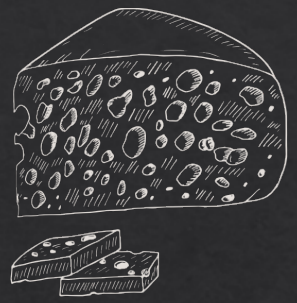
BBC
TP3 – CRISPR Cas

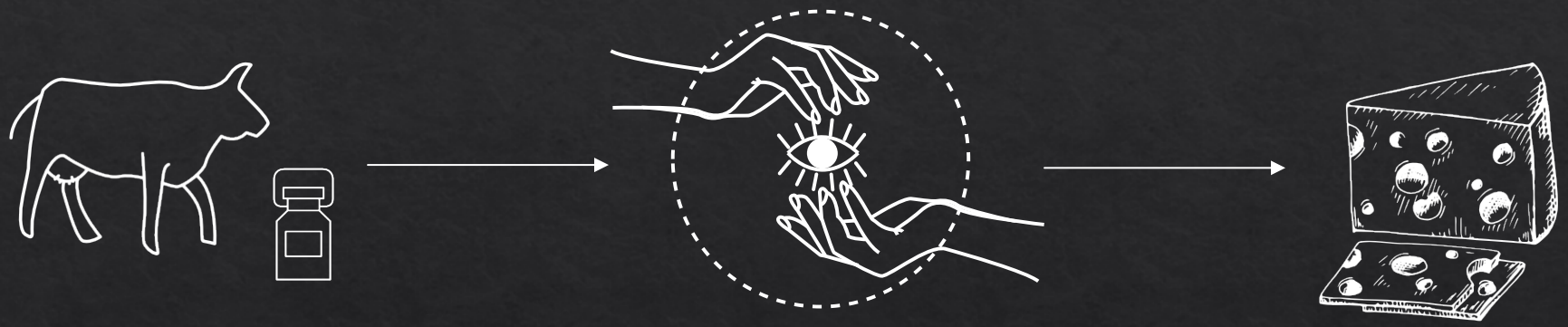
Analyse des défenses immunitaires adaptative des bactéries lactiques du fromage

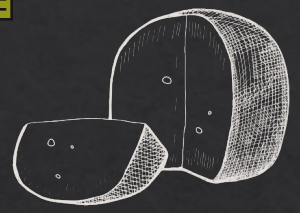
Content

- Introduction
  - Fromage: Théorie
  - Bactéries
  - Bactériophages
- Arsenal de défense bactérienne
- CRISPR Cas
- Intro TP

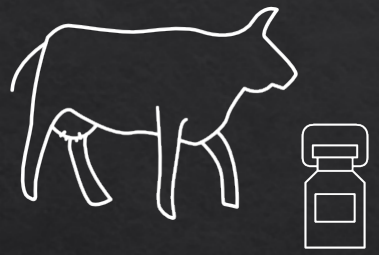


Cheese theory



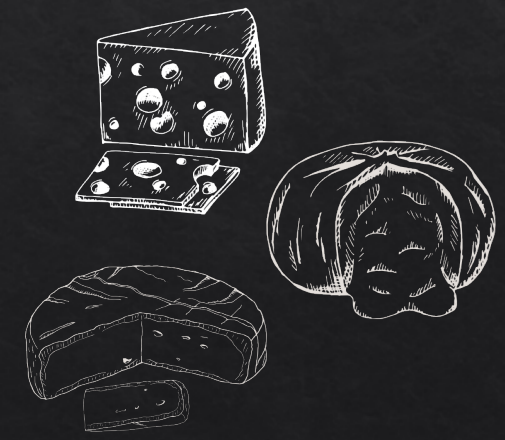


Cheese theory



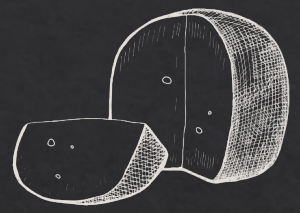
Coagulation

Refining
(Affinage)

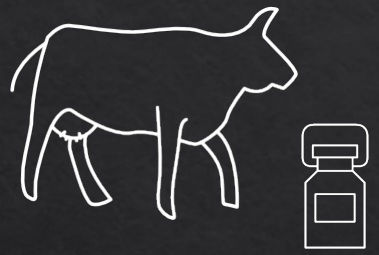


Rennet
(Présure)

LAB
(Lactic Acid Bacteria)



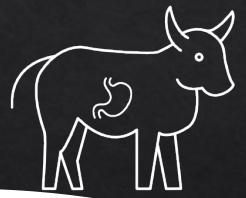
Cheese theory



Coagulation

Refining
(Affinage)

Rennet
(Présure)



LAB
(Lactic Acid Bacteria)

Cheese theory

LAB
(Lactic Acid Bacteria)

- Cultures de bactéries propagées depuis des décennies
- Des communautés simples (peu d’espèces différentes) mais très stables

Eucaryotes vs procaryotes

D’abord répondons à une question simple.

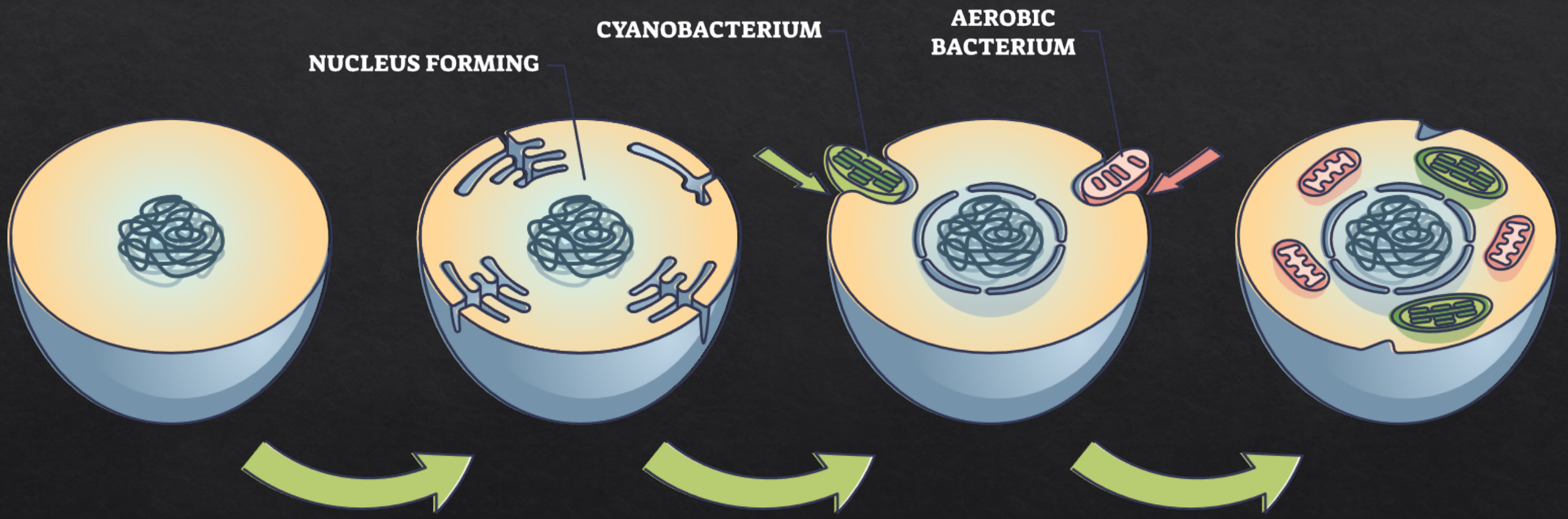
Qu’est-ce que la vie ?

Grands domaines et règnes

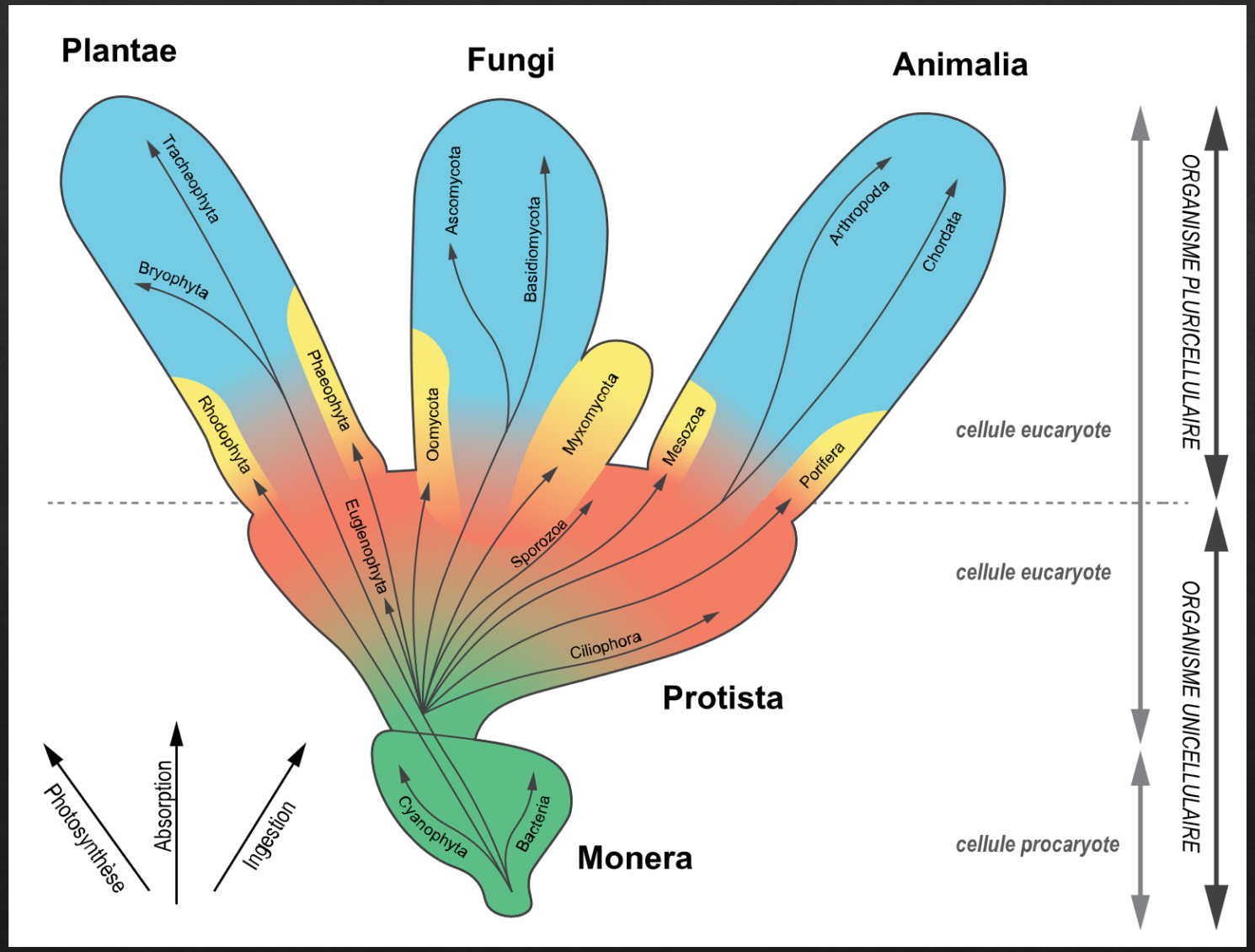
Bacteria **Archaea** **Eukaryota**

Origine des organelles

ENDOSYMBIOSIS



Autre vue du vivant



Revenons à nos bactéries

Qu’est-ce qu’une bactérie ?

Micro-organismes : 1/10 de la taille d’une cellule humaine

Unicellulaires : une seule cellule, se reproduit par mitose / division cellulaire

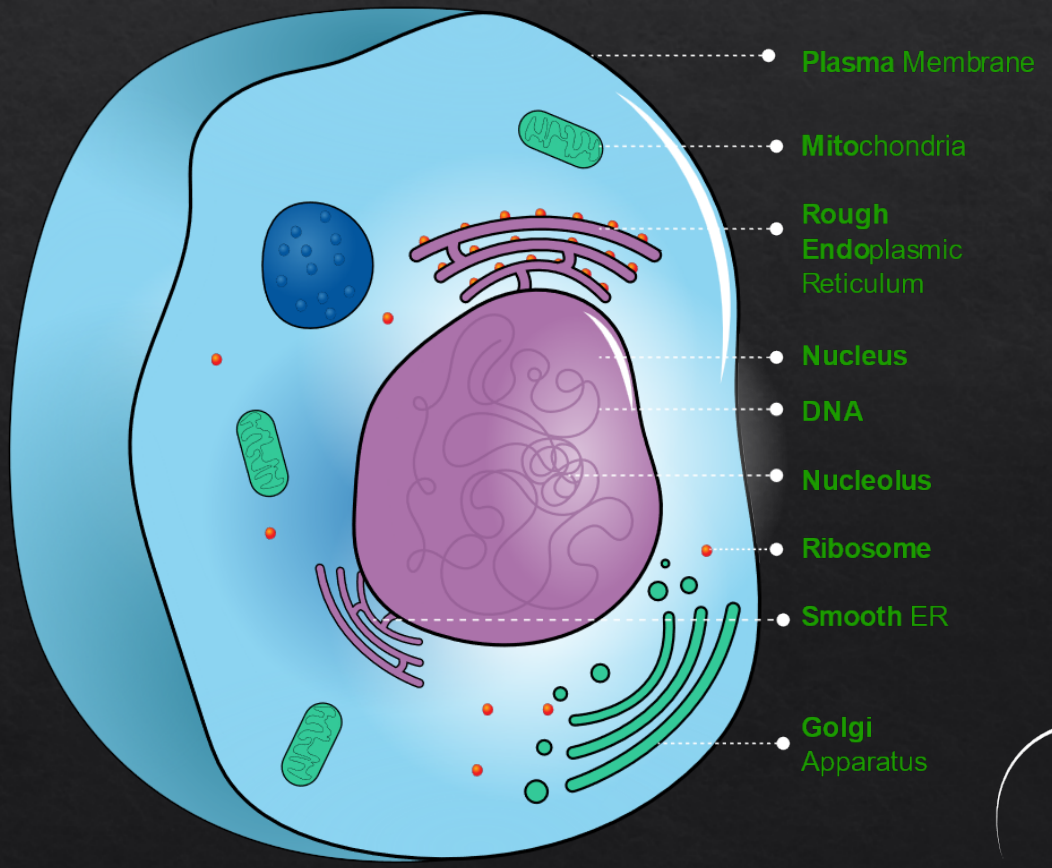
Procaryotes : ne possède pas de noyau pour protéger le matériel génétique

Simplicité: De la transcription au métabolisme, tout est plus simple. Il n’y a pas d’organelles spécialisées comme le réticulum endoplasmique ou les mitochondries.

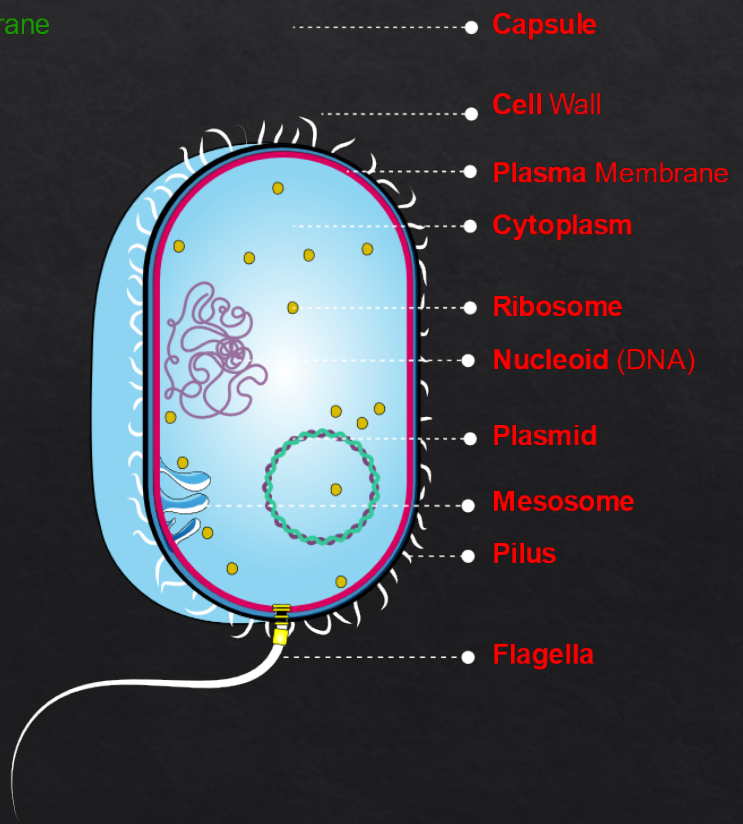
Eucaryotes vs procaryotes

Qu'est-ce qu'une bactérie ?

Eucaryotes

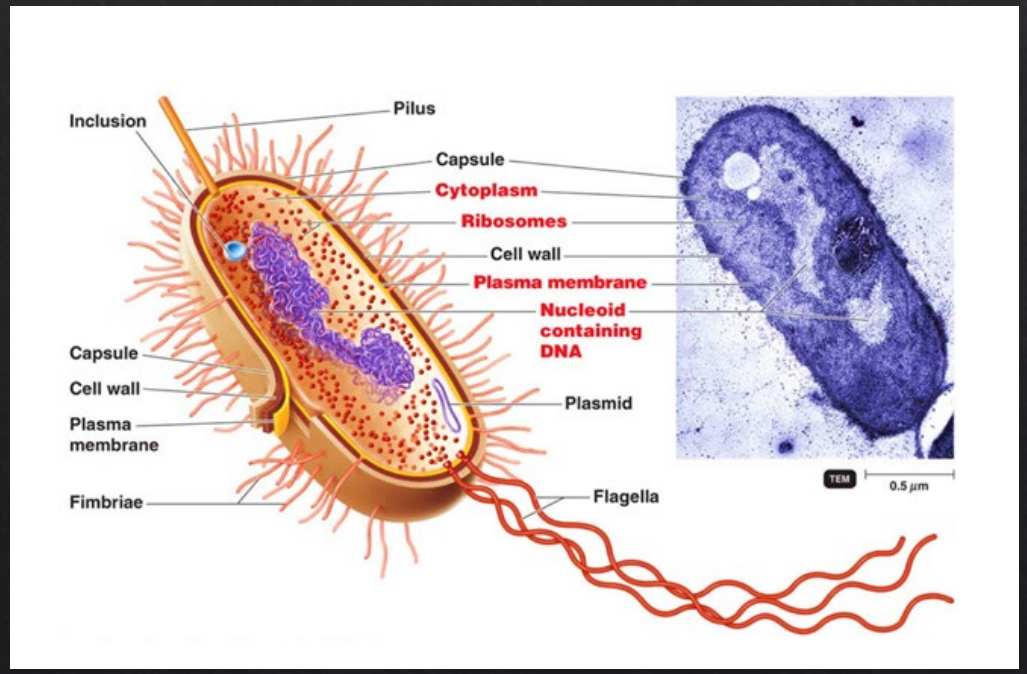


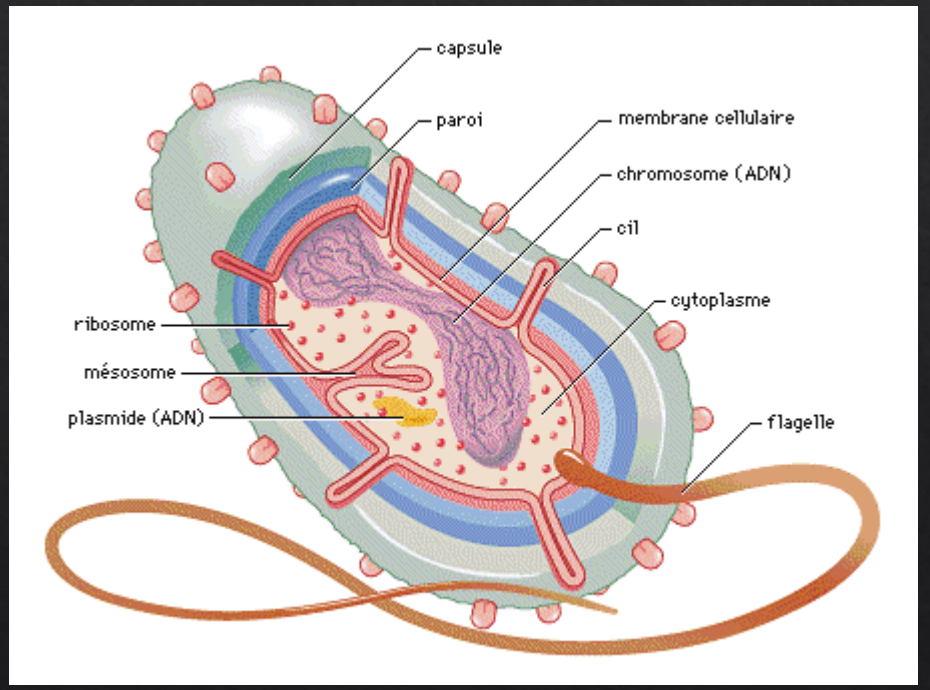
Procaryotes



Bacteria

Qu’est-ce qu’une bactérie ?
Plan de base

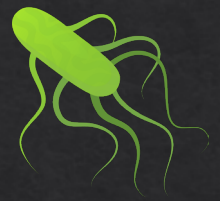




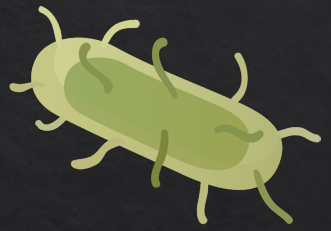
Bacteria

Grande diversité

Helicobacter pylori

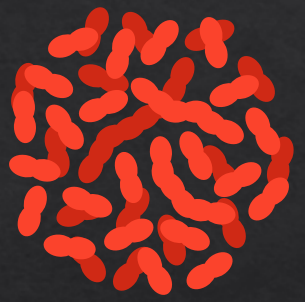


Salmonella



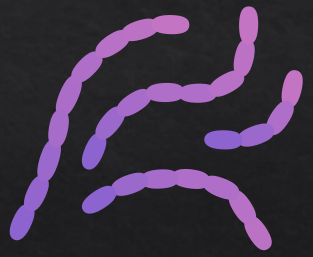
Yersinia pestis

Lactobacillus



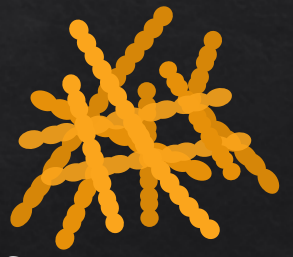
Lactococcus

Mycobacterium tuberculosis



Bacillus anthracis

Bifidobacterium

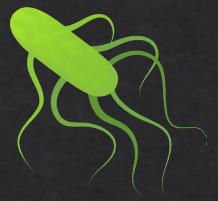


Streptococcus thermophilus

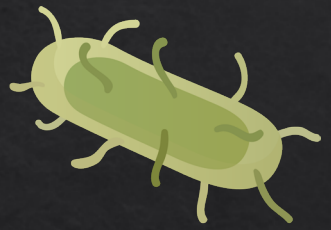
Bacteria

Grande diversité

Helicobacter pylori

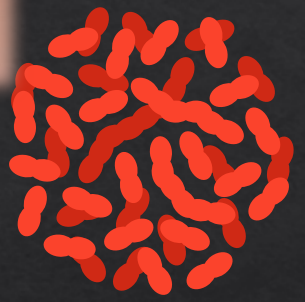


Salmonella



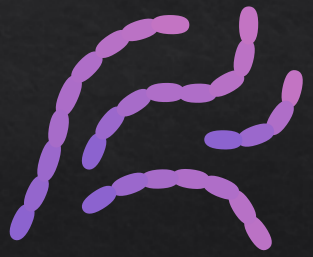
Yersinia pestis

Lactobacillus



Lactococcus

Mycobacterium tuberculosis

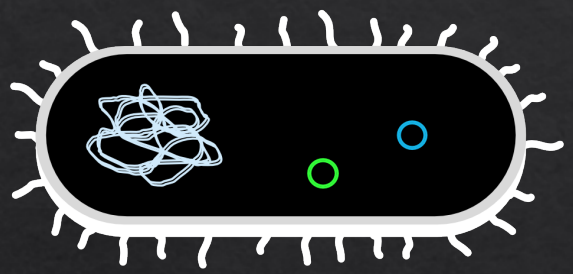


Bacillus anthracis

Bifidobacterium

Streptococcus thermophilus

Bacteria



Core pangenome

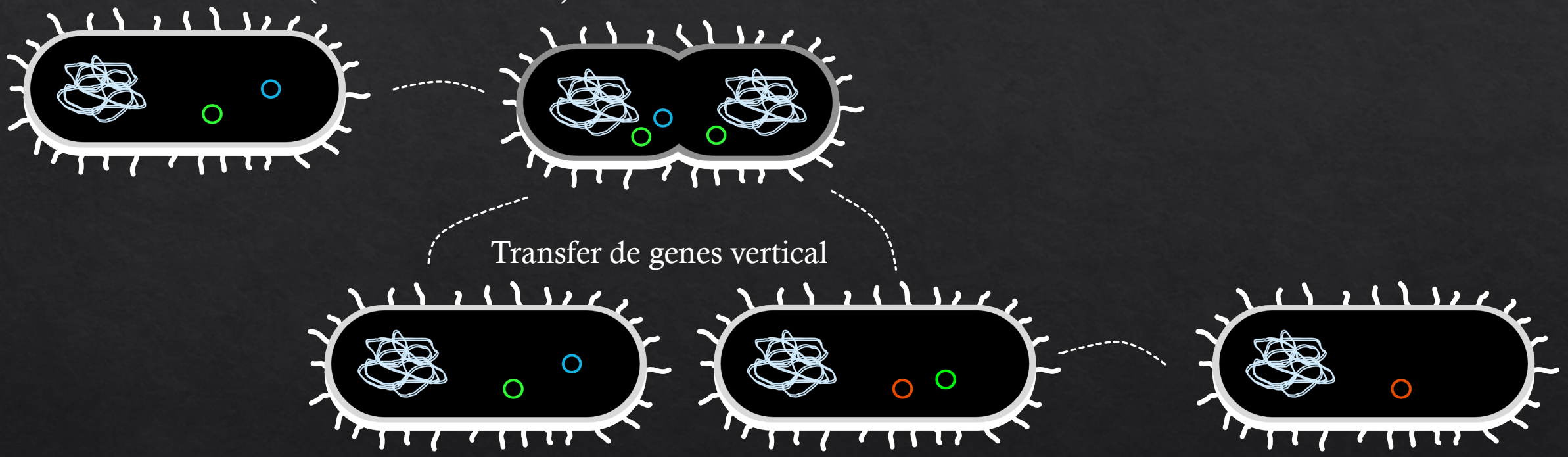
- Fonctions (protéines) vitales
- Presque identique pour toute l'espèce
- Présent sur un « chromosome »

Cloud pangenome / accessory genome

- Souvent sous forme d'ADN circulaire (plasmides)
- Présent de manière sporadique dans la population
- Peut être transmis
- E.G. gènes de résistance à un antibiotique

Bacteria

Mitose
(Division cellulaire)

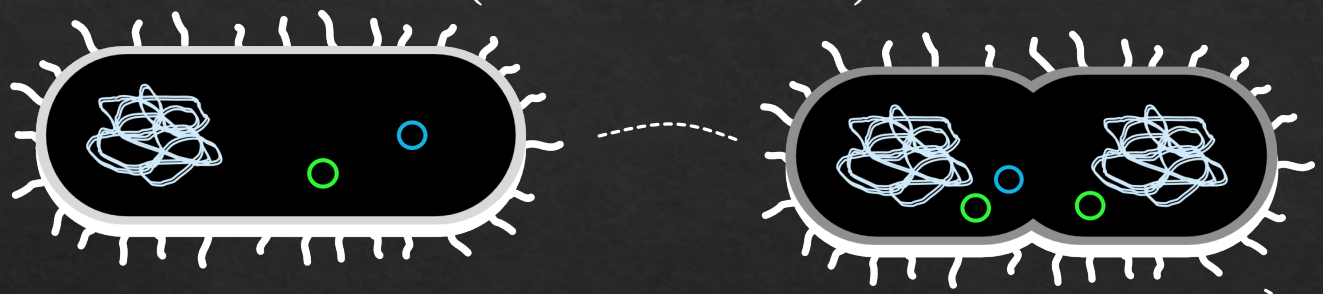


Transfer de genes vertical

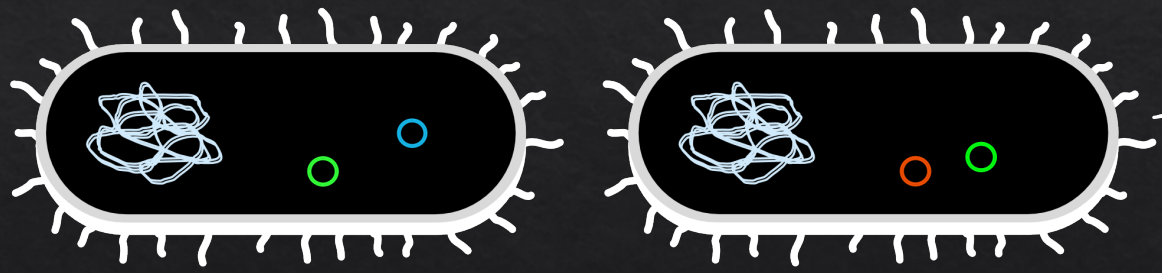
Transfer de genes horizontal

Bacteria

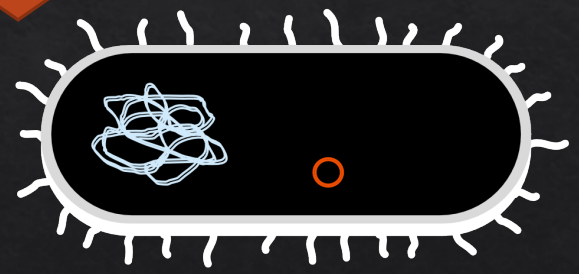
Mitose
(Division cellulaire)



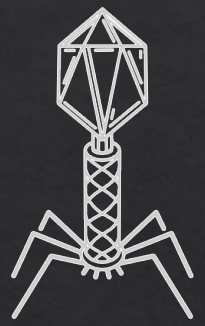
Transfer de genes vertical



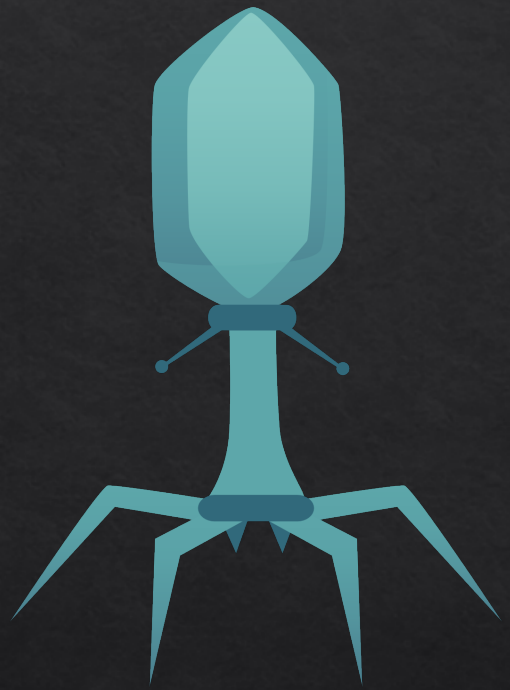
Possibilité d'échanger des gènes: e.g. un gène qui produit une résistance à un antibiotique !



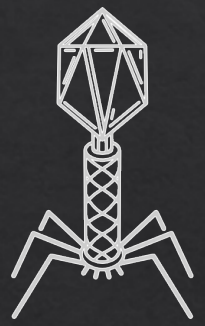
Transfer de genes horizontal



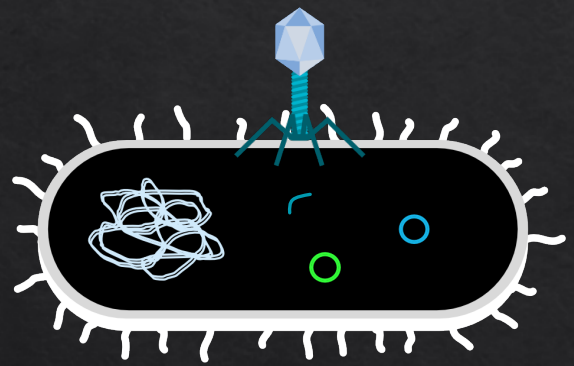
Bacteriophages



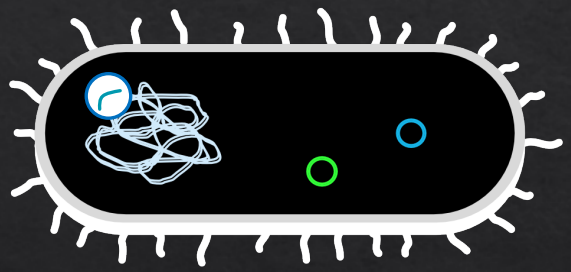
- 1/10 de la taille d'une bactérie
- On en trouve partout où on trouve des bactéries
- Il y a environ 10^{31} phages sur Terre (1 trillion par grain de sable)
- Ils tuent entre 15% et 40% de toutes les bactéries sur terre tous les jours



Bacteriophages infection

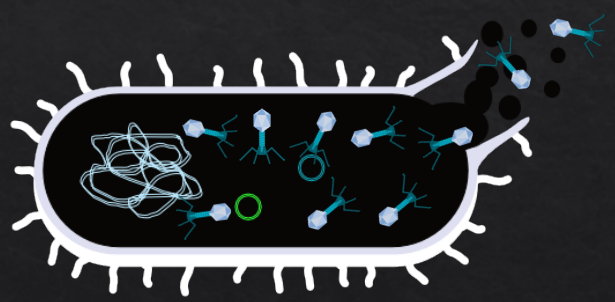


Injection du matériel génétique



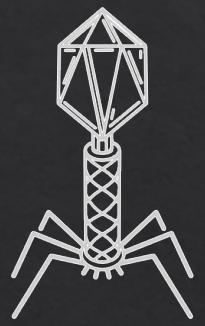
Cycle lysogénique

L'ADN du phage est intégrée au génome de la bactérie en attente d'un signal.



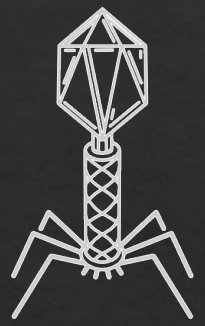
Cycle lytique

La bactérie produit des nouveaux phages jusqu'à la lyse de la cellule (explosion / mort)



Bacteriophages infection

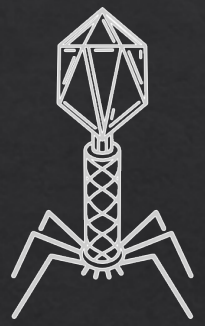
Une infection de bactériophages peut mener à un échec total de la fermentation. Pas de bactéries, pas de fromage.



Bacteriophages infection

Une infection de bactériophages peut mener à un échec total de la fermentation. Pas de bactéries, pas de fromage.

Et ça, c’est triste !



Bacteriophages infection

Une infection de bactériophages peut mener à un échec total de la fermentation. Pas de bactéries, pas de fromage.

Et ça, c'est triste !

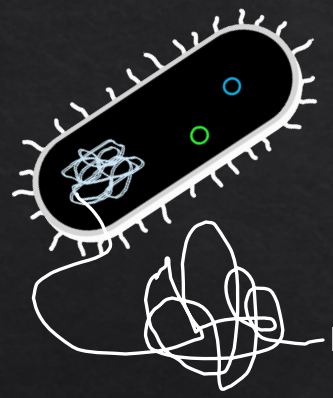
Mais ….

Bacterial defences

- **Mécanismes de défenses innés**
  - **A**bortive **I**nfection (suicide)
  - **R**estriction **M**odification (découpe l'ADN inconnu)
  - **P**hysical barriers (mutation des récepteurs membranaires)
  - **E**nzymes (défense chimique)
  - …et beaucoup d'autres
- **Mécanismes de défenses adaptatifs**
  - CRISPR Cas System (mémorise, reconnais et détruit de l'ADN)

CRISPR Cas

- CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
- Cas : CRISPR associated protein
- La sequence qui contient Repeats et Spacers s'appelle CRISPR Array / Tableau CRISPR



Palindromic Repeats
- Presque identique pour chaque tableau
- Placé entre chaque spacer

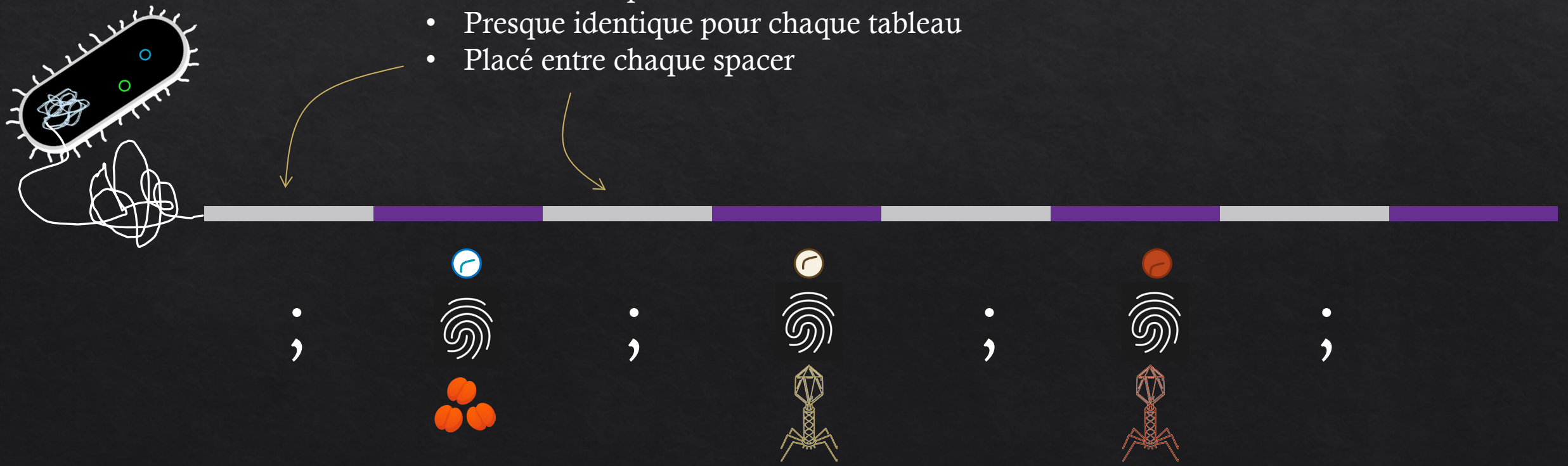
C'est une mémoire !

Spacers
- Séquence d'ADN étranger
- Va servir à reconnaitre et détruire une séquence similaire

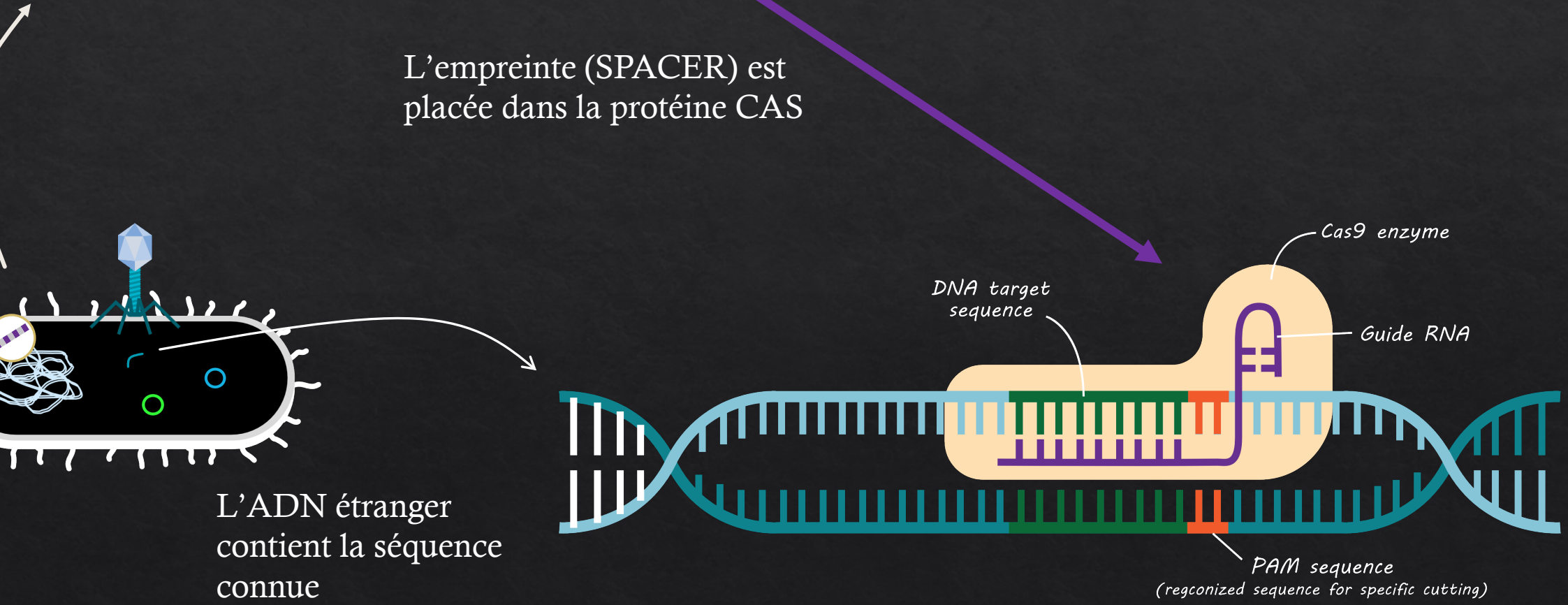
CRISPR Cas

Palindromic Repeats

- Presque identique pour chaque tableau
- Placé entre chaque spacer

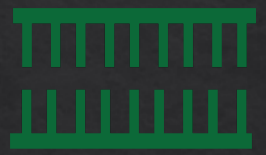


CRISPR Cas


L’empreinte (SPACER) est placée dans la protéine CAS
L’ADN étranger contient la séquence connue
Cas9 enzyme
DNA target sequence
Guide RNA
PAM sequence
(regconized sequence for specific cutting)

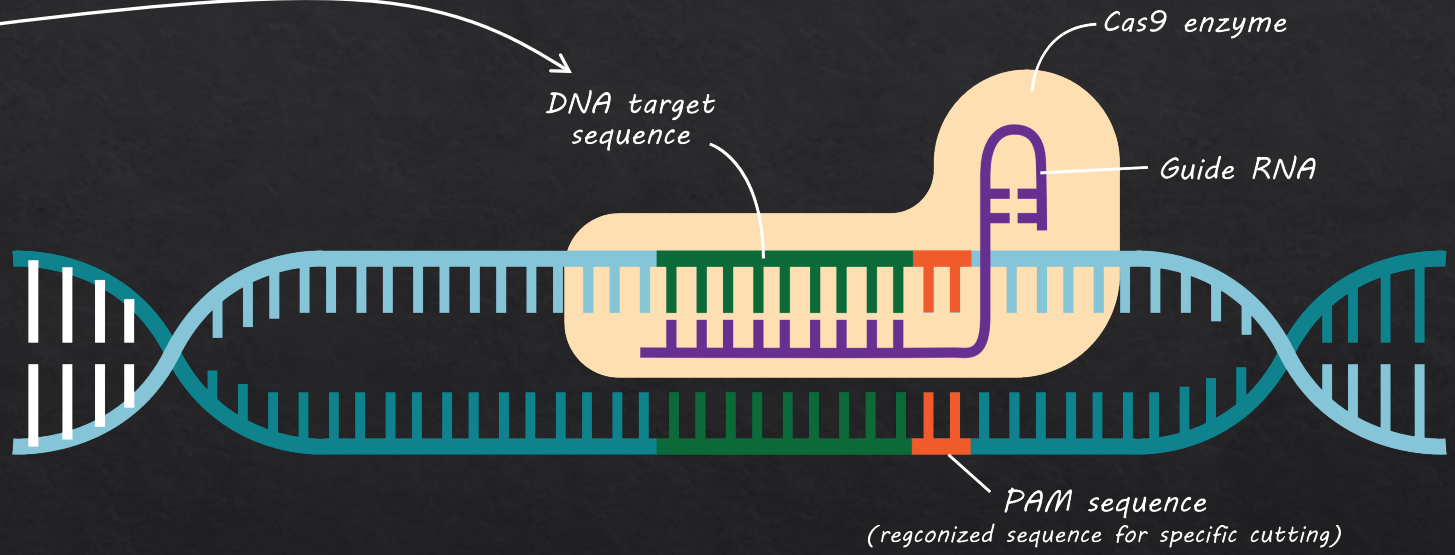
CRISPR Cas

Previously « seen »
foreign DNA



(reverse)

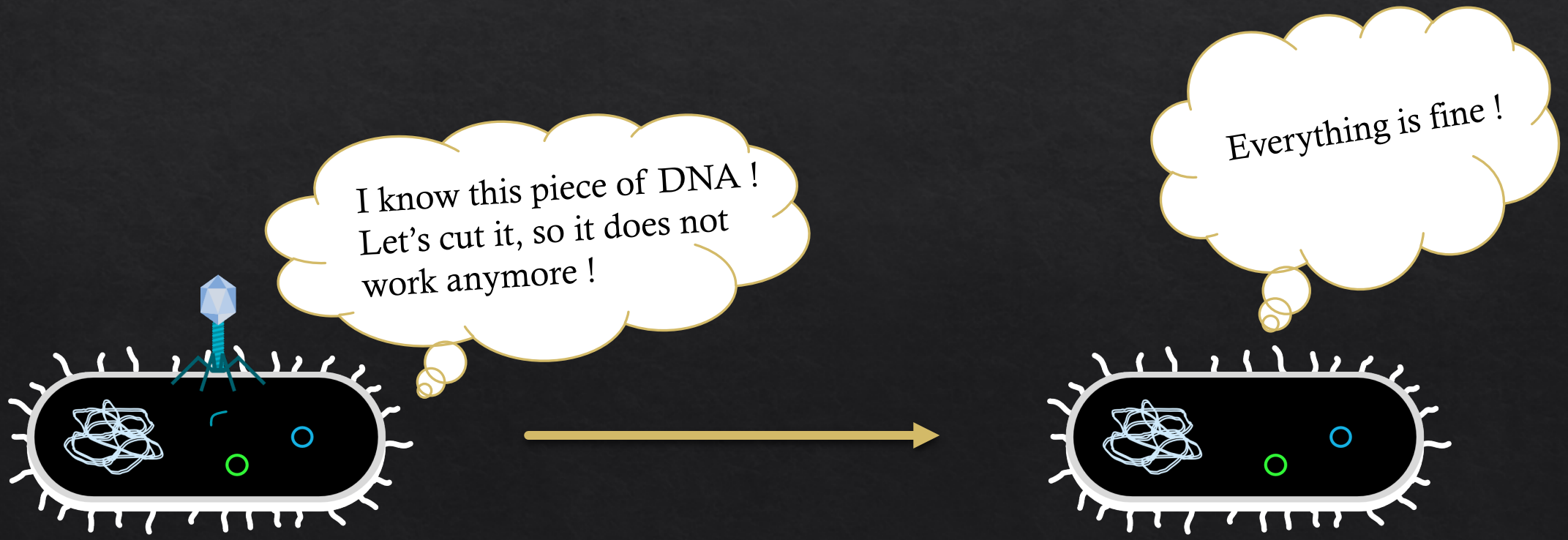
Second visit of
foreign DNA



target sequence
were cut off

CRISPR Cas

- Pour résumer, CRISPR Cas est un ciseau moléculaire permettant de couper l'ADN à un endroit précis
- Vidéo: https://www.youtube.com/watch?v=Aqw4DihmoQY



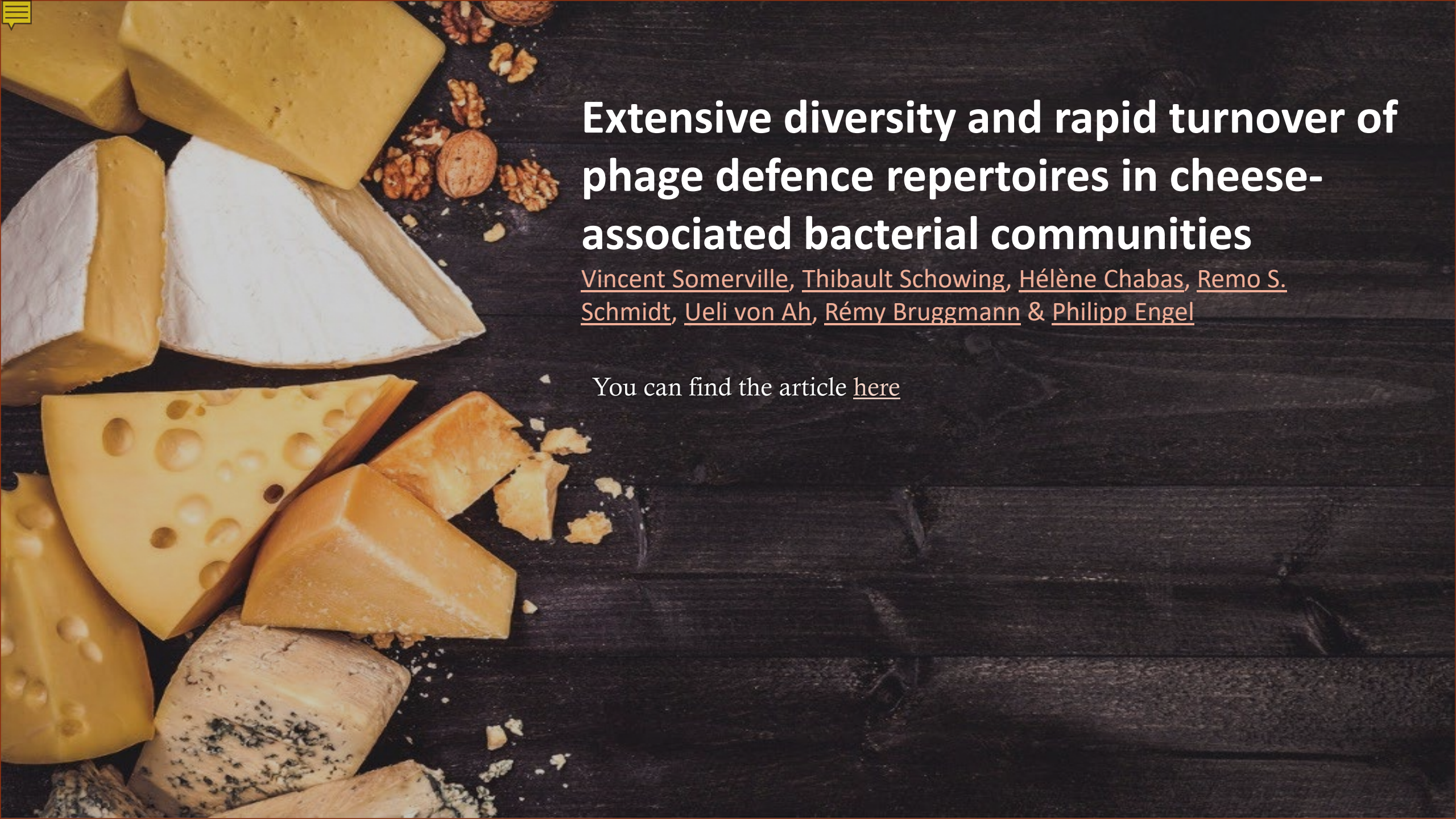
But du TP

- Comprendre les bases de CRISPR-Cas dans les défenses bactériennes

- Utiliser des outils de bioinformatique communs et spécifiques :
  - Obtenir un génome d'intérêt dans une base de données du NCBI
  - Trouver les éléments CRISPR-Cas de ce génome grâce à CRISPRCasFinder
  - Une fois les SPACERS identifiés, trouver quelles sont leurs cibles grâce à BLAST.

A faire

- Récupérez le TP3 sur Cyberlearn (Instructions PDF)
- Suivez les étapes et répondez dans un fichier à part
- Si un service n'est pas disponible, vous pouvez demander les fichiers intermédiaires aux assistants



Extensive diversity and rapid turnover of phage defence repertoires in cheese-associated bacterial communities

Vincent Somerville, Thibault Schowing, Hélène Chabas, Remo S. Schmidt, Ueli von Ah, Rémy Bruggmann & Philipp Engel

You can find the article here

This project's goals

- Understand the repartition of these defence mechanisms within bacterial species and cheese environments
  - Most previous studies have looked at the presence of **defense systems in isolate genomes** and hence it remains elusive how defense systems are distributed within and across closely related communities
  - Cheese-associated bacterial communities harbor only a few bacterial species, have been propagated in relatively stable environments (i.e. cheese or milk) over generations, and are known to be exposed to diverse phages
- Understand the dynamics of bacterial defence adaptivity with the CRISPR Cas systems and horizontal gene transfer

Raw Data

Quick file and data definition:

- **Genome**:
  - Fasta format (.fna, .fasta)
  - One or more contigs

    (continuous string of ACTG's)

```
>NZ_CP018218.1 Lactobacillus delbrueckii subsp. jakobsenii ZN7a-9 = DSM 26046 chromosome, complete genome
TCGAACCCTGGACACCCGGATTAAGAGTCCGGTGCTCTGCCAGCTGAGCTAACGACCCAAATGGGTTTAT
TTAACTAGCGCCAAAGCGCGATGGGTCGTCAGGGGATCGAAACCCTGGACACCCGGATTAAGAGTCCGGTG
CTCTGCCAGCTGAGCTAACGACCCAAGTGGGCTGCGGCTGACTTTCGTCAACCGACATTTATTTATTATA
CAGAAACGCGACAGGAATGCAAGCCTTTTGGCAAGATTTTTGCTCAAAAACTTCCTTTTCCTGTCCAGCC
CGCCCTTTTCTTCTTTATTCTTGCCAAAAAGCACCCCGCGGGGCAACTTCCAGCCGGCGGATTCCATCCT
TTAAACCCGCCTCTCTGACCGCGTCGACAAAGTCGCTGATATGCTGGGGGTCAACCAGGCTCATAATGGT
CGGGCCGGCTCCTGACAGGTAGGTCGCTACCGCCCATGCTCATGGCCGACATTGCGGATGATCTCCAGTT
CGGCACCAATTTGTTCTGTAGCGCTCATGGAAGAGGTCAGCCTCCATTAACCCCGGCCTTTTCATAATCC
TGGGCAAAAAGGGCCGCTACCGCCGTGTTGGCCACGGCACTGGCCTGAACCGCCTCTTTAAAGGAAATCT
GCTTGGGCAAAACGGCCCGGGCATCGCTGGTCTTTAAGTTATAGGCGGGATGTAGGCGATCATGGCAAAA
GGCGGTACCGGGGCCTTGACAGCTGTAAAGTGGCCGTTGACCTTGCAGCCGACAACCAGGCTGCCCAGAA
TGGTCGGGGCTACATTGTCCGGATGCCCTTCCAGTTGGGCCGCGATTTCGATTTTTCATGGGGATCCAGG
```

- **Metagenome**:
  - Short reads, directly out of the sequencer
  - Single or Paired end sequences
  - Fastq format (Fasta with quality score)
  - Much heavier file that genome

Header + Sequence

Header + Q score

```
$ head ERR3340822.fastq
@ERR3340822.1.1 NB501531:11:HY75CBGXY:1:11101:21540:1038 length=151
CACTTNACGAAACCCTACGATGGCGTCTTTATCTTAGGGATTTTACGCGAAACGCCGGCTGCTAAAATGGGGTTGGTCGT
TGGCGATACAATTGTTGAATGTAATGGNGAGGCTGTCTCAAATAACGATAATTTTNNTNGGGCGATTNNNN
+ERR3340822.1.1 NB501531:11:HY75CBGXY:1:11101:21540:1038 length=151
????????????????????????????????????????????????????????????????????????????????
????????????????????????????????????????????????????????????????????????????????
@ERR3340822.1.2 NB501531:11:HY75CBGXY:1:11101:21540:1038 length=151
GTTTAAATCTTGAACCTTTAAATGGCAATACGTTGGCTGACTTTGAATCGCCCGATAAAAATTATCGTTATTTGAGACAN
NCTCGCCATTACATTCAACAATTGTATCGCCAACGACCAACCCCATTTTAGCAGCCGGCGTTTCGCGTAAA
+ERR3340822.1.2 NB501531:11:HY75CBGXY:1:11101:21540:1038 length=151
????????????????????????????????????????????????????????????????????????????????
????????????????????????????????????????????????????????????????????????????????
@ERR3340822.2.1 NB501531:11:HY75CBGXY:1:11101:5051:1040 length=151
GGATCNTCTCTACGGCGAGCTTCATAGCGGTGGTTTTGCCGGCTCCGGCAGGGCCGATGCCGACTGCGAGCACTCTTTCC
CCGCAGGCGAATGCGCGGGCCAGTTCGATCTGGCCGTCGTCGAGAGGTATGGACTGCTCTGCGAGGANNNN
```

Raw Data

Cheese Metagenomes – Starter culture and cheese core

- 188 Shotgun metagenomic samples
- 240 16S-RNA amplicon seq. samples
- Data from publications and FoodMicrobionet database

Genomes Databases - 185 identified species

- RefSeq (Max 500 genome per species)
- Dialact (536 genomes, 9 species, Swiss cheese)

Methods

Now that we have a ton of Genomes and Metagenomes what can we do ?

Using existing tools, we can:

- Extract known innate defence mechanisms (Hidden Markov Models)
- Extract CRISPR Arrays and associated proteins (CRISPRCasFinder / CRASS)
- Pairwise strain comparison (FastANI)

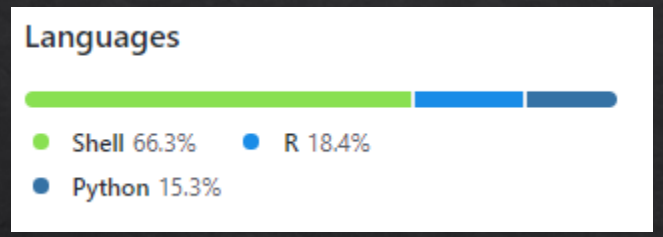
- Sequence clustering (100% and 80% ID) (CD-HIT-EST)

- Mapping spacers to viral and bacterial databases to find targets (BLASTn – megablast)

- Mapping repeat-spacer-repeat sequences to reads to quantify Array VS foreign DNA (BWA, samtools, awk, sed, bedtools)

- Spacer pairwise comparison (6 billion strings comparison with R) (future_map2_dbl)

Methods



- Shell is mainly used for HPC jobs
  - Bash scripts
  - Python for quick files parsing
  - R for specific jobs
- R is used locally once the big jobs have generated usable flat data.