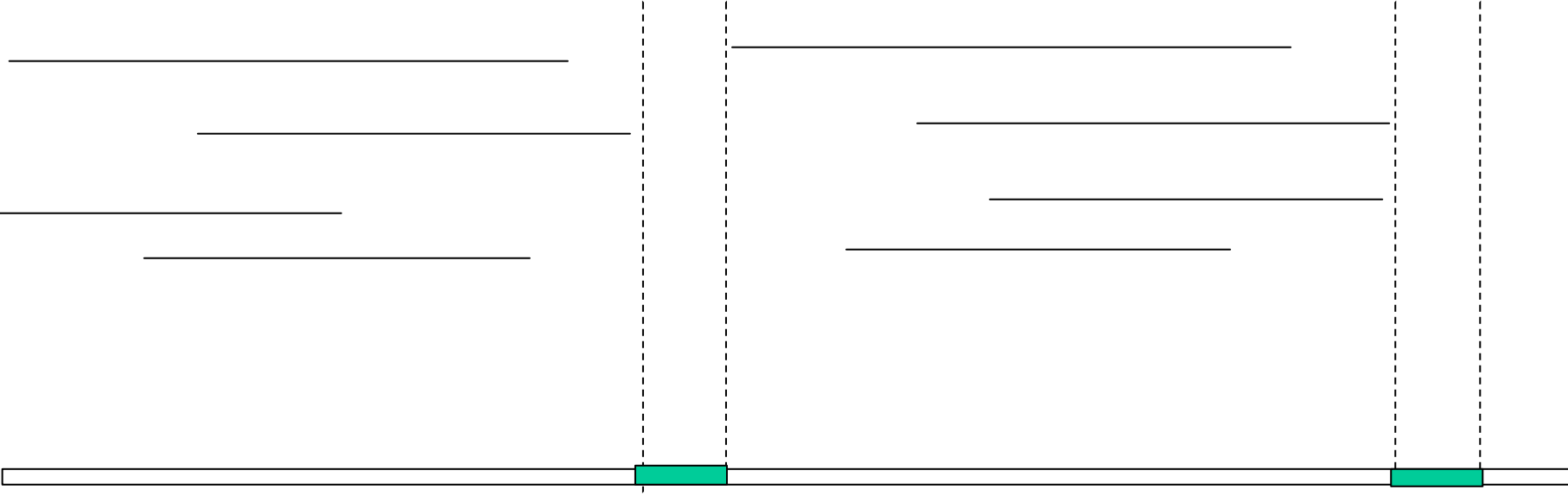
3-7 veces

Assembly:

- Putting sequenced fragments of DNA into their correct chromosomal positions

Contig

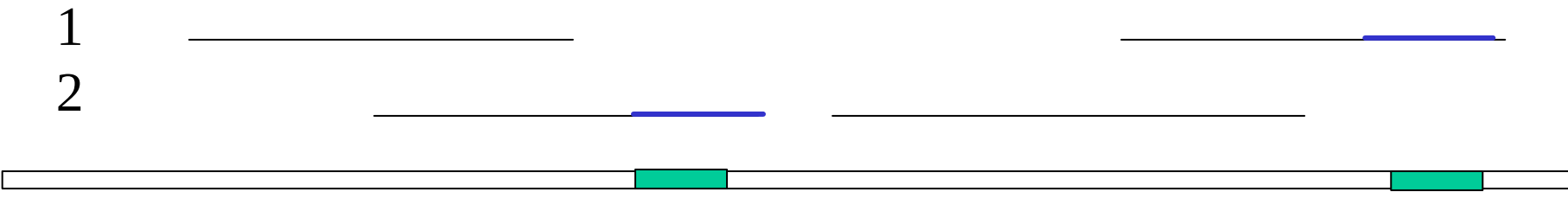
- Contiguous sequence of DNA created by assembling overlapping sequenced fragments of a chromosome (whether natural or artificial, as in BACs)



ecuperación de
ragmentos vía PCR

1

2

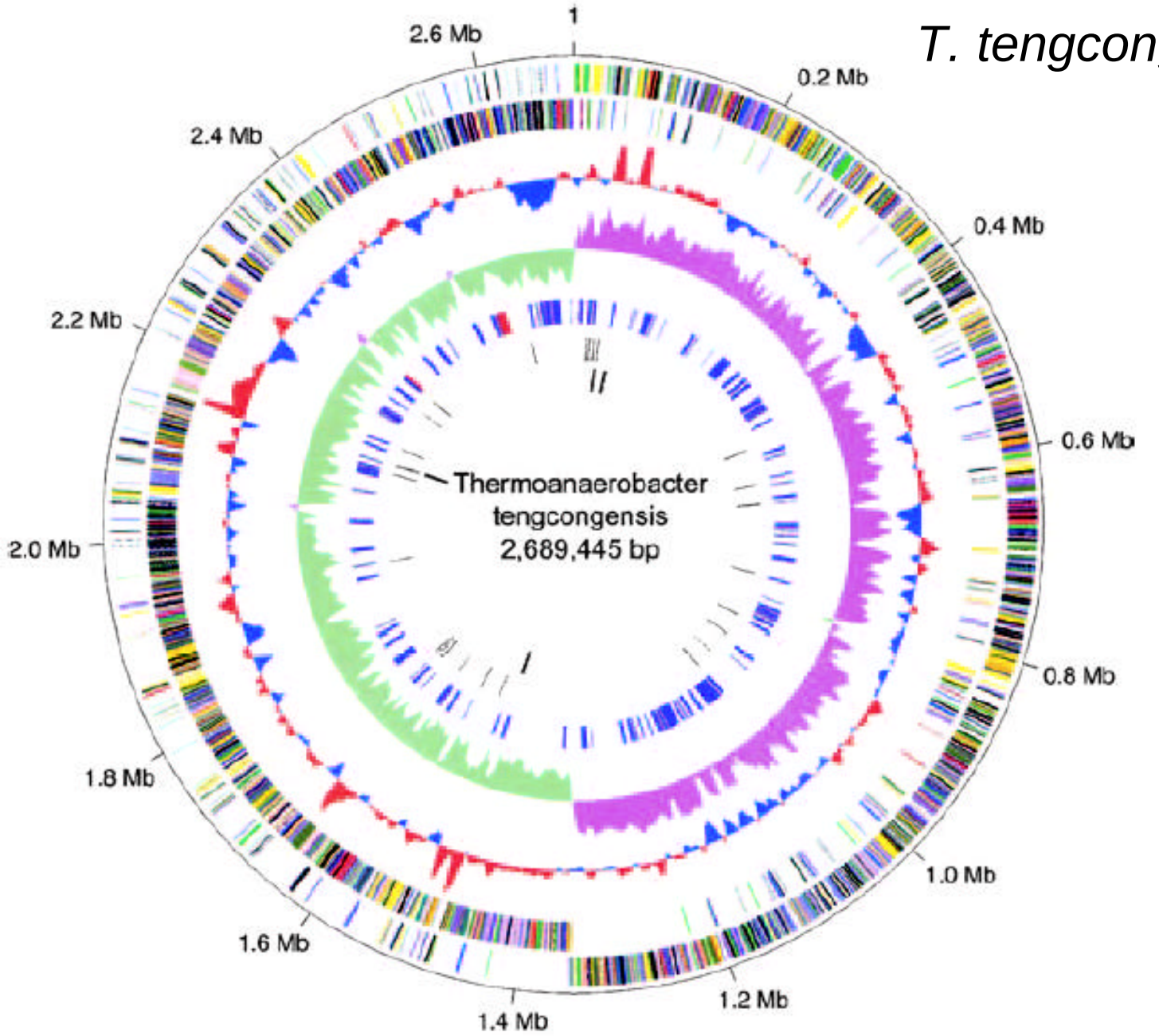


Repetición del protocolo de secuenciación con nuevos clones.

The fragment assembly problem

- Aim: infer the target from the reads
- Difficulties –
  - Incomplete coverage. Leaves *contigs* separated by *gaps* of unknown size.
  - Sequencing errors. Rate increases with length of read. Less than some ε.
  - Unknown orientation. Don’t know whether to use read or its Watson-Crick complement.

T. tengcongensis



Draft sequence

- Sequence with lower accuracy than a finished sequence; some segments are missing or in the wrong order or orientation

Table 1. General Features of the *Thermoanaerobacter tengcongensis* Genome

Genome size (bp)		2,689,445
G+C content		37.6%
Protein coding		87.1%
Average CDS length (bp)		905
Predicted CDS		2,588
	Known proteins	1494 (57.8%)
Homologous to	Protein domains/motifs	270 (10.4%)
	Hypothetical proteins	301 (11.6%)
No homology		523 (20.2%)
Stable RNAs		0.9%
rRNA operons		4
rRNAs		55
Major repetitive elements		
Short noncoding repeats		0.5%
Long coding repeats		8.6%

CDS, coding sequences.

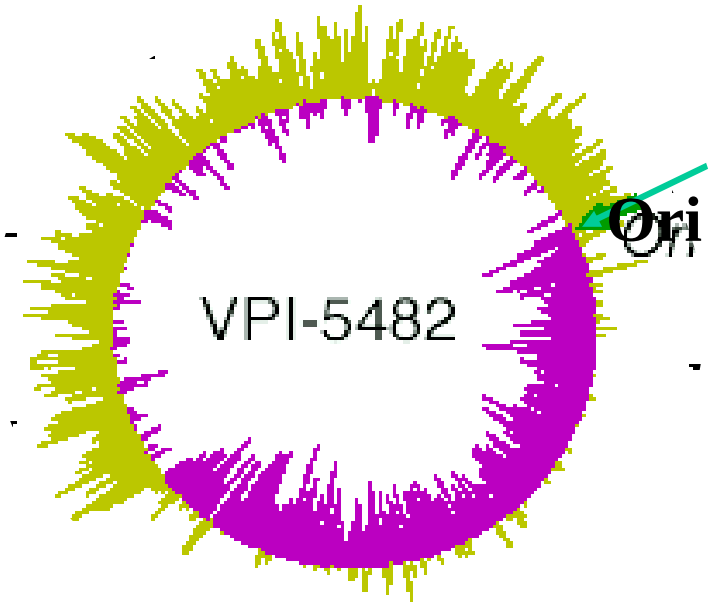
Análisis de:
genoma contra genomas (similitudes globales)
Secuencias repetidas (todos los sabores)

B. thetaiotaomicron

Secuencia lineal de un cromosoma bacteriano

5` 3`

Origin of DNA replication: **GC skew.**

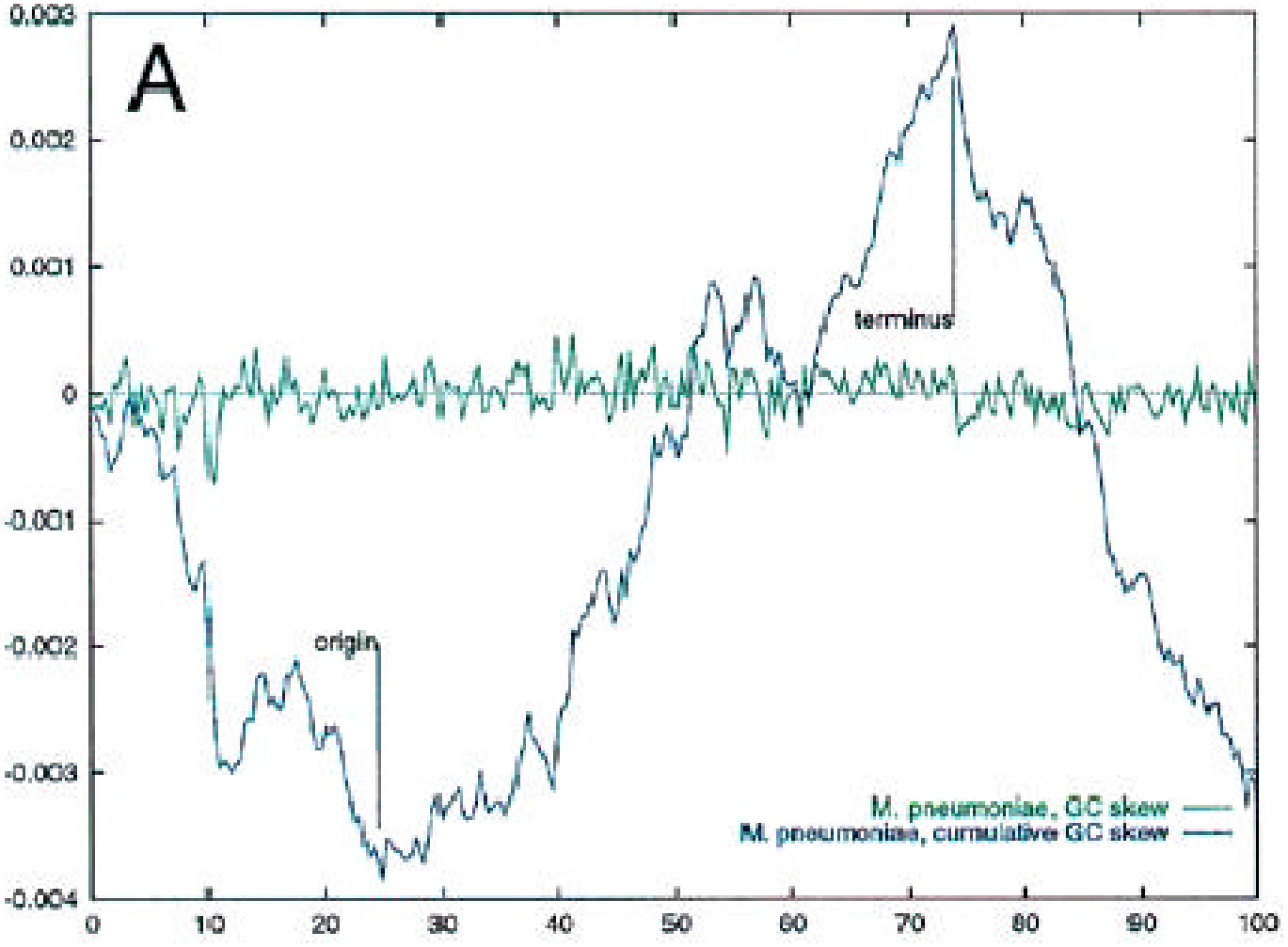


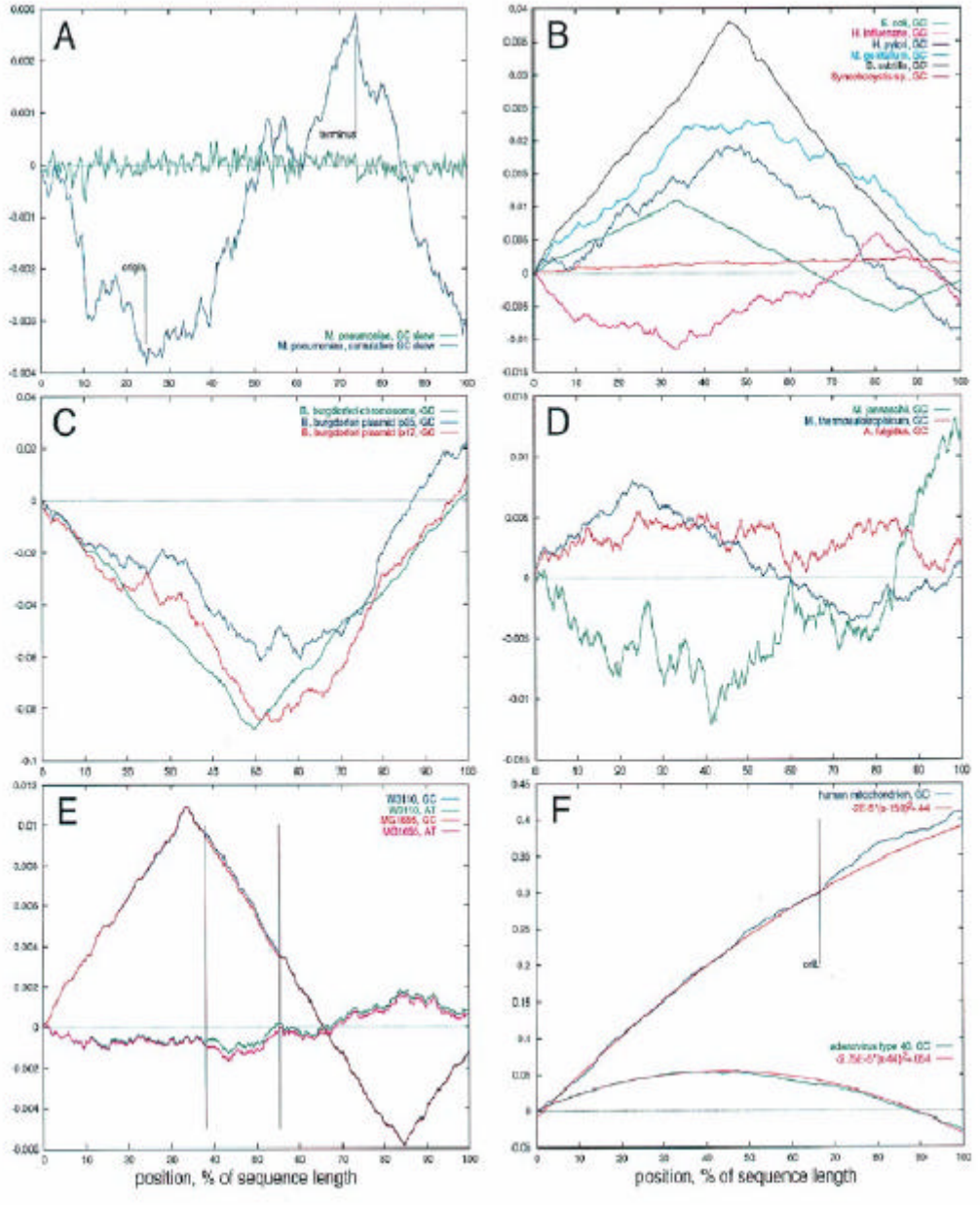
Protocol to find and verify the replicative origin (oriC) of the any bacterial chromosome:

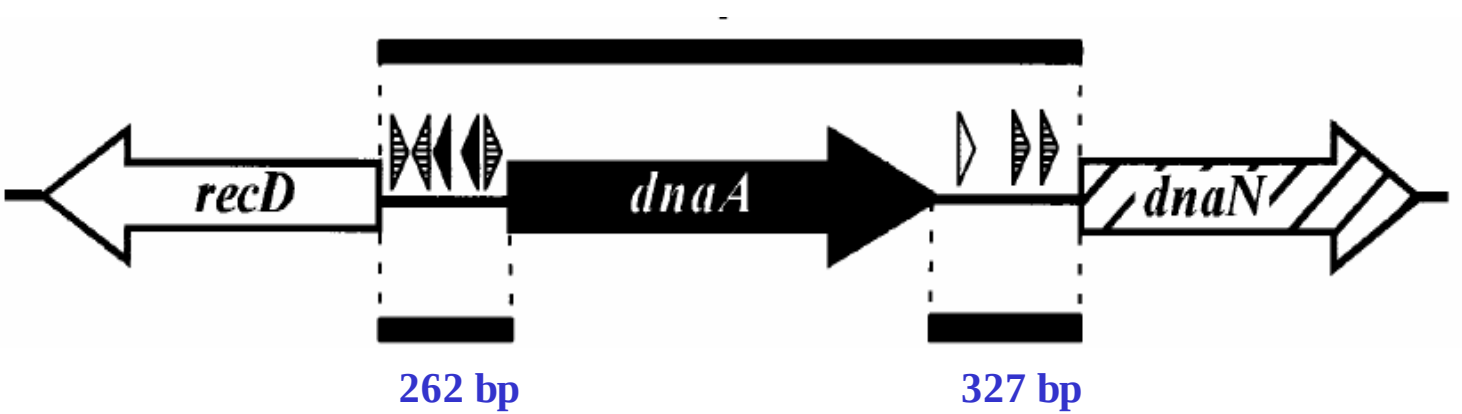
1) GC nucleotide skew [(G-C)/(G+C)] analysis.

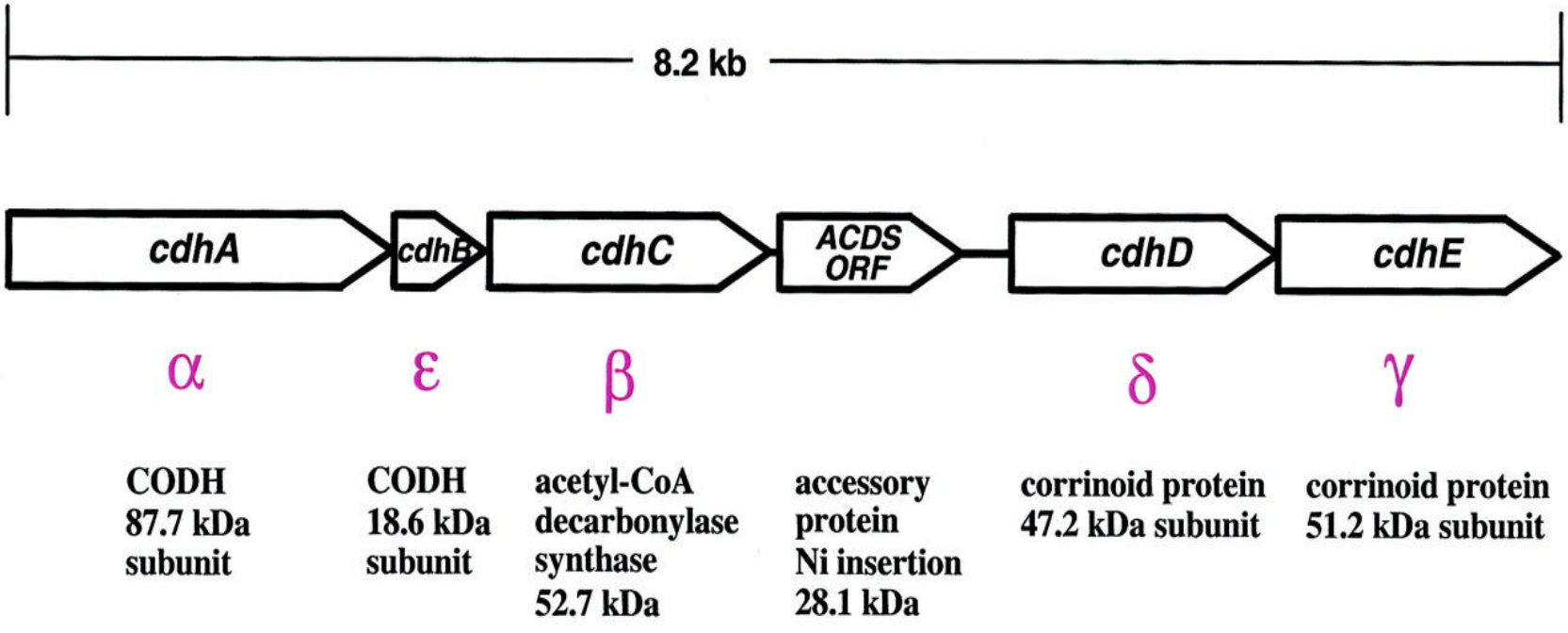
2) Identification by similarity, in this case we evaluated the co-localisation of several genes found around the origin of others bacterias. Typically bacteria contain the genes *dnaA, dnaN, recF and gyrB,* however the *Bacillus subtilis* chromosome include only the genes *dnaA and dnaN while rpmH and recF* are not include (Kuroda et al., 2001. The Lancet, 357: 1225), in *E. coli* dnaN and *gyrB* are present, dnaA is located 600bp away from dnaN – *gyrB*, and *recF* gene is missing. *(*Zawilaketal., 2001. Nucleid Acids Research **29**:2251***).***

3) In addition, we need evaluate if the replication origins region, contain short AT rich sequences and named DnaA box characteristic of bacterial (in *E. coli* and in *H. Pylori* there are 5 DnaA box).


A
0.003
0.002
0.001
0
-0.001
-0.002
-0.003
-0.004
0
10
20
30
40
50
60
70
80
90
100
origin
terminus
M. pneumoniae, GC skew
M. pneumoniae, cumulative GC skew

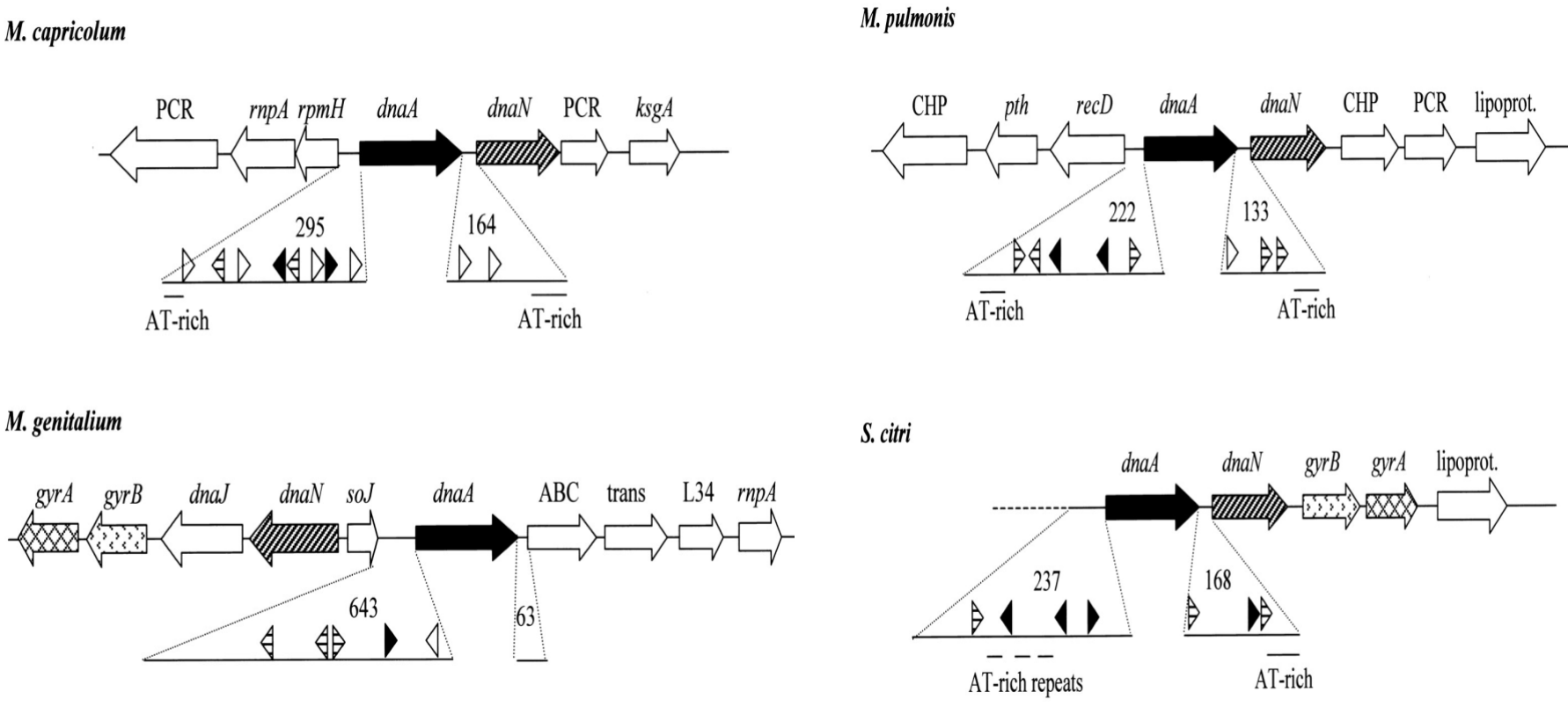

A
terminus
origin
N. pneumoniae, GC skew
M. pneumoniae, complete GC skew
B
C
D
E
W3110, GC
W3110, AT
MG1655, GC
MG1655, AT
F
human mitochondrion, GC
adenovirus type 40, GC
ori
position, % of sequence length
position, % of sequence length


1.9 kbp
recD
dnaA
dnaN
262 bp
327 bp



Organization of the ACDS operon in *M. thermophila* TM-1.

Gene order and putative DnaA boxes within the oriC regions of mollicute genomes.



Orientations of the genes are indicated by rrows; intergenic regions flanking the dnaA genes are magnified with their lengths (in nucleotides) indicated. PCR, putative coding region; CHP, conserved hypothetical protein. Putative DnaA boxes are represented by headless arrows. A match with the consensus defined for mollicutes (TT(A/T)TC(C/A)ACA) is symbolized by the following: black, nine of nine; horizontal stripes, eight of nine; white, seven of nine. Caio et al., J. Bacteriology, 2002, **184:** 5426–5435

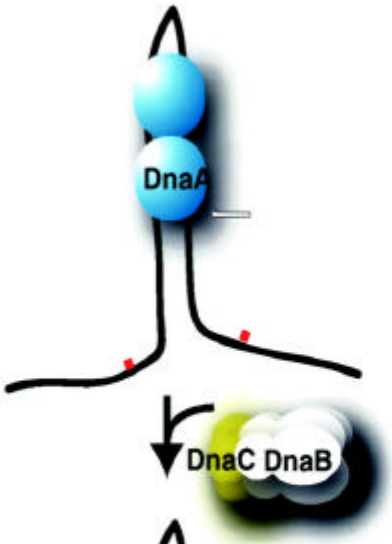
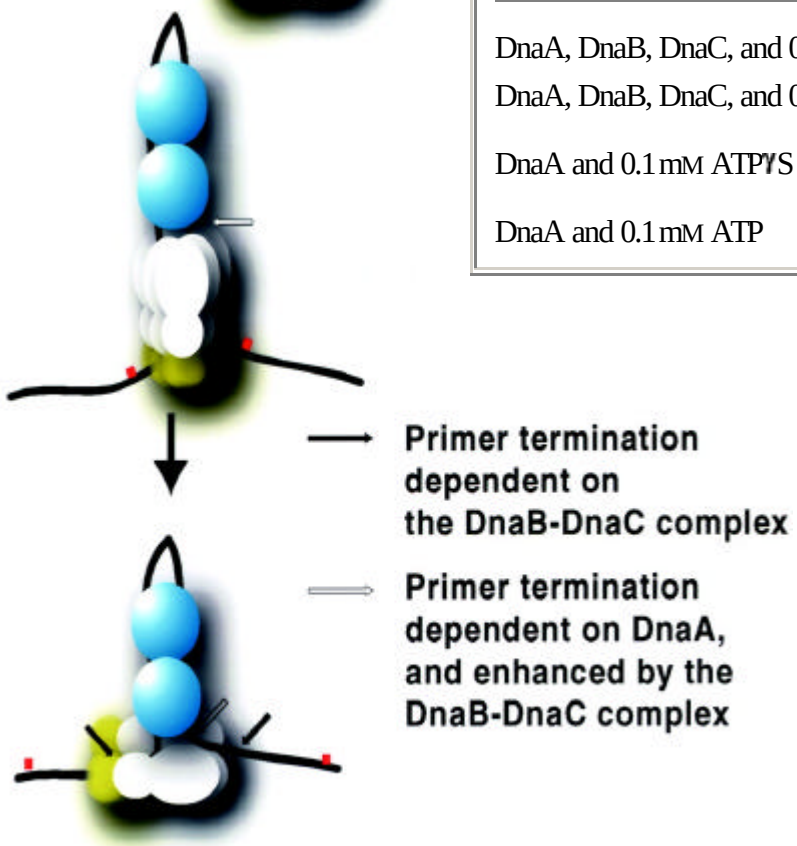


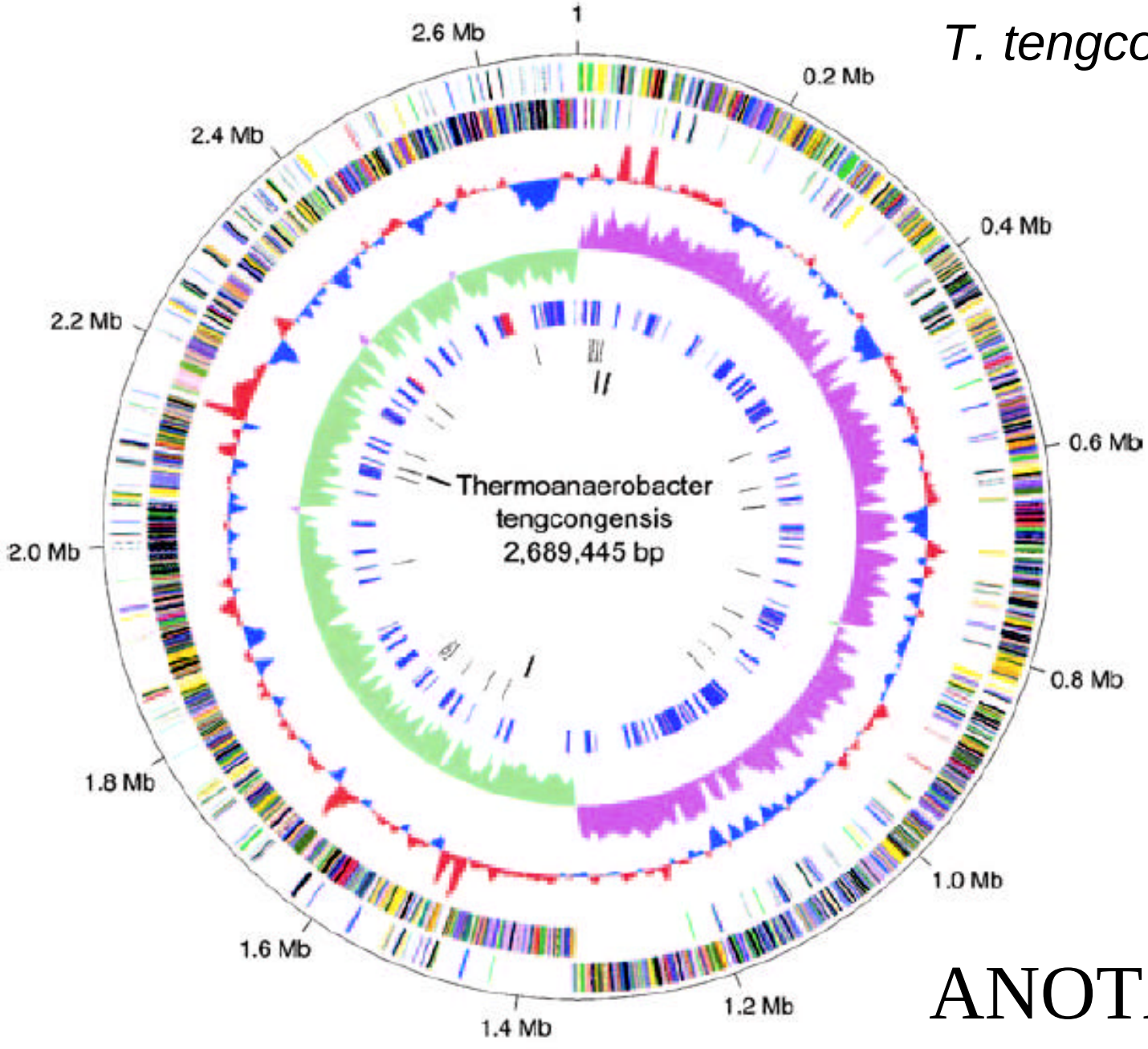
Table I
Stoichiometry of DnaA, DnaB, and DnaC at the DnaA box hairpin
The results from replicate experiments under the indicated experimental conditions were averaged to calculate the S.D. for the stoichiometries of DnaA, DnaB, and DnaC protein bound to the ssDNA. The term *n* refers to the number of replicate experiments under a given condition.

Reaction conditions	DnaA	DnaB	DnaC	*n*
DnaA, DnaB, DnaC, and 0.1 mM ATPγS	3.6 ± 0.8	4.0 ± 0.9	3.8 ± 1.6	3
DnaA, DnaB, DnaC, and 0.1 mM ATP	2.3 ± 0.5	4.2 ± 0.8	None detected	4
DnaA and 0.1 mM ATPγS	3.9 ± 1.9			5
DnaA and 0.1 mM ATP	2.6 ± 1.0			2



Carr and Kaguni. 2002. J. Biol. Chem., Vol. 277: 39815

T. tengcongensis

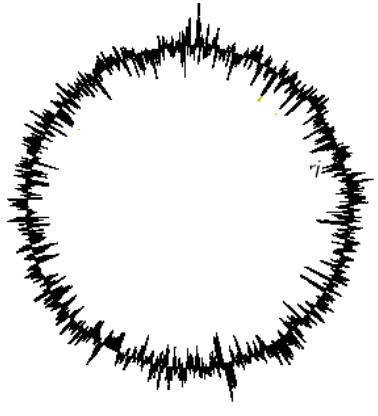


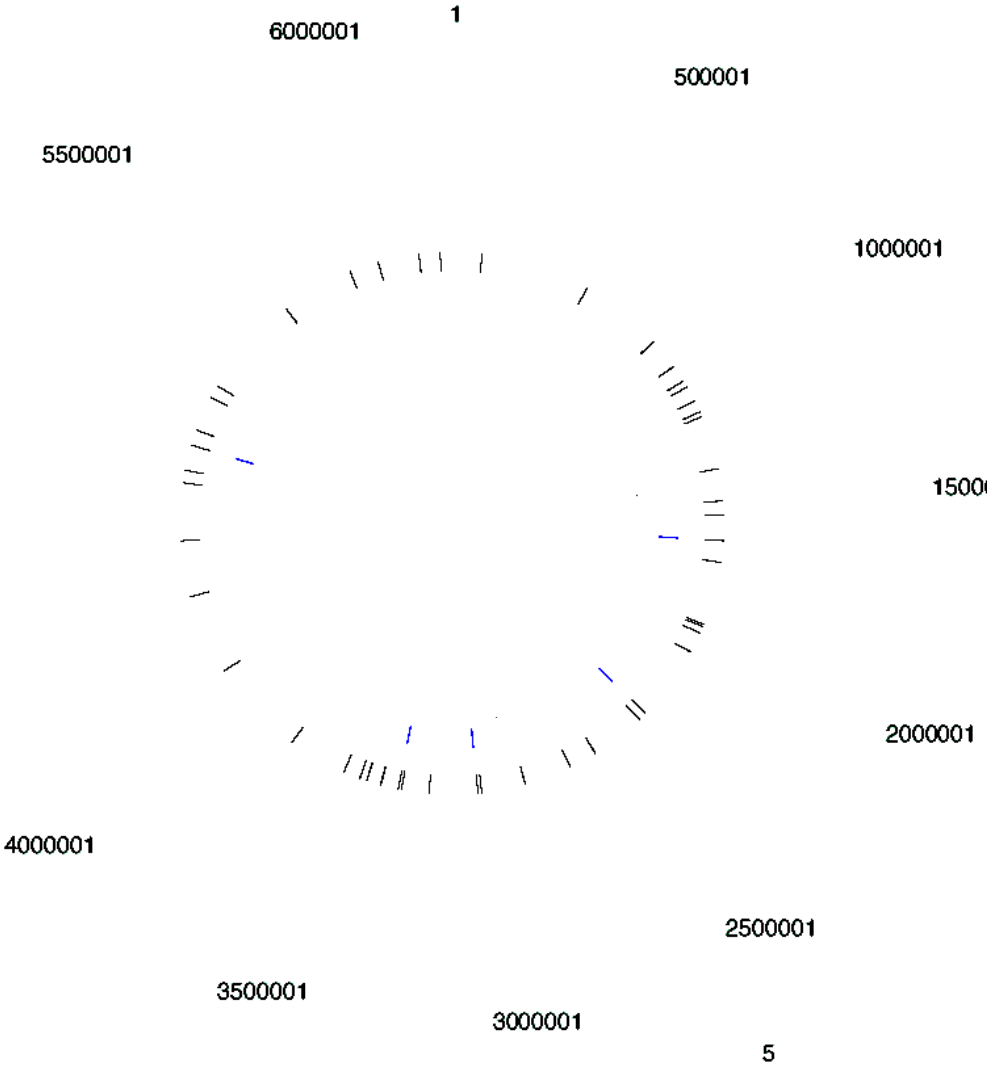
ANOTACION

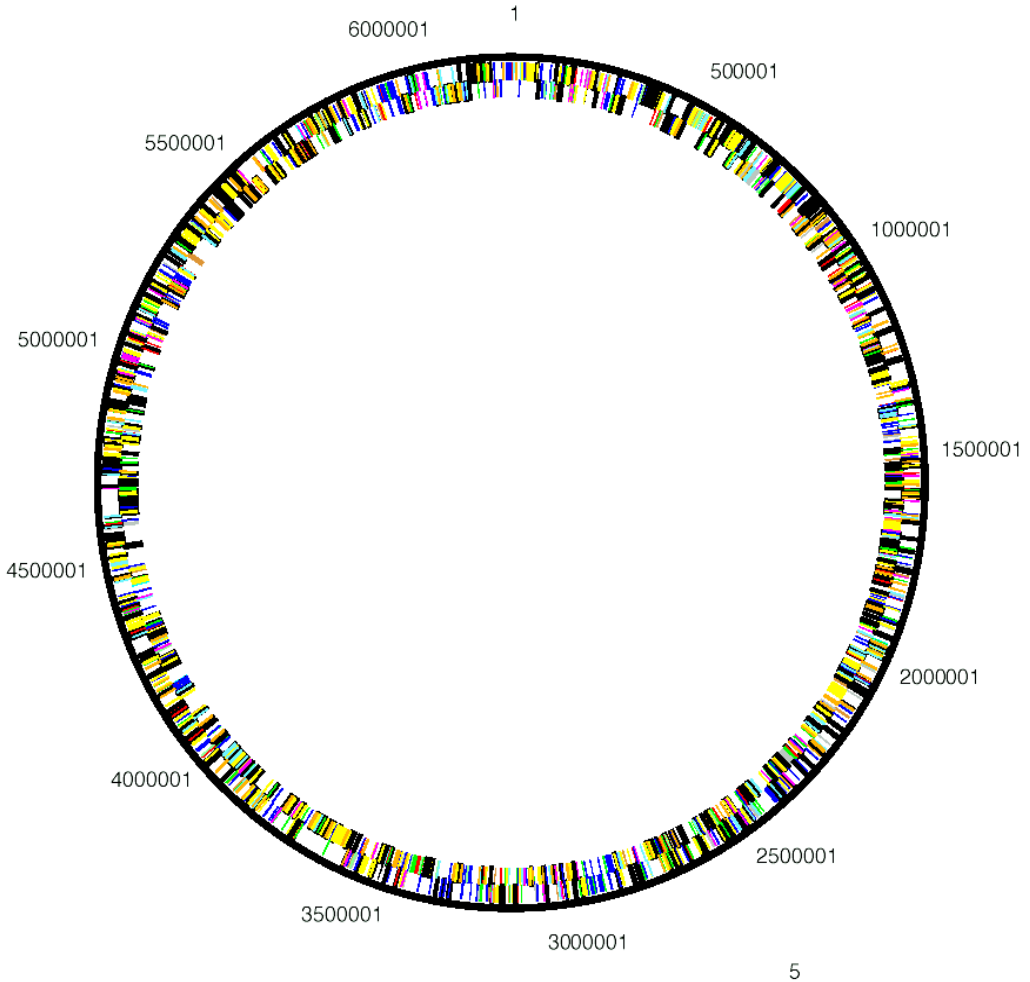
tRNA: BLASTN

tRNA genes: tRNA-Scan

Contenido GC

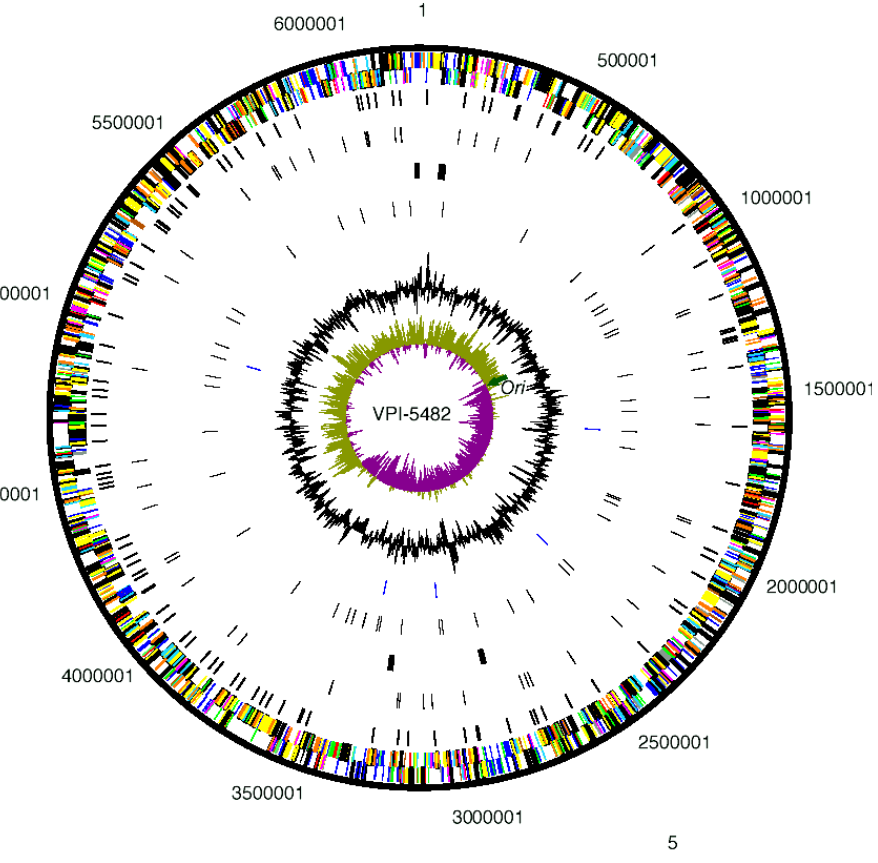




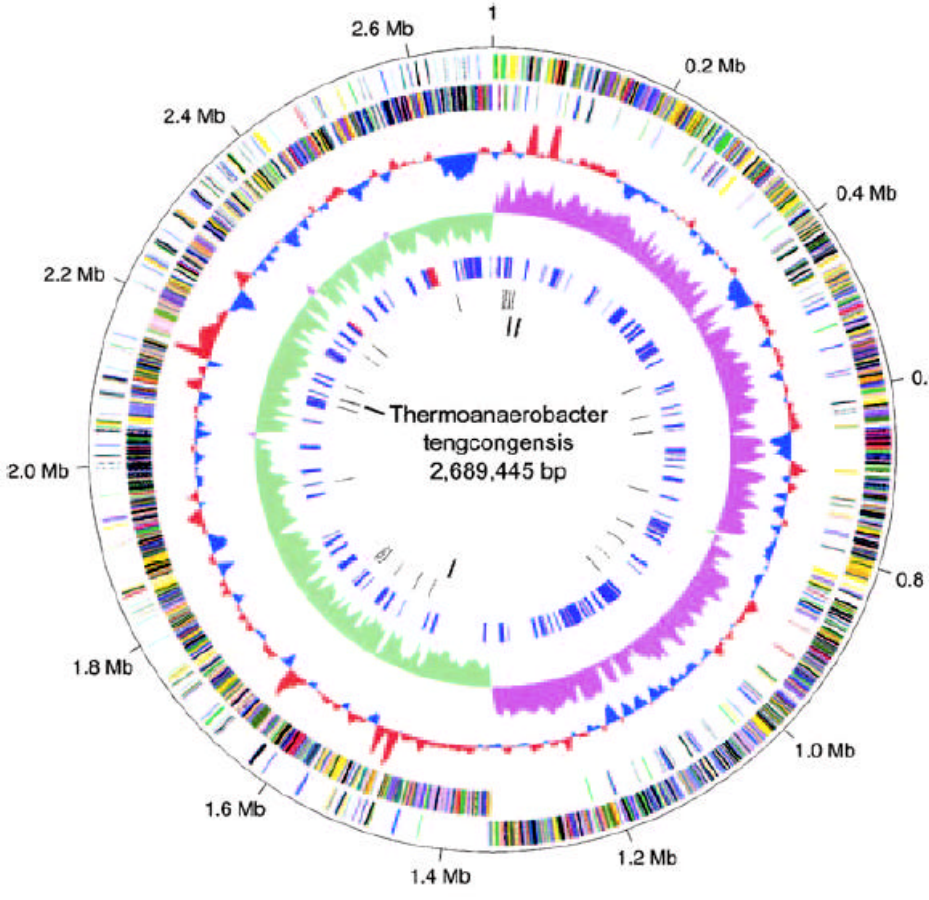


1. **Identificación de marcos de lectura por hebra. (utilizando GLIMMER, ORPHEUS and CRITICA .**
2. **Verificación por blast. % recuperado x homolg.**
3. **Revisión manual**
4. **Genes que codifican para proteinas con <60-100 codones que no presentan homologías con un cierto umbral de *E value* (10^3) deben ser eliminadas.**
5. **Clasificación de funciones (KO; COG) por grandes grupos.**
6. **Mapa físico en megabases**

B. thetaiotaomicron



T. tengcongensis



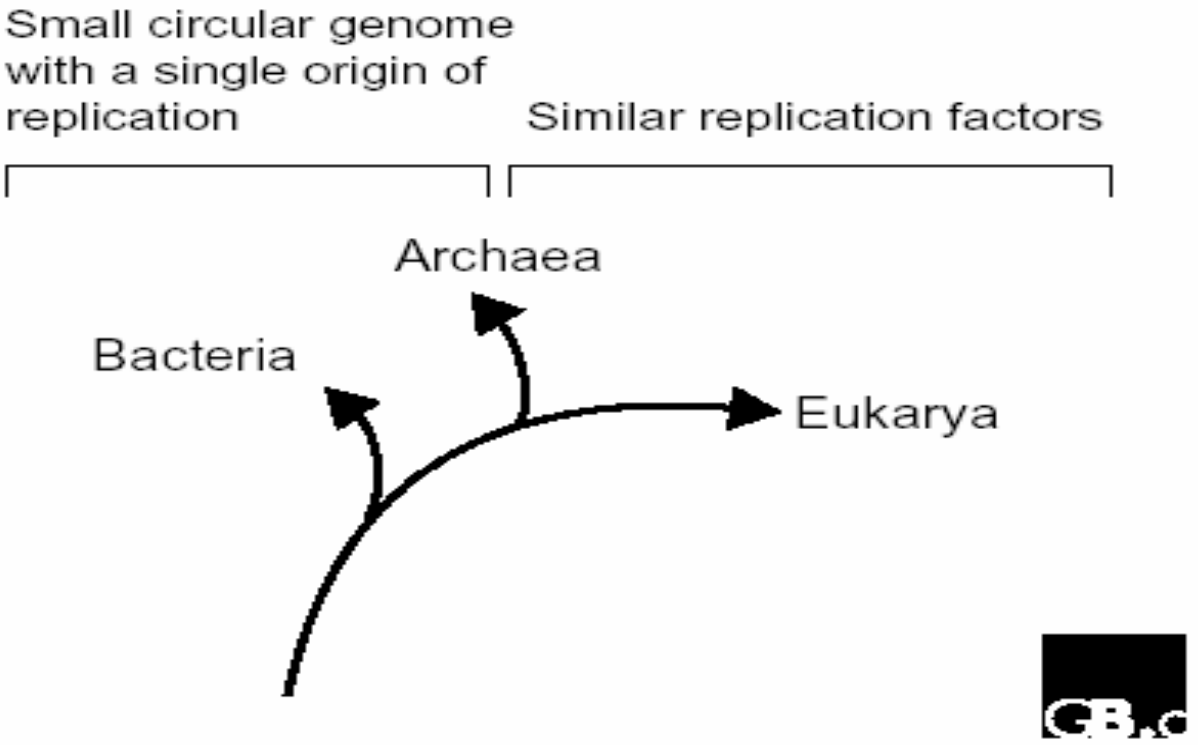


Figure 1
Evolutionary relationships between bacteria, archaea and eukaryotes takes into account the similarities between archaea replication, transcription and translation factors with eukaryotic factors. The report by Myllykallio *et al.* [1] shows that the archaea share chromosome organization and replication pattern with prokaryotes although they use many eukaryotic-like factors to duplicate their chromosomes.

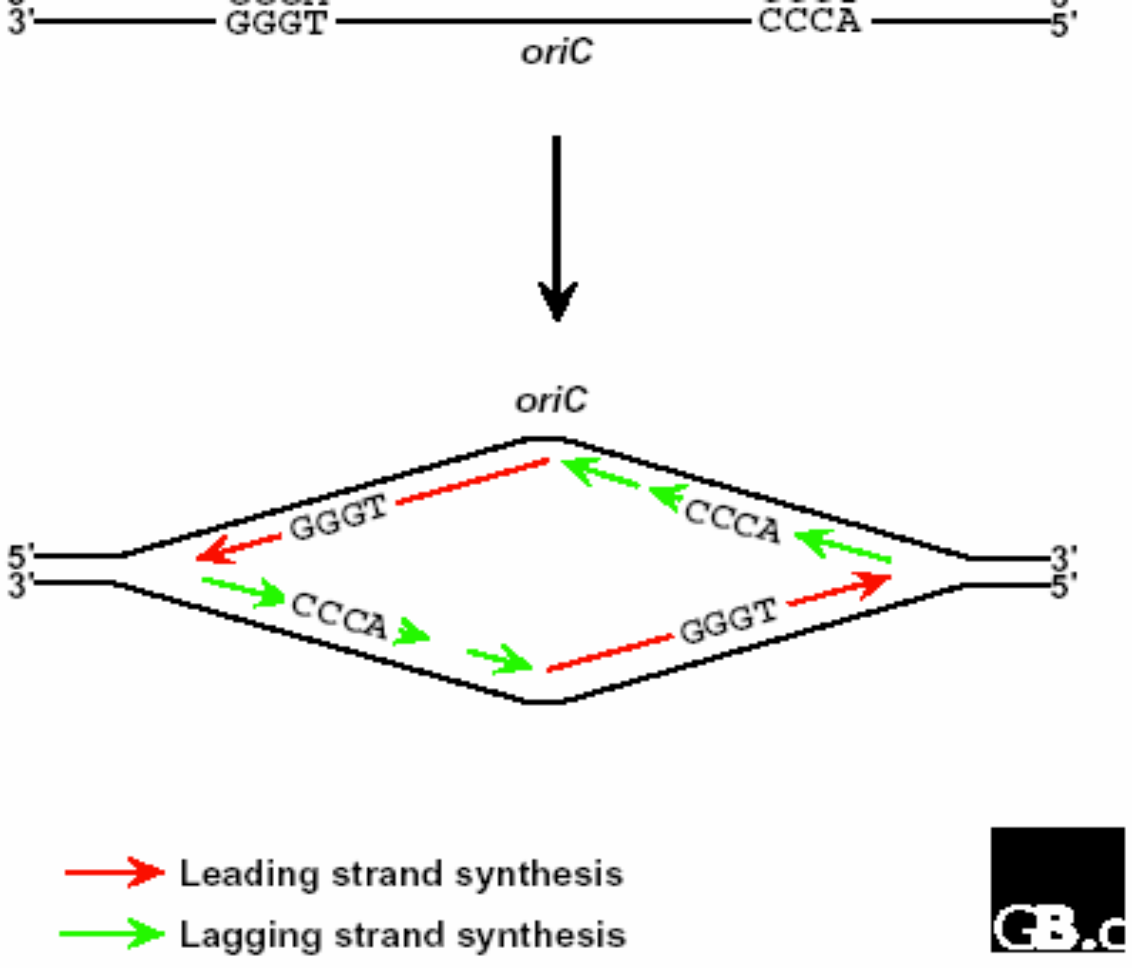


Figure 2
A bi-directional replication fork: DNA replicated by lagging-strand synthesis on one side of the origin will be replicated by leading-strand synthesis on the other side. In bacteria, there is a switch in the strand bias of guanine content at the origin. Myllykallio *et al*. [1] measured the strand bias resulting from GGGT as illustrated.