

# **MOLEKULÁRNÍ BIOLOGIE PROKARYOT (PS 2019)**

*Jiří Doškař, Ivana Mašlaňová*

# **Sylabus předmětu Molekulární biologie prokaryot**

- 1. Úvodní přednáška (obsah, literatura, význam prokaryot)**
  - Genom prokaryot a charakteristika jeho složek**
- 2. Plazmidy I**
- 3. Plazmidy II**
- 4. Prokaryotické transponovatelné elementy**
- 5. Využití transpozonů ke studiu prokaryotického genomu, transpozonová mutagenese**
- 6. HGT - Transformace**
- 7. HGT - Transdukce**
- 8. HGT - Konjugace**
- 8. Mutace bakterií**
- 9. Reparace mutačních poškození DNA u bakterií**
- 10. Obranné mechanismy bakterií vůči externí DNA (RM systémy)**
- 11. Obranné mechanismy bakterií vůči externí DNA (CRISPR-Cas)**
- 12. Evoluce bakteriálních genomů**

# DOPORUČENÁ LITERATURA

- **Maloy S.R. et al.:** Microbial Genetics, 5. vydání, Jones and Barlett Publ, Boston-London, 2010.
- **Birge E.A.:** Bacterial and bacteriophage genetics. Springer 2000, 2006.
- **Snyder L., Peters JE, Henkin TM, Champness W.:** Molecular genetics of bacteria, 4nd ed. , ASM Press 2013.
- **Miller R. V., Day M. J.:** Microbial evolution – Gene establishment, Survival and Evolution. ASM Press, Washington, D.C., 2004.
- **Funnell B.E., Phillips G.J.:** Plasmid Biology, ASM Press, Washington, D.C., 2004, 2015
- **Persing D.H.** ed. Molecular Microbiology – Diagnostic Principles and Practice, ASM Press, Washington, 2011
- **Snustad P., Simmons MJ.:** Genetika, český překlad, MU 2009, **2017.**
- **M.P. Francino:** Horizontal gene transfer in Microorganisms. Caister Acad. Press, Norfolk, UK 2012.
- **Alberts et al.:** Molecular biology of the cell. Garland Publ. 2015.

# Informace ke zkoušce

Zkouška má dvě části:

**A. Písemná příprava** – student si připraví odpovědi na 7 otázek, pokrývajících vybrané tématické okruhy z probírané látky.

**Příklady otázek:**

- 1. Znázorněte schematicky a popište strukturu Ti-plazmidu a funkce jeho genů**
- 2. Znázorněte schematicky, jak byla experimentálně prokázána konzervativní transpozice Tn10**
- 3. Uveďte základní rysy přirozené transformace (charakter přijímané DNA, průběh přenosu, způsob začlenění DNA do genomu recipienta)**
- 4. Vysvětlete princip SOS-mutagenese**
- 5. Shrňte rozdíly v procesu konjugace u G- a G+ bakterií.**
- 6. Znázorněte schematicky a popište fungování systému CRISPR/Cas**
- 7. Vysvětlete pojem abortivní transdukce**

**B. Ústní část** - student doplňuje informace k jednotlivým odpovědím na základě dotazů zkoušejících.

Písemná příprava trvá 60 minut, ústní část 10-30 minut.

# **VÝZNAM PROKARYOT Z HLEDISKA MOLEKULÁRNÍ BIOLOGIE A GENETIKY**

- Modelové organismy pro studium základních biologických procesů**
- Využití bakterií v základním výzkumu (klonování genů, studium jejich exprese, atd)**
- Výhody**
  - krátká generační doba, vysoký počet jedinců**
  - snadná kultivace na definovaných půdách, znalost metabolických drah**
  - existence dobře definovaných mutant**
  - dobře přístupné studiu, snadná izolace biol. makromolekul a struktur**
  - studium evoluce - možnost sledování genetických změn v krátkých časových intervalech**

# **PRAKTICKÝ VÝZNAM PROKARYOTICKÝCH ORGANISMŮ**

- patogenita řady druhů (parazitismus - symbioza)**
- využití v tradičních biotechnologiích (kvasný průmysl, příprava nápojů a potravin)**
- producenti významných látek (antibiotika, enzymy, farmaka, sekundární metabolity)**
- producenti cizorodých látek připravených metodami GI (např. inzulin, hormony, protilátky aj.)**
- příprava probiotických preparátů (úprava mikrobiomu, mikrobiota)**
- příprava vakcín genetickou modifikací („infektomika“)**
- využití extrémofilů a jejich produktů (bakterie a archea)**

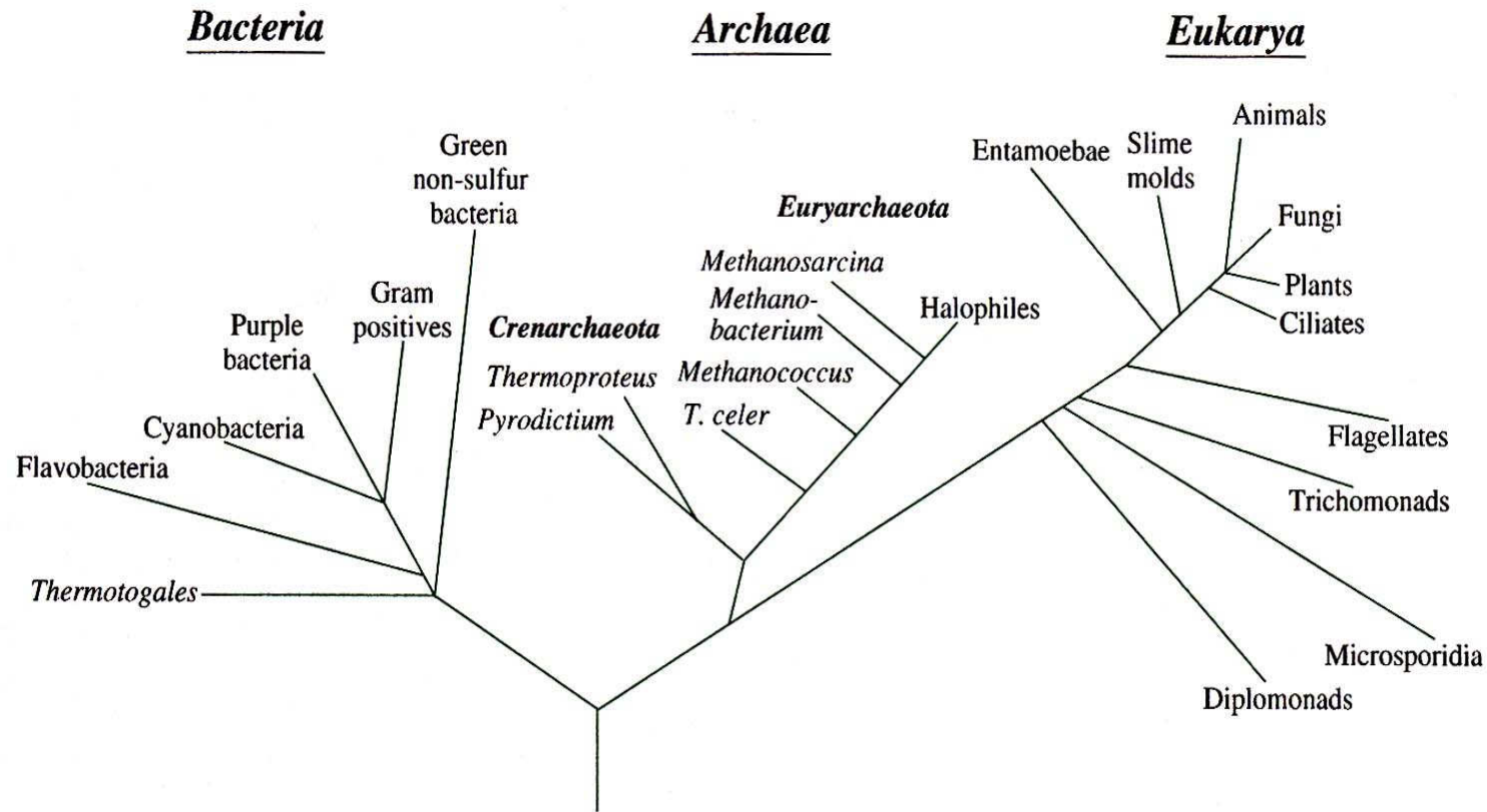
# Současné trendy v oblasti molekulární biologie prokaryot

## A. poznání struktury genomu a funkce jeho složek

- vytváření cílených změn genomu za účelem navození nových vlastností:
- - příprava produkčních kmenů v biotechnologiích
- - **Interakce bakterií s hostitelem (patogenní a potenciálně patogenní druhy) – antigenní struktury x imunitní odpověď**
- - příprava vakcín – metody genového inženýrství
- - vyhledání buněčných cílů pro nová terapeutika

## B. objasnění mechanismů evoluce

- charakter (makro)mutačních změn genomů
- úloha horizontálního přenosu genů v evoluci genomů (objasnění vzniku kmenů rezistentních k léčivům nebo kmenů se zvýšenou virulencí, vznik nových druhů)



UNIVERSAL PHYLOGENETIC TREE.

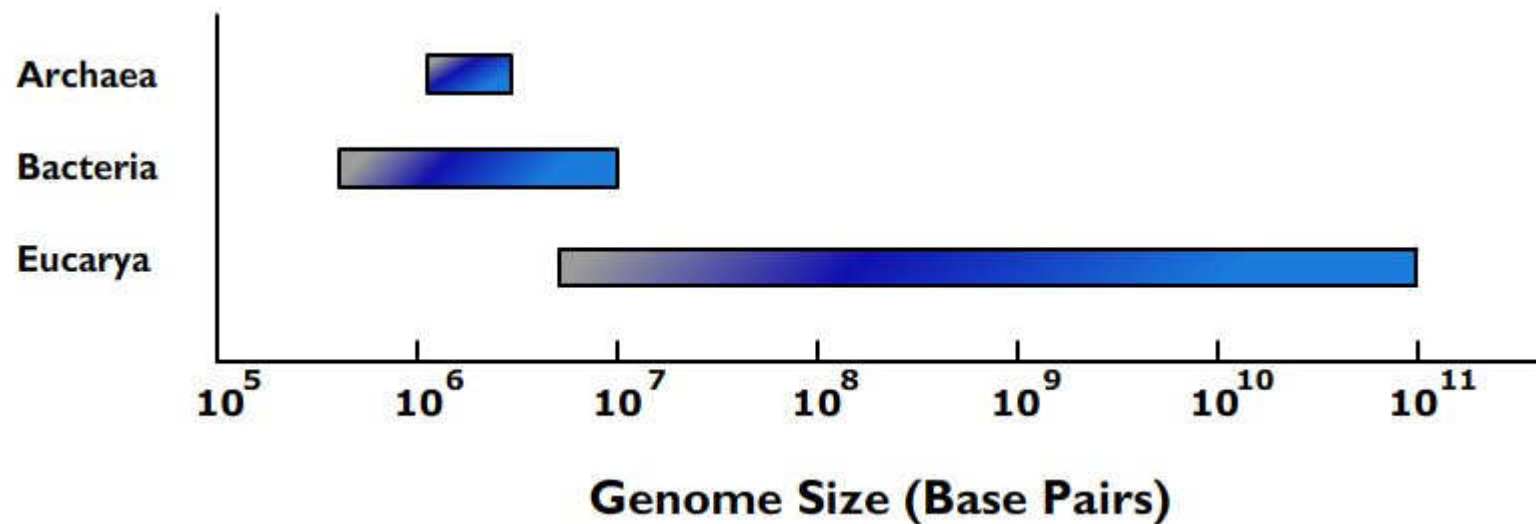
## Více než 30 000 popsaných bakteriálních druhů

According to a new estimate, **there** are about one trillion **species** of microbes on Earth, and 99.999 percent of them have yet to be discovered.

**10,000,000,000,000,000,000 (10<sup>19</sup>) tj. deset TRILIONŮ**

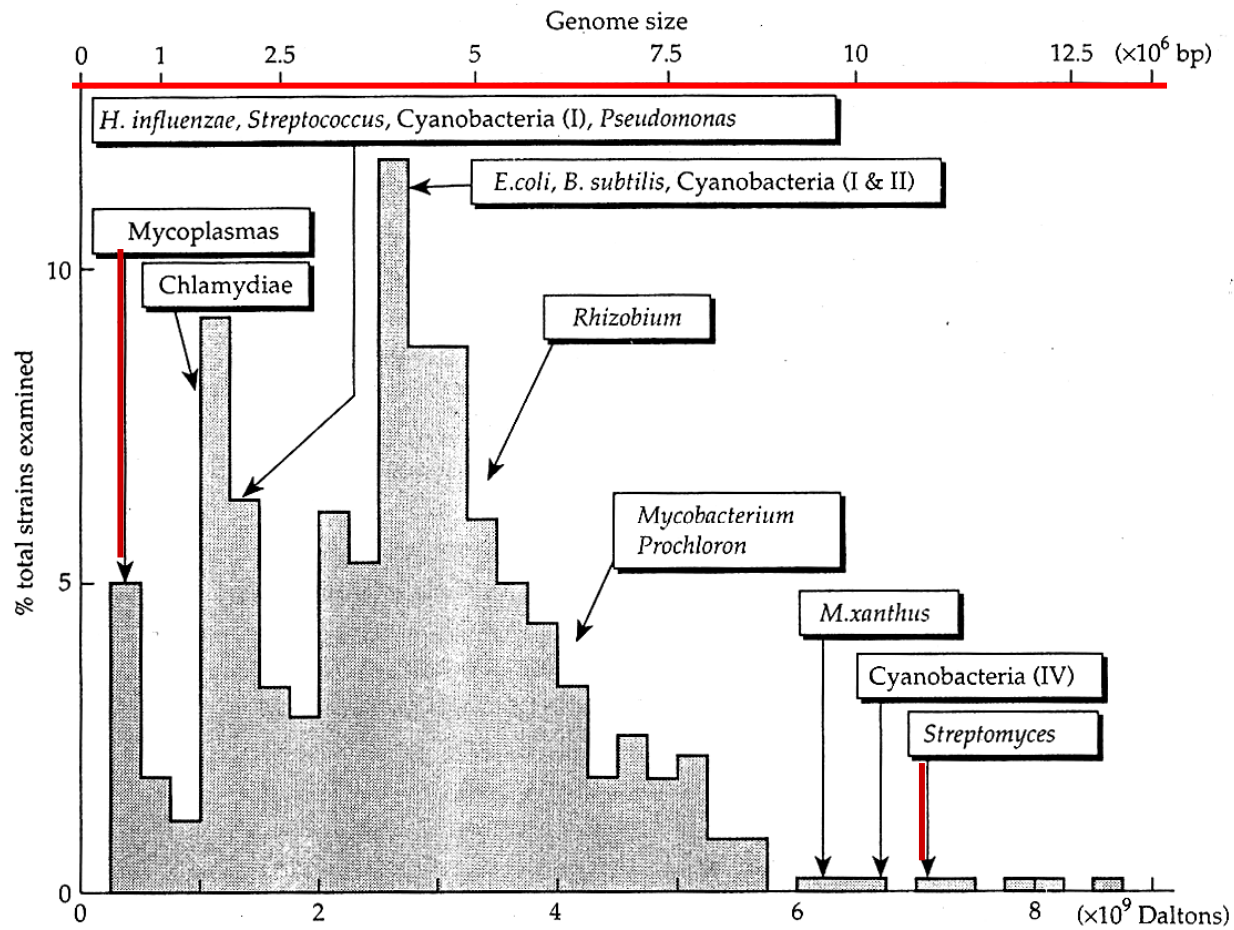


# SROVNÁNÍ VELIKOSTI GENOMŮ PROKARYOT A EUKARYOT



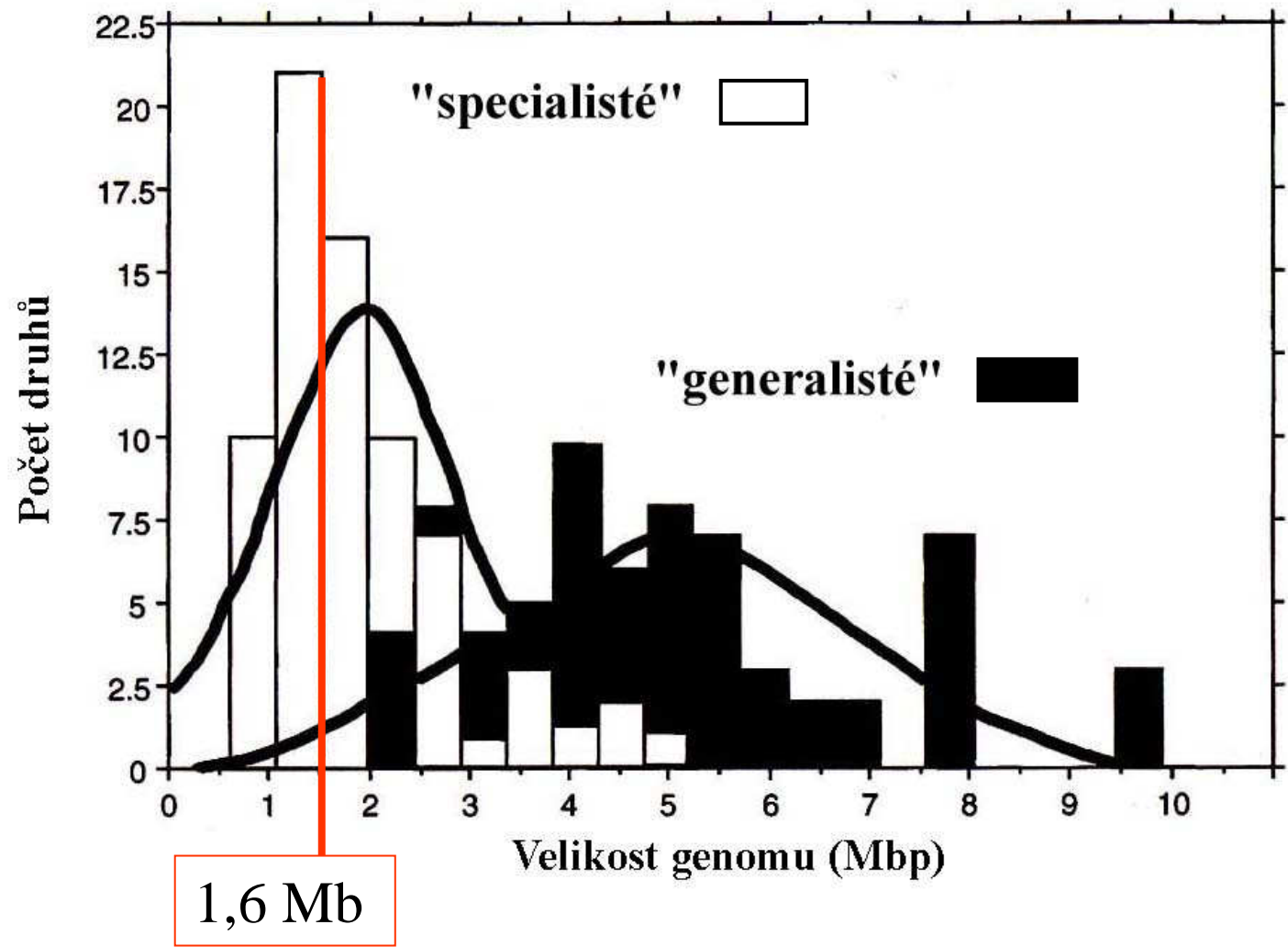
**U prokaryot neplatí paradox hodnoty C**

# Velikost genomu vybraných druhů bakterií

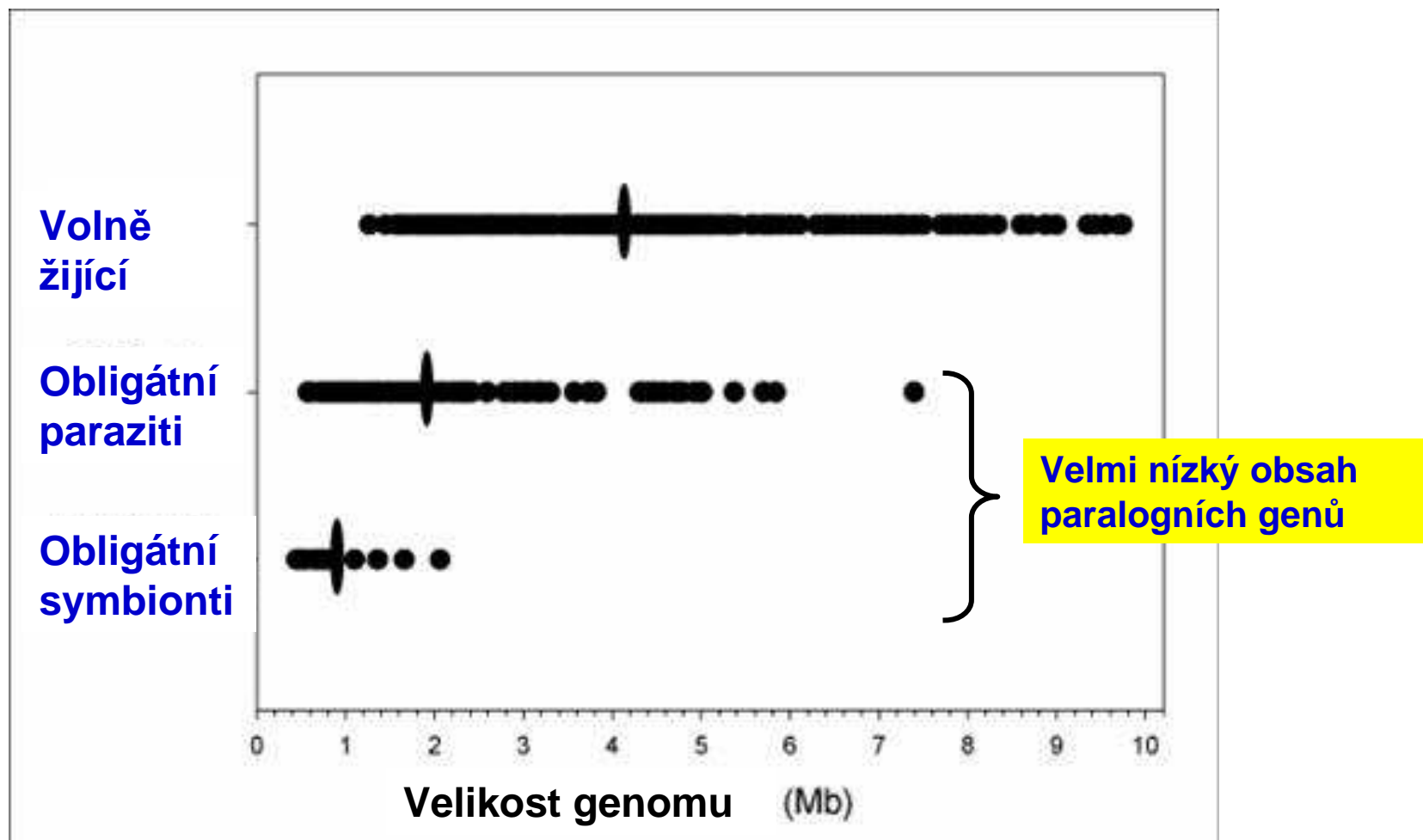


**Obsah GC = 22-75 mol%**

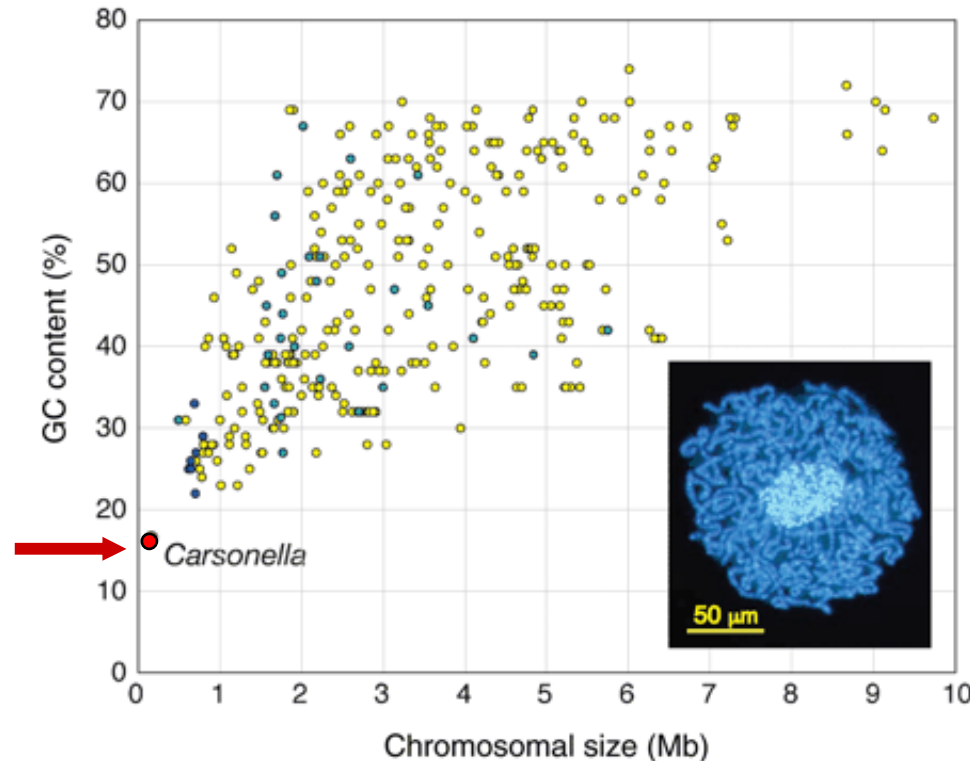
# Velikost genomu prokaryot



## Velikost genomů bakterií podle jejich životního stylu



Symbiont *Carsonella ruddii*, obsahuje kružnicový chromozom o velikosti 159,662 bp, má 182 ORFs, obsah GC 16.5%. Její genom je zatím nejmenší z popsaných, má nejvyšší obsah AT. 97% genomu má kódující kapacitu, jeho geny se překrývají a mají redukovanou velikost. Geny pro translaci a biosyntézu aminokyselin jsou početné, ale řada genů považovaných za esenciální chybí. *Carsonella* se tak podobá organelám.



Relationship between genome sizes and GC content of 358 complete genomes from Bacteria and Archaea: red indicates *Carsonella*; blue represents endosymbionts *Buchnera*, *Blochmannia*, *Wigglesworthia*, and *Baumannia*; yellow, other Bacteria; and green, Archaea

**Genome reduction appear to reach limits of about 400 kb and about 20% GC, which are believed to be the minimal limits for cellular organisms**

*Carsonella ruddii* (nebo též *Candidatus C. ruddii*, protože nebyla kultivována) je endosymbiotická proteobakterie s nejmenším genomem ze všech buněčných organismů.[1] *Carsonella ruddii* žije uvnitř těla určitého druhu mer, drobného hmyzu. Má jen 159 662 párů bází a její genom je složen pouze z 182 genů, což je méně, než bylo vydáváno za minimální nutné množství genů k životu.[1] Mnoho důležitých genů patrně chybí, a tak se zdá, že se *Carsonella* dostává v těle svého hostitele do pozice „organely“.[2]

***Na hmyzím jedinci jsou patrné žluté skvrny, jsou to místa, kde endosymbiotické bakterie jsou součástí buněk těla svého hostitele. Jeden bez druhého se již nemůže obejít. Jsou to svědci evoluce, kdy se z bakterie stává organela. Útvaru ve kterém hostitel svého hosta hýčká, se říká **bakteriom**.***



## *Nanoarchaeum equitans*

*Nanoarchaeum equitans* je neobvyklý druh [archea](#), objevený v roce 2002. Většinou je řazen do samostatného kmene [Nanoarchaeota](#), přestože některé studie ho řadí k [Euryarchaeota](#). Dosahuje i na [prokaryotické](#) poměry velmi drobných rozměrů. Byl objeven v mořích poblíž [Islandu](#) mikrobiologem [Karlem Stetterem](#) a jeho kolegy.

### Popis

*Nanoarchaeum* byl objeven v horké vodě poblíž [hydrotermálních průduchů](#) v mořích u Islandu. Jelikož zde voda dosahuje bodu varu, jedná se o [hypertermofila](#). Zajímavou skutečností je fakt, že zde *Nanoarchaeum* žije v symbiotickém vztahu s jiným archeonem, a to z rodu [Ignicoccus](#). Tento vztah je některými vědci považován za [parazitismus](#) a *Nanoarchaeum* je tedy zatím jediným známým [parazitem](#) z domény [archea](#).<sup>[6]</sup> *Nanoarchaeum* je extrémně drobný, jeho [kokovité](#) buňky mají v průměru asi 400 [nm](#). Také **genom je velmi malý a je vůbec nejmenší mezi všemi buněčnými organismy: má jen 490 885 párů [bází](#)**. Tato DNA je však velmi efektivní a 95% jsou kódující sekvence (kódují [proteiny](#) či některé druhy [RNA](#)). Na druhou stranu však nebyly nalezeny geny pro [biosyntézu lipidů](#), [kofaktorů](#), [aminokyselin](#) či [nukleotidů](#).

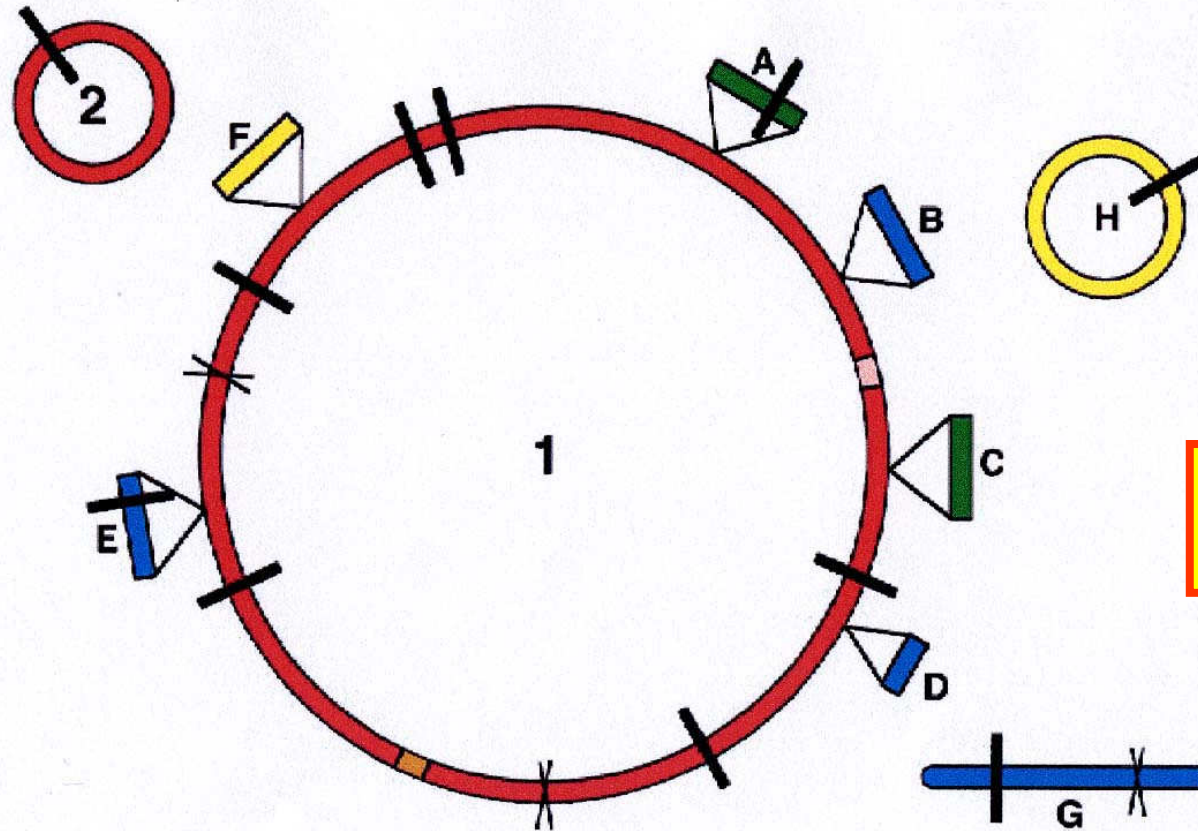


*Nanoarchaeum equitans*

# STRUKTURA PROKARYOTICKÝCH GENOMŮ

**Mozaikový charakter**

**Vysoká dynamika a plasticita**




**Rozdíly v obsahu VGE u různých druhů**

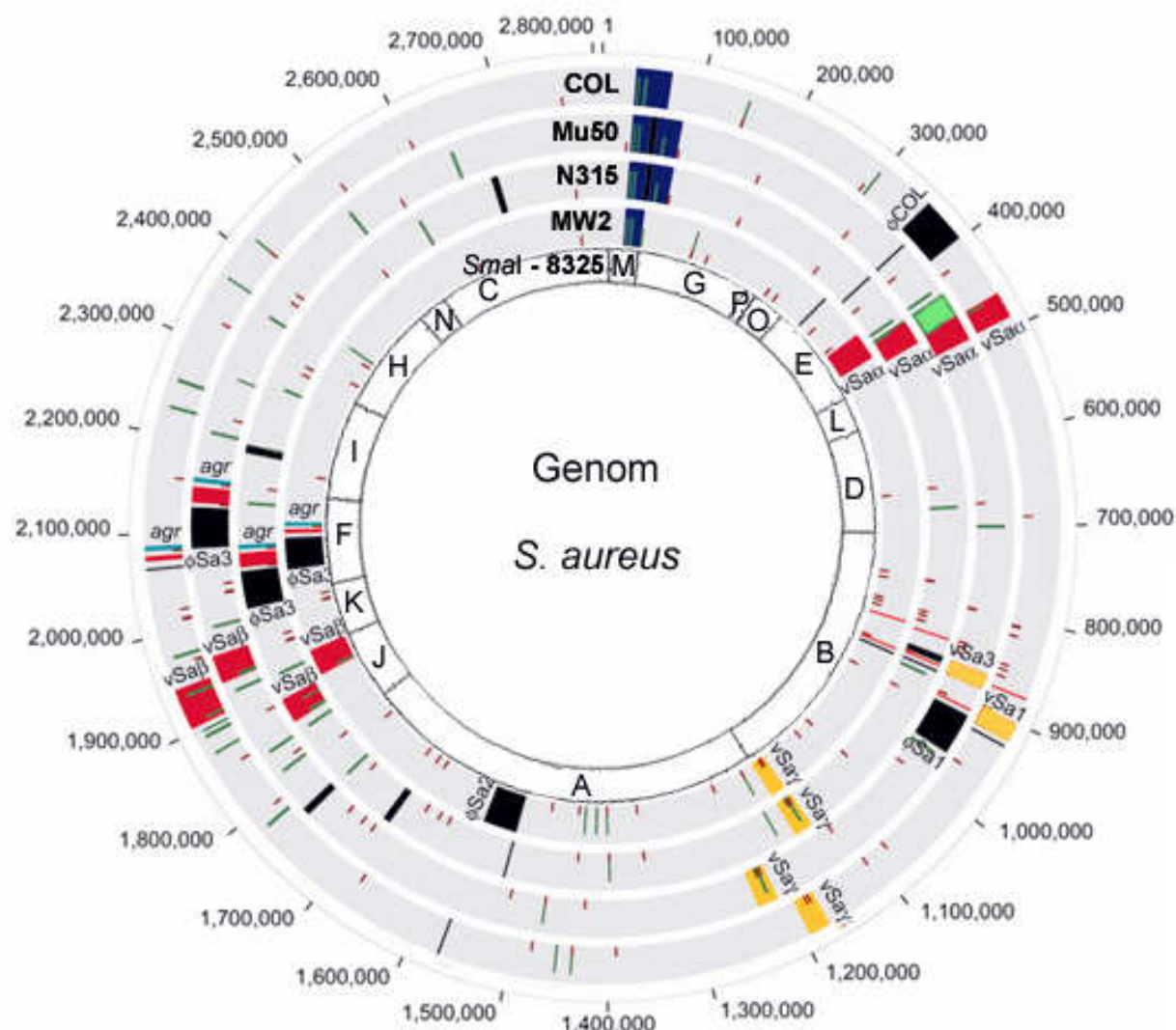
- Endochromozomy (1,2)**  
**Přídavné genetické elementy (A-H)**
- Plazmidy (H, F)
  - Profágy a defektní profágy (B, D, E, G)
  - Genomické ostrovy (A, C)
  - IS-elementy, transpozony, integrony (I)

**Variabilní složka genomu – adaptace na prostředí**



# Variabilní genetické elementy (VGE) (20 % genomu *S. aureus*)

-  Chromozomové kazety rezistence k meticilinu (SCCmec)
-  Jedinečné ostrovy patogenity (SaPI)
-  Společné genomické ostrovy
-  Profágy
-  Lokus přídatného genového regulátoru (*agr*)
-  Integrované plazmidy
-  Repetitivní elementy (STAR)
-  Inzerční sekvence a transpozony



# Pangenom (supra-genom)

Soubor genů u daného druhu bakterií nebo archeí

- odpovídá počtu genů zjištěných u kmenů daného druhu – obsah a forma genů u různých kmenů se výrazně liší.

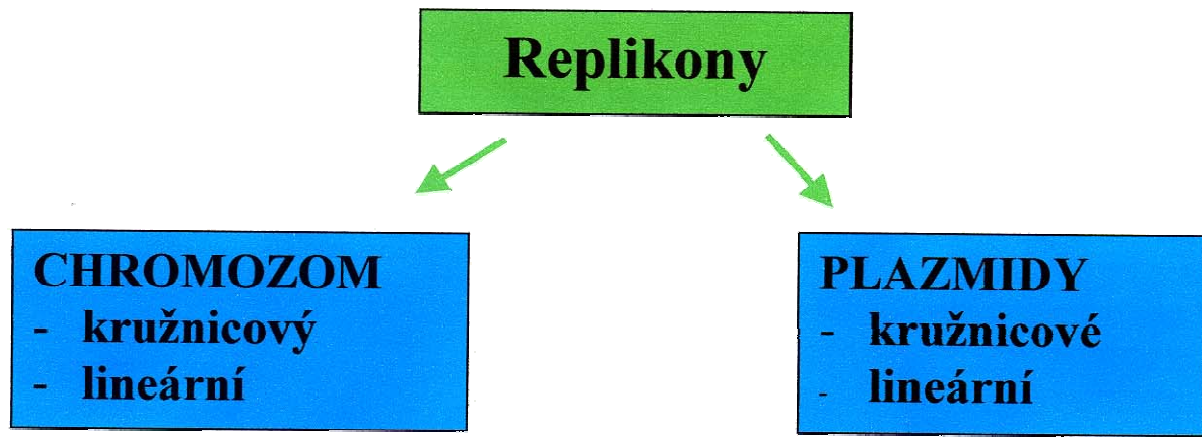
**„Core“ genom** – geny přítomné u všech kmenů

**Postradatelná část genomu** – geny přítomné jen u dvou nebo více kmenů

**Jedinečné geny** – geny vyskytující se pouze u jednoho kmene

Pangenom určitého bakteriálního druhu může být až řádově vyšší než jsou genomy jednotlivých kmenů

# TOPOLOGIE PROKARYOTICKÉHO GENOMU



| DRUH                             | CHROMOZOMY             |            | PLAZMIDY              |            |
|----------------------------------|------------------------|------------|-----------------------|------------|
|                                  | KRUŽNICOVÉ             | LINEÁRNÍ   | KRUŽNICOVÉ            | LINEÁRNÍ   |
| <i>Escherichia coli</i>          | 1 (4,6 Mb)             |            | několik               |            |
| <i>Bacillus cereus</i>           | 1 (5 Mb)<br>1 (2,4 Mb) |            | 0<br>několik (2,6 Mb) |            |
| <i>Vibrio cholerae</i>           | 2<br>(2,9+1,1 Mb)      |            |                       |            |
| <i>Borrelia burgdorferi</i>      |                        | 1 (0,9Mb)  | 9<br>(kb)             | 12<br>(kb) |
| <i>Agrobacterium tumefaciens</i> | 1 (2,8 Mb)             | 1 (2,1 Mb) | 2 (0,75 Mb)           |            |
| <i>Streptomyces lividans</i>     | 1 (8 Mb)               |            |                       | 1 (50 kb)  |

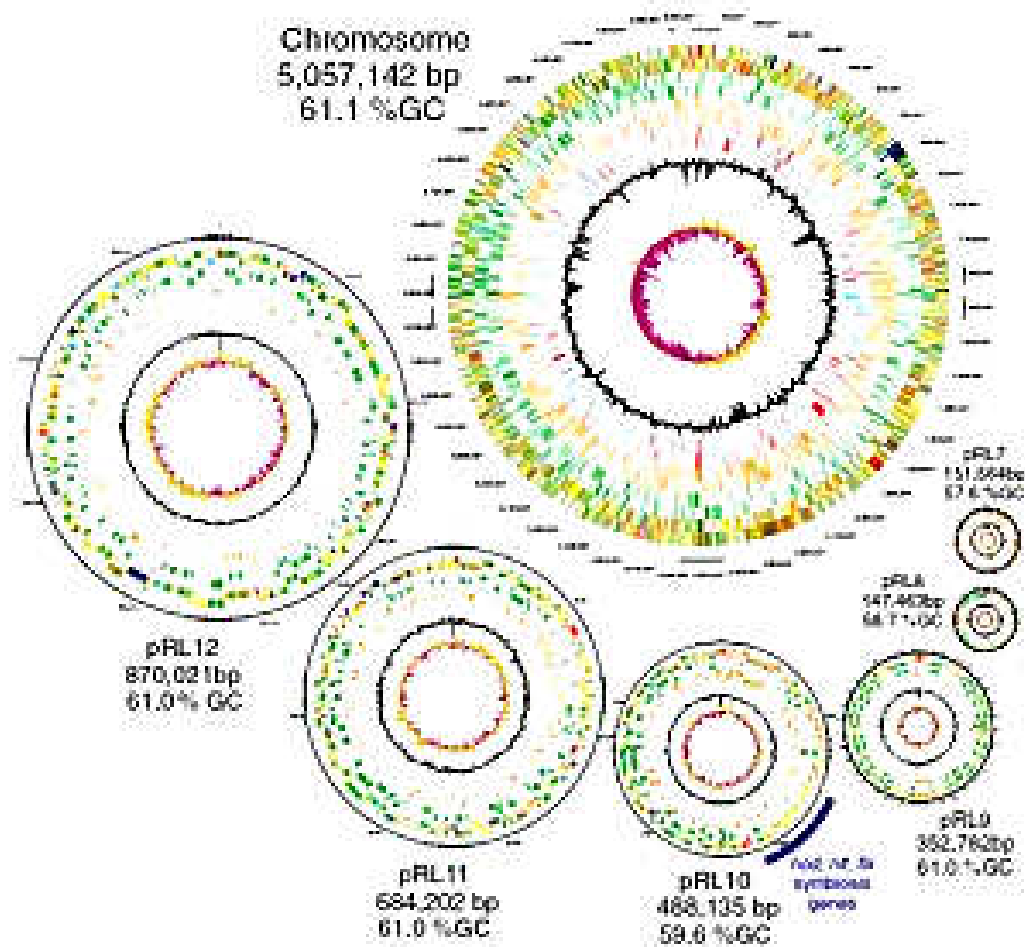
## Příklady organizace bakteriálních genomů

| Bacteria                             | Chromosome(s)                               | Plasmid(s)                            |
|--------------------------------------|---|---------------------------------------|
| <i>Agrobacterium tumefaciens</i>     | one linear (2.1 Mb) + one circular (3.0 Mb) | two circular (450 + 200 Kb)           |
| <i>Bacillus subtilis</i>             | one circular (4.2 Mb)                       |                                       |
| <i>Bacillus thuringiensis</i>        | one circular (5.7 Mb)                       | six (each >50 Kb)                     |
| <i>Borrelia</i>                      | one linear (0.91 Mb)                        | multiple circular + linear (5-200 Kb) |
| <i>Bradyrhizobium japonicum</i>      | one circular (8.7 Mb)                       |                                       |
| <i>Brucella melitensis</i>           | two circular (2.1 + 1.2 Mb)                 |                                       |
| <i>Brucella suis</i> biovars 1, 2, 4 | two circular (1.0 + 2.0 Mb)                 |                                       |
| <i>Brucella suis</i> biovar 3        | one circular (3.1 Mb)                       |                                       |
| <i>Buchnera</i> sp. strain APS       | one circular (640 Kb)                       | two circular (<7.8 Kb each)           |
| <i>Deinococcus radiodurans</i>       | two circular (2.6 + 0.4 Mb)                 | two circular (177 + 45 Kb)            |
| <i>Escherichia coli</i> K-12         | one circular (4.6 Mb)                       |                                       |
| <i>Leptospira interrogans</i>        | two circular (4.7 + 0.35 Mb)                |                                       |
| <i>Paracoccus denitrificans</i>      | three circular (2.0 + 1.1 + 0.64 Mb)        |                                       |
| <i>Pseudomonas aeruginosa</i>        | single circular (6.3 Mb)                    |                                       |
| <i>Rhizobacterium meliloti</i>       | two circular (3.4 + 1.7 Mb)                 | one circular megaplasmid (1,400 Kb)   |
| <i>Rhodobacter sphaeroides</i>       | two circular (3.0 + 0.3 Mb)                 |                                       |
| <i>Ureaplasma urealyticum</i>        | one circular (0.75 Mb)                      |                                       |
| <i>Vibrio cholerae</i>               | two circular (2.9 + 1.1 Mb)                 |                                       |
| <i>Vibrio parahaemolyticus</i>       | two circular (3.2 + 1.9 Mb)                 |                                       |
| <i>Xylella fastidiosa</i>            | one circular (2.7 Mb)                       | two circular (51 + 1.3 Kb)            |

**Počet kopií chromozomu: 1-40/buňku (haploidie, ~polyploidie)**

**MCLR – multiple chromosome-like replicons – známy u 44 druhů**

Hlavní chromozom *Rhizobium leguminosarum* o velikosti 5 Mb a soubor plazmidů o velikostech od 0.87 do 0.15 Mb, které dohromady představují 35% celého genomu



## Příklady bakteriálních druhů obsahujících více chromozomů

| Bacterium                                   | Strain/origin<br>(T = type strain) | No.<br>of<br>replicons <sup>a</sup> | Sizes of<br>replicons<br>(Mb) <sup>b</sup>                          | Genome<br>size<br>(Mb) | Method<br>of<br>analysis <sup>c</sup> |
|---|------------------------------------|-------------------------------------|---|------------------------|---------------------------------------|
| <i>Agrobacterium radiobacter</i> bv. 1      | CFBP 2414 <sup>T</sup>             | 3(2)                                | 3.0*, 2.1(L)*, 0.68   | 5.78                   | PFGE                                  |
| <i>A. tumefaciens</i> bv. 1                 | ATCC 23308 <sup>T</sup>            | 4(2)                                | 3.0*, 2.1(L)*, 0.55, 0.25   | 5.90                   | PFGE                                  |
| <i>A. tumefaciens</i>                       | C58                                | 4(2)                                | 3.0* <sup>2</sup> , 2.0* <sup>2</sup> (L), 0.55, 0.21               | 5.70                   | Genome                                |
| <i>Azospirillum brasilense</i>              | FP2                                | 7(4)                                | 2.5*, 1.72*, 0.81(L)*,<br>0.7(L), 0.63(L)*, 0.17, 0.15              | 6.70                   | PFGE                                  |
| <i>A. lipoferum</i>                         | Sp59b                              | 7(6)                                | 2.6*, 1.8*, 1.38*,<br>1.18(L)*, 0.97(L)*,<br>0.71(L), 0.65(L)*      | 9.70                   | PFGE                                  |
| <i>Brucella melitensis</i>                  | 16M                                | 2(2)                                | 2.1* <sup>2</sup> , 1.2*  | 3.30                   | Genome                                |
| <i>B. abortus</i>                           | 544                                | 2(2)                                | 2.1* <sup>2</sup> , 1.2*  | 3.30                   | PFGE                                  |
| <i>B. suis</i>                              | 1330                               | 2(2)                                | 2.1* <sup>2</sup> , 1.2*  | 3.30                   | PFGE                                  |
| <i>Mycoplana dimorpha</i>                   | ATCC 4279 <sup>T</sup>             | 3(NE)                               | 3.2, 0.3, 0.15  | 3.15                   | PFGE                                  |
| <i>Ochrobactrum anthropi</i>                | ATCC 49188 <sup>T</sup>            | 4(2)                                | 2.7*, 1.9*, 0.15, 0.1   | 4.85                   | PFGE                                  |
| <i>Paracoccus denitrificans</i>             | Pd1222                             | 3(NE)                               | 2.0, 1.1(L), 0.64(L)  | 3.74                   | PFGE                                  |
| <i>Phyllobacterium myrsinacearum</i>        | ATCC                               | 5(1)                                | 3.5*, 0.68, 0.5,<br>0.365, 0.285                                    | 5.33                   | PFGE                                  |
| <i>Rhizobium leguminosarum</i> bv. trifolii | ATCC 14480 <sup>T</sup>            | 4(1)                                | 4.6*, 1.1, 0.65, 0.45   | 6.80                   | PFGE                                  |
| <i>Rhizobium</i>                            | ATCC 14482 <sup>T</sup>            | 4(NE)                               | 4.6, 1.1, 0.45, 0.285   | 6.435                  | PFGE                                  |
| <i>R. capsulatus</i>                        | SB1003                             | 2(1)                                | 3.8*, 0.134   | 3.93                   | Genome                                |
| <i>R. sphaeroides</i> <sup>d</sup>          | 2.4.1                              | 7(2)                                | 3.110*, 0.9* <sup>2</sup> , 0.114,<br>0.113, 0.104, 0.101,<br>0.037 | 4.48                   | Genome                                |
| <i>Sinorhizobium</i>                        | ATCC 35423 <sup>T</sup>            | 3(1)                                | 4.0*, 2.2, 0.45   | 6.65                   | PFGE                                  |

| Bacterium  | Strain/origin<br>(T = type strain)        | No.<br>of<br>replicons <sup>a</sup> | Sizes of<br>replicons<br>(Mb) <sup>b</sup> | Genome<br>size<br>(Mb) | Method<br>of<br>analysis <sup>c</sup> |
|--|---|-------------------------------------|--|------------------------|---------------------------------------|
| <i>Burkholderia cepacia</i> <sup>e</sup>                   | ATCC 25416                                | 4(3)                                | 3.7* <sup>4</sup> , 3.2*, 1.1*, 0.2        | 8.20                   | PFGE                                  |
| <i>B. cepacia</i> <sup>e</sup>                             | C1 (CF patient, Wales)                    | 3(NE)                               | 3.1, 1.9, 0.1                              | 5.10                   | PFGE                                  |
| <i>B. cepacia</i> <sup>e</sup>                             | NCPFB 2993 (rotting onion, USA)           | 3(NE)                               | 3.4, 3.1, 1.1, 0.2                         | 7.80                   | PFGE                                  |
| <i>B. cepacia</i> <sup>e</sup>                             | NCIMB 9087 (rotting tree trunk, Trinidad) | 3(NE)                               | 3.2, 3.0                                   | 6.20                   | PFGE                                  |
| <i>B. gladioli</i> <sup>e</sup>                            | NCPFB 2478                                | 2(NE)                               | 3.2, 3.0                                   | 5.20                   | PFGE                                  |
| <i>B. glumae</i> <sup>e</sup>                              | PG1                                       | 2(NE)                               | 3.8, 3.0                                   | 6.80                   | PFGE                                  |
| <i>Ralstonia (Alcaligenes) eutropha</i>                    | H16                                       | 3(2)                                | 4.1*, 2.9*, 0.44                           | 7.10                   | PFGE                                  |
| <i>Ralstonia solanacearum</i>                              | GMI1000                                   | 2(2)                                | 3.7* <sup>3</sup> , 2.1*                   | 5.80                   | Genome                                |
| <i>Buchnera aphidicola</i> sp. APS                         |   | 3(1)                                | 0.64*, 0.078, 0.073                        | 0.80                   | Genome                                |
| <i>Vibrio cholerae</i>                                     | El Tor N16961                             | 2(1)                                | 3.0* <sup>8</sup> , 1.0                    | 4.00                   | Genome                                |
| <i>V. parahaemolyticus</i>                                 | AQ4673                                    | 2(2)                                | 3.2* <sup>4</sup> , 1.9*                   | 5.10                   | PFGE                                  |
| <i>V. mimicus</i>  | 2031                                      | 2(NE)                               | 2.4, 1.6                                   | 4.00                   | PFGE                                  |
| <i>Deinococcus radiodurans</i>                             | R1  | 4(1)                                | 2.6* <sup>6</sup> , 0.4, 0.2, 0.045        | 3.30                   | Genome                                |
| <i>Leptospira interrogans</i> serovar ictero-haemorrhagiae | Verdun                                    | 2(1)                                | 4.6*, 0.35                                 | 5.00                   | PFGE                                  |

<sup>a</sup> This column gives the number of replicons that make up the genome of each organism. In parentheses are the numbers of replicons that hybridize or are shown to have 16S ribosomal RNA (*rm*) genes through sequencing. If 16S ribosomal RNA hybridization was not carried out, this is indicated by NE (not examined).

<sup>b</sup> The sizes of the replicons are given in megabase pairs. (L) indicates that a replicon is linear rather than circular. An asterisk (\*) indicates that a 16S *rm* gene or a *rm* operon has been detected by hybridization or sequencing. A number after an asterisk indicates the number of *rm* hybridization signals or operons that were detected on the replicons by hybridization or sequencing, e.g., *A. tumefaciens* C58 has the following entry; 3.0\*<sup>2</sup>, 2.0\*<sup>2</sup>(L), 0.55, 0.21, meaning that there are two *rm* operons on the 3.0-Mb replicon, two *rm* operons on the 2.0-Mb linear replicon, and no *rm* operons on the smaller replicons.

<sup>c</sup> Indicates whether pulsed-field gel electrophoresis (PFGE) or genome sequencing was the method used to determine the size and number of the replicons and the locations of their ribosomal RNAs.

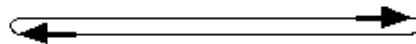
<sup>d</sup> In addition to *R. sphaeroides* 2.4.1 at least 23 other strains are known to contain multiple chromosomes.

<sup>e</sup> In addition to the four strains of *B. cepacia* given as examples in the table, 25 other strains are known to contain multiple large replicons. Two other *Burkholderia* species, *B. glathei* and *B. pickettii*, have also been shown to contain multiple replicons greater than 1.0 Mb in size.

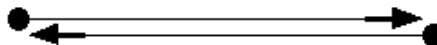
# PŘÍKLADY PROKARYOT S LINEÁRNÍM CHROMOZOMEM

- *Streptomyces* (*S. ambofaciens*, *S. lividans*) 10 Mb
- *Borrelia burgdorferi* (*B. hermsii*) 0.95 Mb
- *Coxiella burnetii* 2,1 Mb
- *Paracoccus denitrificans* - tři molekuly DNA
  - 2 Mb; 1,1 Mb; 0,64 Mb, dvě jsou lineární
- *Agrobacterium tumefaciens* - tři molekuly DNA
  - 3,0 Mb; 2,1 Mb; 0,55 Mb; 0,25 Mb

Vlásenkové telomery



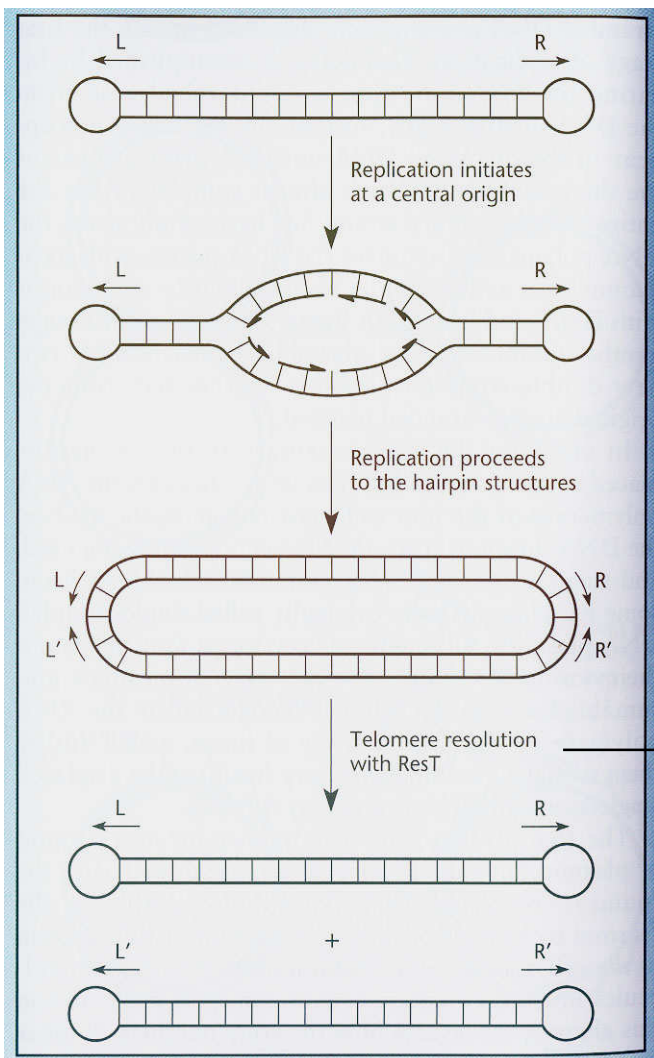
Invertronové telomery



dNTP vázající se na TP (telomerový protein) poskytuje 3'-OH



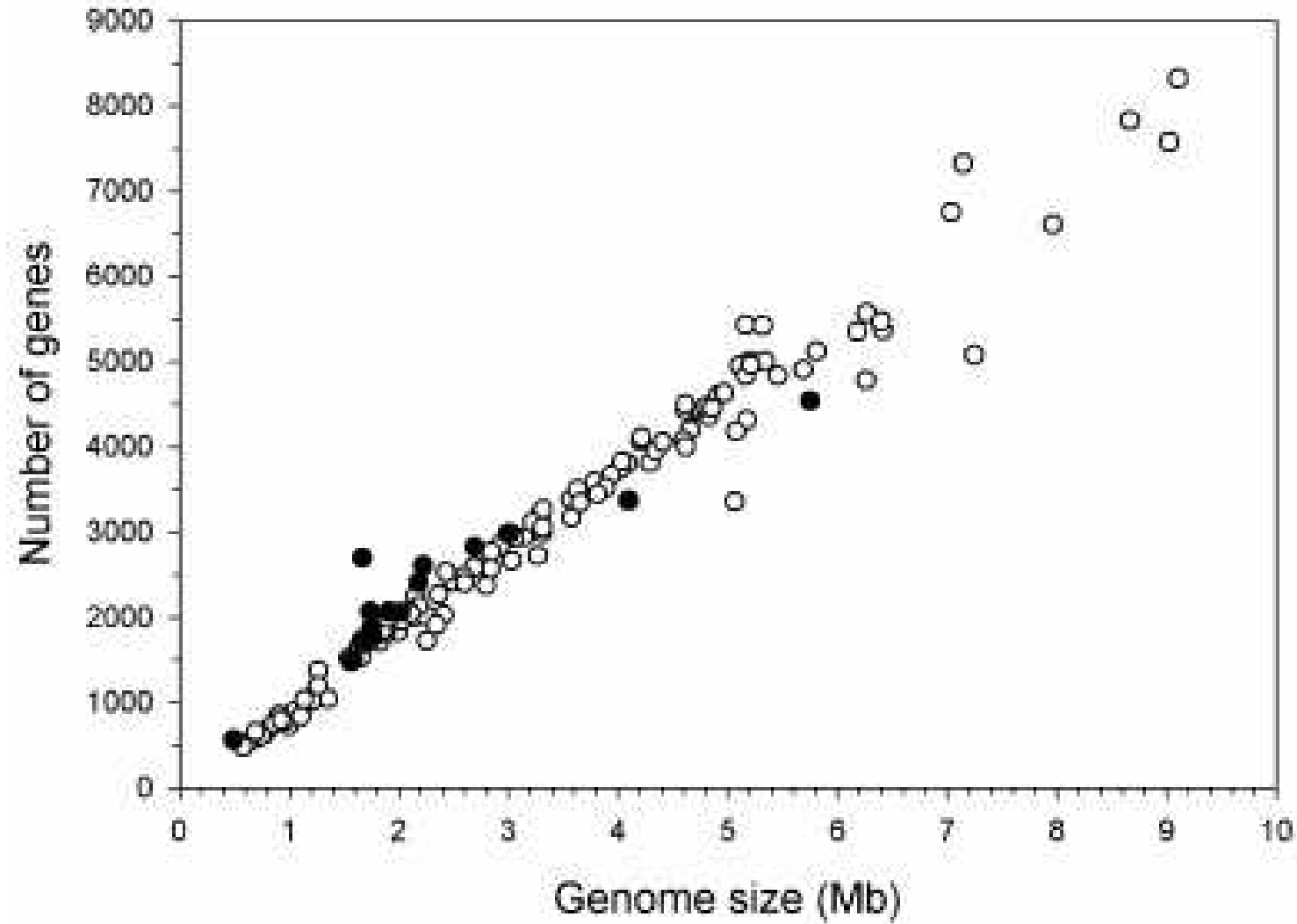
# Replikace lineárního chromozomu u bakterií r. *Borrelia*



**Protein ResT (telomere resolvase protein) vytvoří zlomy v obou řetězcích a znovuspojí je za vzniku vlásenek**

**Bakteriofág N15 – koncová oblast *fos* jeho lineárního genomu + protein TelN – přenos těchto elementů do *E. coli* vede k linearizaci jejího chromozomu**

## *Velikost genomu a počet genů u bakterií a archeí*



# SEKVENOVANÉ GENOMY PROKARYOT

<http://www.sanger.ac.uk/resources/downloads/bacteria/>



| Species                                   | Strain    | Type   | bp        | Genes |
|---|-----------|--|-----------|-------|
| Acidobacteria bacterium                   | Ellin345  | <a href="#">Acidobacteria</a>                                      | 5,650,368 | 4,777 |
| <a href="#">Acinetobacter</a> sp.         | ADP1      | <a href="#">Proteobacteria</a> <a href="#">Gammaproteobacteria</a> | 3,598,621 | 3,325 |
| <a href="#">Agrobacterium tumefaciens</a> | C58       | Proteobacteria <a href="#">Alphaproteobacteria</a>                 | 2,841,581 | 2,722 |
| <a href="#">Anabaena</a> nostoc           | PCC7120   | <a href="#">Cyanobacteria</a> Nostocales                           | 6,413,771 | 5,368 |
| Anabaena variabilis                       | ATCC29413 | Cyanobacteria Nostocales   | 6,365,727 | 5,039 |
| Anaeromyxobacter dehalogenans             | 2CP-C     | Proteobacteria delta-epsilon                                       | 5,013,479 | 4,346 |
| <a href="#">Anaplasma</a> marginale       | StMaries  | Proteobacteria Alphaproteobacteria                                 | 1,197,687 | 949   |
| <a href="#">Anaplasma phagocytophilum</a> | HZ        | Proteobacteria Alphaproteobacteria                                 | 1,471,282 | 1,264 |
| <a href="#">Aquifex aeolicus</a>          | VF5       | <a href="#">Aquificae</a> Aquificae                                | 1,551,335 | 1,522 |
| Azoarcus sp.                              | EbN1      | Proteobacteria <a href="#">Betaproteobacteria</a>                  | 4,296,230 | 4,128 |
| <a href="#">Bacillus anthracis</a>        | Ames      | <a href="#">Firmicutes</a> Bacilli                                 | 5,227,293 | 5,311 |
| Bacillus anthracis                        | Sterne    | Firmicutes Bacilli   | 5,228,663 | 5,287 |
| <a href="#">Bacillus cereus</a>           | ATCC10987 | Firmicutes Bacilli   | 5,224,283 | 5,603 |

**Současný stav: několik tisíc kompletně sekvenovaných genomů, u některých druhů více kmenů: značné rozdíly ve velikosti a struktuře**

# SEKVENOVANÉ GENOMY: ARCHAEA



| Species   | Strain    | bp        | Genes |
|---|-----------|-----------|-------|
| <a href="#"><u>Aeropyrum pernix</u></a>                     | K1        | 1,669,695 | 2,694 |
| <a href="#"><u>Archaeoglobus fulgidus</u></a>               | DSM4304   | 2,178,400 | 2,407 |
| <a href="#"><u>Haloarcula marismortui</u></a>               | ATCC43049 | 3,131,724 | 3,131 |
| <a href="#"><u>Haloferax volcanii</u></a>                   | DS2       |           |       |
| <a href="#"><u>Halobacterium</u> species</a>                | NRC1      | 2,014,239 | 2,058 |
| <a href="#"><u>Methanobacterium thermoautotrophicum</u></a> | delta-H   | 1,751,377 | 1,869 |
| <a href="#"><u>Methanocaldococcus jannaschii</u></a>        | DSM2661   | 1,664,970 | 1,715 |
| <a href="#"><u>Methanococcoides burtonii</u></a>            | DSM6242   | 2,575,032 | 2,273 |
| <a href="#"><u>Methanococcus maripaludis</u></a>            | S2        | 1,661,137 | 1,722 |

**Současný stav: zhruba 30 kompletně sekvenovaných genomů**

# GENETICKÁ ORGANIZACE PROKARYOTICKÉHO GENOMU

- velmi kompaktní genom s malými mezerami mezi geny
- většina genomu je obsazena strukturními geny (převážně jednoduché geny, u  $\frac{1}{4}$  genomů byly nalezeny introny II sk
- malá část (10%) je tvořena nekódující DNA
- operonové uspořádání genů (většinou funkčně příbuzné)
- pořadí genů není u prokaryot konzervováno

# TYPY SEKVENCÍ PROKARYOTICKÉHO GENOMU

- **Kódující oblasti (na chromozomu n. plazmidech)**
    - Geny (ORF) se známou nebo neznámou funkcí
    - Inzerční sekvence a transpozony
    - Profágy (a defektní profágy)
  - **Sekvence přepisované do RNA (rRNA, tRNA)**
  - **Repetice**
    - Polynukleotidové sekvence a tandemové repetice
    - Krátké roztroušené repetitivní sekvence
    - Dlouhé roztroušené repetitivní elementy
    - Mosaikové repetitivní elementy
    - **Chi-místa a jim podobné sekvence**
    - **Rhs elementy**
- 
- 90%
- 10%

## Charakteristika předpovězených kódujících sekvencí (ORF) v sekvencovaných genomech

| Organismus                                   | VELIKOST GENOMU<br>(Mbp) | CELKOVÝ POČET<br>ORF | NEZNÁMÁ FUNKCE<br>(%) | JEDINEČNÉ ORF<br>(%) |
|--|--------------------------|----------------------|-----------------------|----------------------|
| <i>Archaoglobus fulgidus</i>                 | 2.18                     | 2437                 | 1315 (54)             | 641 (26)             |
| <i>Methanobacterium thermotautotrophicum</i> | 1.75                     | 1855                 | 1010 (54)             | 496 (27)             |
| <i>Methanococcus jannaschii</i>              | 1.66                     | 1749                 | 1076 (62)             | 525 (30)             |
| <i>Pyrococcus horikoshii</i>                 | 1.74                     | 2061                 | 859 (42)              | 453 (22)             |
| <i>Aquifex aeolicus</i>                      | 1.50                     | 1521                 | 663 (44)              | 407 (27)             |
| <i>Bacillus subtilis</i>                     | 4.20                     | 4100                 | 1722 (42)             | 1053 (26)            |
| <i>Borrelia burgdorferi</i>                  | 1.44                     | 1751                 | 1132 (65)             | 682 (39)             |
| <i>Chlamydia pneumoniae</i>                  | 1.23                     | 1073                 | 437 (40)              | 186 (17)             |
| <i>C. trachomatis</i>                        | 1.04                     | 894                  | 290 (32)              | 255 (28)             |
| <i>Deinococcus radiodurans</i>               | 3.28                     | 3193                 | 1515 (47)             | 1001 (31)            |
| <i>Escherichia coli</i>                      | 4.60                     | 4288                 | 1632 (38)             | 1114 (26)            |
| <i>Haemophilus influenzae</i>                | 1.83                     | 1692                 | 592 (35)              | 237 (14)             |
| <i>Helicobacter pylori</i>                   | 1.66                     | 1657                 | 744 (45)              | 539 (33)             |
| <i>Mycobacterium tuberculosis</i>            | 4.41                     | 3924                 | 1521 (39)             | 606 (15)             |
| <i>Mycoplasma genitalium</i>                 | 0.58                     | 470                  | 173 (37)              | 7 (2)                |
| <i>Mycobacterium pneumoniae</i>              | 0.81                     | 677                  | 248 (37)              | 67 (10)              |
| <i>Rickettsia prowazekii</i>                 | 1.11                     | 834                  | 311 (38)              | 207 (25)             |
| <i>Synechocystis sp.</i>                     | 3.57                     | 3168                 | 2384 (75)             | 1426 (45)            |
| <i>Thermotoga maritima</i>                   | 1.86                     | 1877                 | 863 (46)              | 373 (20)             |
| <i>Treponema pallidum</i>                    | 1.14                     | 1040                 | 461 (44)              | 28 (27)              |
| CELKEM                                       | 41.6                     | 40,261               | 18,948 (47)           | 10,303 (26)          |

**Hustota genů je zhruba stejná: 1 gen ~ 1 kb**

# POČTY PARALOGNÍCH GENŮ V GENOMECH BAKTERIÁLNÍCH DRUHŮ

Se zvětšováním genomu se zvyšuje obsah paralogních genů

HGT, genové duplikace

| Organismus             | Velikost genomu (Mbp) | Počet ORF | Počet paralogů |
|------------------------|-----------------------|-----------|----------------|
| <i>T. pallidum</i>     | 1.14                  | 1 040     | 129 (12%)      |
| <i>B. burgdorferi</i>  | 1.44                  | 1 751     | 707 (40%)      |
| <i>H. pylori</i>       | 1.66                  | 1 657     | 266 (16%)      |
| <i>A. fulgidus</i>     | 2.18                  | 2 437     | 719 (30%)      |
| <i>B. subtilis</i>     | 4.20                  | 4 100     | 1 947 (47%)    |
| <i>M. tuberculosis</i> | 4.41                  | 3 924     | 2 000 (51%)    |
| <i>E. coli</i>         | 4.60                  | 4 288     | 2 272 (53%)    |



# TYPY REPETICÍ V PROKARYOTICKÝCH GENOMECH

## A. Polynukleotidové sekvence a tandemové repetice

- **Trinukleotid TGG** – nejčastější trinukleotid *E. coli* (součást penta nebo oktanukleotidů)
- **Nonamer AAGTGCGGT (uptake signal sequence – USS) *H. influenzae*** - 1465 kopií
- **Tandemově opakované polynukleotidové sekvence (GTG)<sub>n</sub> nebo (GCC)<sub>n</sub>** - vysoce repetitivní u *E. coli*, *S. typhimurium* a *Shigella* sp.
- **Short tandemly repeated repetitive (STRR) sequences** - heptanukleotidová opakování u sinice *Calothrix*
- **Major polymorphic tandem repeat (MPTR)** - polymorfní 10-bp DR u *Mycobacterium tuberculosis* a dalších mykobakterií (podobnost s místy **Chi** a roztroušenými **REP** elementy)

# TYPY REPETICÍ V PROKARYOTICKÝCH GENOMECH

## B. Krátké roztroušené repetitivní sekvence (kratší než 50 bp)

- **REP (repetitivní extragenové palindromatické sekvence)** 38 bp, **500 REP** u *E. coli*
- **PU (palindromic units)** u *E. coli* a *S. typhimurium*.
- **Mnohokopiový 26-mer (Ngrep)** u *Neisseria gonorrhoeae* a *N. meningitidis*
- **Mnohokopiový 24-mer DR element** u *Mycobacterium bovis* (38 kopií)

# TYPY REPETICÍ V PROKARYOTICKÝCH GENOMECH

## C. Dlouhé roztroušené repetitivní elementy (delší než 50 bp)

- intergenic repeat unit (IRU)
- enterococcal repetitive intergenic consensus (ERIC) 126 bp nebo zkrácené formy. Chromozomová lokalizace se liší u kmenů a druhů
- ERIC-like sekvence 127 bp – enterobakterie, celá bakteriální říše.
- RLEP (545 - 1063 bp) u *Mycoplasma leprae* 28 (0,6% genomu)
- Mx-rep u *Myxococcus xanthus* - 87 pb jádrová sekvence
- Dr-rep (SARK) u *Deinococcus radiodurans* Element o variabilní délce (150-192 bp)
- RepMP2, RepMP1, SDC1 (150 bp – 1 kb) u *Mycoplasma pneumoniae*, v genomu 8-10 kopií.
- RepMP2-like u *Staphylococcus*

# TYPY REPETICÍ V PROKARYOTICKÝCH GENOMECH

## D. Mosaikové repetitivní elementy

- **Bacterial Interspersed Mosaic Elements (BIME)** – (kombinace REP a sedmi dalších repetitivních motivů).  
500 u *E. coli* a dalších G- enterobakterií (*Klebsiella pneumoniae*, *S. typhimurium*)
- **REP MP** 1 300 bp element ohraničený kratšími repetitivními sekvencemi u *M. pneumoniae*.
- **BOX elementy** – rozptýlené repetitivní elementy u G+ (*Streptococcus pneumoniae*).

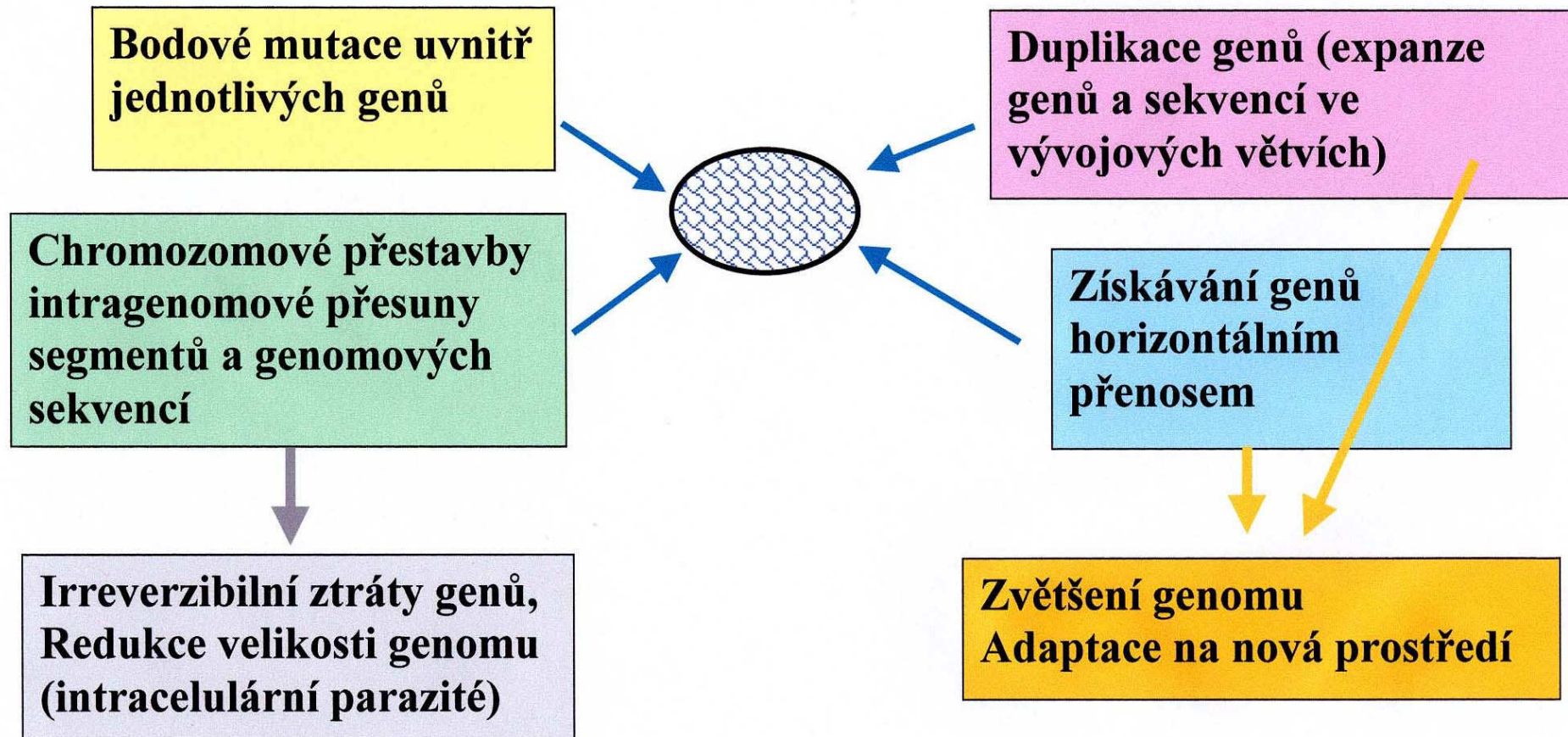
„**Genomic signature**“ - charakteristika frekvence jednotlivých typů oligonukleotidů v genomech. Genomic signature je u fylogeneticky příbuzných genomů podobná.

*Signature genomics* – vyhledávání mutací zodpovědných za choroby

# ZÁVĚRY ZE SROVNÁNÍ SEKVENOVANÝCH GENOMŮ

- **Počet informačních genů je v každém genomu zhruba stejný, i když se jejich velikosti liší až 5x.**
- **Počet genů ostatních funkčních kategorií je mnohem variabilnější a má tendenci se zvyšovat**
- **Se zvětšováním velikosti genomu přibývá paralogních genů a zvětšuje se též biochemická komplexita organismu.**
- **Jedna čtvrtina ORF u každého druhu je jedinečná a nemá významnou sekvenční homologii k žádné známé nukleotidové nebo proteinové sekvenci.**

## Podstata změn v obsahu genomu prokaryot v průběhu evoluce



**Důsledek: Pořadí genů je zachováno jen u velmi blízké příbuzných druhů**

# HORIZONTALNÍ PŘENOS GENŮ

- **Často přenášené: operační geny**  
(metabolismus a regulace, buněčné struktury)
  
- **Zřídka přenášené: informační geny**  
(transkripce, translace)
  
- **Horizontálně jsou přenášeny variabilní genetické elementy (a geny na nich lokalizované)**
  - profágy,
  - plazmidy,
  - IS-elementy,
  - transpozony,
  - Integrony
  - Genomické ostrovy

# POČET HORIZONTÁLNĚ PŘENESENÝCH GENŮ U VYBRANÝCH DRUHŮ BAKTERIÍ A ARCHEÍ

| Druh                                  | Velikost genomu (Mbp) | Počet ORF | Horizontálně přenesené ORF |       |
|---------------------------------------|-----------------------|-----------|----------------------------|-------|
|                                       |                       |           | počet                      | %     |
| <b><u>Proteobacteria</u></b>          |                       |           |                            |       |
| <i>Escherichia coli</i>               | 4,64                  | 4289      | 381                        | 9,6   |
| <i>Haemophilus influenzae</i>         | 1,83                  | 96        | 96                         | 6,2   |
| <i>Helicobacter pylori</i>            | 1,67                  | 1553      | 89                         | 6,4   |
| <i>Rickettsia prowazekii</i>          | 1,11                  | 834       | 28                         | 3,6   |
| <b><u>Gram-pozitivní bakterie</u></b> |                       |           |                            |       |
| <i>Bacillus subtilis</i>              | 4,21                  | 4100      | 537                        | 14,5  |
| <i>Mycoplasma genitalium</i>          | 0,58                  | 480       | 67                         | 14,5  |
| <i>Mycoplasma pneumoniae</i>          | 0,82                  | 677       | 39                         | 5,9   |
| <i>Mycobacterium tuberculosis</i>     | 4,41                  | 3918      | 187                        | 5,0   |
| <b><u>Spirochaete</u></b>             |                       |           |                            |       |
| <i>Borrelia burgdorferi</i>           | 0,91                  | 850       | 12                         | 1,56  |
| <i>Treponema pallidum</i>             | 1,14                  | 1031      | 77                         | 8,3   |
| <b>Chlamydiae</b>                     |                       |           |                            |       |
| <i>Chlamydia trachomatis</i>          | 1,04                  | 894       | 36                         | 4,3   |
| <i>Deinococcus radiodurans</i>        | 2,65                  | 2580      | 95                         | 3,92  |
| <i>Synechocystis sp.</i>              | 3,57                  | 3169      | 219                        | 7,5   |
| <i>Thermotoga maritima</i>            | 1,86                  | 1846      | 198                        | 11,63 |
|                                       |                       |           |                            |       |
| <b><u>Archaea</u></b>                 |                       |           |                            |       |
| <i>Aeropyrium pernix</i>              | 1,67                  | 2694      | 370                        | 14,0  |
| <i>Methanobacterium therm.</i>        | 1,75                  | 1869      | 179                        | 10,3  |
| <i>Methanococcus jannaschii</i>       | 1,66                  | 1715      | 77                         | 5,0   |
| <i>Pyrococcus abyssi</i>              | 1,76                  | 1765      | 124                        | 7,35  |



## Bakterie s umělým genomem

Tým čtyřadvaceti genetiků v čele s Craigen Venterem syntetizoval v laboratořích *The J. Craig Venter Institute* dědičnou informaci jednoduché bakterie ***Mycoplasma mycoides*** složenou z více než milionu písmen (1 Mbp) genetického kódu. Tak dlouhou molekulu nelze vyrobit najednou. Vědci nejprve syntetizovali její kratší části a ty pak skládali do větších celků. Nakonec byly součásti genomu tak velké, že si s nimi už laboratorní technika neporadila. Finální kompletaci bakteriální DNA proto genetici svěřili buňkám kvasinek. V těch pak získali úplný genom bakterie, který přenesli do příbuzné bakterie ***Mycoplasma capricolum***.

Gibson DG, et al.: Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*. 2010 Jul 2;329(5987):52-6.

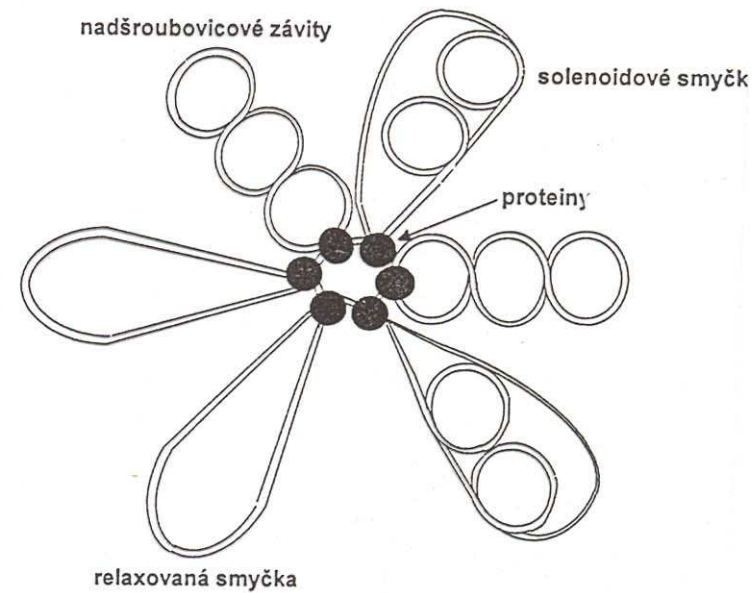
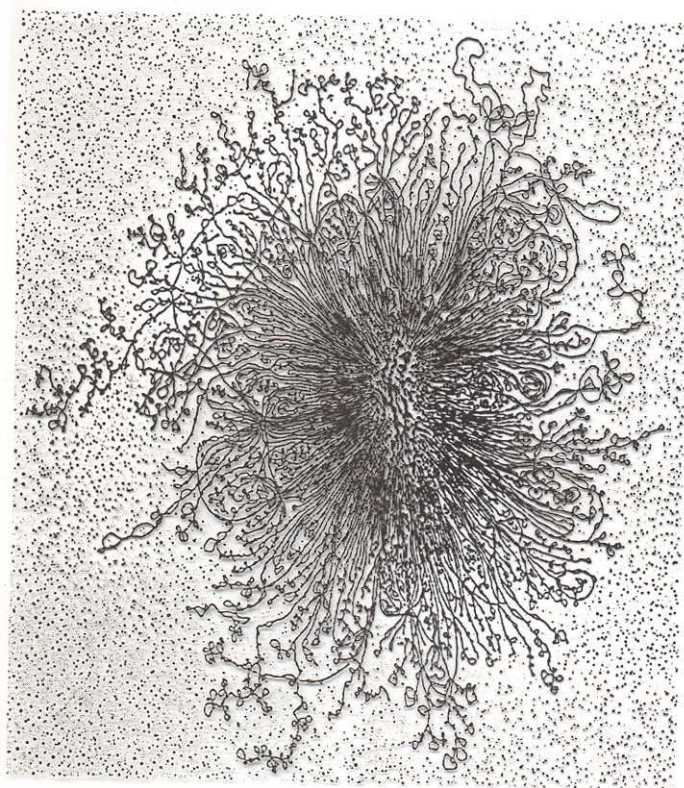
# FYZIKÁLNÍ ORGANIZACE BAKTERIÁLNÍHO CHROMOZOMU

- **dsDNA** - není obalená jadernou membránou
- vazba malých proteinů typu **HU** (histon-like)  
- 105 molekul, vázajících se sekvenčně nespecificky, ovlivňujících flexibilitu DNA (ohýbání DNA do smyček) („architektura“ nukleoidu).  
Účast při replikaci, transkripci a rekombinaci
- Proteiny **HN-S** vázají se na AT-bohaté oblasti DNA, zvyšuje její teplotní stabilitu a brání transkripci. Možná má podobnou funkci jako histony.
- Protein **Fis** – ovlivňuje topologii a superhelicitu DNA
- Vazba proteinu **IHF** (integration host factor), vázající se sekvenčně specificky na DNA, zejména na regulační oblasti
- 50 superhelikálních smyček, bez účasti RNA
- kontrola stavu superhelicity: topoizomeráza, gyráza – změny vinutí při replikaci, přichycení na membránu

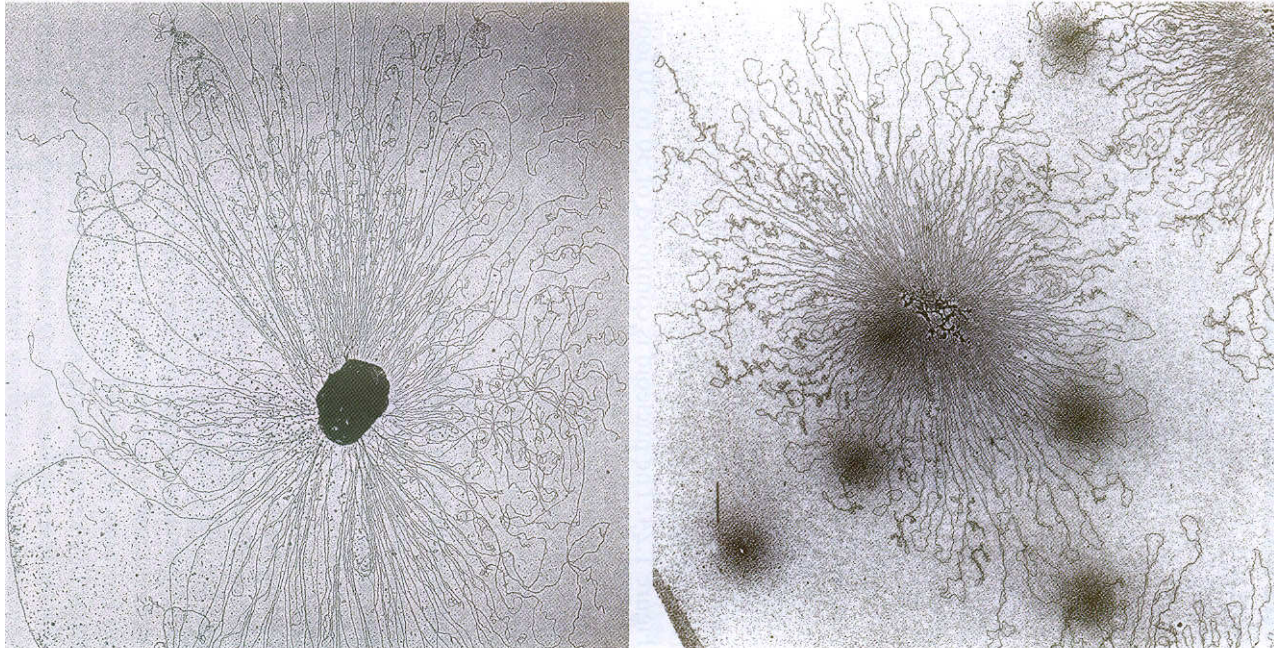
Archea: mají rudimentární histony příbuzné histonům eukaryot

# STRUKTURA CHROMOZOMU *E. coli*

- 4,6 Mb, 50% GC, 1  $\mu\text{m}$

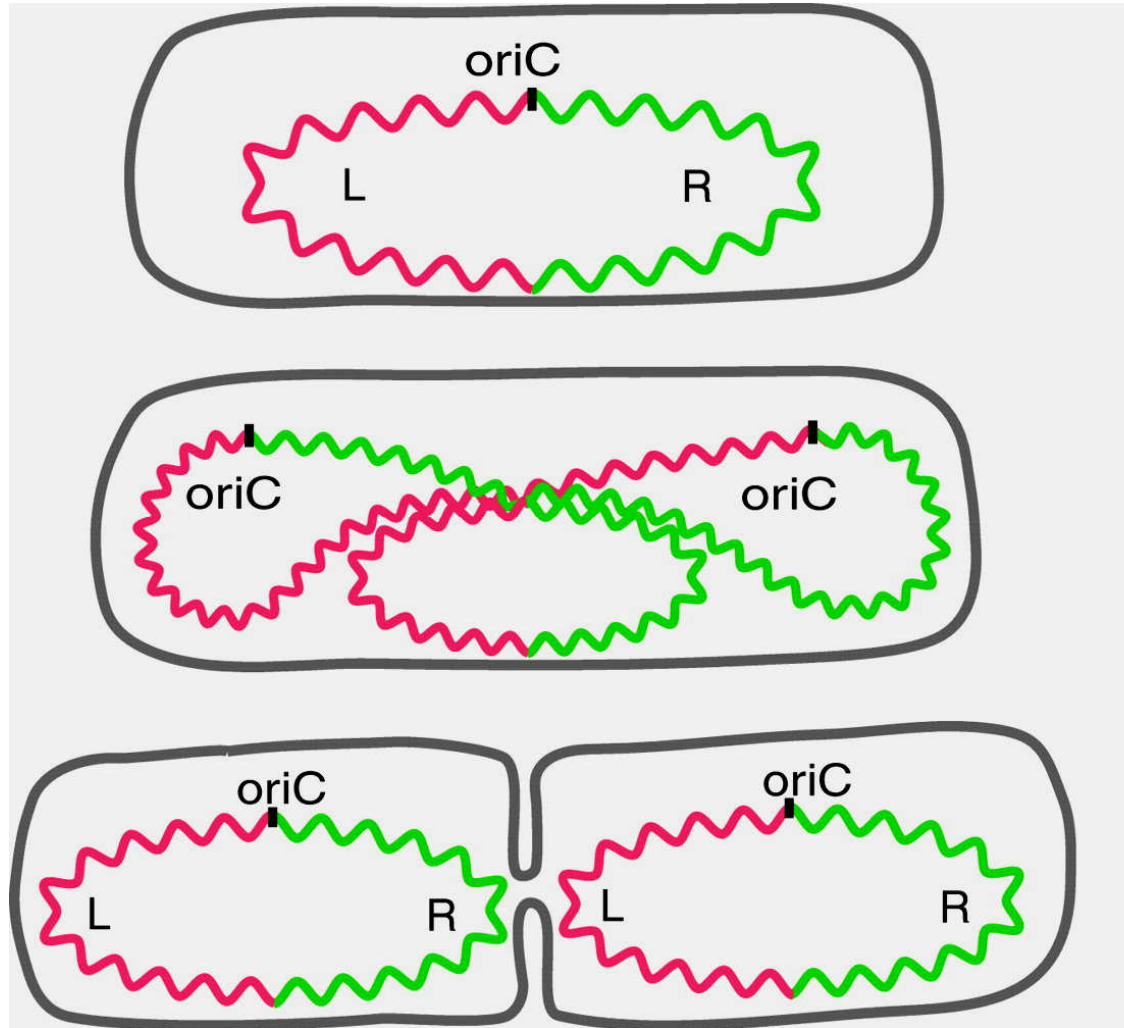


## Bakteriální nukleoid

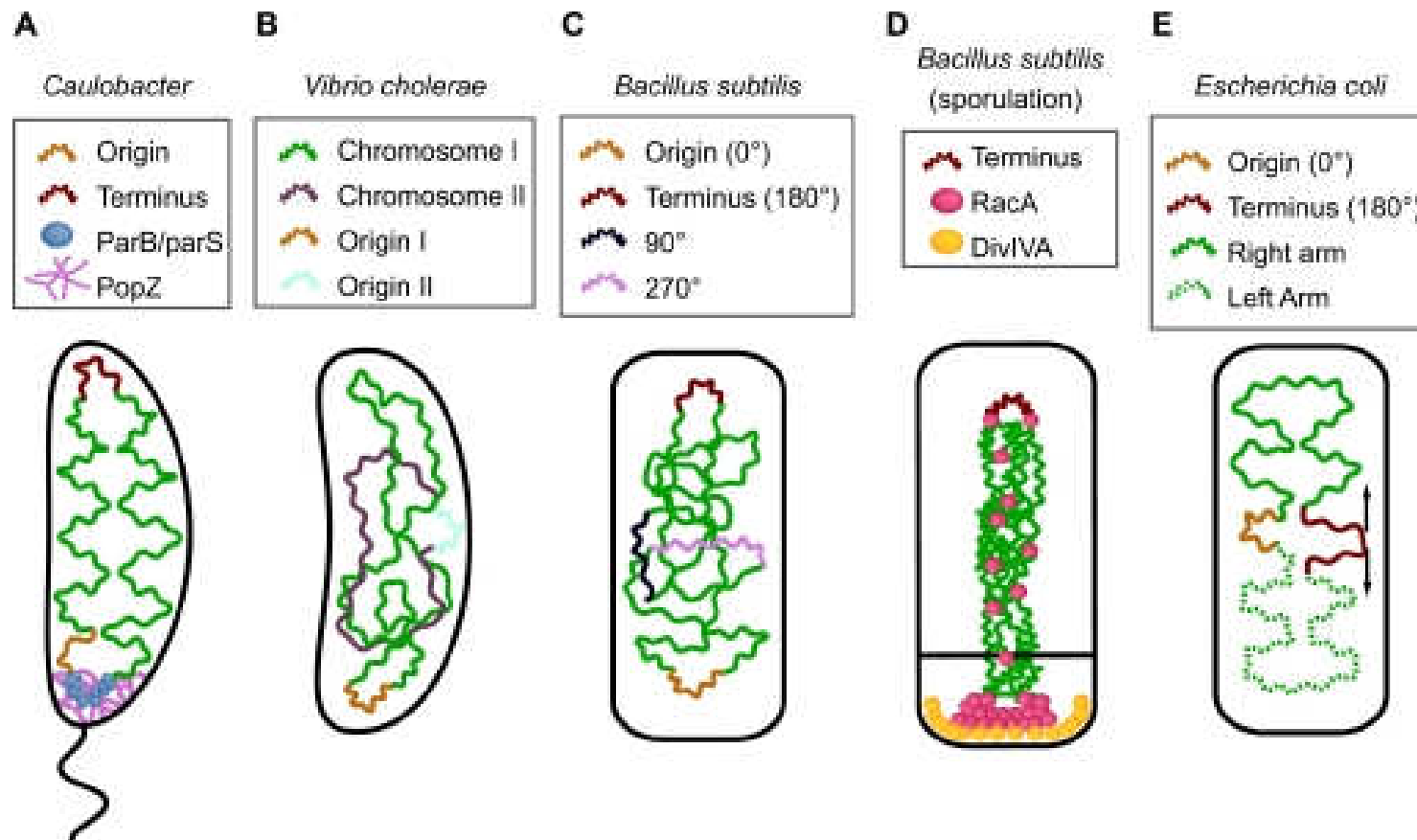


**...molekulární mechanismy udržující nukleotid v buňce jsou neznámé, ale jistě existují specifické systémy, které zajišťují jeho replikaci, transkripci, rekombinaci a další funkce...**

## Organizace chromozomu u *E. coli*.



## Organizace chromozomu u modelových bakterií



(A) The *Caulobacter* chromosome is linearly organized, and anchored to the flagellated pole via *parS*/ParB/PopZ. (B) In *Vibrio cholerae*, the origin region of the larger chromosome (chromosome I) is localized to the cell pole, whereas the origin of the smaller chromosome is localized to the cell center. The organization of the bulk of the chromosomes, as well as their separation or intermingling, are currently unknown. (C) Four loci have been localized in vegetative cells of *Bacillus subtilis*, and their organization is reminiscent of the linear order seen in *Caulobacter*. Although the origin region is localized near to one pole, it appears not to be anchored to the cell membrane. (D) Sporulating cells of *B. subtilis*, however, do anchor the origin region, through RacA/DivIVA, to the negatively curved membrane at the pole. RacA also binds all along the chromosome, compacting it into a long "axial filament" before sporulation. (E) The *E. coli* origin localizes to mid-cell, and the two replichores are separated into opposite cell halves. The terminus is broadly localized (arrows), and may be found on either side of the cell center

