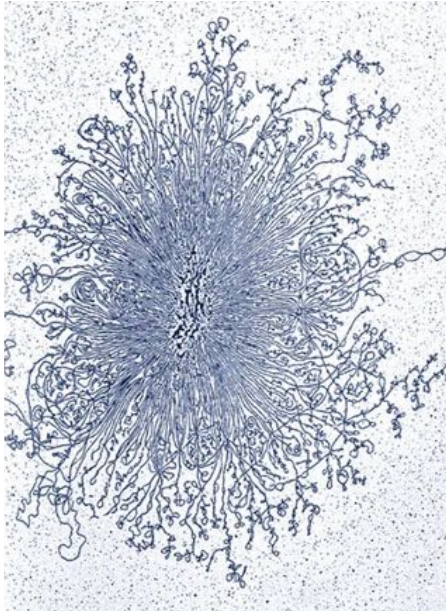


1. Introduction to prokaryotic genomes annotation

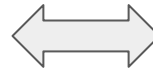
Guillaume GAUTREAU - 28/09/2022

Prokaryotic genomes...

DNA/chromosome : molecular support



E. coli nucleoid (Wang *et al.*, 2013)



Genome : the information, the sequence from the chromosome

```
GTGAAGGGCAATCAGCTGTTGCCGTCTCACTGGTGAAAAGAAAAAC
CACCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCC
GATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGG
GCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCA
CCCAGGCTTTACACTTTTATGCTTCCGGCTCGTATGTTGTGTGGAATT
GTGAGCGGATAACAATTTACACAGGAAACAGCTATGACCATGATTA
CGGATTACTIONGCGCGTCTTTTACAACGTCGTGACTGGGAAAACCT
GGCGTTACCCAACCTAATCGCCTTGCGACATCCCCCTTTCCGCAG
CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAG
TTGCGCAGCCTGAATGGCGAATGGCGCTTTGCTGGTTTCCGGCAC
CAGAAGCGGTGCCGAAAGCTGGCTGGAGTGCGATCTTCTGAGGC
CGATACTGTCGTCGTCCCCTCAAACCTGGCAGATGCACGGTTACGATG
CGCCCATCTACCAACGTGACCTATCCATTACGGTCAATCCGCCG
TTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAAT
GTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGA
TGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTC
GGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCG
CATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTGCG
CTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATG
AGCGGCATTTTCCGTGACGTCTG.....
```

...annotation

According to Cambridge Dictionary



annotate

verb [T] • formal

UK  /ˈæn.ə.teɪt/ US  /ˈæn.ə.teɪt/



[often passive]

to add a short explanation or opinion to a text or image:

- *Annotated editions of Shakespeare's plays help readers to understand old words.*
- *an annotated bibliography/manuscript/edition*
- *His great-granddaughter has painstakingly transcribed and annotated his wartime diaries.*
- *The students annotate their photos, saying why they are relevant.*



COMPUTING, LANGUAGE • specialized

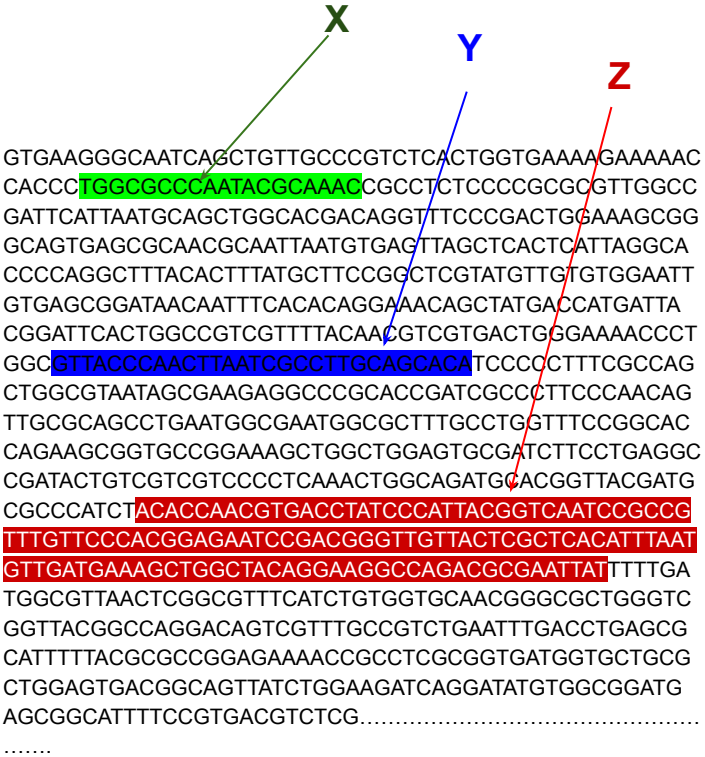
to add a description or piece of information to data, for example a label saying whether a word is a noun, a verb, etc.:

- *After the corpus was collected we annotated it.*
- *Textual or numerical data can be copied into databases, annotated, and linked to other data.*

Prokaryotic genome annotation

It means the combination of:

1. a genomic sequence location (described using coordinates)
2. a biological meaning for this sequence:
 - Is it a gene ?
 - what the function of that gene ?
 - Is it a coding gene or a non coding gene ?
 - Is it a set of co-transcribed genes ?
 - Is it regulated by a common factor ?
 - Is it a restriction site ?
 - Is it a integration hotspot ?
 - Is it a replication origin ?
 -



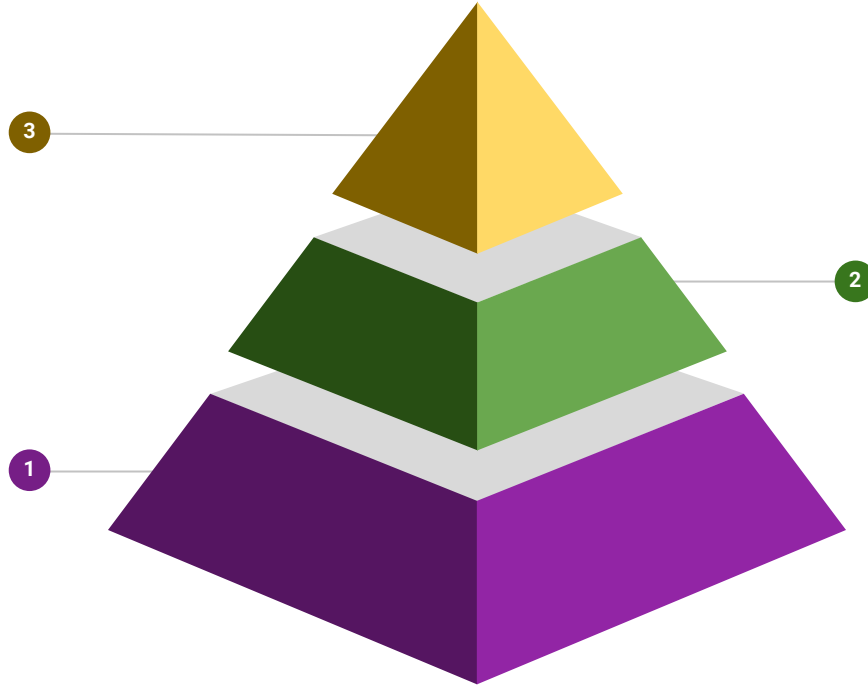
Cumulative levels of genomic annotation

Relational annotation

Which genes are associated in a common biological process?
(how?)
Lesson 9.

Syntactic annotation

Detect the location of the genes (where?)
Lesson 1.



Functional annotation

Find the biological functions of the genes (what?)
Lesson 5.

Quiz on prokaryotic genome characteristics

To make sure we are all clear about the type of data to be studied

Quizz on prokaryotic genome characteristics

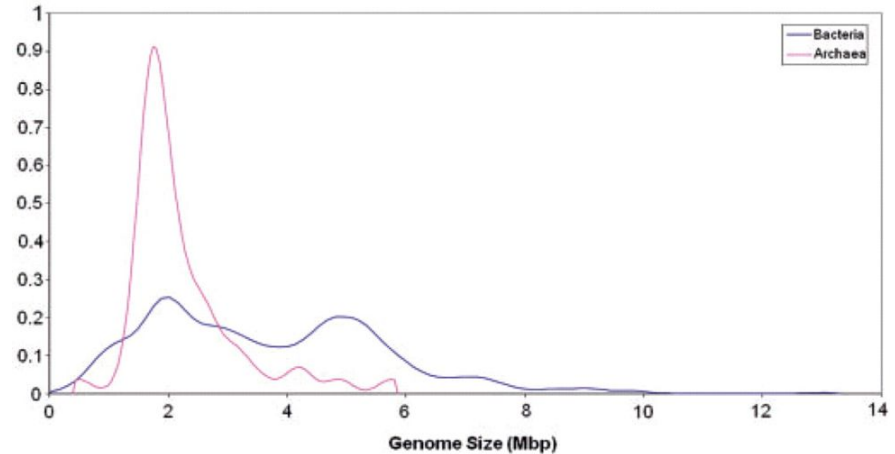
- Genome sizes?
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- Procaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?

Quizz on prokaryotic genome characteristics

- **Genome sizes?**
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- Procaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?

Quizz on prokaryotic genome characteristics

- **Genome sizes?**
 - Minimum: *Carsonella ruddii* : 160 Kb,
 - Maximum: *Sorangium cellulosum* : 15 Mb
 - Usually: *Escherichia coli* : 4,6 Mb
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- Procaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?



Koonin and Wolf, 2008

Quizz on prokaryotic genome characteristics

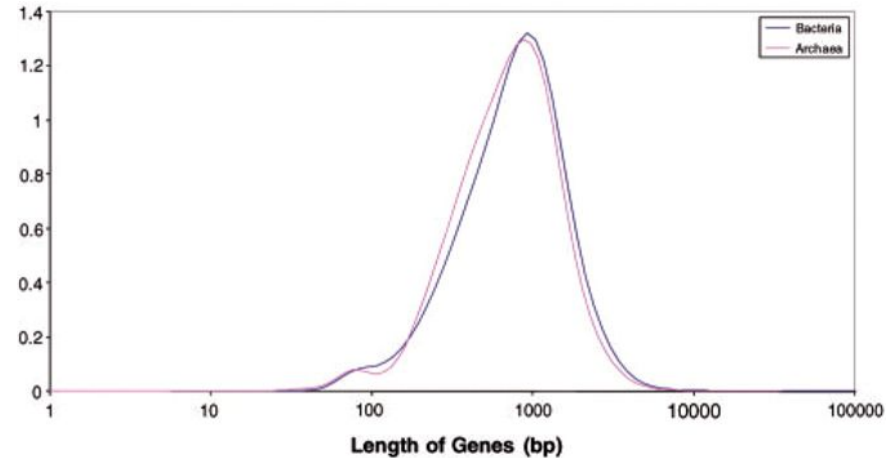
- Genome sizes?
- **Number of genes?**
- Gene sizes?
- Sizes of intergenic spaces?
- Procaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?

Quizz on prokaryotic genome characteristics

- Genome sizes?
- **Number of genes?**
 - Minimum: *Carsonella ruddii* : 224
 - Maximum: *Sorangium cellulosum* : 9 367
 - Usually: *Escherichia coli* : 4 400
- Gene sizes?
- Sizes of intergenic spaces?
- Procaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?

Quizz on prokaryotic genome characteristics

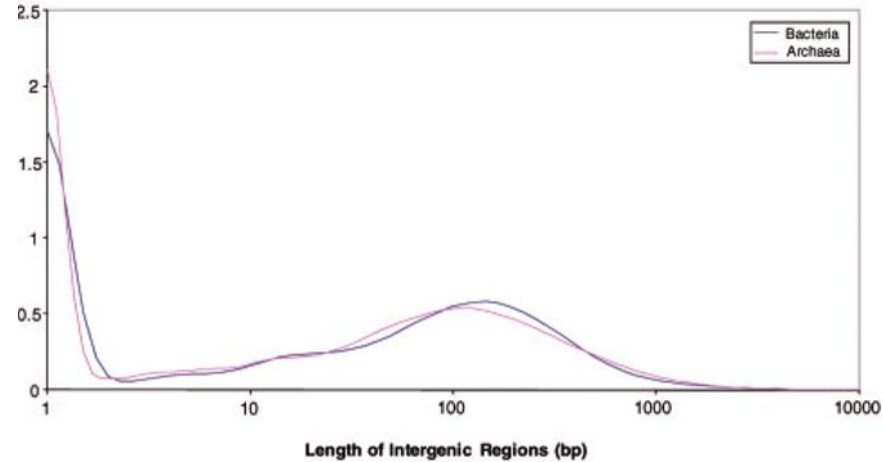
- Genome sizes?
- Number of genes?
- **Gene sizes?**
 - Minimum: sORFs (<50 aa)
 - Maximum: Repeat domain proteins (37Kaa)
 - Usually: LacZ (1024aa)
- Sizes of intergenic spaces?
- Procaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?



Koonin and Wolf, 2008

Quizz on prokaryotic genome characteristics

- Genome sizes?
- Number of genes?
- Gene sizes?
- **Sizes of intergenic spaces?**
 - median : 100pb
 - genes can overlap too
- Procaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?



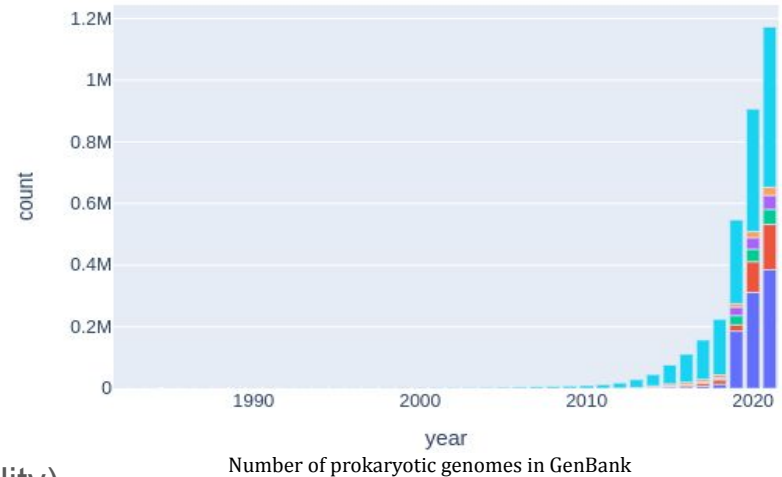
Koonin and Wolf, 2008

Quizz on prokaryotic genome characteristics

- Genome sizes?
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- **Procaryotic species?**
 - morphologic criteria
 - physiologic criteria
 - pathogenicity criteria
 - serogrouping criteria
 - **genomic criteria (next lesson n°2)**
- Number of available genomes?
- Genome evolution?
- Genome structure?
- Genome organization?

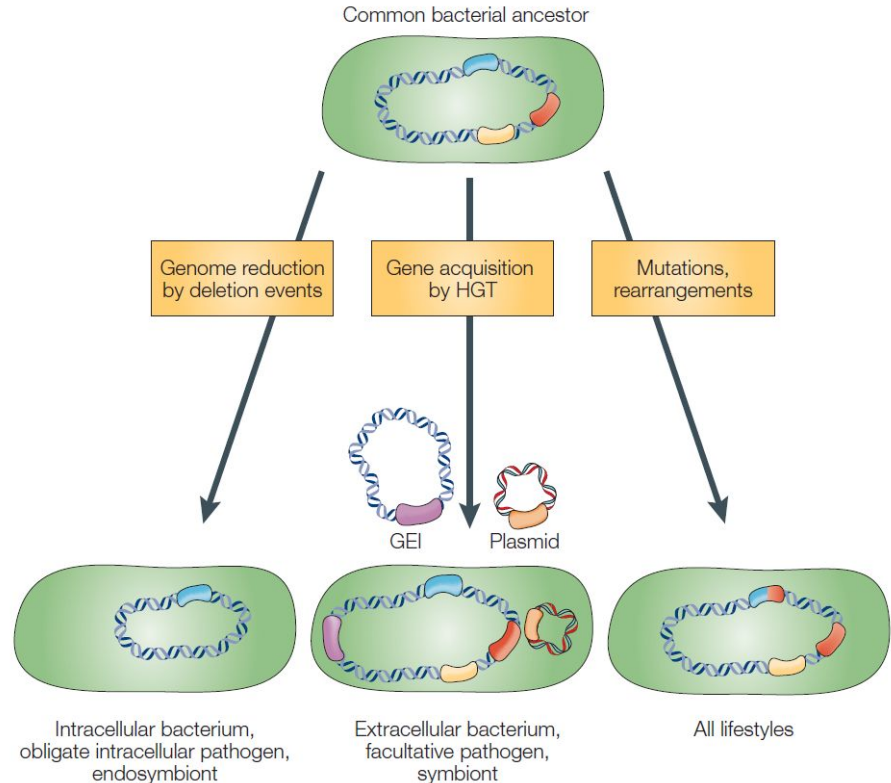
Quizz on prokaryotic genome characteristics

- Genome sizes?
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- Prokaryotic species?
- **Number of available genomes?**
 - species of medical interest : several thousands
 - cultivable species : dozens
 - hardly cultivable species: none or few (or low quality)
- Genome evolution?
- Genome structure?
- Genome organization?



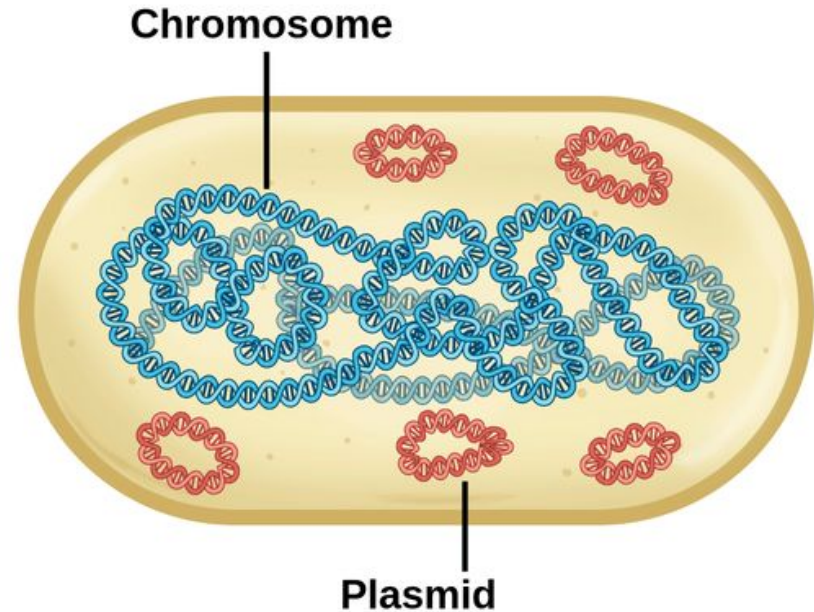
Quizz on prokaryotic genome characteristics

- Genome sizes?
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- Prokaryotic species?
- Number of available genomes?
- **Genome evolution?**
- Genome structure?
- Genome organization?



Quizz on prokaryotic genome characteristics

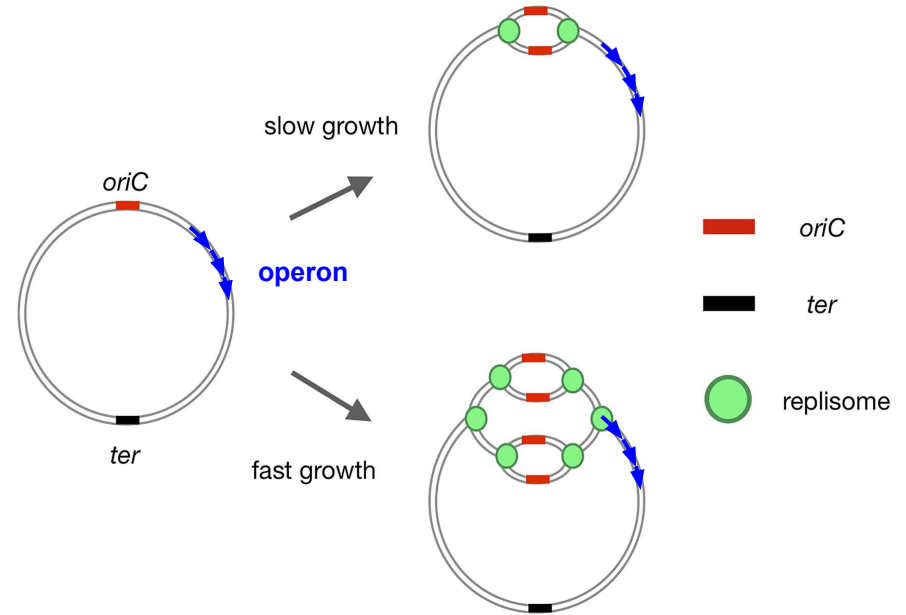
- Genome sizes?
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- Prokaryotic species?
- Number of available genomes?
- Genome evolution?
- **Genome structure?**
- Genome organization?



credit : https://theory.labster.com/bacterial_genome/

Quizz on prokaryotic genome characteristics

- Genome sizes?
- Number of genes?
- Gene sizes?
- Sizes of intergenic spaces?
- Prokaryotic species?
- Number of available genomes?
- Genome evolution?
- Genome structure?
- **Genome organization?**



modified from Trojanowski *et al.*, 2018

Challenge about genome annotation

- Annotation needs to be automated (no longer possible to be performed “manually”)
- Data bank are crucial to transfer known annotations between genomes
 - Findable
 - Accessible
 - Interoperable
 - Reusable
- But there is a risk of propagation of poor annotations and errors

Detailed program

<https://docs.google.com/document/d/1BGZvJDNgxOEXCPBg2KRPIK13K4LoSqkYb7CibyKWxYc/edit?usp=sharing>