

DE LA RECHERCHE À L'INDUSTRIE



université  
PARIS-SACLAY



[www.cea.fr](http://www.cea.fr)

# Des pangénomes des microorganismes aux microbiotes

David Vallenet

Laboratoire d'Analyses Bioinformatiques pour la Génomique et le Métabolisme

Séance publique de l'Académie d'Agriculture de France  
13/04/2022



**UMR GENOSCOPE**  
**METABOLIC GENOMICS**

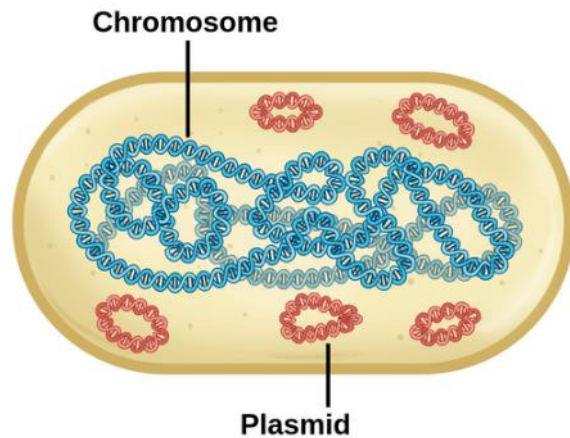
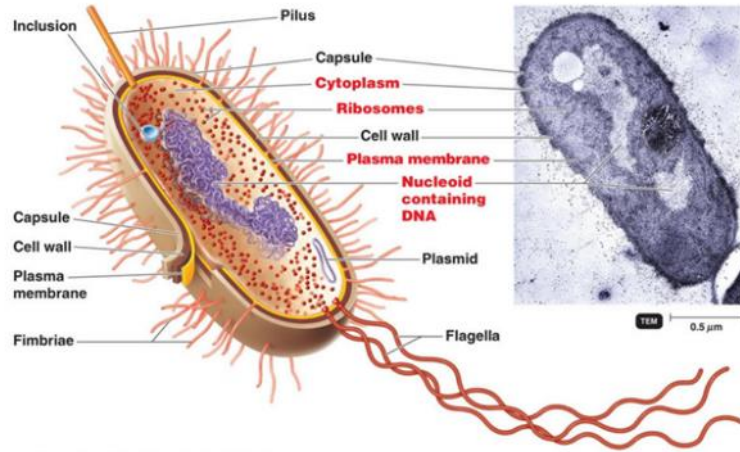


RESEARCH UNIT METABOLIC GENOMIC UMR8030

CEA . DRF . IBFJ . GENOSCOPE

2 RUE GASTON CRÉMIEUX . CP 5706 ÉVRY CEDEX . FRANCE

# Les microorganismes procaryotes



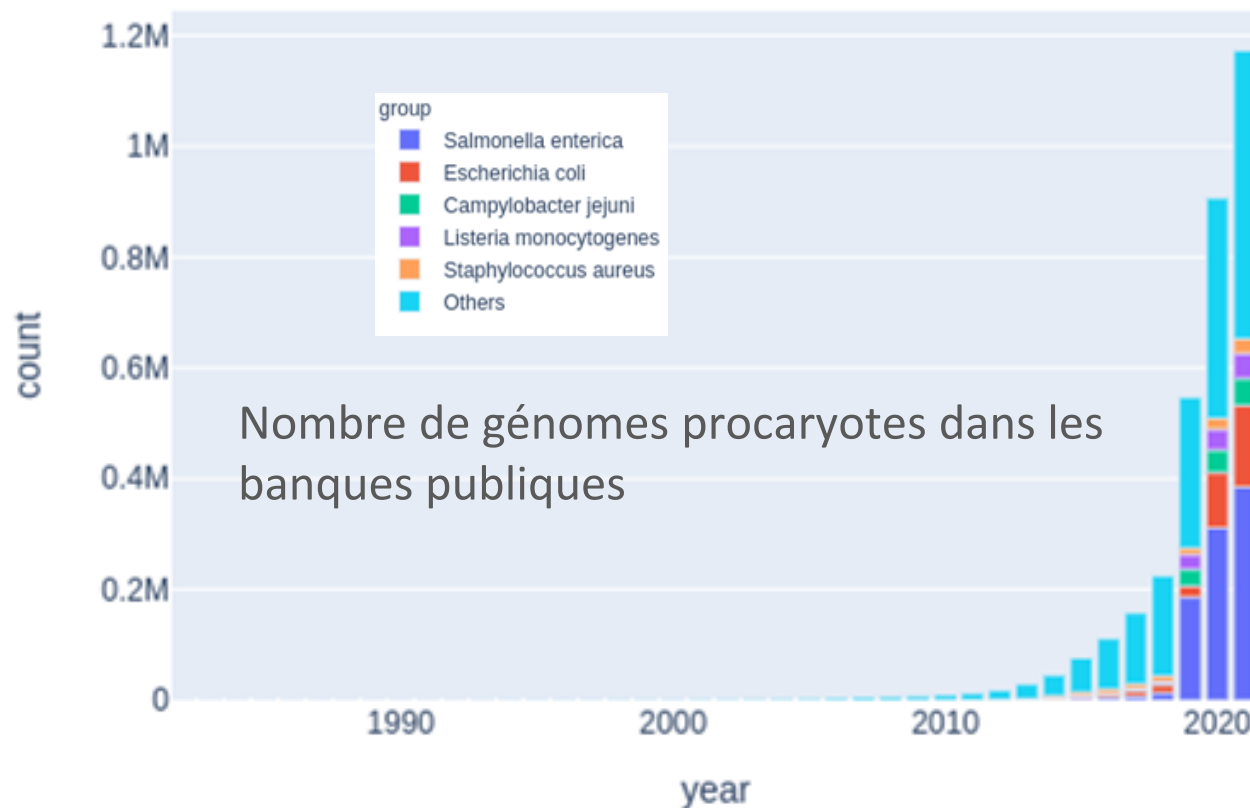
## Bactéries et Archées

- formes de vie les plus anciennes et les plus répandues
- une grande diversité - plusieurs millions d'espèces
- présents dans la plupart des écosystèmes
- à l'origine de l'essentiel des cycles biogéochimiques
- indispensables à la survie des autres organismes  
⇒ concept d'holobionte

## Génome

- un à quelques chromosomes + éventuels plasmides
- dense en gènes : 1 gène / kb ⇒ 5 Mb -> 5 000 gènes

# Des millions de séquences de génomes ...



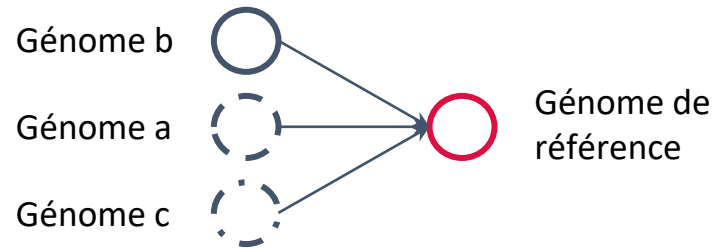
- >1 million de génomes en 2022
- fort biais dans les bactéries d'intérêt médical (>350k *Salmonella enterica*)
- ~30 000 espèces représentées par au moins un génome d'isolat

# Comment comparer ces génomes ?

Des centaines voire milliers de génomes d'une même espèce à comparer

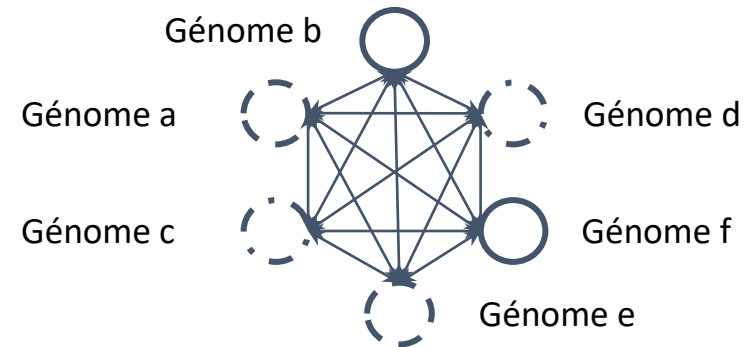
- une grande diversité du contenu en gènes issus de transferts horizontaux (5% à 40% de gènes variables)
- un haut niveau de polymorphisme (95% d'identité en nucléotides moyenne)

## Utilisation d'une référence



⇒ Problème d'exhaustivité

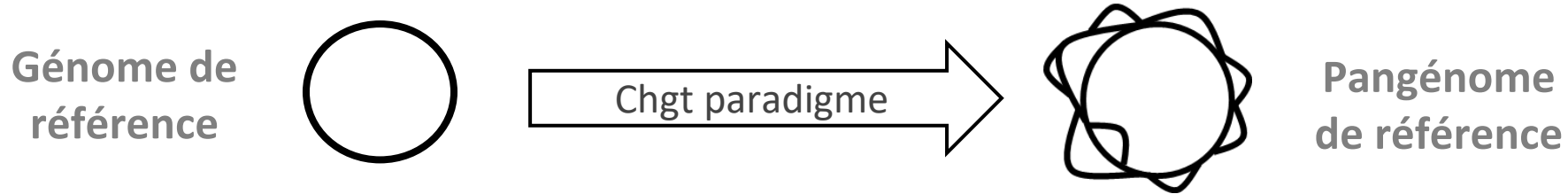
## Sans référence



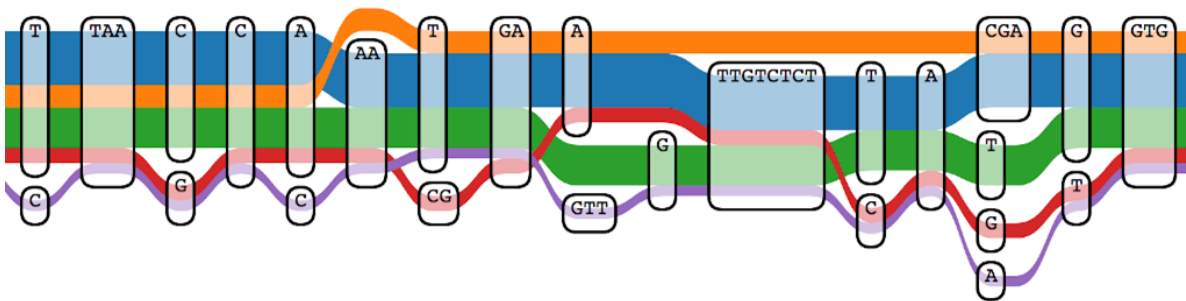
⇒ Problème de combinatoire et d'interprétation

# Concept de pangénomes

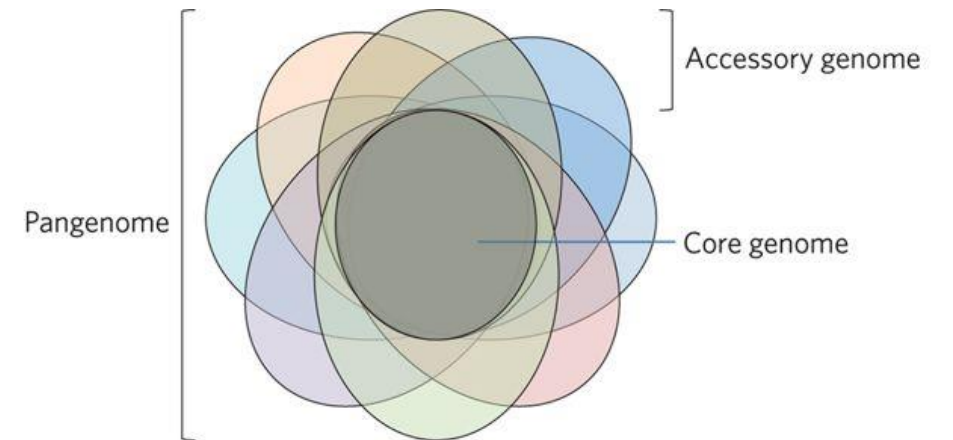
**Pangénome** : union des séquences représentant l'intégralité de la diversité d'un groupe de génomes d'intérêt généralement de la même espèce



Pangénome à l'échelle des variations nucléotidiques



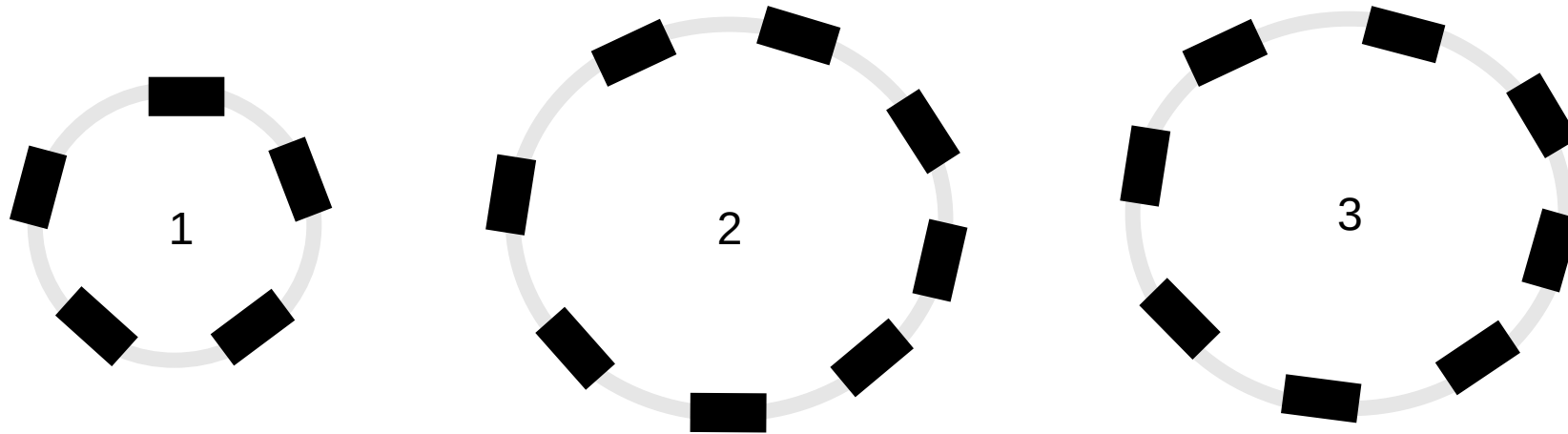
Pangénome à l'échelle du contenu en gènes



# Construction d'un pangénome au niveau des familles de gènes

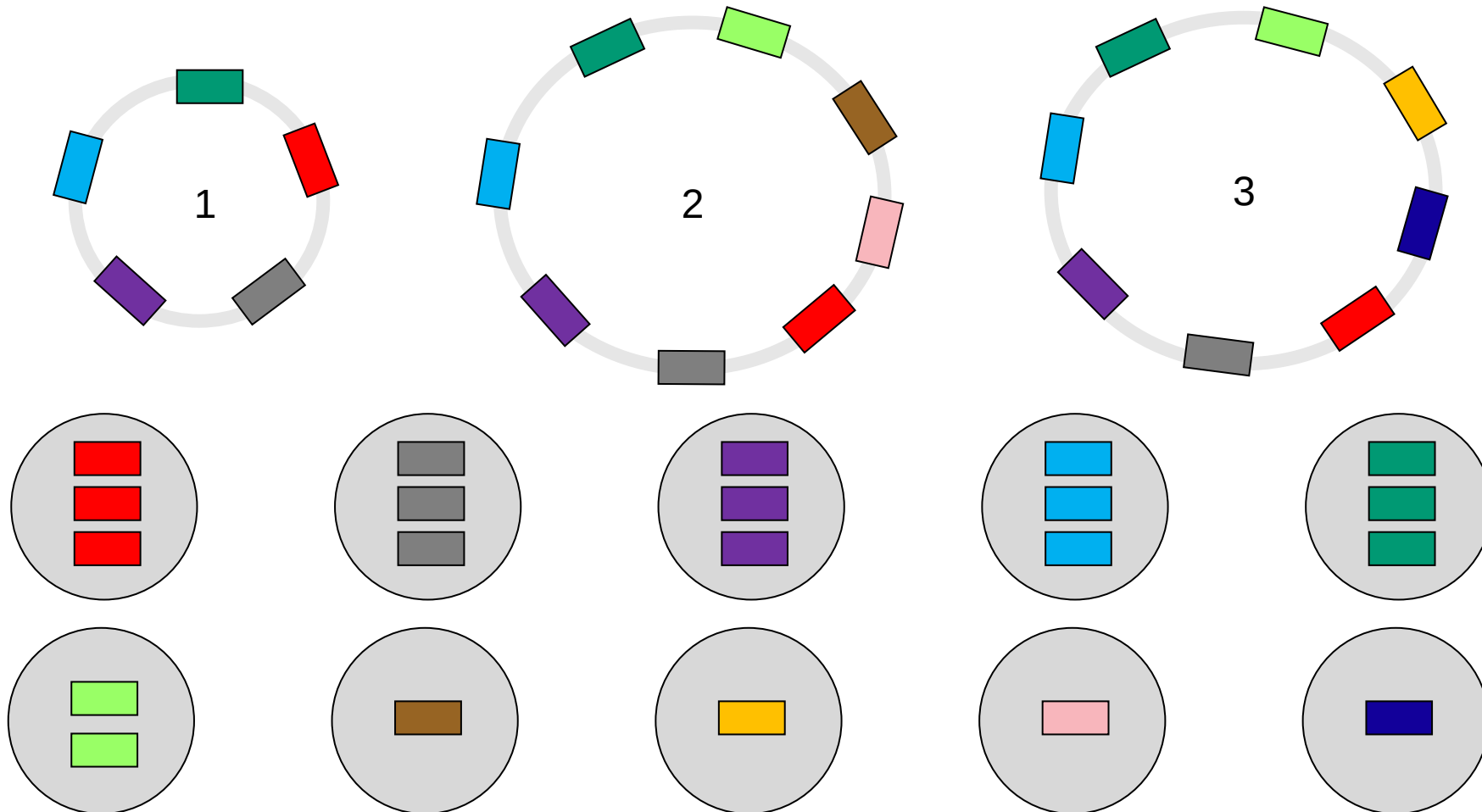
---

Exemple sur 3 génomes bactériens



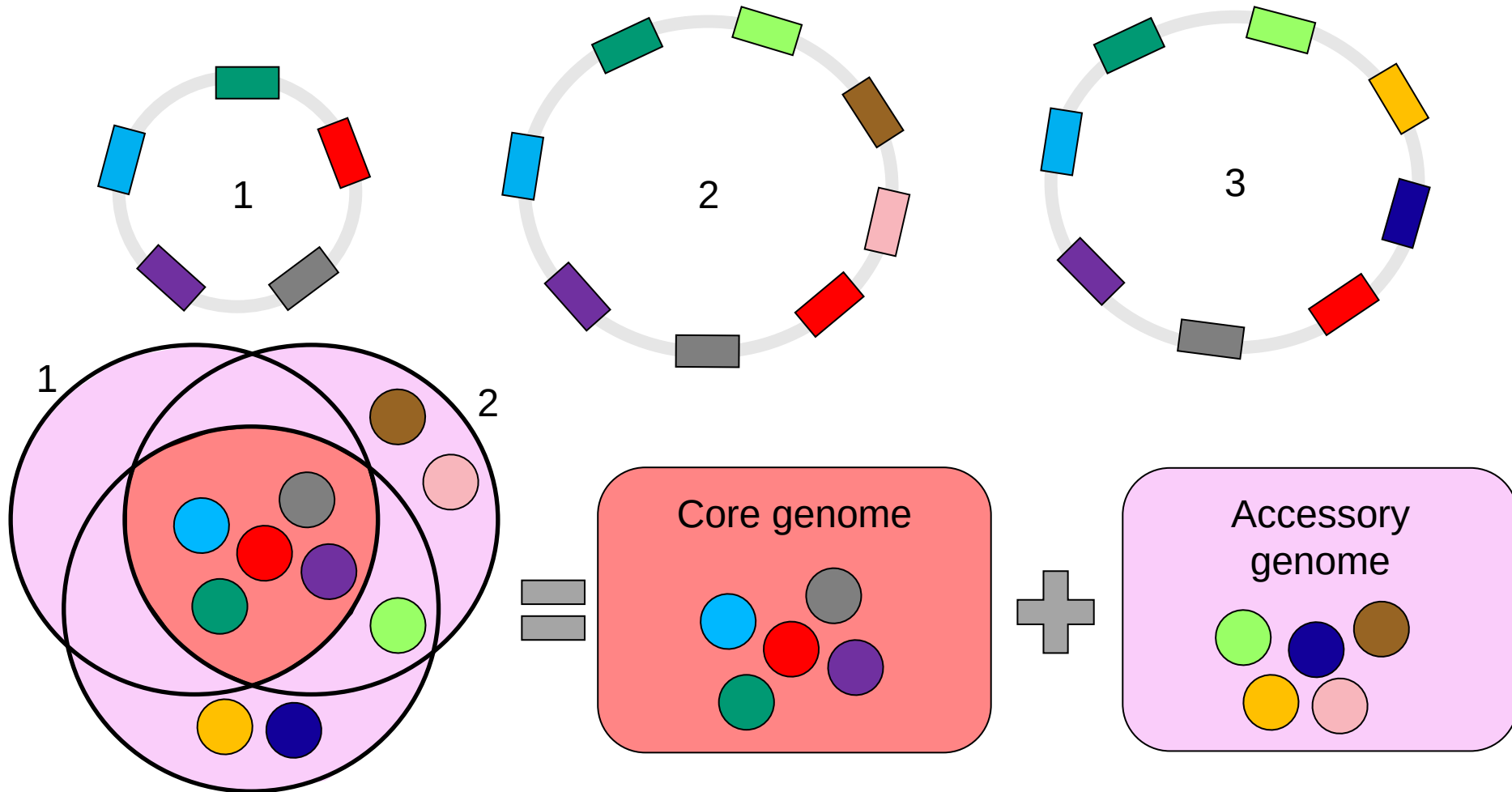
# Construction d'un pangénome au niveau des familles de gènes

Construire des familles de gènes homologues



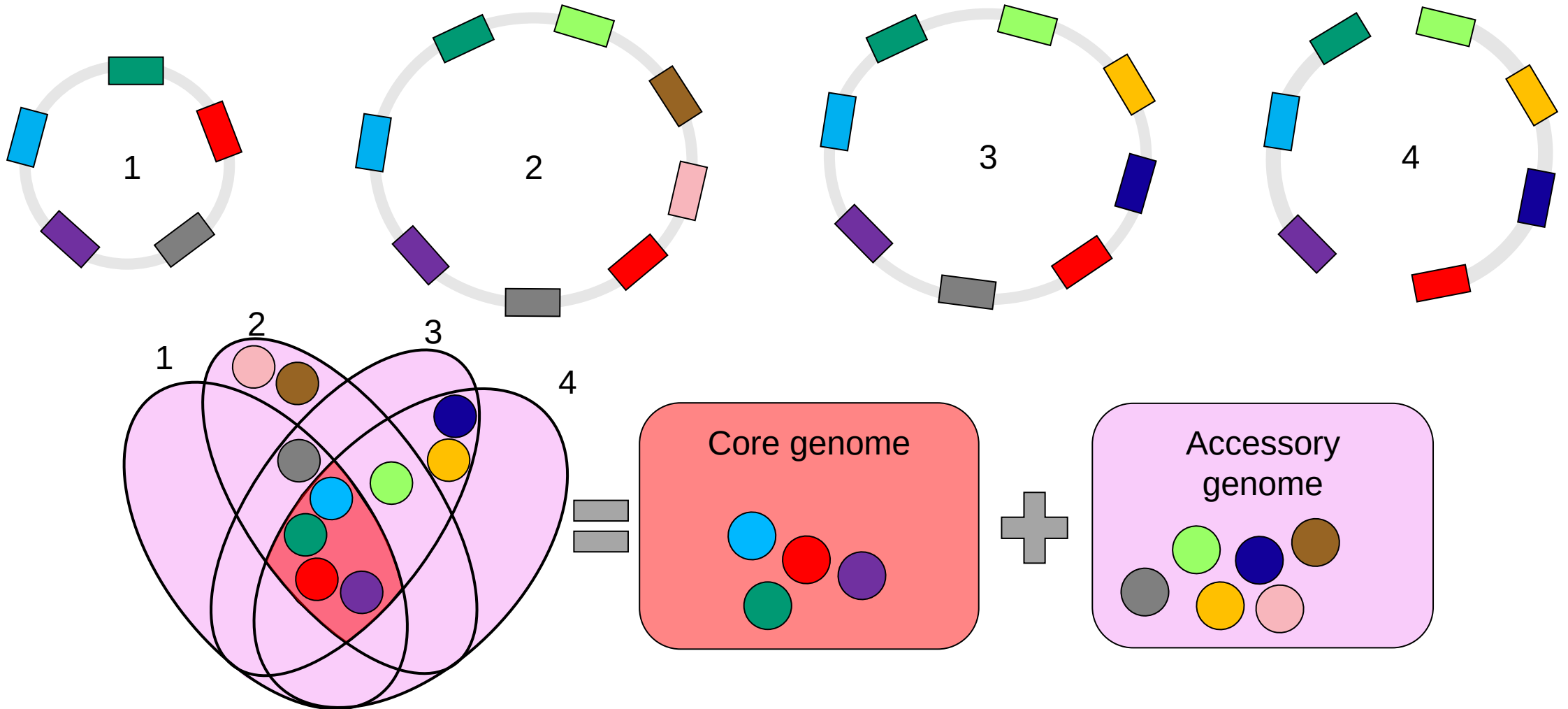
# Construction d'un pangénome au niveau des familles de gènes

Classifier les familles en éléments conservés (core genome)  
et variables (accessory genome)

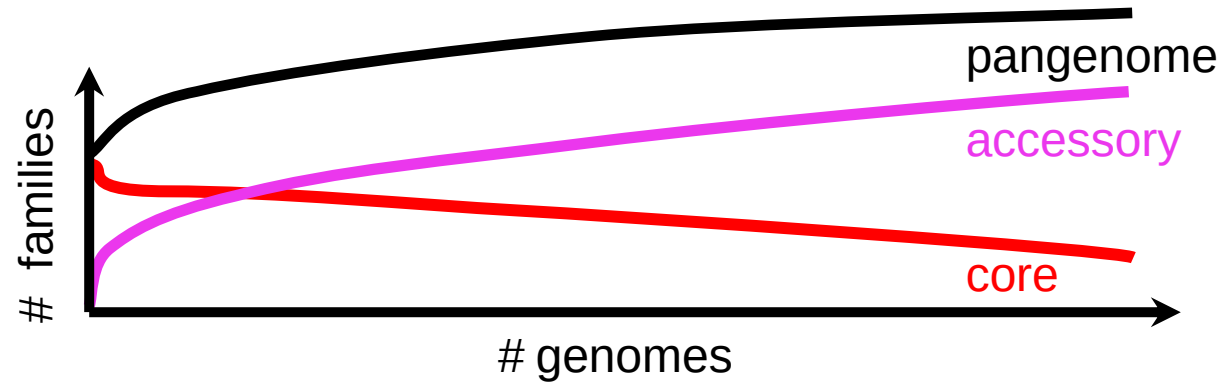


# Construction d'un pangénome au niveau des familles de gènes

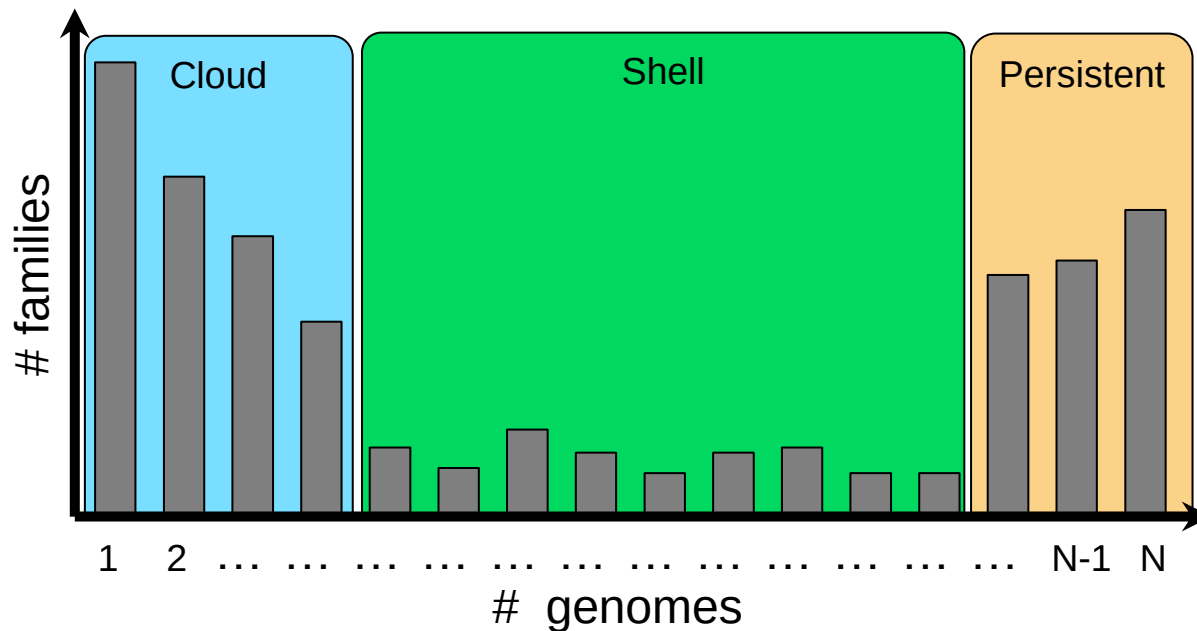
Que se passe-t-il si on ajoute un nouveau génome ?



# Comment subdiviser un pangénome ?



Définition classique du **génom** **coeur** et **accessoire trop restrictive** : pas de prise en compte des artéfacts d'assemblage et pertes ponctuelles de gènes

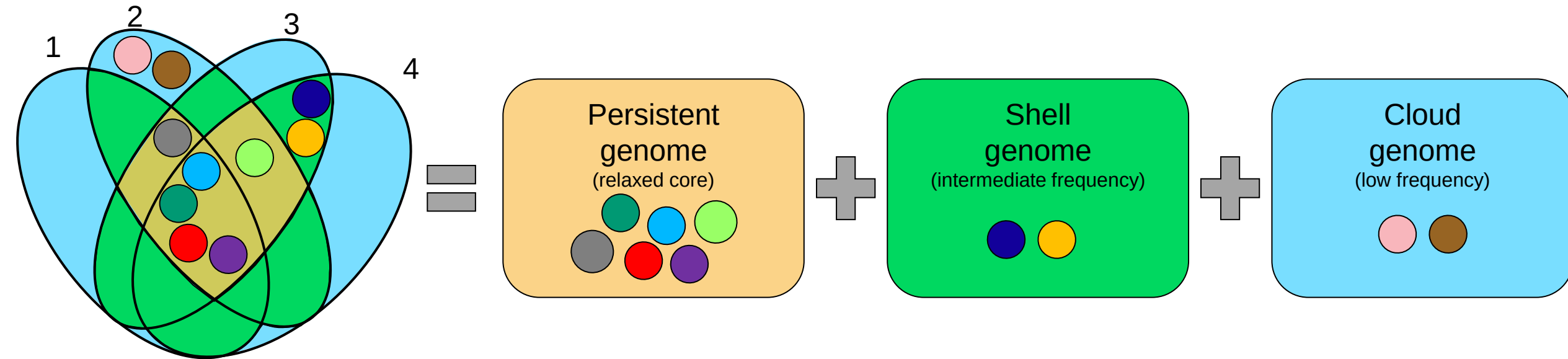


**génom** ***persistent*** : gènes conservés dans presque tous les individus

**génom** ***shell*** : gènes de fréquence intermédiaire issus de transferts horizontaux et potentiellement associés à l'adaptation environnementale

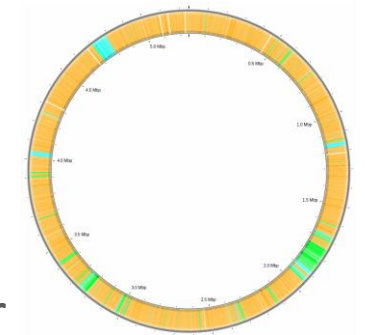
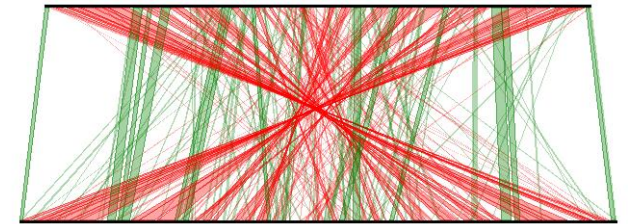
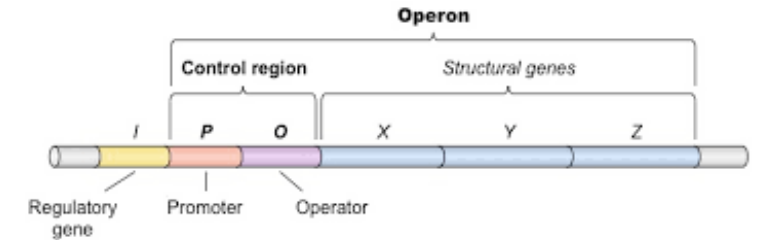
**génom** ***cloud*** : gènes de très basse fréquence issus de transferts récents et/ou peu partagés

# Pangénome en trois parties



# Importance du contexte génomique dans les génomes microbiens

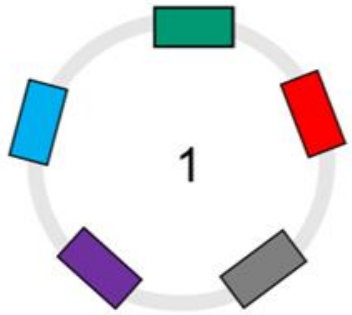
- structure opéronique, le principal système de régulation de l'expression des gènes chez les procaryotes
- les gènes *persistent* tendent à être maintenus regroupés sur les chromosomes - synténie conservée
- les flux de gènes issus de transferts horizontaux se concentrent dans certains points chauds des génomes



⇒ **graphe de pangénome** : une structure de données permettant de condenser l'information de voisinage issue de n génomes

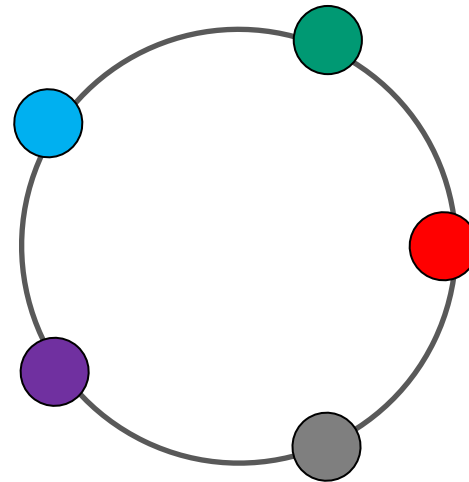
# Graphe de pangénome

---

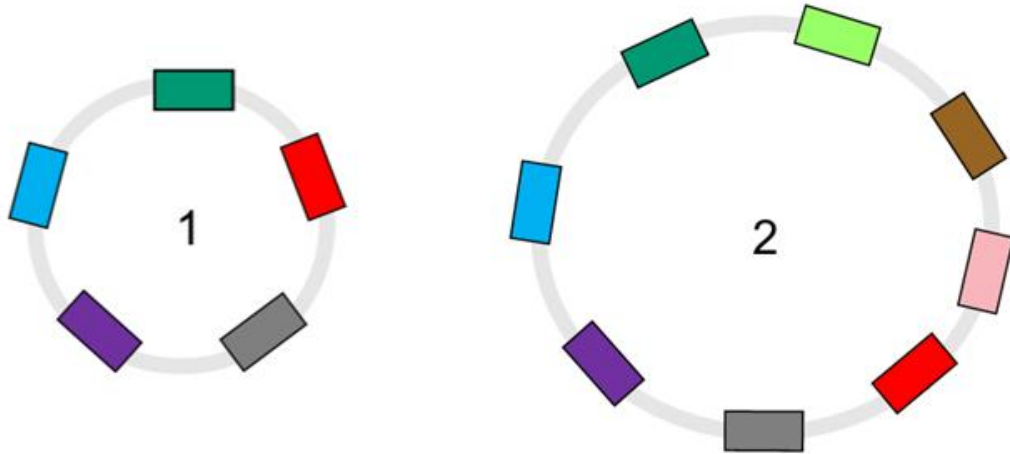


**Noeuds** : familles de gènes

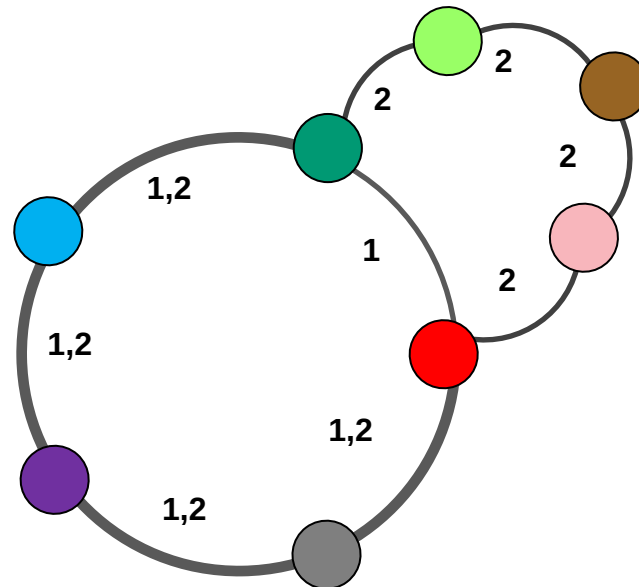
**Arêtes** : voisinage génomique des gènes  
composant les familles



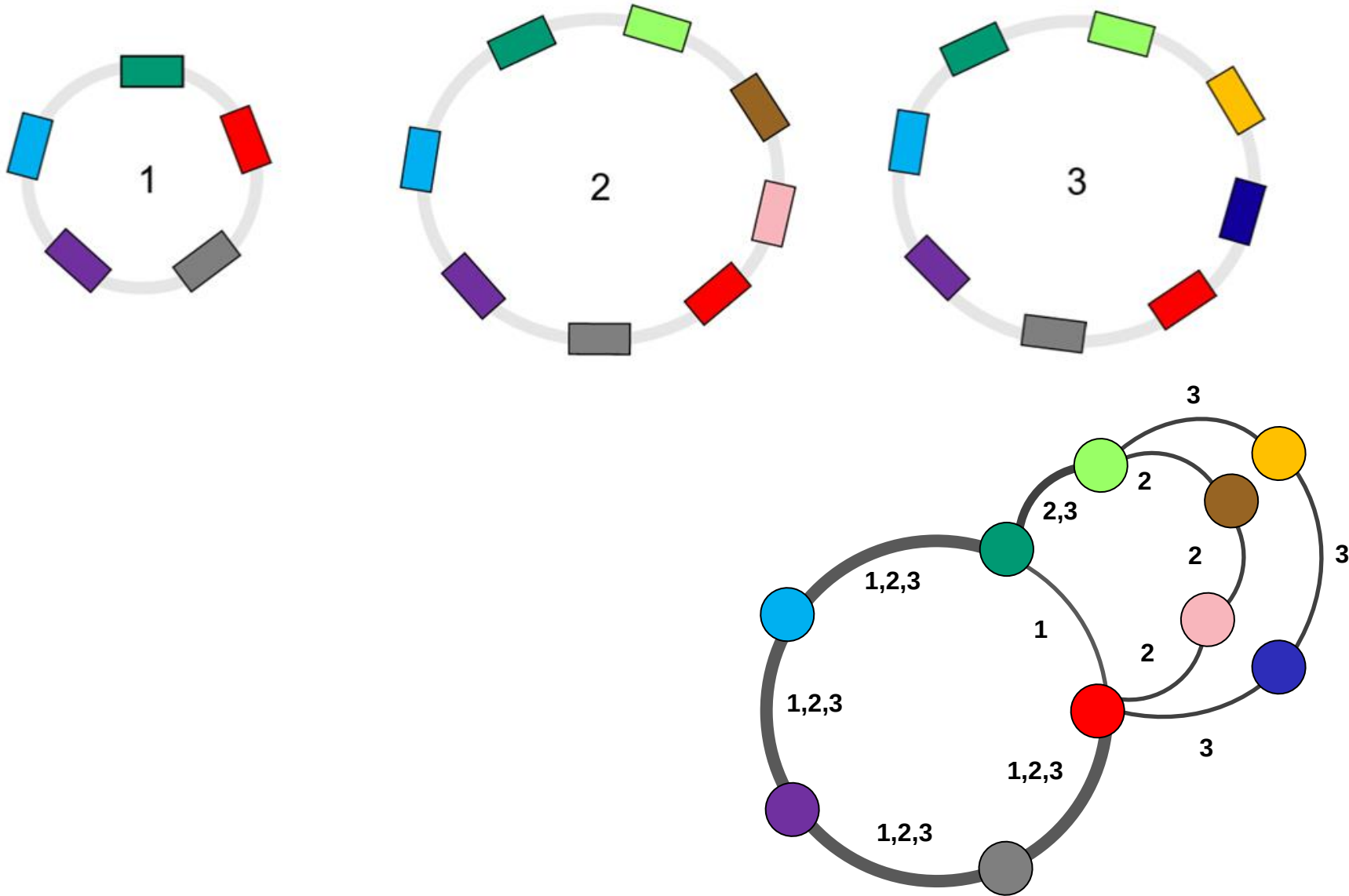
# Graphe de pangénome



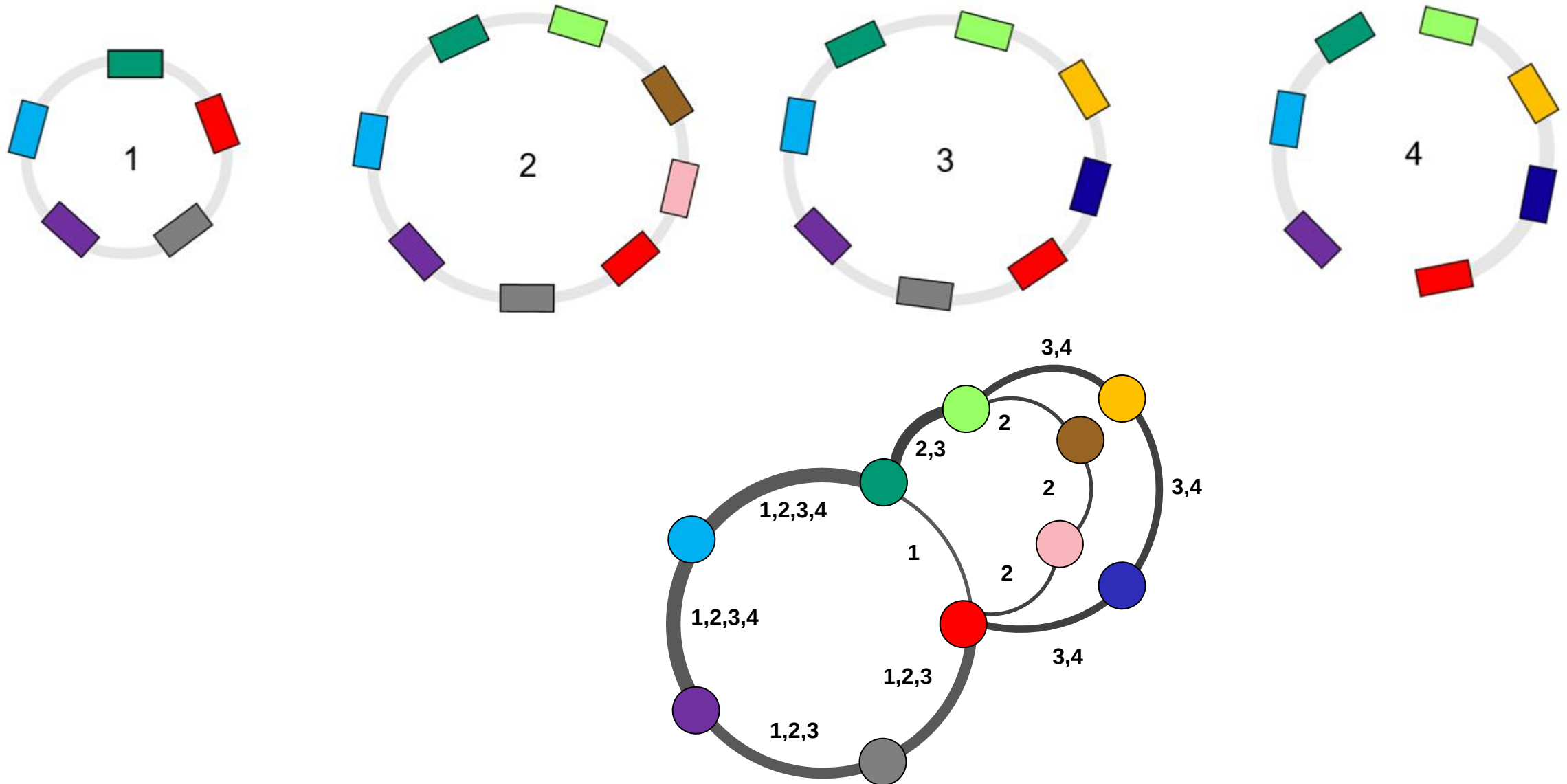
Arêtes étiquetées avec les identifiants d'organismes partageant le voisinage



# Graphe de pangénome



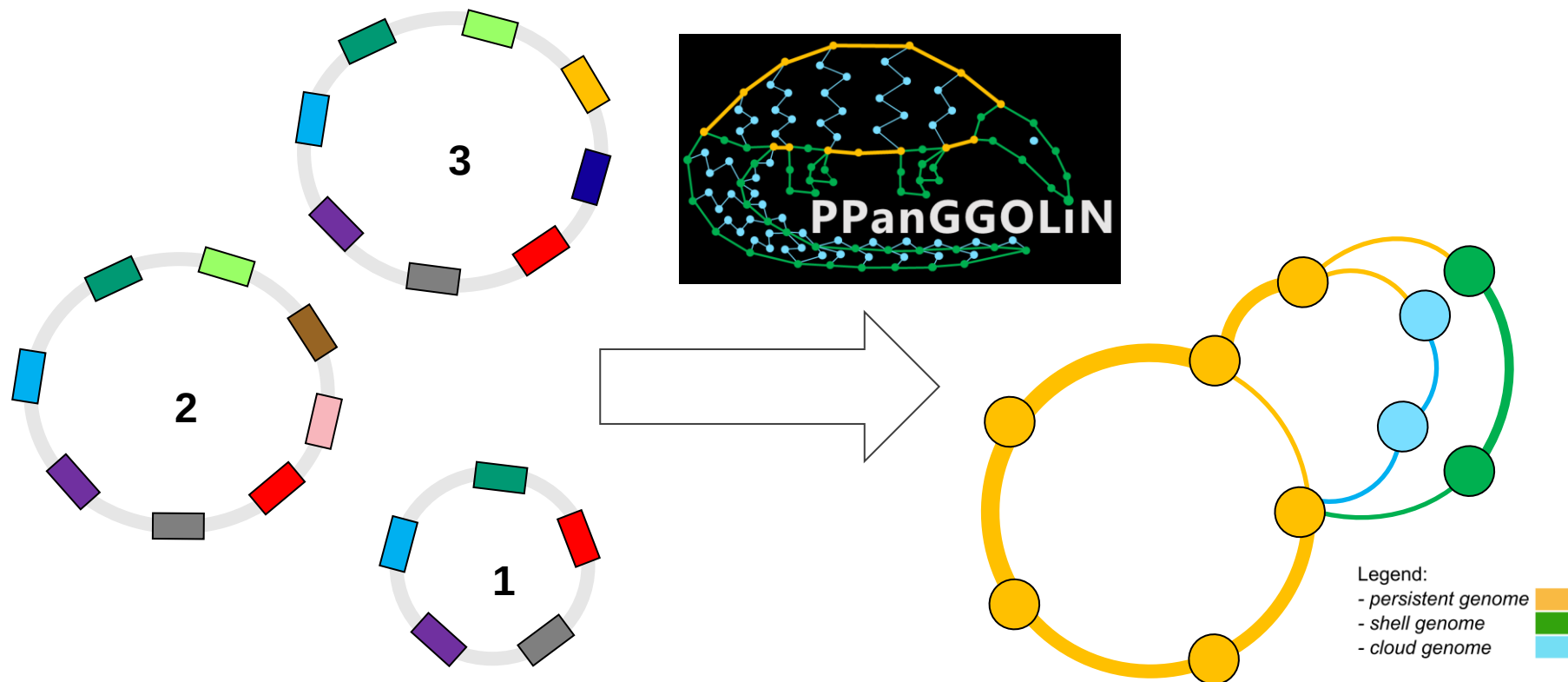
# Graphe de pangénome



# PPanGGOLiN : partitionnement d'un graphe de pangénome

Une méthode pour représenter les pangénomes en graphe et classer les familles suivant :

- la matrice de présence/absence des familles
- l'information de voisinage des familles dans le graphe

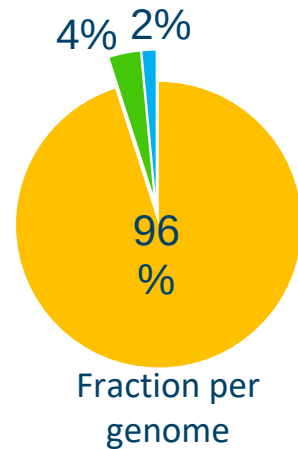


## 160 génomes de *Chlamydia trachomatis*

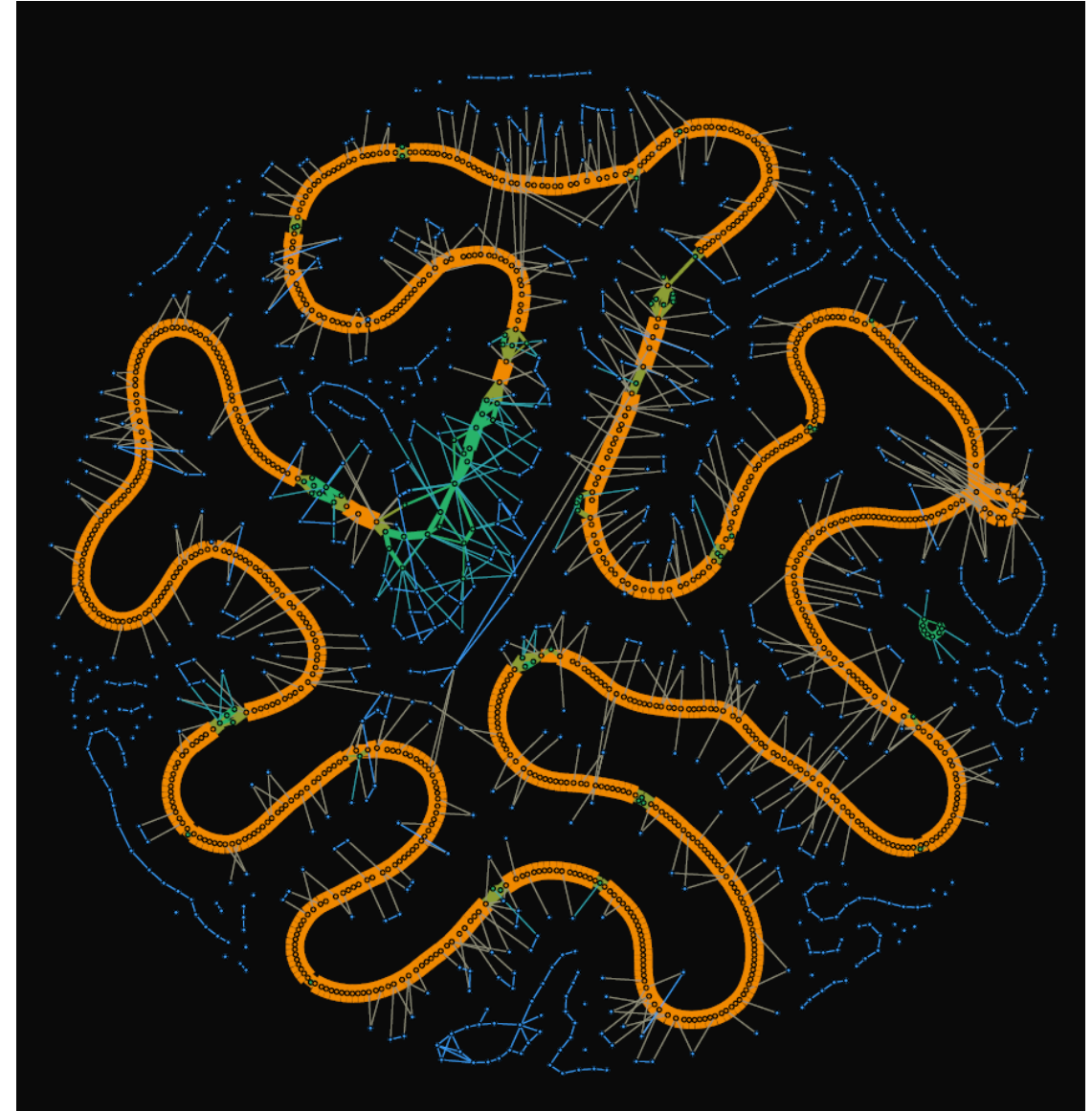
~890 gènes par génome  
core exact : 564 familles

### Pangénome :

- **persistent** : 861 familles
- **shell** : 84 familles
- **cloud** : 889 familles



Shell très petit avec une seule grande région variable. Cloud composé principalement de plasmides.

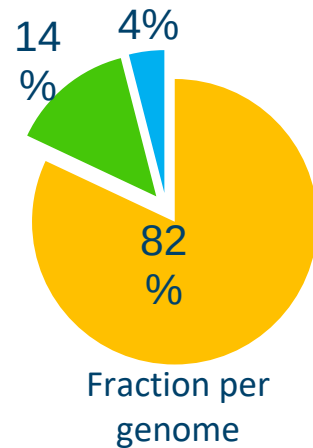


## 3117 génomes de *Acinetobacter baumannii*

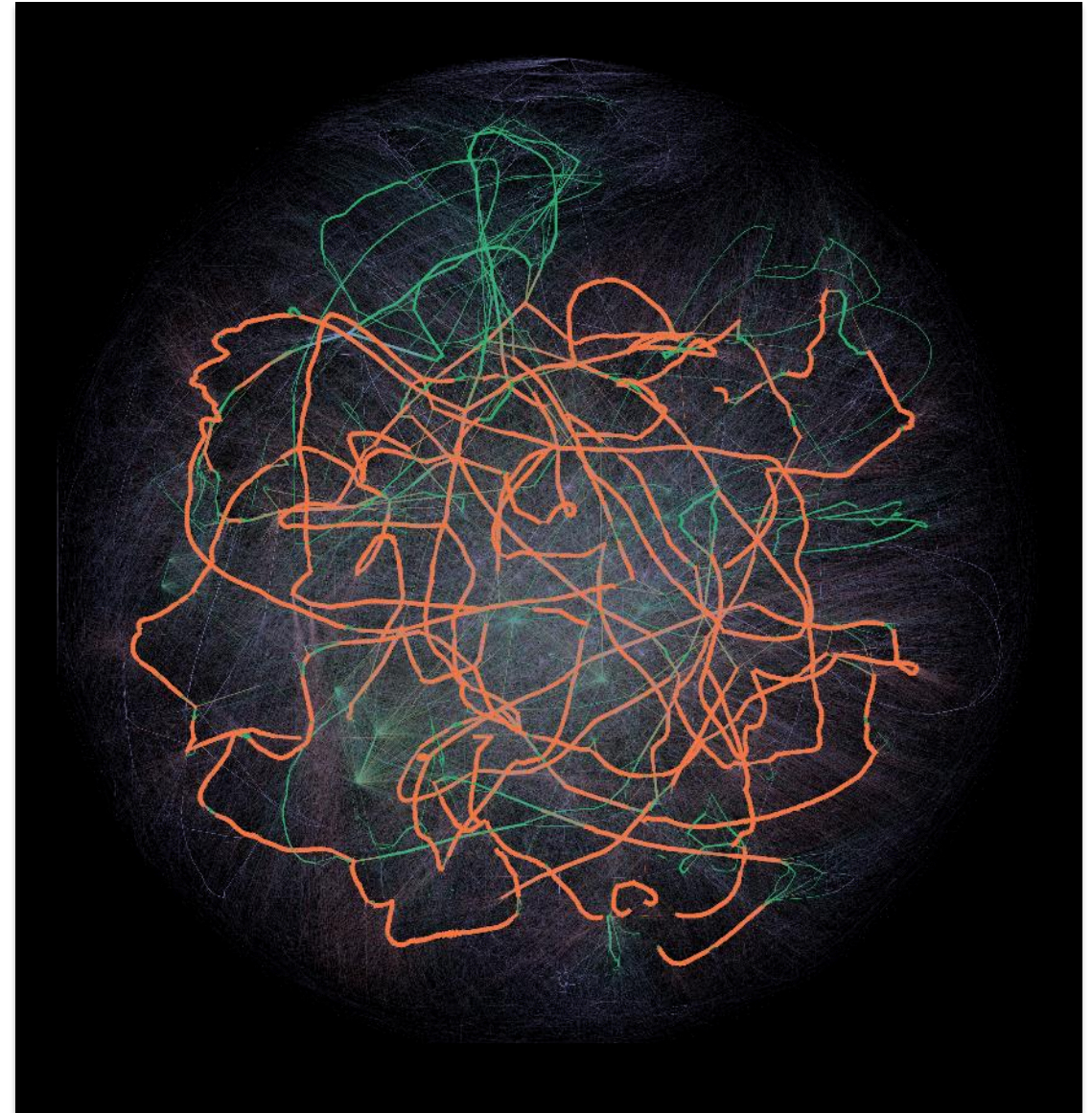
~3700 gènes par génome  
core exact : 0 familles

### Pangénome :

- **persistent** : 3080 familles
- **shell** : 1534 familles
- **cloud** : 64871 familles



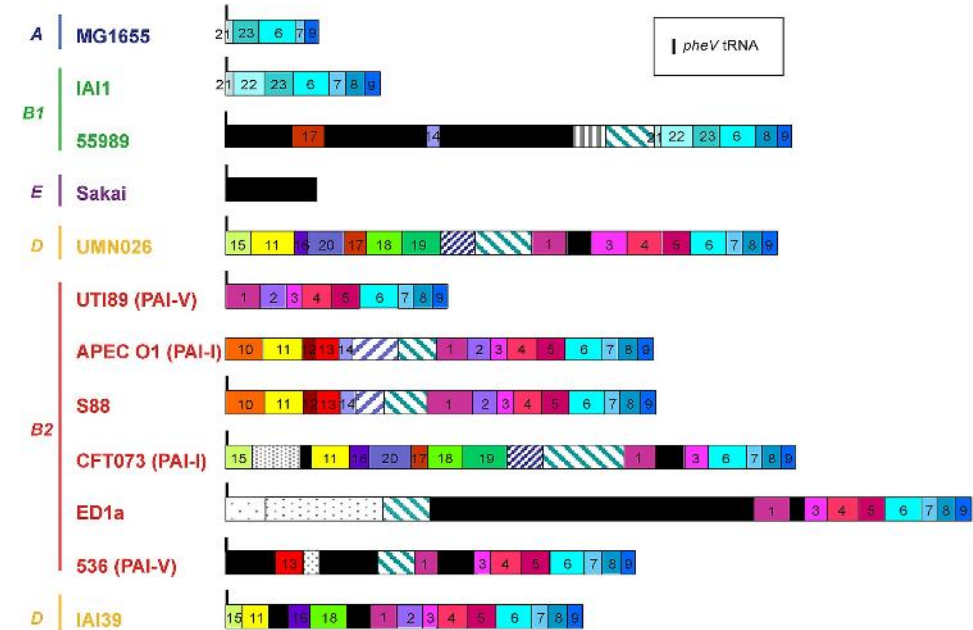
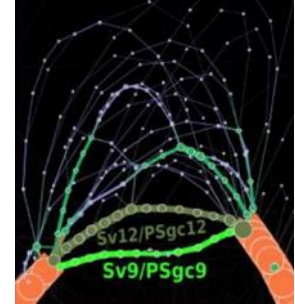
Shell important (îlots cataboliques,  
sidérophores, gènes d'antibiorésistance ...)



# Exploration des graphes de pangénomes

## Etude de la variabilité intra- et inter-spécifique

- méthode panRGP : détection des régions de plasticité génomique (îlots génomiques) et des points chauds d'insertion
- méthode panModule : détection de familles co-localisées et co-occurentes  
=> modules génomiques échangés par transferts horizontaux
- méthode panorama : comparaison de pangénomes  
=> flux de gènes entre espèces

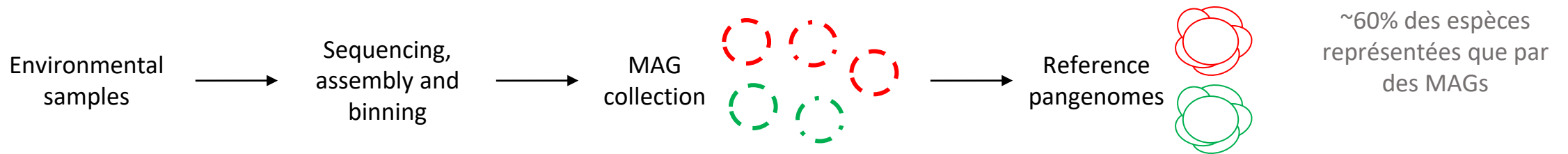


Îlot génomique modularisé chez *Escherichia coli*

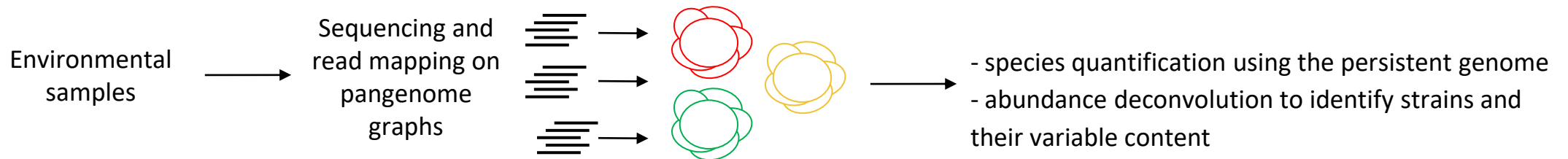
# Les pangénomes à l'échelle des écosystèmes

## Approches méta-pangénomiques pour accéder à la diversité des microbiotes

- reconstruction de pangénomes d'espèces non encore isolées/cultivées à partir de Metagenome-Assembled Genomes (MAGs)



- analyses métagénomiques en utilisant les graphes de pangénome comme référence



⇒ comprendre la dynamique d'évolution des espèces en interaction dans un écosystème et leur adaptation à différents environnements

# Pour conclure

---

## Début de maturité dans les méthodes utilisées en pangénomique mais reste à ...

- combiner l'information de présence/absence de gènes et de variants nucléotidiques
- gérer les biais de représentativité des génomes
- développer des méthodes de comparaison de pangénomes
- adapter les méthodes d'annotation fonctionnelle
- visualiser et naviguer dans les graphes ....

## Applications de la génomique microbienne pour l'agriculture



- microbiote des animaux d'élevage et des plantes
- santé : antibiorésistance (one health)
- agroécologie : monitoring du sol
- biocontrôle / biotech

# Remerciements



Alexandra Calteau

Guillaume Gautreau (PhD 2016-2020)

Adelme Bazin (PhD 2018-2022)

Jérôme Arnoux (PhD 2022-2025)



Eduardo Rocha



Catherine Matias

Christophe Ambroise

