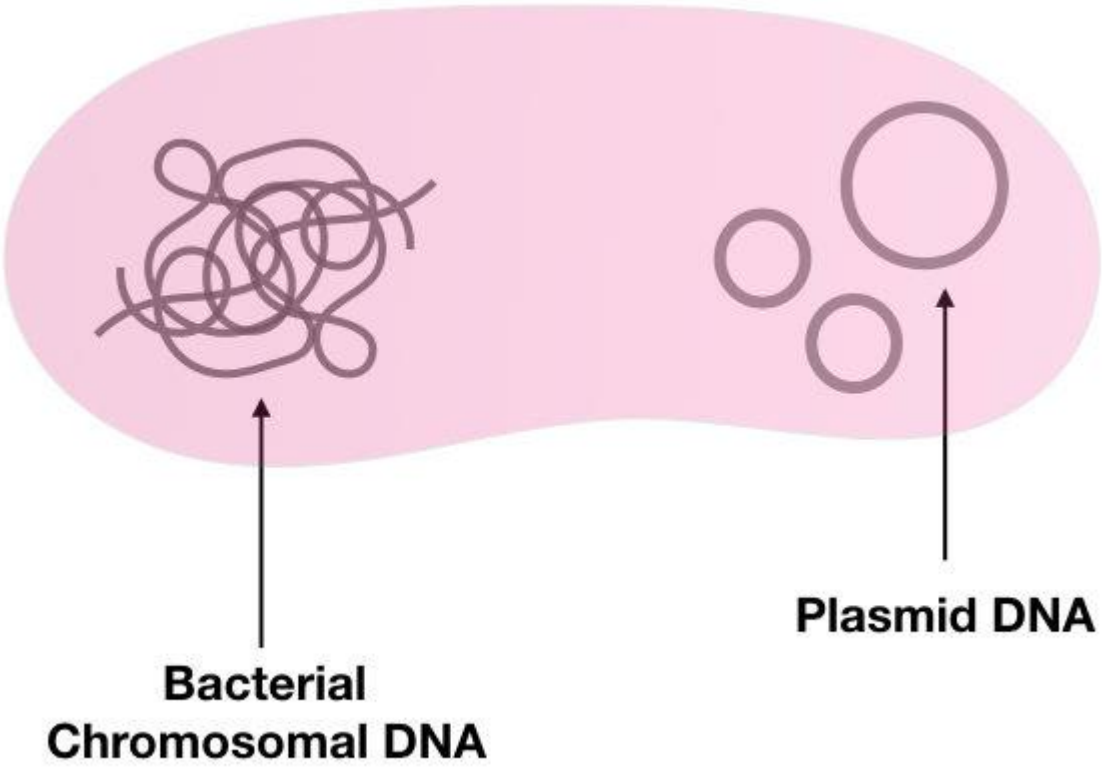
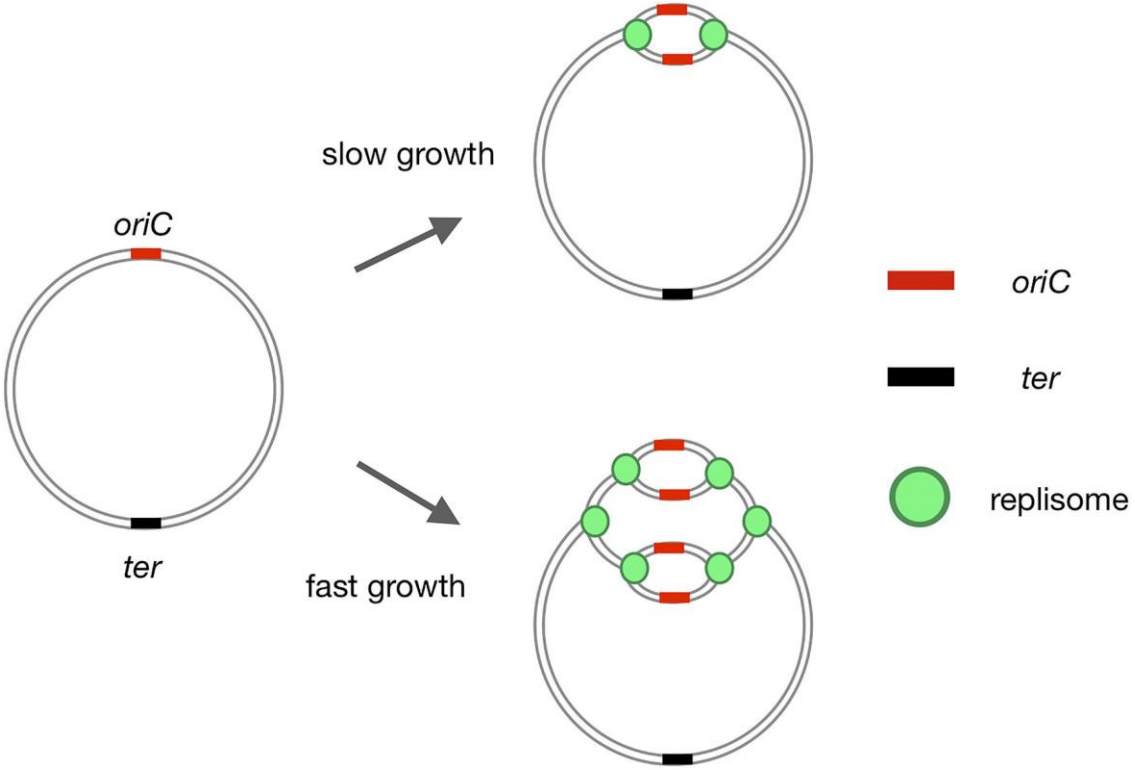
# Genome organization in prokaryotes



**Chromosomal replication**

the cell doubles in 20 min, which is less than the time required to replicate the chromosome (45 to 60 min), starting from one replicative fork.
Replication of the bacterial chromosome is initiated at **a single oriC region, proceeds in both directions, and terminates at the ter region.**

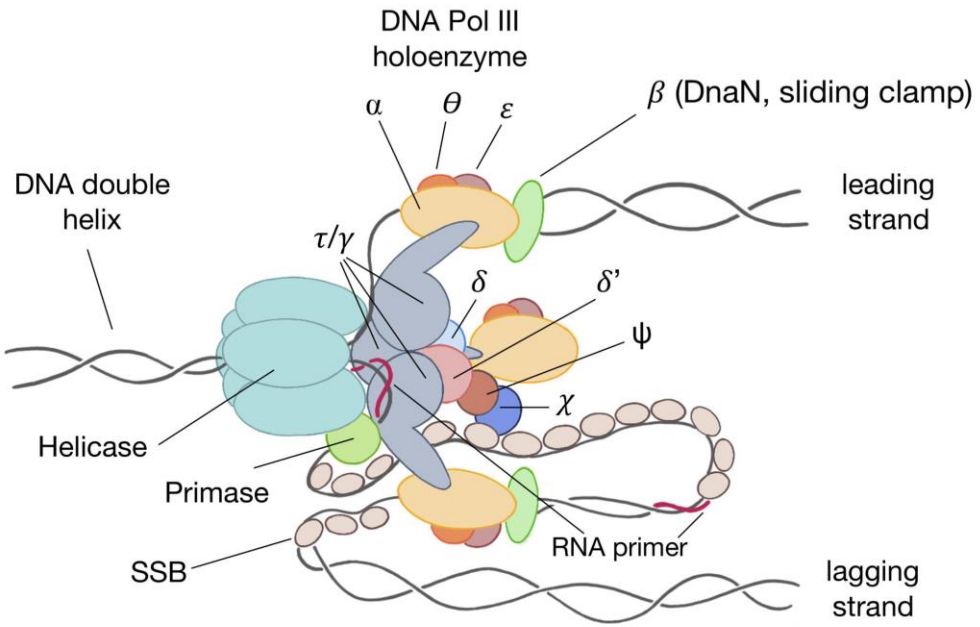



During slow growth, replication is initiated once per cell cycle. In fast growers under optimal conditions, another round of replication is initiated before the previous round has been completed, resulting in the inheritance by daughter cells of partially replicated chromosomes.

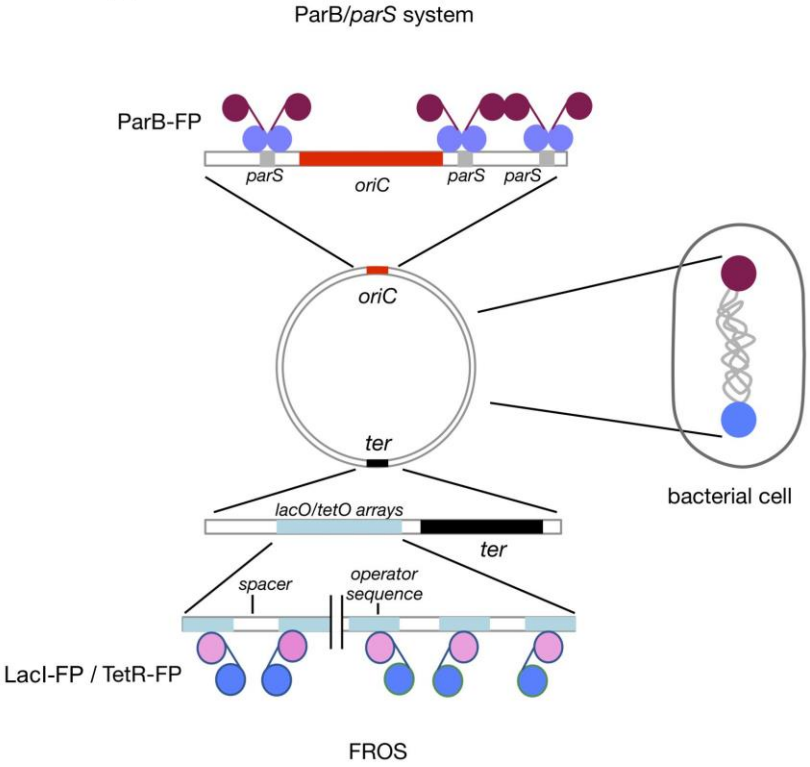
Trojanowski D, Hołówka J, Zakrzewska-Czerwińska J. Where and When Bacterial Chromosome Replication Starts: A Single Cell Perspective. Frontiers in Microbiology 9, 2018 10.3389/fmicb.2018.02819

# Replisome structure and localization

A




B



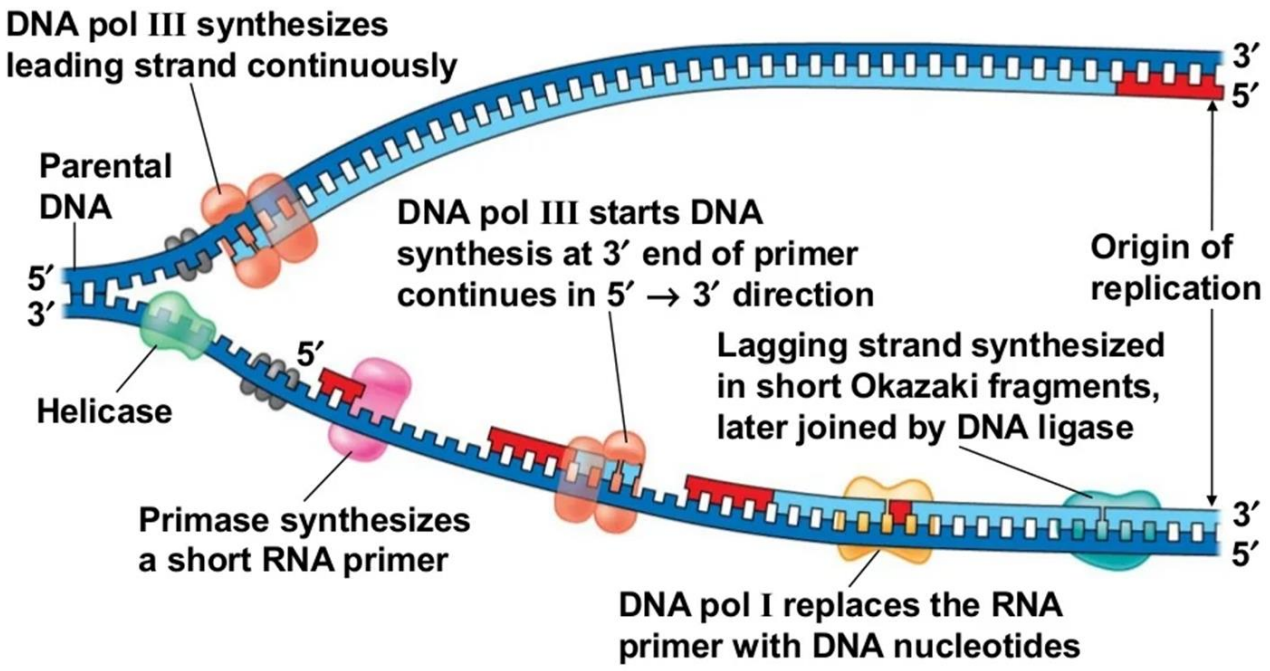

(A) A replisome is a multiprotein complex

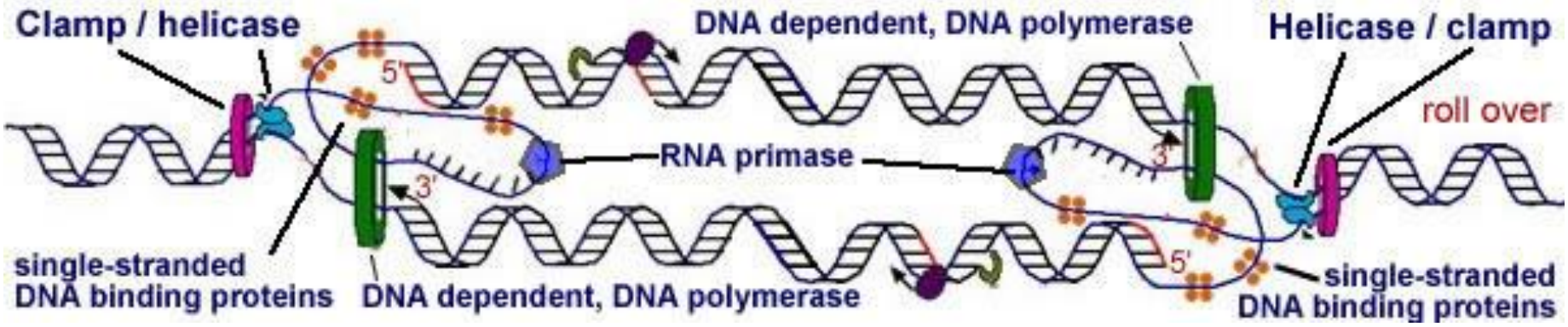
- a **helicase** unwinds the chromosome, separating the two single-stranded DNA strands
- the **three core polymerases** are loaded into each replication fork by
- the **clamp loader** and bind to the sliding clamp.

The leading strand is synthesized continuously, while the lagging strand is synthesized in approximately 1 kbp fragments, starting from the short primers added by the primase.

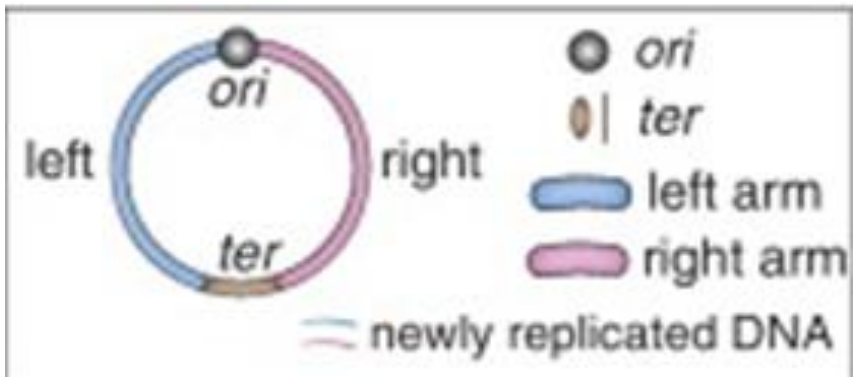
(B) Schematic localization of chromosomal loci using **ParB/parS and FROS system**. ParB binds to parS sequences (purple) in the oriC region
The ter region consists of two components: operator sequences (usually lacO or tetO arrays repeated up to several hundred times in tandem) and a repressor protein (LacI or TetR), which binds to the operator sequences.

•Replicative helicase: An enzyme which catalyses continuous unwinding of the parental duplex DNA at the replication fork.
•Replication fork: The site of DNA replication where two replicating single-stranded DNA separates.
•Primer RNA: A short stretch of RNA, the 3'-terminus of which is utilised by DNA polymerases for DNA elongation.
•Primosome: A name given to the protein complex capable of duplex DNA unwinding and primer RNA synthesis at the replication fork.




Clamp / helicase
DNA dependent, DNA polymerase
Helicase / clamp
roll over
5'
3'
RNA primase
3'
5'
single-stranded
DNA binding proteins
DNA dependent, DNA polymerase
single-stranded
DNA binding proteins

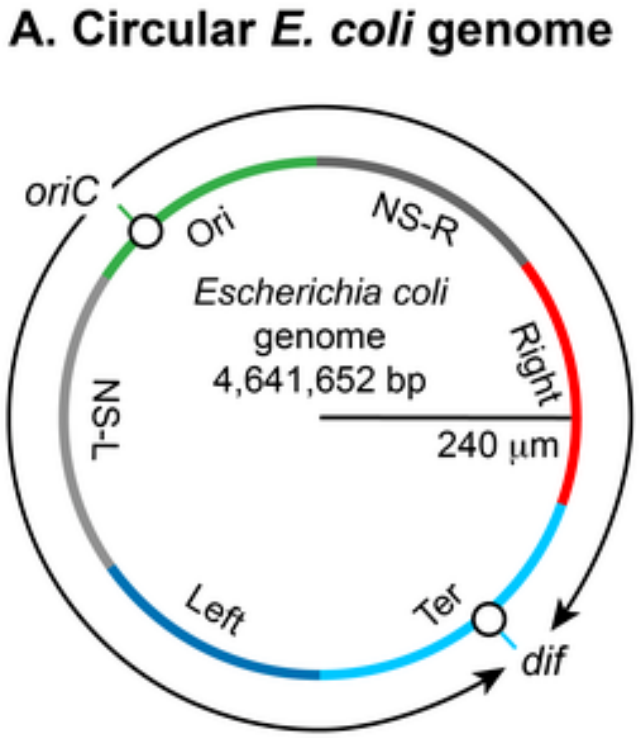
Localization of the replication machinery at the beginning of DNA synthesis determines the spatial arrangement of the chromosome.
*oriC* and *ter* regions positioned at the mid-cell, the intervening chromosomal regions (i.e., the left and right chromosomal arms) are stretched out toward opposite cell poles, creating a *left-ori-right* pattern (Wang and Rudner, 2014).



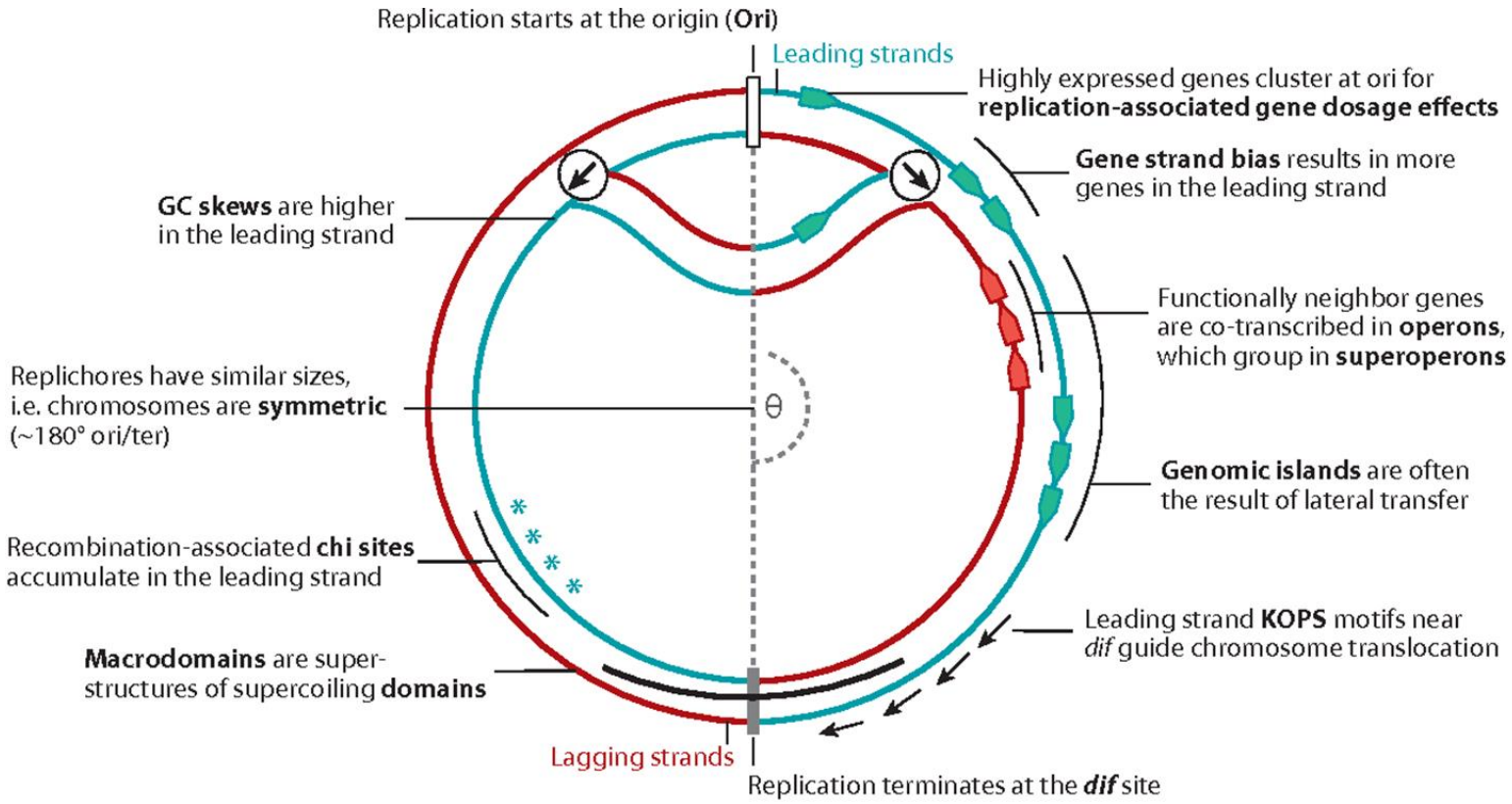
Nonrandom gene-distribution patterns of genetic elements in the chromosome

In *E. coli*, the chromosome is organized into four macrodomains Ori, Ter, Right and Left

and two large unstructured regions **NS-L and NS-R**



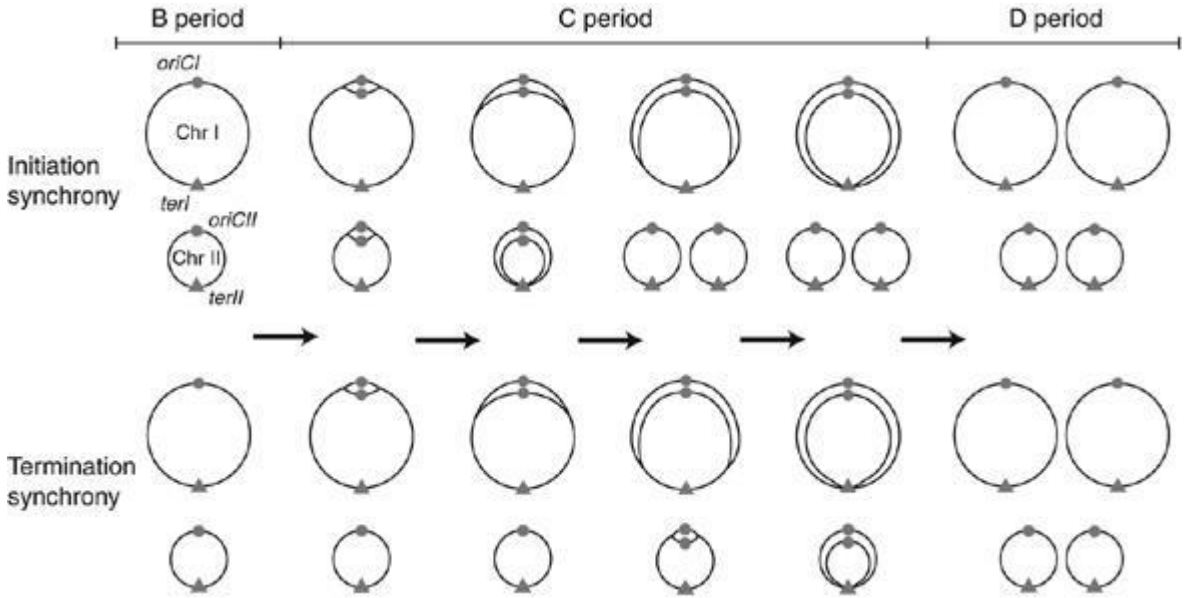
In newly replicated cells, the macrodomains around the origin (Ori) and terminus (Ter) of replication are localized near opposite cell poles leading to a linear arrangement of the genetic information in the cell




The link between genome organization and chromosome structure might also drive the evolutionary rate of genes, because the density of point mutations is higher in the regions of higher superhelicity of the *E. coli* chromosome, and horizontally transferred genes accumulate in Ter-proximal macrodomains.

Some bacterial species contain more than one type of chromosome. Such genomes typically have chromosomes of very different sizes. The larger chromosome encodes most essential and highly expressed genes. Smaller chromosomes are also called secondary chromosomes or chromids.

The systematic presence of two chromosomes in the genus of *Vibrio*, typically very fast-growing bacteria, led to suggestions that multiple chromosomes facilitate rapid bacterial growth and the management of chromosome dimerization in large genomes.

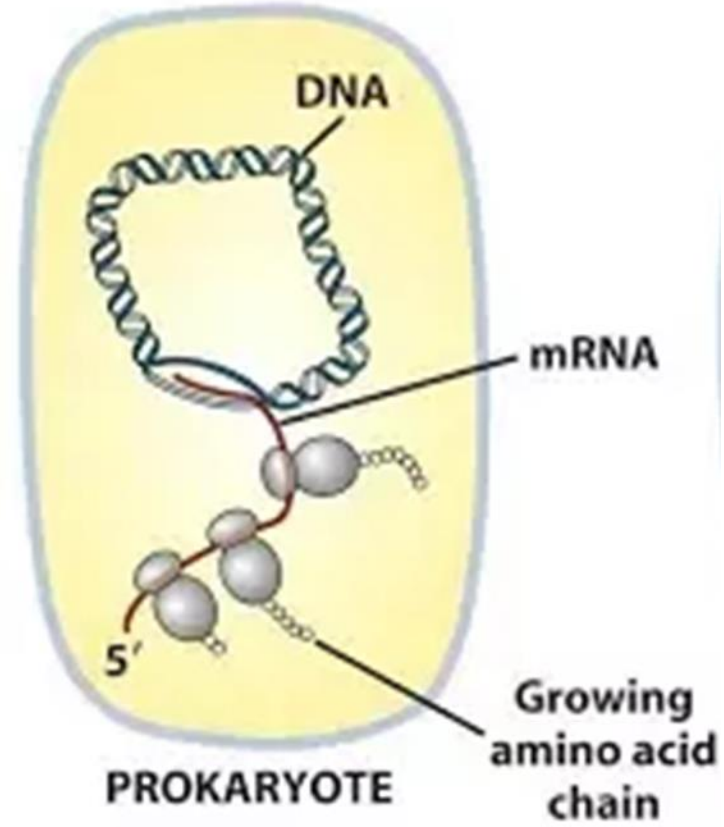



DnaA controls replication initiation of chromosome I and RctB for chromosome II.

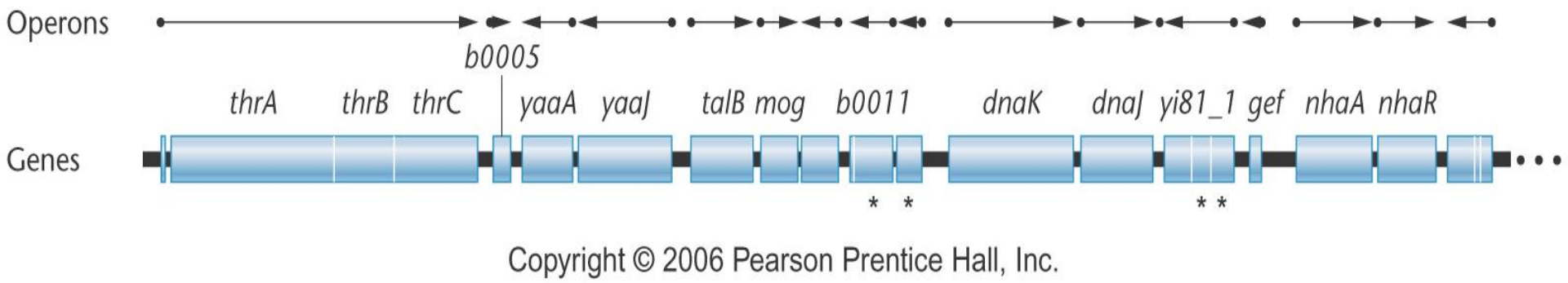
Experiments suggest a model where approximately two-thirds of the large chromosome I is replicated before the smaller chromosome II is initiated, leading to approximately simultaneous termination of replication of the two chromosomes

The EMBO Journal.The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle
Tue Rasmussen, Rasmus Bugge Jensen, and Ole Skovgaard

As replication proceeds, DNA regions are under different states of replication and are being segregated in function of the growing multiple division septa. In prokaryotes, nascent transcripts are immediately translated by multiple ribosomes and, for certain membrane proteins, integration in the membrane takes place before the end of transcription and translation. Hence, transcription, translation, and protein localization are tightly linked.



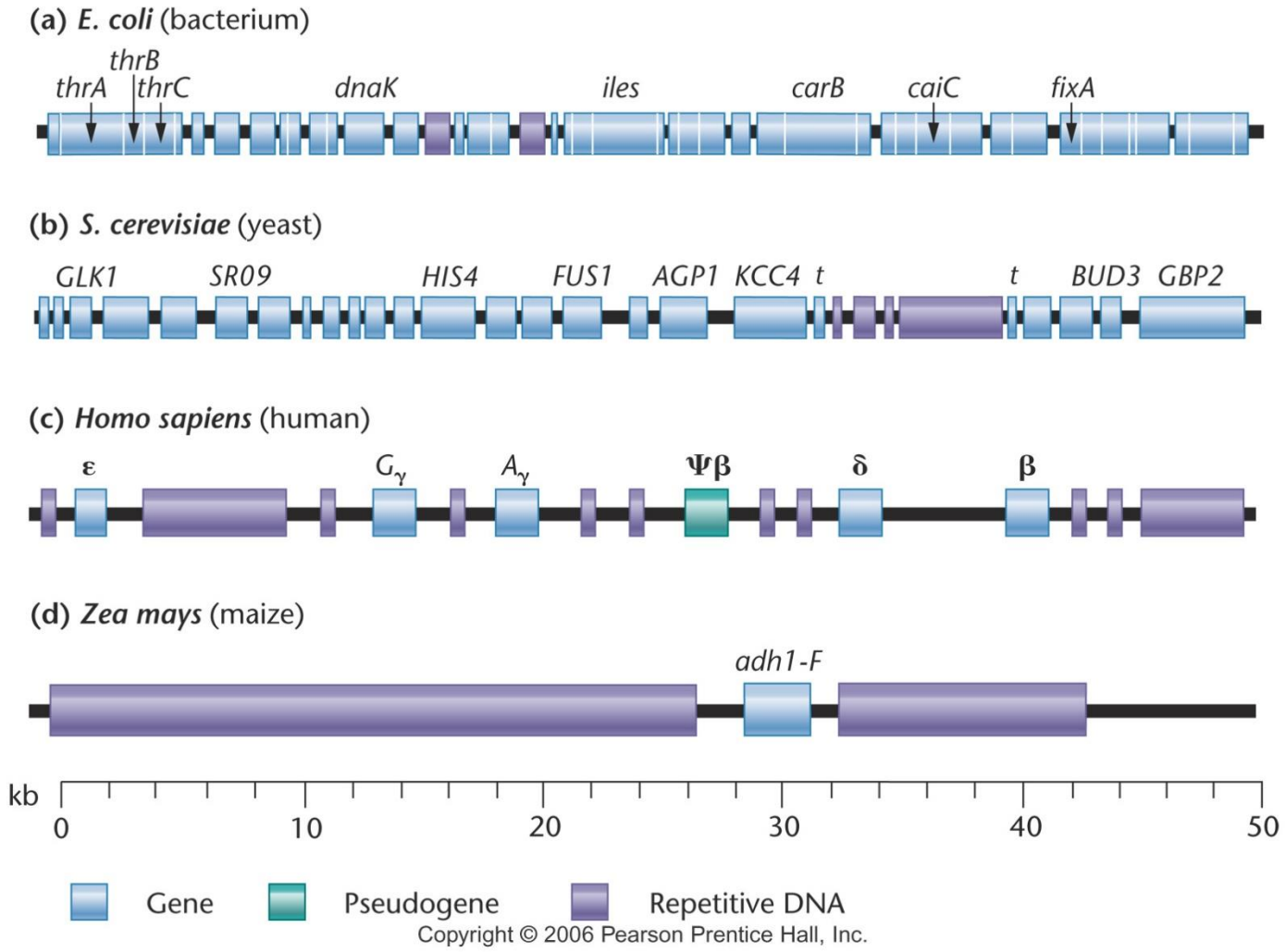
# *Genetic organization in Escherichia coli*



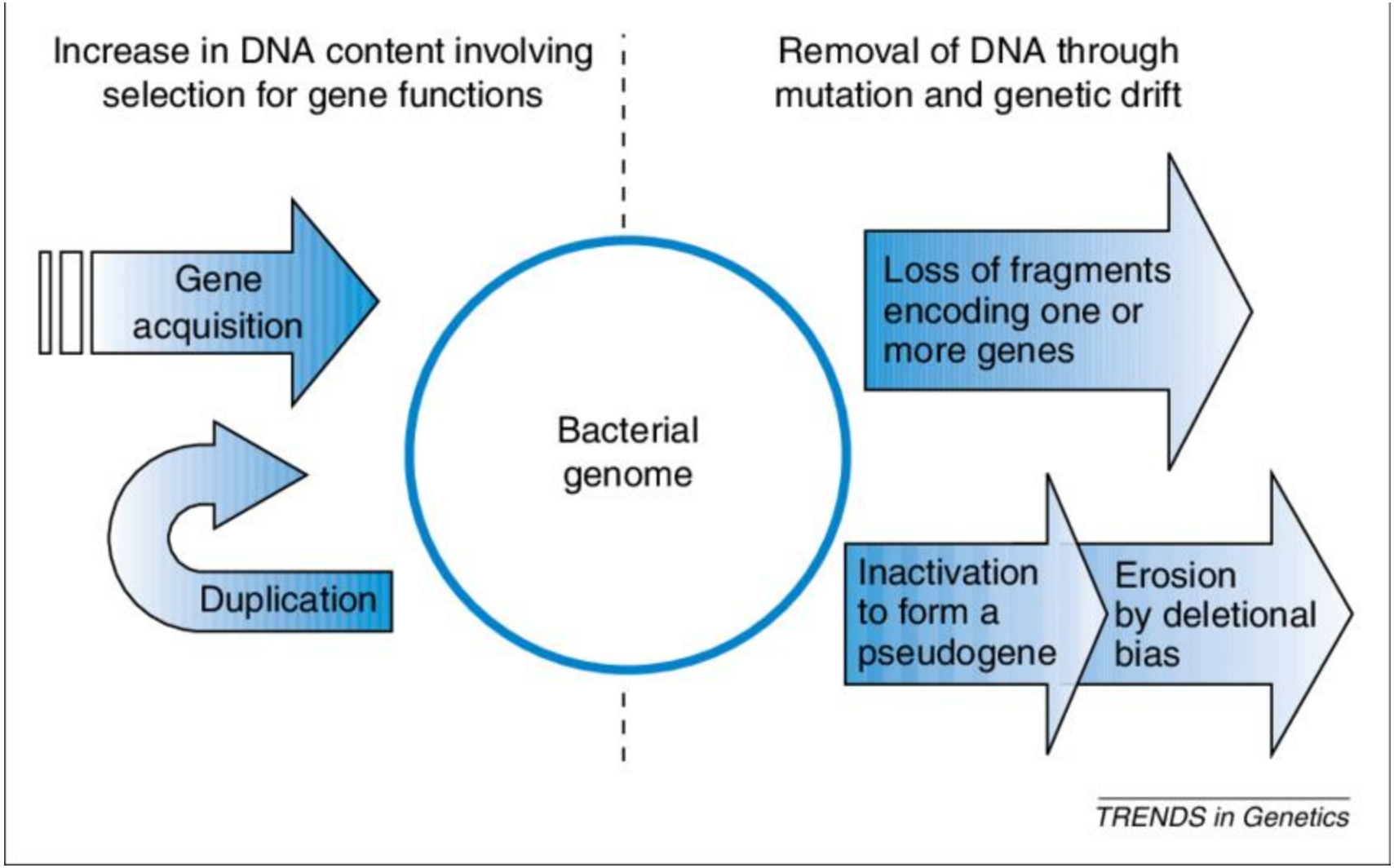



- *E. coli* has >600 operons

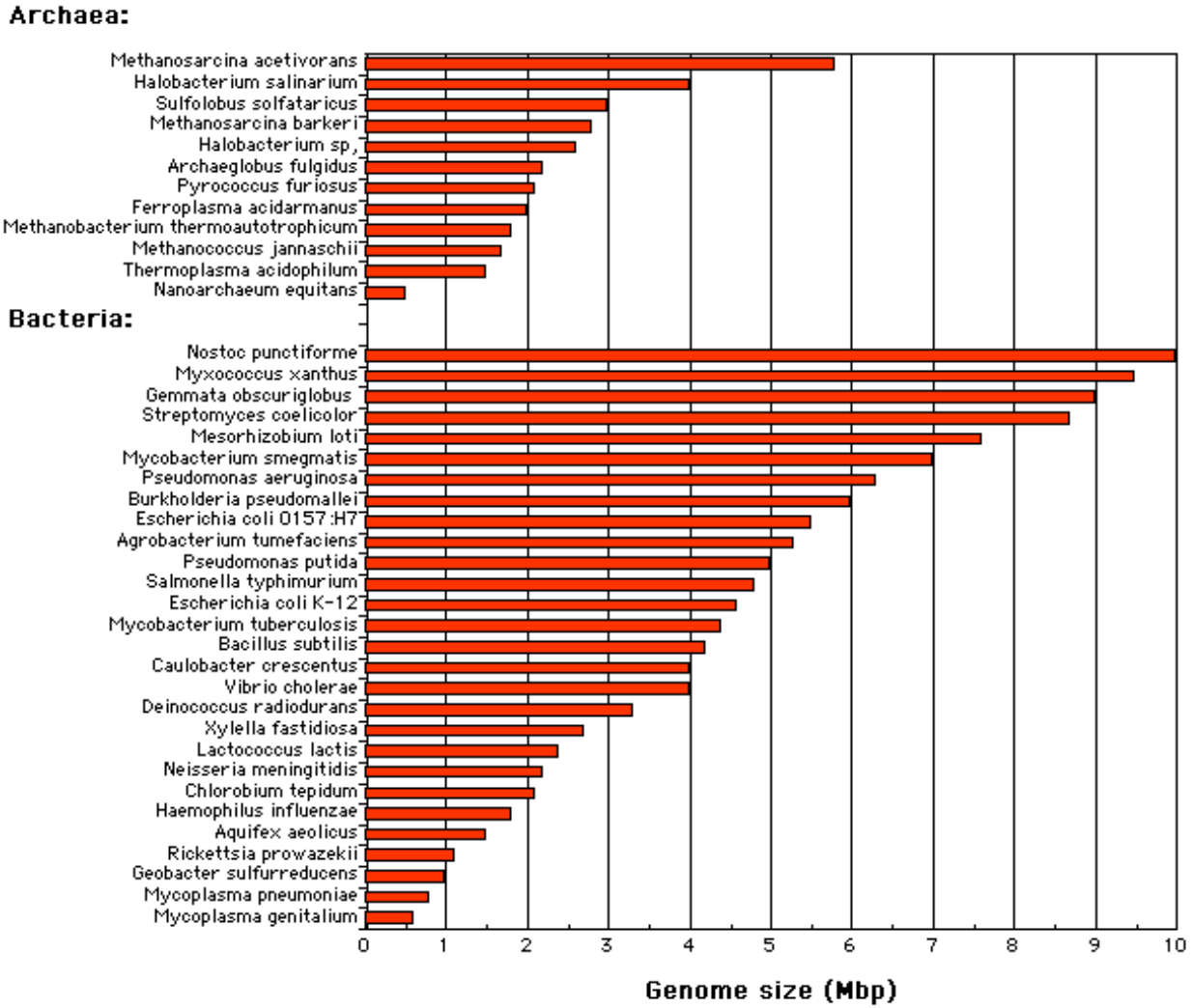
# Gene density



Evolution!!!!


Increase in DNA content involving selection for gene functions
Removal of DNA through mutation and genetic drift
Gene acquisition
Duplication
Bacterial genome
Loss of fragments encoding one or more genes
Inactivation to form a pseudogene
Erosion by deletional bias
TRENDS in Genetics

The size of prokaryotic genomes ranges from around 60 kb to more than 13 Mb.
There is a direct proportionality between the size of the genome and the number of encoded proteins.
The smallest genomes (<500 kb) correspond to obligatory endosymbionts that have arisen by reduction of larger genomes of free-living bacteria.




Touchon M, Rocha EP. Coevolution of the Organization and Structure of Prokaryotic Genomes. Cold Spring Harb Perspect Biol. 2016 Jan 4;8(1):a018168. doi: 10.1101/cshperspect.a018168. PMID: 26729648; PMCID: PMC4691797.
Cole, S., and I. Saint-Girons. 1999. Bacterial genomes -- all shapes and sizes. In R. Charlebois (ed.), Organization of the prokaryotic genome, pp. 35-62. ASM Press, Washington DC

Homologous genes can be divided in

## ORTHOLOGOUS genes

they derive from a common ancestor and encode proteins with the same function in different species

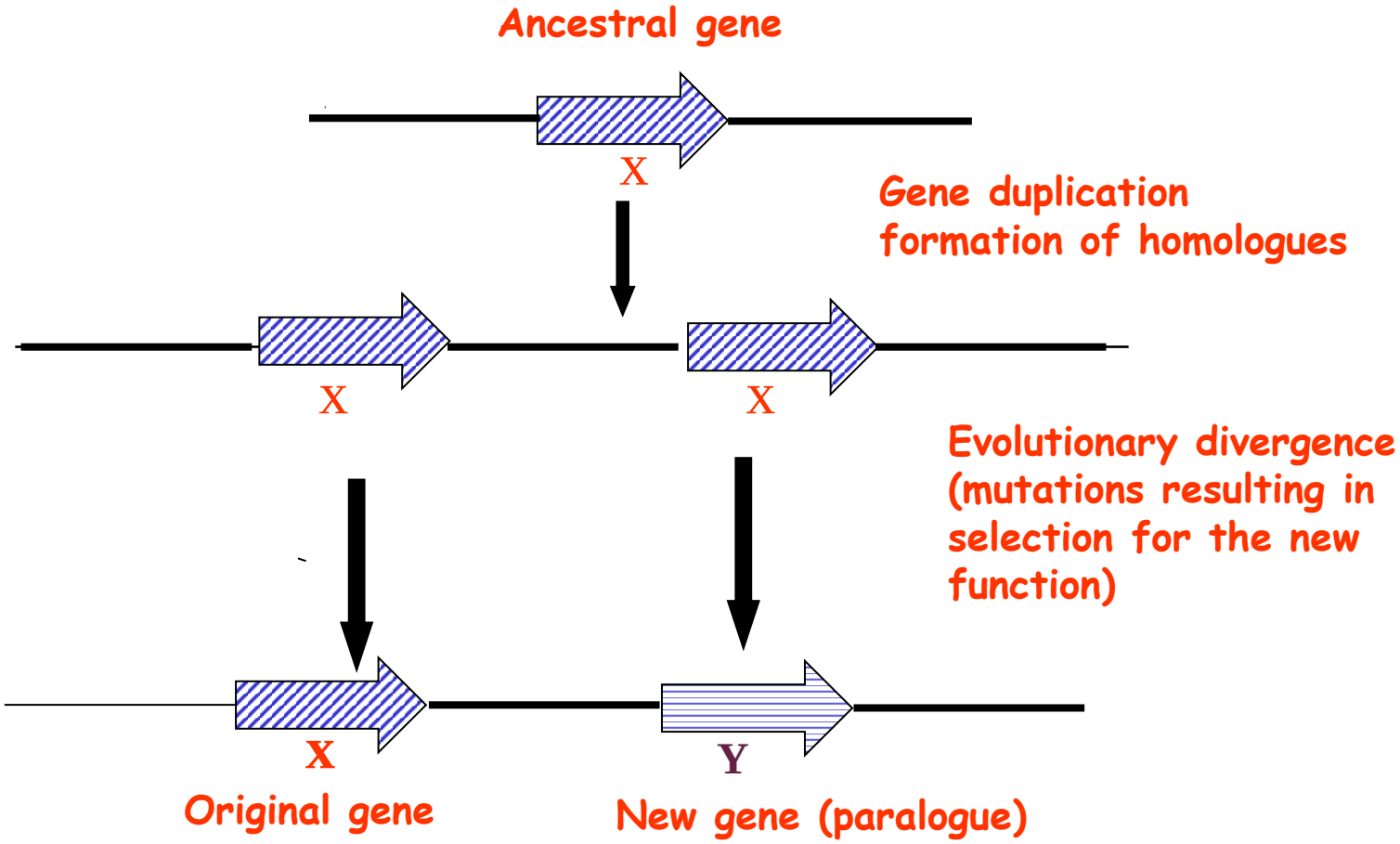
The history of these genes reflects the history of the species

## PARALOGOUS genes

they evolved by gene duplication and can encode proteins with similar but not identical function

Homology is the result of duplication rather than speciation

# Gene duplication and divergence: formation of paralogues



# What is the role of gene duplication?

Gene duplication allowed:

- a rapid diversification of enzymatic reactions
- an increase in the size of the genome with potential for the development of new enzymatic properties

following the duplication, one of the two genes can become DISPENSABLE and therefore can undergo a series of mutations or rearrangements

Generally, the paralogous genes perform different, albeit similar, functions within the same microorganism.

However, gene duplication can generate copies of a gene that maintains the same function allowing the production of large quantities of RNA or proteins (rRNA)

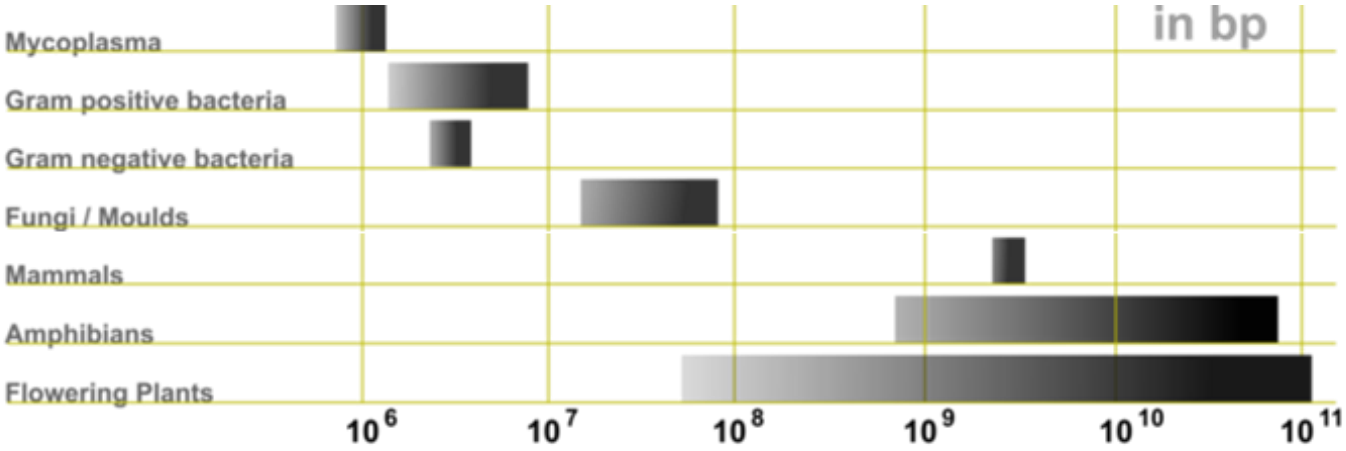
Two paralogues can then in turn give rise to duplication forming a family of paralogous genes

**Key Points**

•Genomes tend to increase in size over time, however exceptions occur.

•Genome reduction is observed in species that depend on a host for survival, the most extreme examples are eukaryotic organelles that have bacterial origins such as mitochondria.

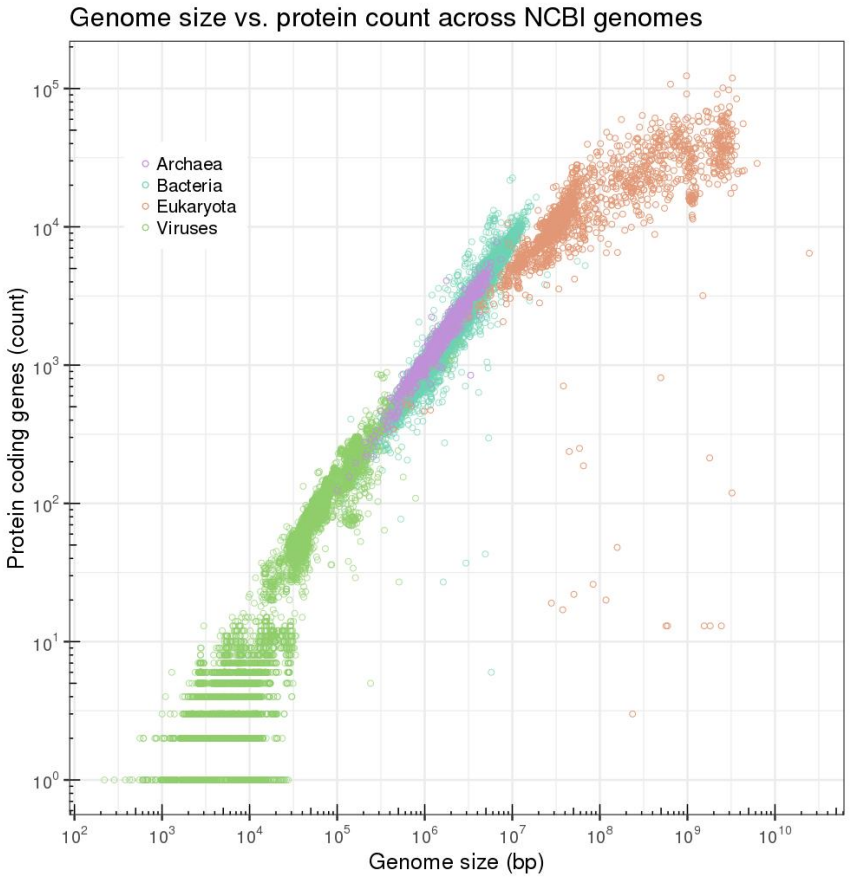
•As an endosymbiont becomes dependent on its host, it becomes an obligate endosymbiont; during this time the genome is reduced due to deletions of genes not needed to live in the host. Parts of the genome can be transferred to the host as well, leading to further genome reduction.



# How many genes?

| Species | Approximate genome size (Mb) | Approximate no. of genes | Approximate GC content | Lifestyle |
|---|---|---|---|---|
| *E. coli** | 4.7* | 4500 | 50% | Free-living and host-associated |
| *B. subtilis* | 4.2 | 4100 | 43% | Free-living |
| *S. mutans* | 1.6 | 1950 | 37% | Host-associated |
| *S. pyogenes* | 1.8 | 1800 | 39% | Host-associated |
| *S. agalactiae* | 2.2 | 2200 | 35% | Host-associated |
| *S. pneumoniae* | 2.1 | 2100 | 40% | Host-associated |
| *Staphylococcus aureus* | 2.8 | 2850 | 33% | Host-associated |
| *Listeria monocytogenes* | 2.9 | 2800 | 38% | Free-living and host-associated |

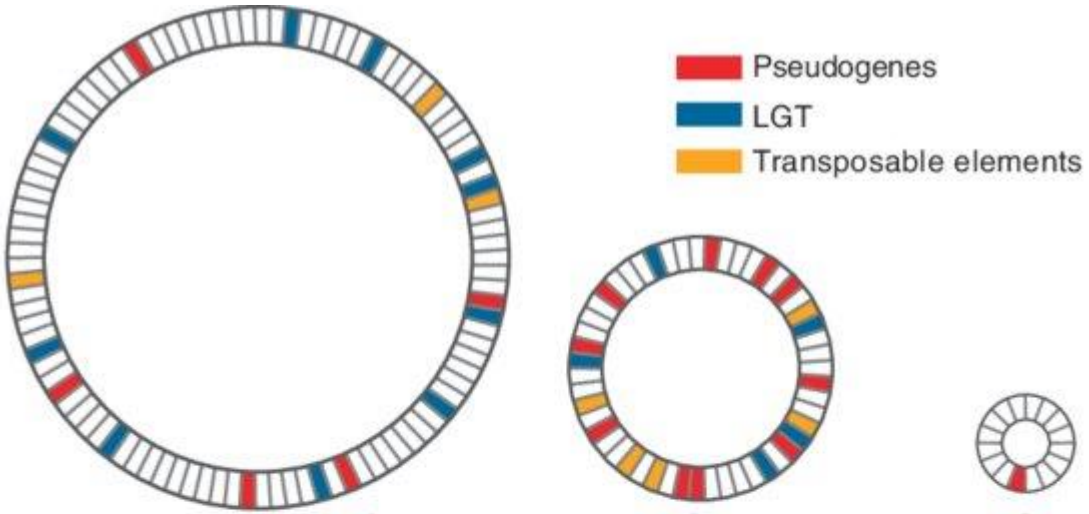
*Sequenced genomes range in size from 4.6 to 6.2 Mb.




parasitic bacteria have 500–1200 genes, free-living bacteria have 1500–7500 genes, and archaea have 1500–2700 genes

Smaller genomes lack sensory, transport, communication, and regulatory functions. Larger genomes engage much more frequently horizontal gene transfer and encode more transposable elements.

It is thus generally thought that larger genomes correspond to more versatile prokaryotes that are less sexually isolated and in which selection is more efficient.




| | Free-living | Recent or facultative pathogen | Obligate symbiont or pathogen |
|---|---|---|---|
| **Genome size** | Large (5-10 MB) | Intermediate (2-5 MB) | Small (0.5-1.5 MB) |
| **Number of pseudogenes** | Few | Many | Rare |
| **Incidence of LGT** | Frequent | Frequent to rare | Rare to none |
| **Selfish genetic elements** | Few | Common | Rare |
| **Genome organization** | Stable or unstable | Unstable | Stable |
| **Effective population size** | Large | Small | Small |

Bacteria with small genomes depend on the host for numerous functions: glycolysis, biosynthesis of amino acids and purines and pyrimidines

# *Microorganisms with large genomes*

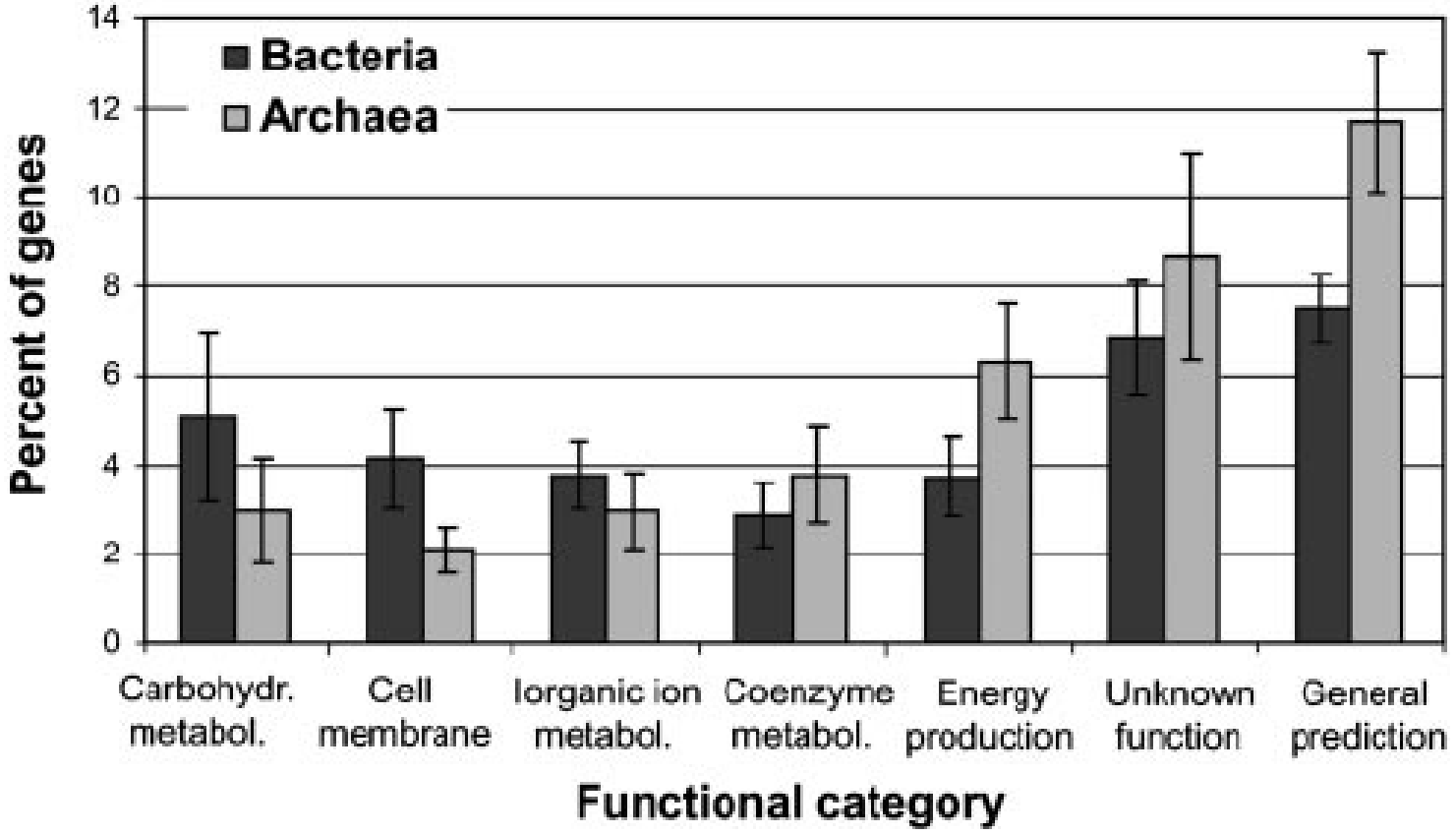
Organisms with large genomes possess the ability to code for many genes involved in both metabolism and regulatory processes.

The regulatory mechanisms allow the cell to respond better to the availability of different substrates through the activation of specific genes.

ll microorganisms with genomes larger than 6 Mbp are soil microorganisms.

Soil is a habitat in which carbon and energy sources are often scarce or available in different forms and often usable intermittently.

# Variation of gene categories in Bacteria and Archaea




General prediction = genes encoding hypothetical proteins that they may or may not exist

# The genomes of the Archaea are characterized by:

- higher percentage of genes involved in energy production
- greater number of coenzymes (especially in methanogenic Archaea)

- large number of genes of unknown function
- large number of genes encoding hypothetical proteins

- reduced number of genes involved in carbohydrate metabolism
- reduced number of genes linked to cell membrane functions such as membrane transport and biosynthesis

## The genomes of bacteria are characterized by:

- a large number of genes for carbohydrate metabolism
- a significant number of genes for membrane-related functions
- a large number of genes still of unknown function
- a large number of hypothetical proteins

Compared to the Archea
- fewer genes for coenzyme metabolism
- fewer genes for energy production

**In addition to the size of the genome, membership in the Domain (Bacteria or Archaea) also seems to influence the functional categorization of genes in prokaryotes**

# The research article selection by the Prof.

RESEARCH ARTICLE

SYNTHETIC BIOLOGY

## Design and synthesis of a minimal bacterial genome

Clyde A. Hutchison III,[1*†] Ray-Yuan Chuang,[1†‡] Vladimir N. Noskov,[1] Nacyra Assad-Garcia,[1] Thomas J. Deerinck,[2] Mark H. Ellisman,[2] John Gill,[3] Krishna Kannan,[3] Bogumil J. Karas,[1] Li Ma,[1] James F. Pelletier,[4§] Zhi-Qing Qi,[3] R. Alexander Richter,[1] Elizabeth A. Strychalski,[4] Lijie Sun,[1||] Yo Suzuki,[1] Billyana Tsvetanova,[3] Kim S. Wise,[1] Hamilton O. Smith,[1,3] John I. Glass,[1] Chuck Merryman,[1] Daniel G. Gibson,[1,3] J. Craig Venter[1,3*]

We used whole-genome design and complete chemical synthesis to minimize the 1079–kilobase pair synthetic genome of *Mycoplasma mycoides* JCVI-syn1.0. An initial design, based on collective knowledge of molecular biology combined with limited transposon mutagenesis data, failed to produce a viable cell. Improved transposon mutagenesis methods revealed a class of quasi-essential genes that are needed for robust growth, explaining the failure of our initial design. Three cycles of design, synthesis, and testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531 kilobase pairs, 473 genes), which has a genome smaller than that of any autonomously replicating cell found in nature. JCVI-syn3.0 retains almost all genes involved in the synthesis and processing of macromolecules. Unexpectedly, it also contains 149 genes with unknown biological functions. JCVI-syn3.0 is a versatile platform for investigating the core functions of life and for exploring whole-genome design.

tial genes. These results showed that it should be possible to produce a minimal genome that is smaller than any found in nature, but that the minimal genome would be larger than the common set of 256 genes. At that time, we proposed to create and test a cassette-based minimal artificial genome (*5*). We have been working since then to produce the tools needed to accomplish this. We developed methods to chemically synthesize the *M. genitalium* genome (*7*). However, *M. genitalium* grows very slowly, so we turned to the faster-growing *M. mycoides* genome as our target for minimization. We developed the method of genome transplantation, which allowed us to introduce *M. mycoides* genomes, as isolated DNA molecules, into cells of a different species, *M. capricolum* (*8*, *9*). In this process, the *M. capricolum* genome is lost, resulting in a cell containing only the transplanted genome. In 2010, we reported the complete chemical synthesis and installation of the genome of *M. mycoides* JCVI-syn1.0 [1,078,809 base pairs (bp) (*10*); hereafter abbreviated syn1.0). This genome was an almost exact copy of the wild-type *M. mycoides* genome, with the addition of a few watermark and vector sequences.

Genome reduction in bacteria such as *E. coli* and *B. subtilis* has previously been achieved by a series of sequential deletion events (*11*, *12*). After each deletion, viability, growth rate, and other phenotypes can be determined. In contrast to this approach, we set out to design a

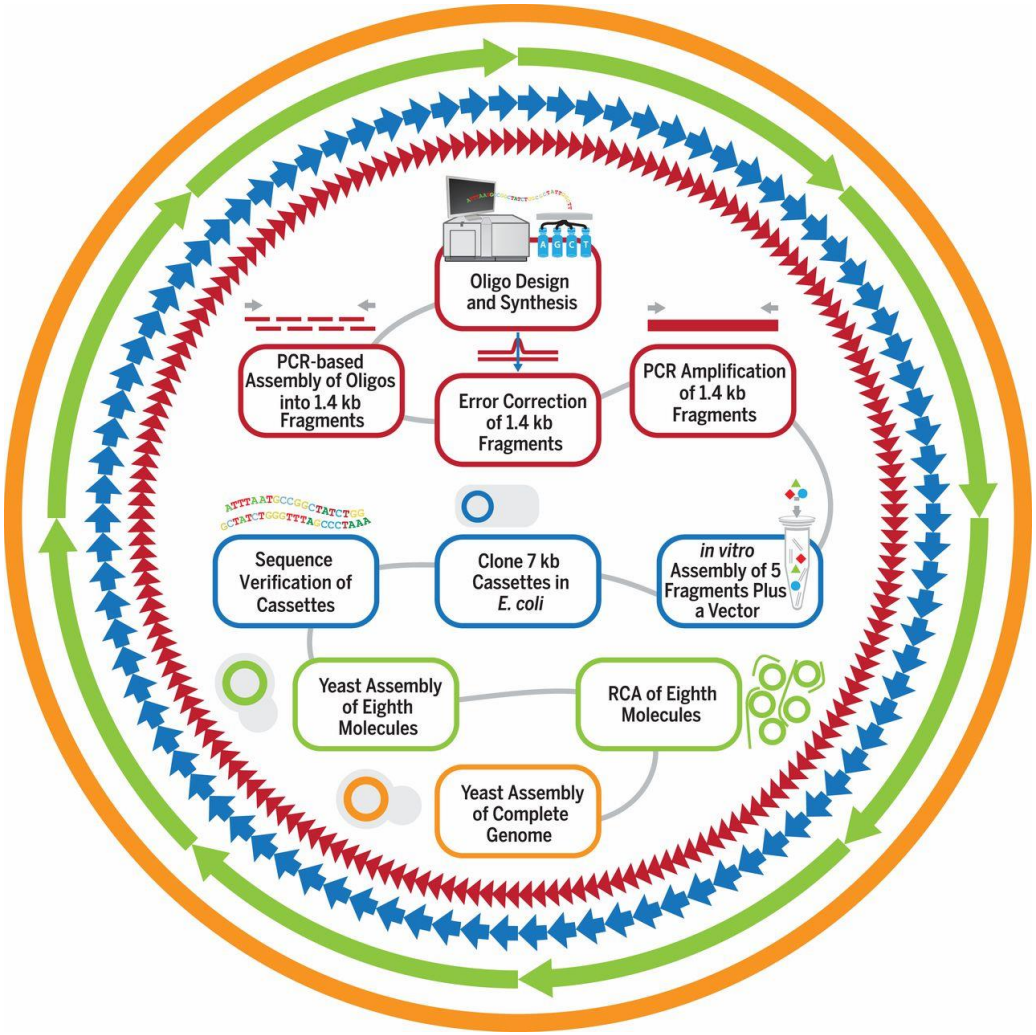


Hutchison CA 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC. Design and synthesis of a minimal bacterial genome. Science. 2016 Mar 25;351(6280):aad6253. doi: 10.1126/science.aad6253.
Erratum in: ACS Chem Biol. 2016 May 20;11(5):1463. PMID: 27013737.

# Design and synthesis of a minimal bacterial genome

A goal in biology is to understand the molecular and biological function of every gene in a cell. One way to approach this is to build a minimal genome that includes only the genes essential for life. In 2010, a 1079-kb genome based on the genome of *Mycoplasma mycoides* (JCV-syn1.0) was chemically synthesized and supported cell growth when transplanted into cytoplasm. Hutchison III *et al.* used a design, build, and test cycle to reduce this genome to 531 kb (473 genes). The resulting JCV-syn3.0 retains genes involved in key processes such as transcription and translation, but also contains 149 genes of unknown function.

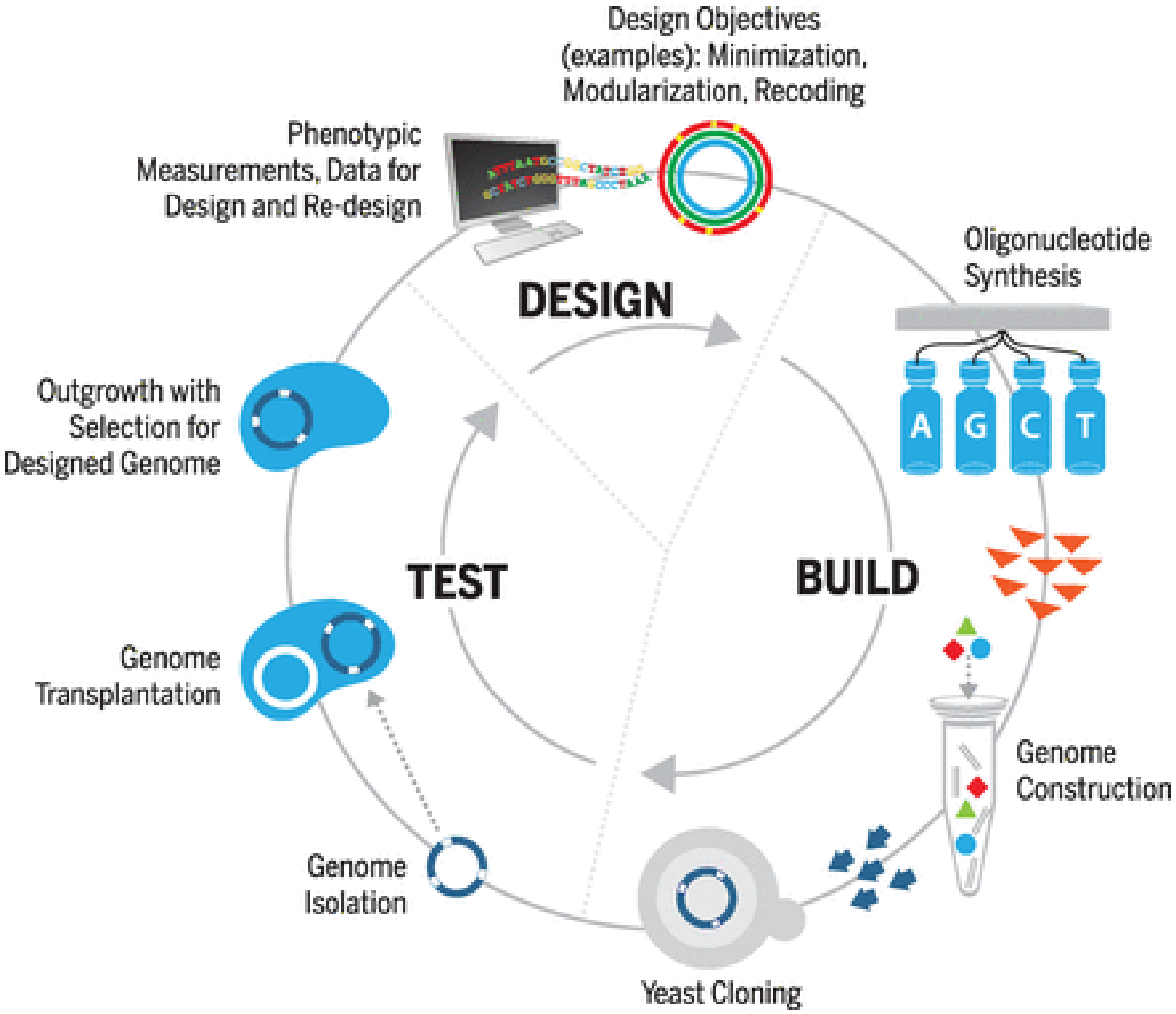
**Strategy for whole-genome synthesis.**




Overlapping oligonucleotides (oligos) were designed, chemically synthesized, and assembled into 1.4-kbp fragments (red). After error correction and PCR amplification, five fragments were assembled into 7-kbp cassettes (blue). Cassettes were sequence-verified and then assembled in yeast to generate one-eighth molecules (green). The eight molecules were amplified by RCA and then assembled in yeast to generate the complete genome (orange).

**Clyde A. Hutchison III et al. Science 2016;351:aad6253**

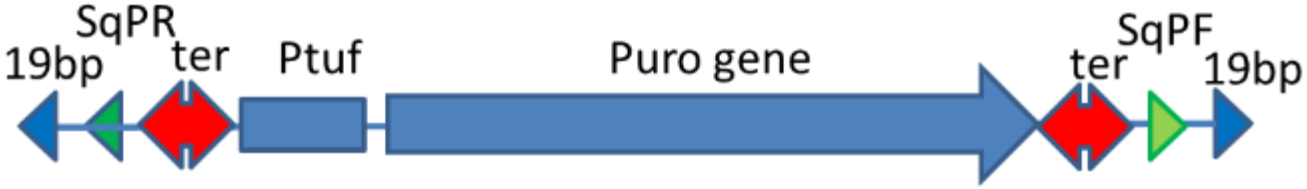
# DBT: Design- Build- Test



“We started with syn1.0 and used information from the biochemical literature, as well as some transposon mutagenesis data, to produce a rational design. Genes that could be disrupted by transposon insertions without affecting cell viability were considered to be nonessential. Based on ~16,000 transposon 4001 (Tn*4001*) and Tn*5* insertions into the syn1.0 genome, we were able to find and delete a total of 440 apparently nonessential genes from the syn1.0 genome. The resulting HMG design was 483 kbp in size and contained 432 protein genes and 39 RNA genes.

All eight HMG segments were tested in a syn1.0 background, but only one of the segment designs produced viable colonies (HMG segment 2), and the cells grew poorly.”

**Step 1.** Construct Tn5 transposon containing 19 bp mosaic ends, sequencing primer sites, terminator sequences, and a selectable marker (puromycin-resistance gene). Bind Tn5 transposase (Epicentre) to the 19bp termini to form the active transposome.




**Step 2.** Introduce transposome into *Mycoplasma mycoides* JCVI-syn1.0 R-M (minus) strain by polyethylene glycol (PEG) transformation method. Collect puromycin-resistant colonies and serially propagate to eliminate slow growers.

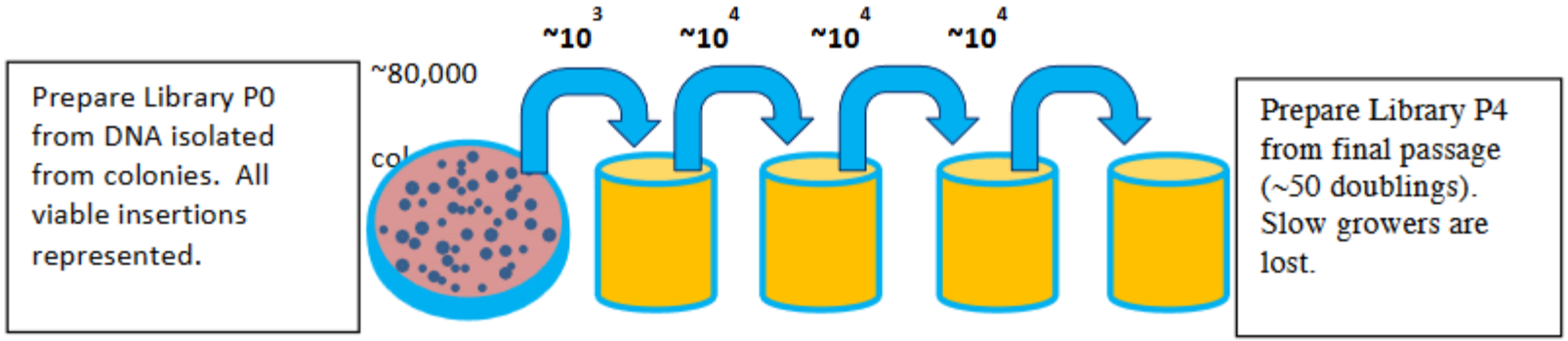
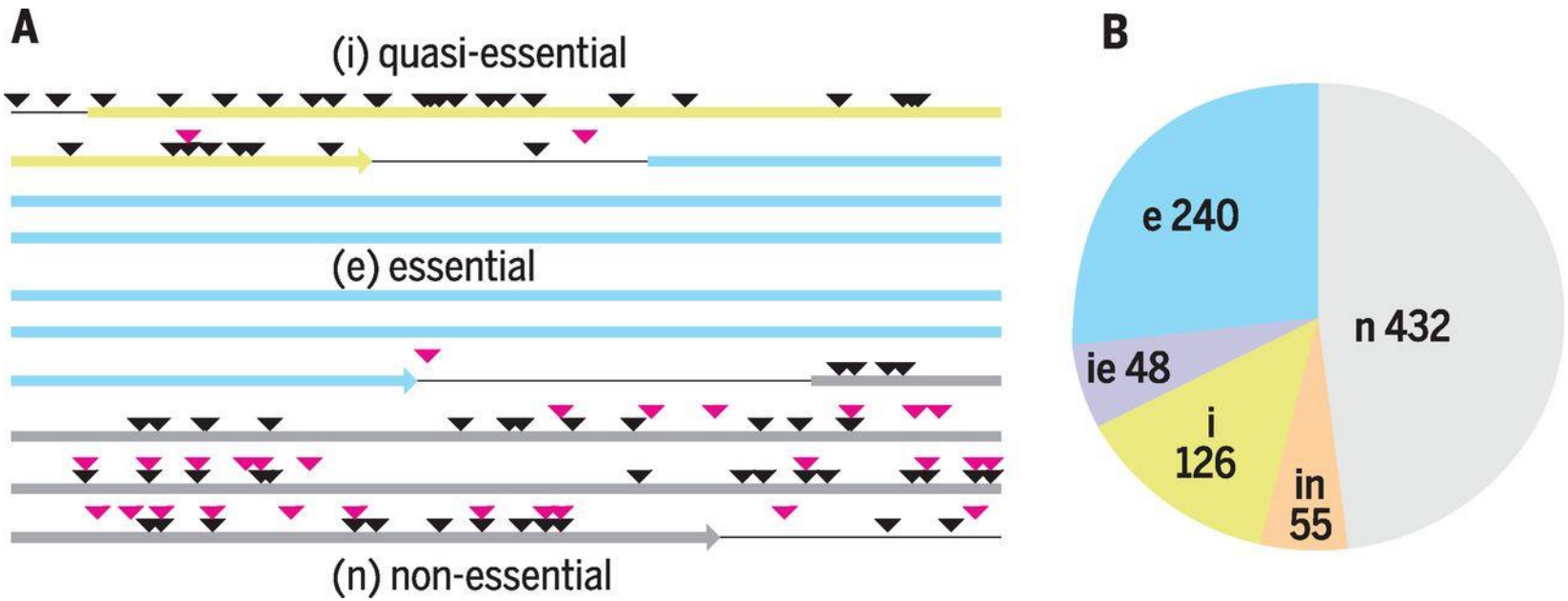
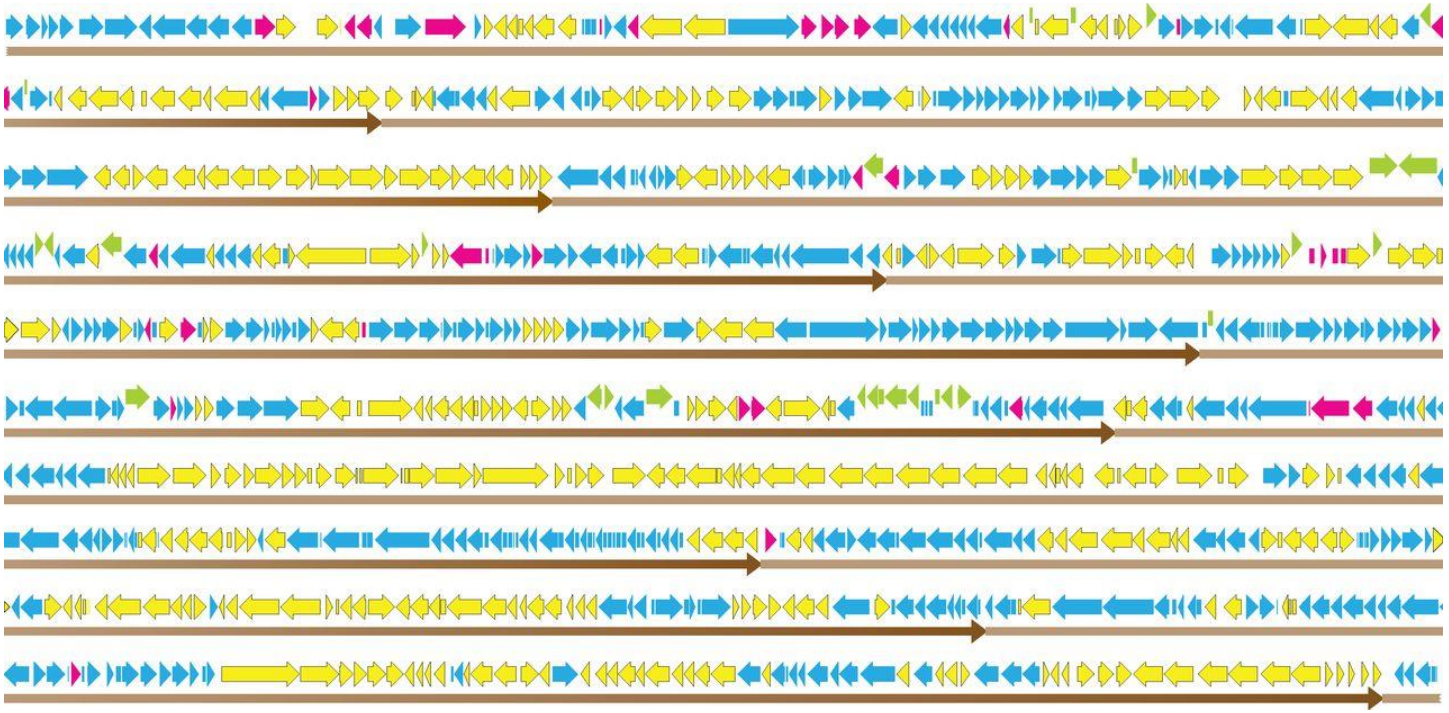


**Fig. 3 Classification of gene essentiality by transposon mutagenesis.**




Clyde A. Hutchison III et al. Science 2016;351:aad6253



The three DBT cycles involved in building syn3.0.
This detailed map shows syn1.0 genes that were deleted or added back in the various DBT cycles leading from syn1.0 to syn2.0 and finally to syn3.0 (compare with fig. S7). The long brown arrows indicate the eight NotI assembly segments. **Blue arrows represent genes that were retained** throughout the process. **Genes that were deleted** in both syn2.0 and syn3.0 **are shown in yellow**. **Green arrows (slightly offset) represent genes that were added back**. **The original RGD1.0 design was not viable, but a combination of syn1.0 segments 1, 3, 4, and 5 and designed segments 2, 6, 7, and 8 produced a viable cell, referred to as RGD2678**. Addition of the genes shown in green resulted in syn2.0, which has eight designed segments. Additional deletions, shown in magenta, produced syn3.0 (531,560 bp, 473 genes). The directions of the arrows correspond to the directions of transcription and translation.

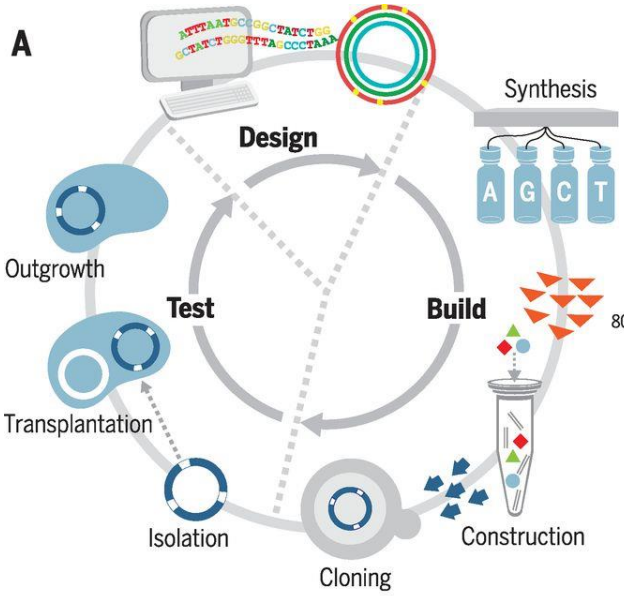
# Fig. 1 The JCVI DBT cycle for bacterial genomes.

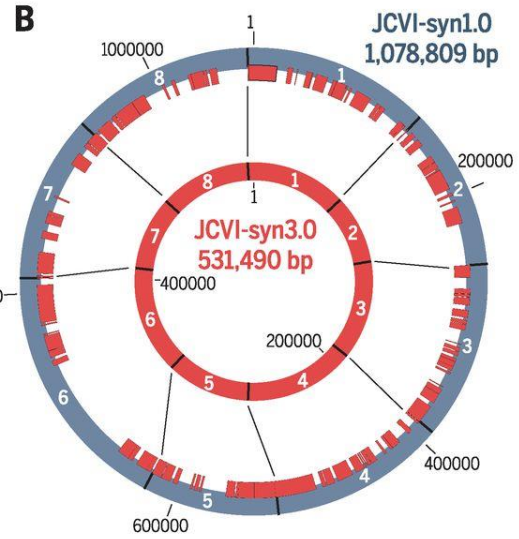
Four design-build-test cycles produced JCVI-syn3.0.

(A) The cycle for genome design, building by means of synthesis and cloning in yeast, and testing for viability by means of genome transplantation. After each cycle, gene essentiality is reevaluated by global transposon mutagenesis.

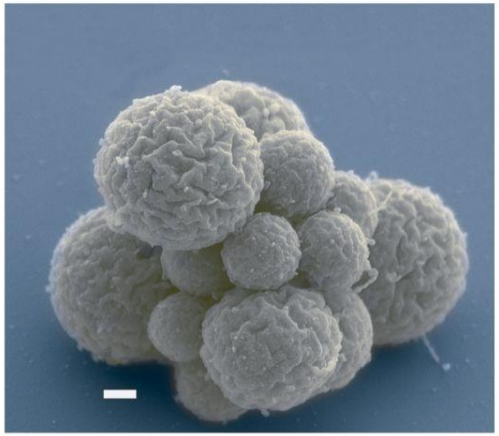
(B) Comparison of JCVI-syn1.0 (outer blue circle) with JCVI-syn3.0 (inner red circle), showing the division of each into eight segments. The red bars inside the outer circle indicate regions that are retained in JCVI-syn3.0.

(C) A cluster of JCVI-syn3.0 cells, showing spherical structures of varying sizes (scale bar, 200 nm).



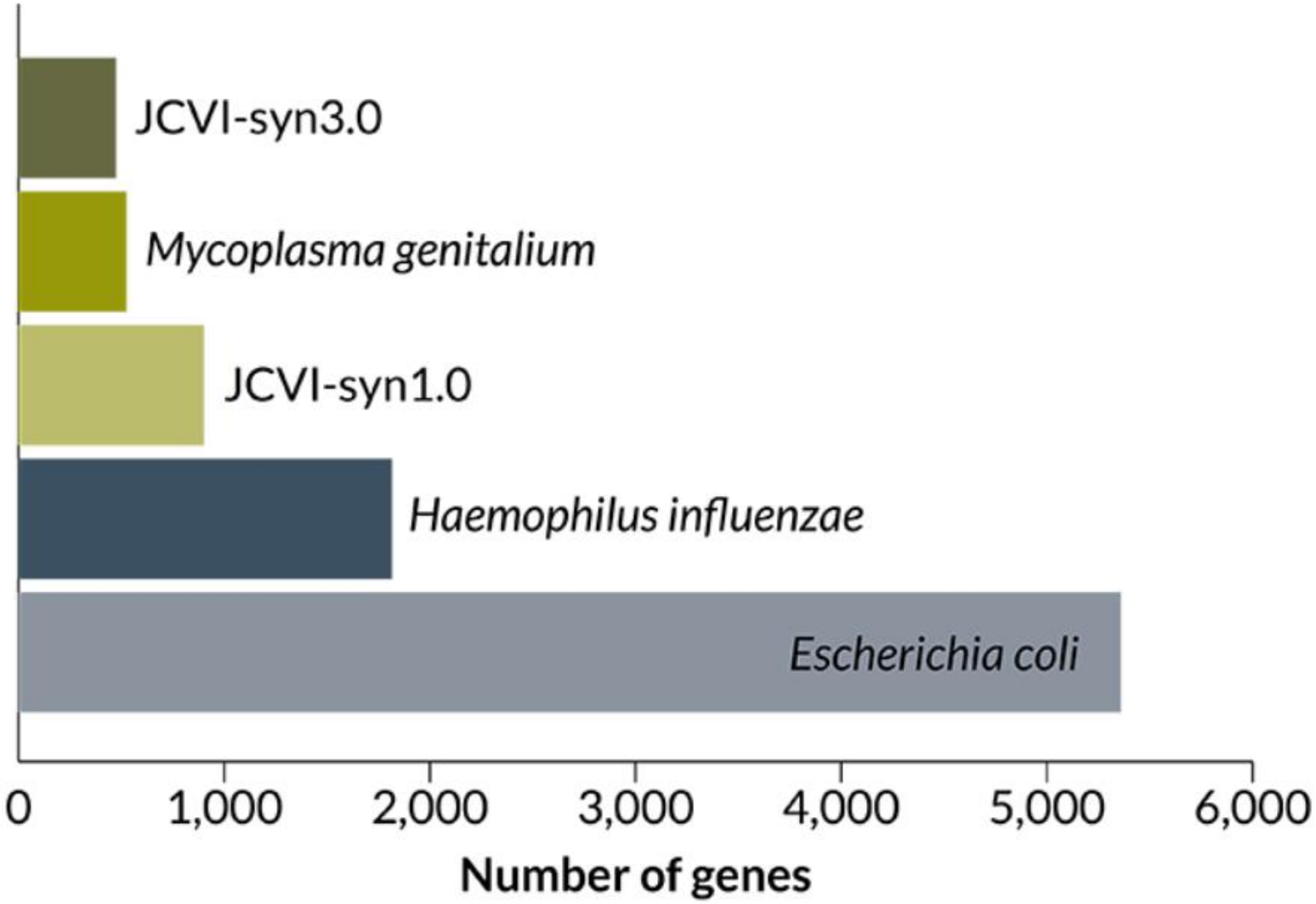

C



**DBT: Design- Build- Test**

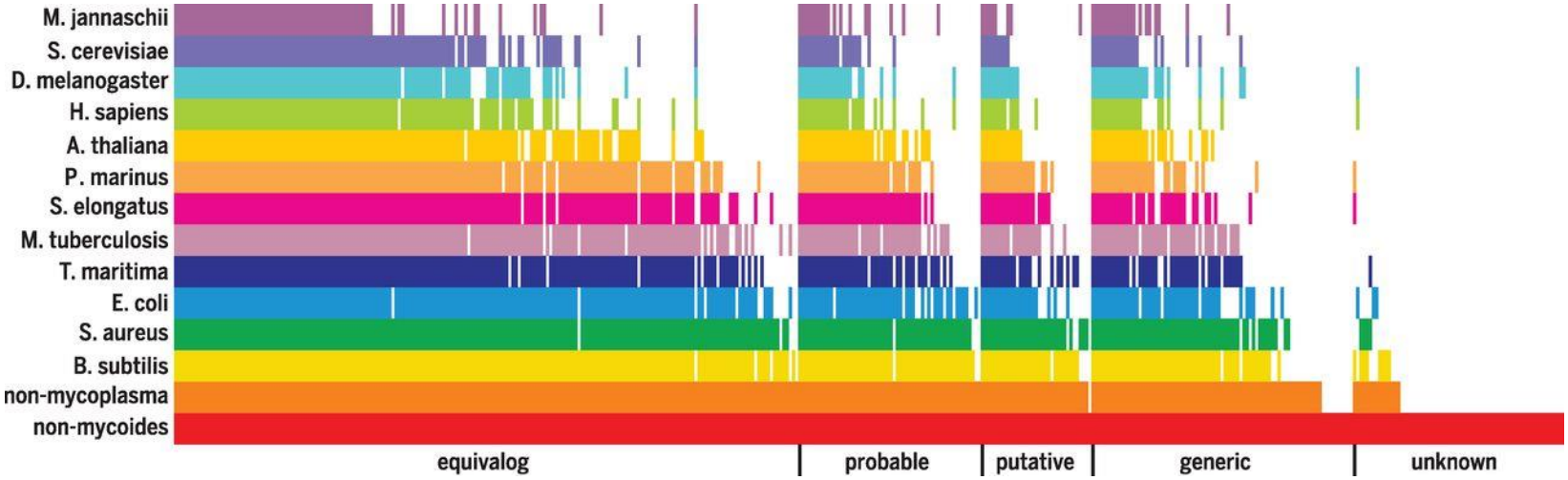
Published by AAAS

Clyde A. Hutchison III et al. Science 2016;351:aad6253



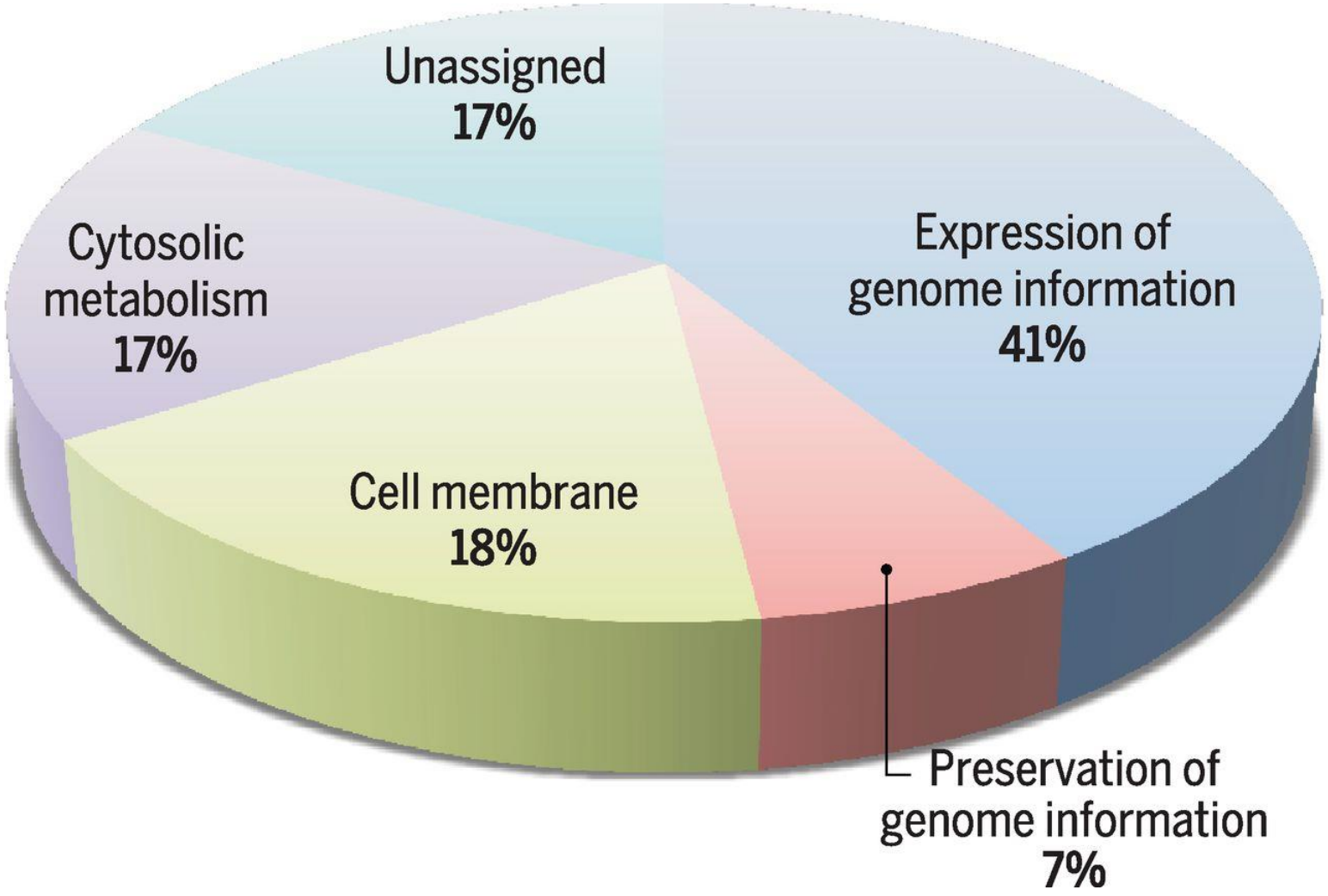

Clyde A. Hutchison III et al. Science 2016;351:aad6253

# In syn3.0, 149 genes cannot be assigned a specific biological function




Clyde A. Hutchison III et al. Science 2016;351:aad6253

**Fig. 6 Partition of genes into four major functional groups.**




Clyde A. Hutchison III et al. Science 2016;351:aad6253

# Syn3.0 has a doubling time of 3 hours and is polymorphic in appearance

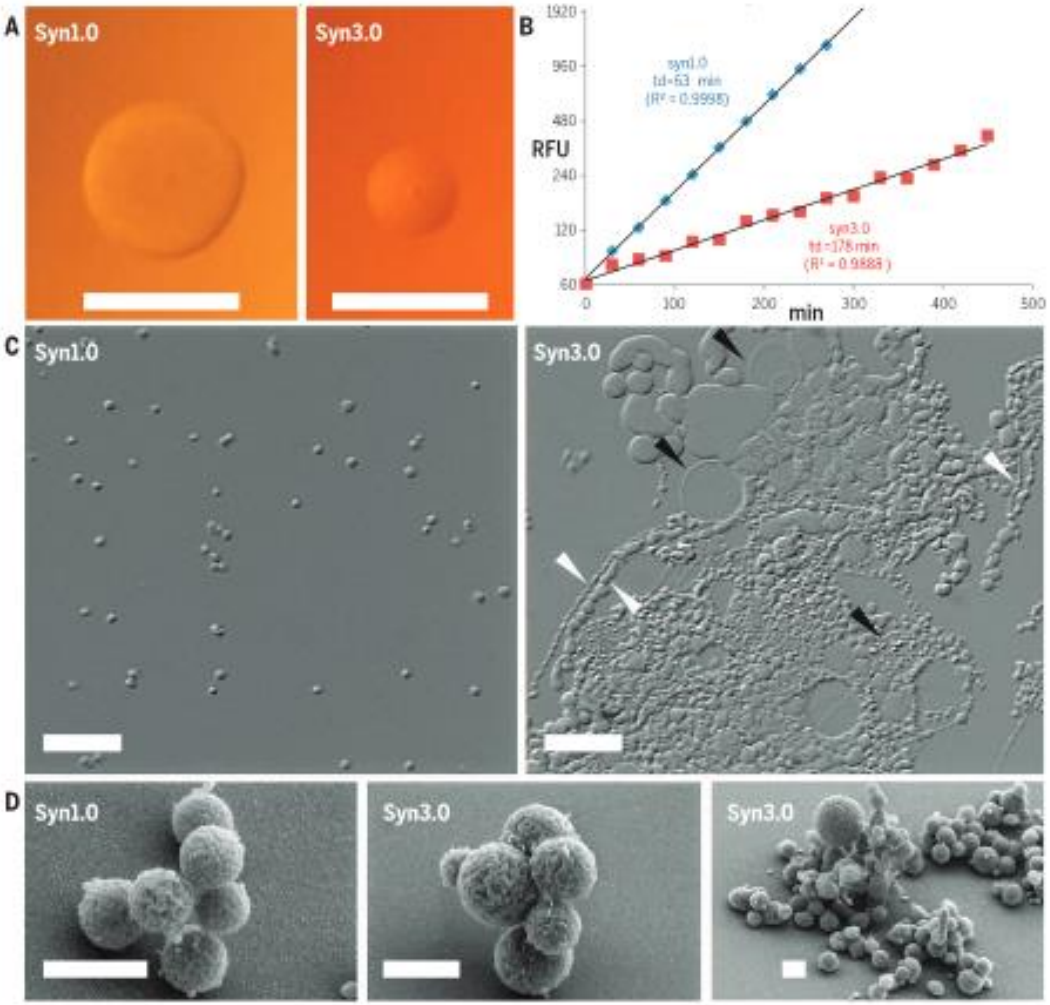



**Fig. 7. Comparison of syn1.0 and syn3.0 growth features.** (**A**) Cells derived from 0.2 μm–filtered liquid cultures were diluted and plated on agar medium to compare colony size and morphology after 96 hours (scale bars, 1.0 mm). (**B**) Growth rates in liquid static culture were determined using a fluorescent measure (relative fluorescent units, RFU) of double-stranded DNA accumulation over time (minutes) to calculate doubling times (td). Coefficients of determination ($R^2$) are shown. (**C**) Native cell morphology in liquid culture was imaged in wet mount preparations by means of differential interference contrast microscopy (scale bars, 10 μm). Arrowheads indicate assorted forms of segmented filaments (white) or large vesicles (black). (**D**) Scanning electron microscopy of syn1.0 and syn3.0 (scale bars, 1 μm). The picture on the right shows a variety of the structures observed in syn3.0 cultures.

Clyde A. Hutchison III et al. Science 2016;351:aad6253

“A minimal cell is usually defined as a cell in which all genes are essential. This definition is incomplete, because the genetic requirements for survival, and therefore the minimal genome size, depend on the environment in which the cell is grown. The work described here has been conducted in medium that supplies virtually all the small molecules required for life. A minimal genome determined under such permissive conditions should reveal a core set of environment-independent functions that are necessary and sufficient for life. Under less permissive conditions, we expect that additional genes will be required.
There is a large body of literature concerning the minimal cell concept and minimal sets of essential genes in a number of organisms [for a review, see (22)]. Work in the area has focused on comparative genomic analyses and on experiments in which genes are individually knocked out or disrupted by transposon insertion. **Such studies identify a core of essential genes, often about 250 in number. But this is not a set of genes that is sufficient to constitute a viable cellular genome, because redundant genes for essential functions are scored as nonessential in these studies.”**

**Clyde A. Hutchison III et al. Science 2016;351:aad6253**