



Unité des Bactéries Lactiques
et pathogènes Opportunistes

MOSAIC

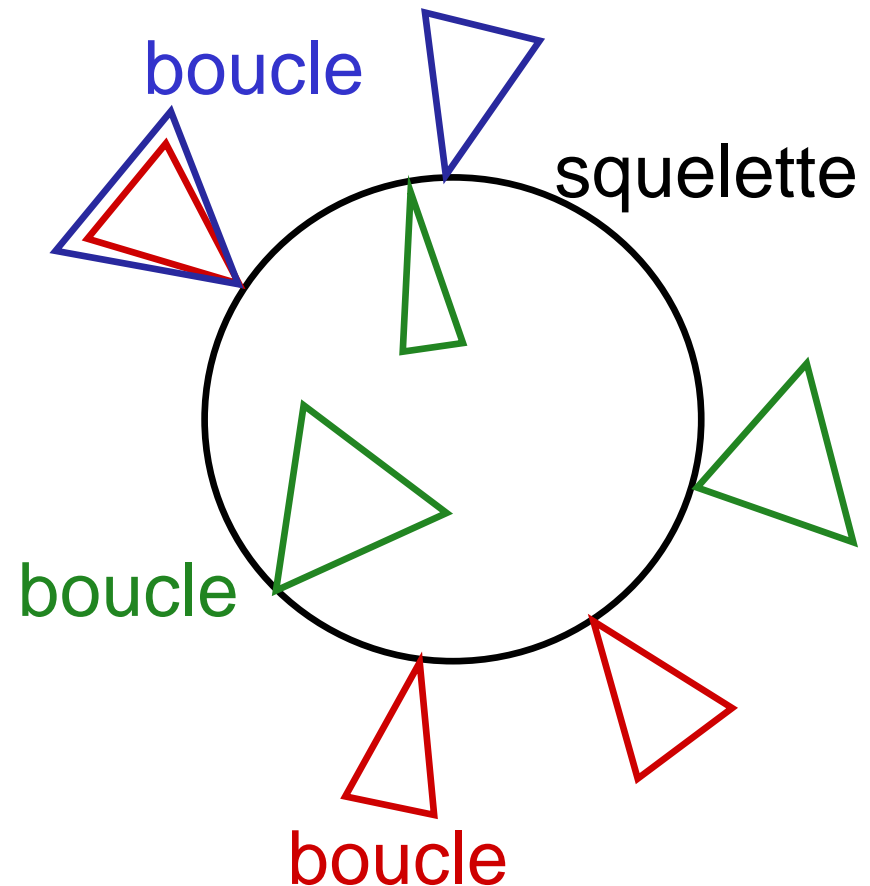
**Détermination de la structure squelette/boucles des
génomés bactériens et application à la prédiction de motifs
stimulant la réparation de l'ADN**

meriem.el_karoui@jouy.inra.fr

Hypothèse de travail

Les génomes ont une structure mosaïque visible à l'échelle micro-évolutive

- une partie conservée (le squelette) qui correspond à l'ancêtre commun
- des parties variables (les boucles)



Axes de recherche

Les génomes ont une structure mosaïque visible à l'échelle micro-évolutive

- une partie conservée (le squelette) qui correspond à l'ancêtre commun
- des parties variables (les boucles)

AXE A

- **Stratégie de segmentation squelette/boucles**
- **Base de données MOSAIC**

AXE B

- **Analyse des propriétés du squelette**
- **Typologie et mécanismes d'acquisition des boucles**

Axe A : stratégie de segmentation et base de donnée MOSAIC

Etape 1 :

Choix d'un outil de comparaison de génomes

- **Alignement multiple de génomes complets**
- **Présence ou non de réarrangements**

Etape 2 :

Segmentation squelette/boucles

- **Jeu de paramètres validé sur alignement de référence**
- **Post-traitement de l'alignement**

Etape 3 :

Intégration dans la base de donnée MOSAIC

- **Choix des génomes à comparer**
- **Automatisation du processus**

MGA: Multiple Genome Aligner

Höhl, Kurtz and Ohlebusch, *Bioinformatics* 2002
<http://bibiserv.techfak.uni-bielefeld.de/mga>

- **Alignement global multiple de n ($n \geq 2$) génomes proches**
 - **Principe de l'ancrage** par détection de MEMs (Maximal Exact Matches) (**paramètre l_1**)
 - **Fermeture des gaps entre les MEMs**
 - ✓ Récursivité sur la taille des MEMs (**paramètre l_2**)
 - ✓ Alignement clustalw des gaps de taille inférieure à un seuil (**paramètre gl**).
- **Traite les génomes colinéaires sans réarrangements majeurs**

Procédure de définition des coordonnées squelette/boucles

Jeu de données de référence : boucles annotées manuellement par Hayashi (comparaison *E. coli* K12/Sak)
[*E.coli* O157:H7 Sakai Genome Information, <http://genome.gen-info.osaka-u.ac.jp>]



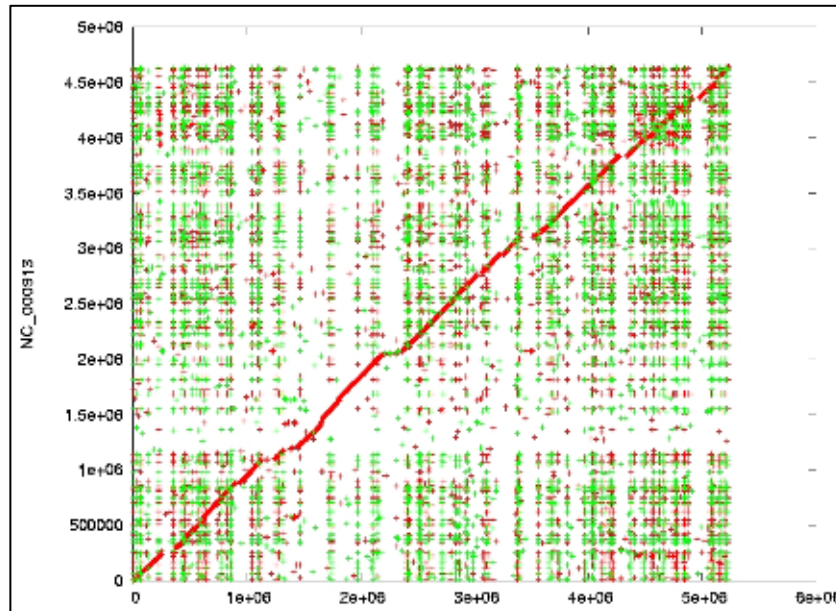
Test de différents jeux de paramètres MGA (l_1 , l_2 , gl)
Introduction d'un paramètre aditionnel (P_{id}) pour traiter les sorties clustalw.



Choix du jeu de paramètres MGA le plus proche des données validées par Hayashi :
 $l=50-20$; $gl=3000$; P_{id} AlignedGaps=76%

Mosaic : segmentation des génomes non réarrangés

MGA L=50-20, gl=3000, Pid=76%



*E. coli K12 x O157:H7
Sakai*

Squelette : 4,1 Mbase

- 98,3 % d'identité
- 81,5 % de couverture

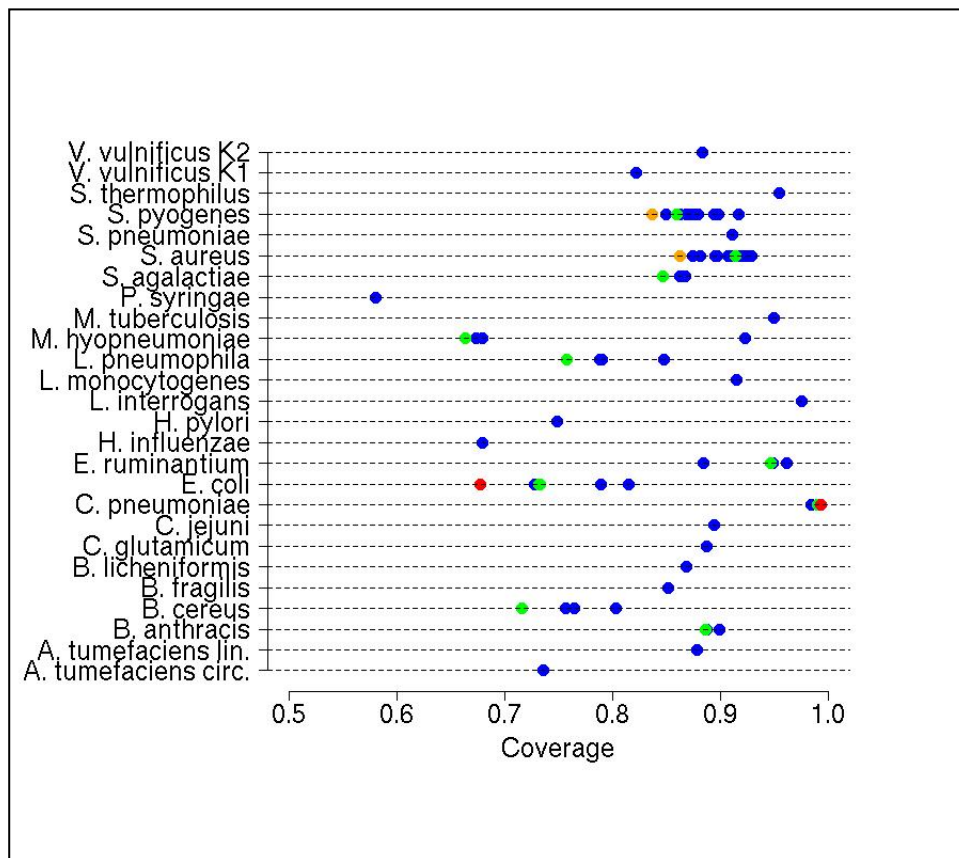
Boucles

- 566 K-boucles (536 kb)
- 520 S-boucles (1395 kb)

MOSAIC version 3 (Octobre 2005)

<http://genome.jouy.inra.fr/mosaic>

- 86 comparaisons (2 six-genome, 2 quadruples ; 11 triples ; 71 doubles)
- 195 genomes segmentés, représentant 24 espèces bactériennes



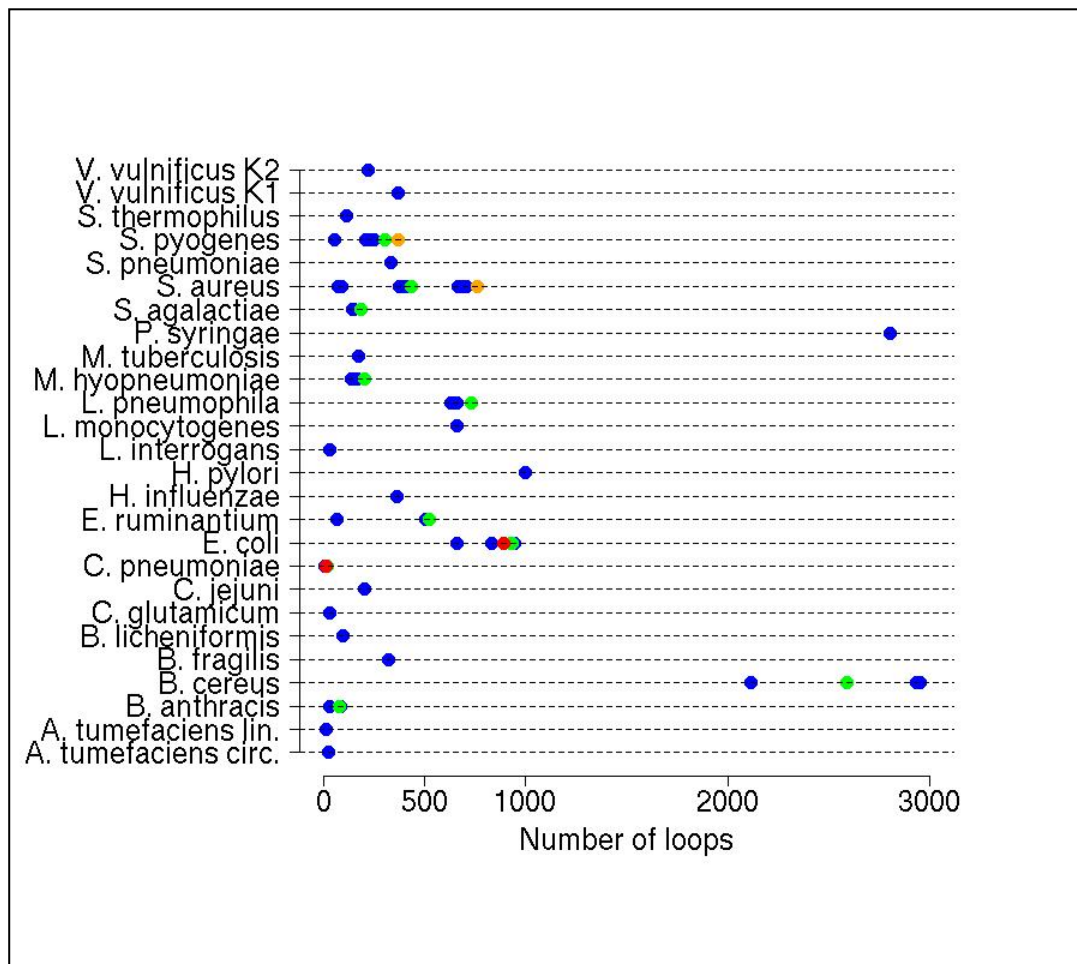
Couverture moyenne par espèce

- **min : 58 %**
(*P. syringae* pv. *syringae* B728 X pv. *phaseolicola* 1448A)
- **médiane : 88 %**
- **max : 99 %**
(*C. pneumoniae* AR39 x CWL29 x J138 x TW183)

Chiapello et al. *BMC Bioinformatics*, 2005

MOSAIC version 3

Nombre moyen de boucles par espèce



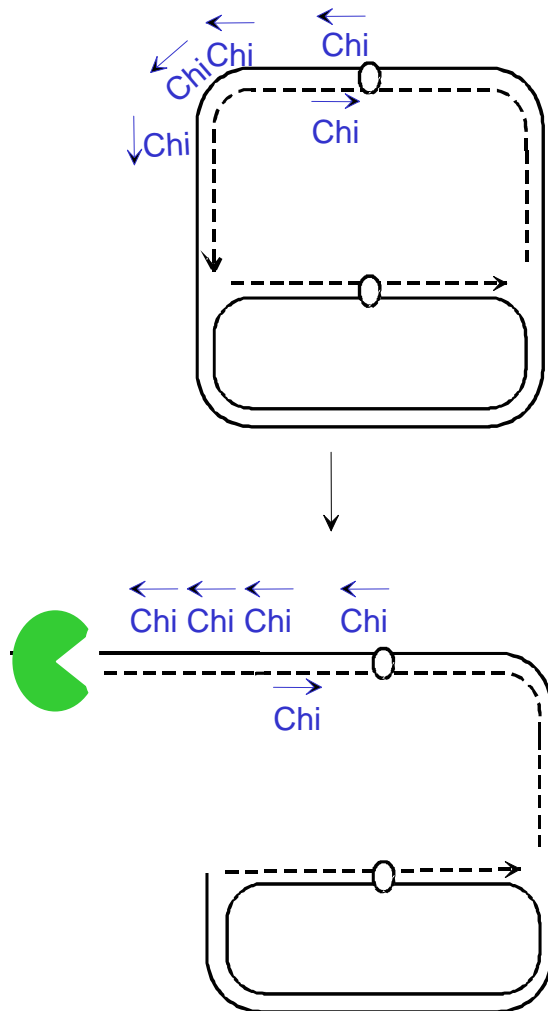
- **min : 6** (*C. pneumoniae* CWL29 x TW183)

- **médiane : 248**

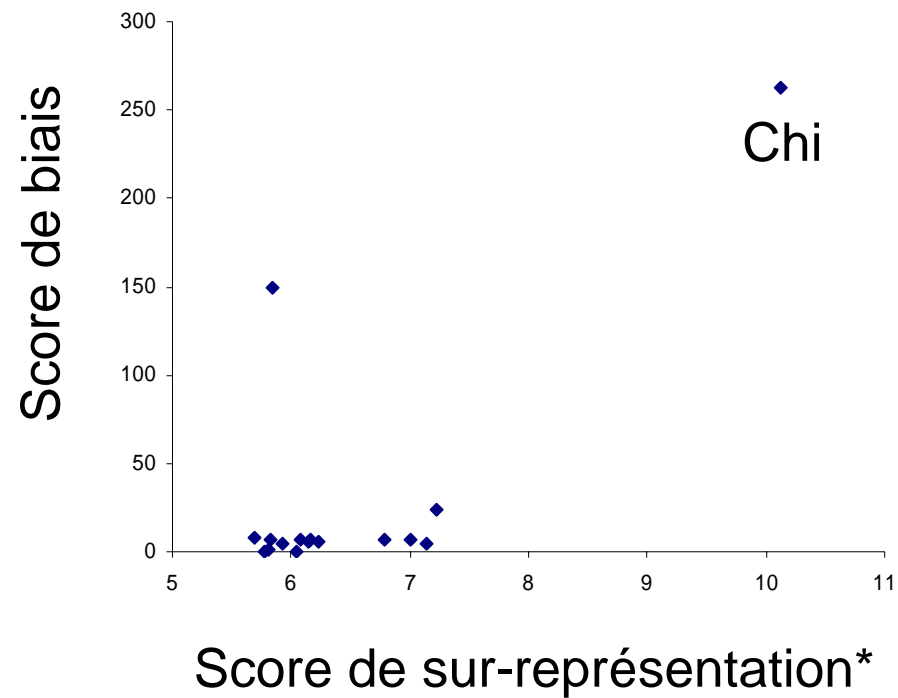
- **max : 2951** (*B. cereus* ATCC14579 x ARCC10987)

Axe B : propriétés du squelette

Chi : réparation des cassures double-brin de l'ADN

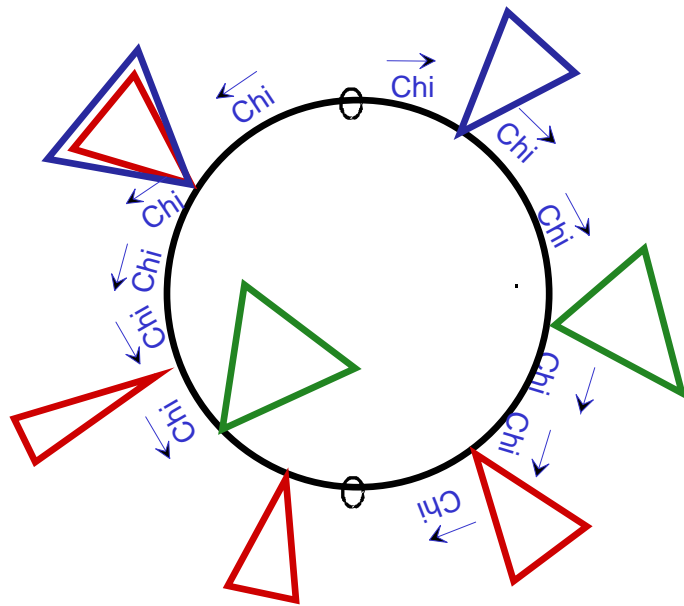


Squelette (comparaison triple) : 3,7 Mb; couverture 80%

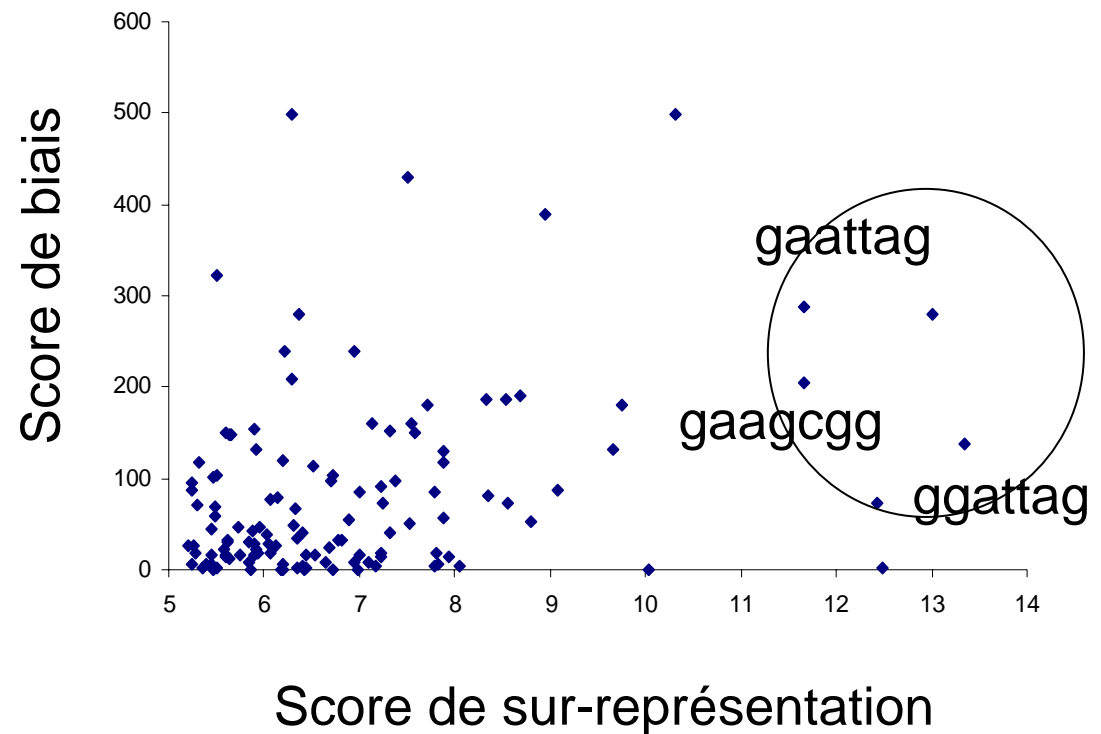


* R'mes, Schbath et al. 2002

Prédiction chez *Staphylococcus aureus*



Squelette (comparaison sextuple) :
2,47Mb; couverture 87%



MOSAIC : la suite....

- **Alignements des génomes**
 - Traitement des génomes réarrangés (MAUVE)
 - Evaluation de la qualité de l'alignement
- **Stratégie de comparaison**
 - Choix des génomes à comparer sur la base d'un critère de distance
 - Segmentation inter-espèces
- **Prédiction d'autres motifs structurant le chromosome**
- **Typologie détaillée des boucles chez *E. coli* et *S. aureus*, afin de comprendre les mécanismes d'acquisition**



Centre de recherche
Jouy en Josas

M.-A. Petit

D. Halpern

A. Gruss

Unité des Bactéries

Lactiques et pathogènes

Opportunistes

ACI IMPBio (2003)

H. Chiapello

C. Caron

S. Schbath

A. Gendrault-Jacquemard

Mathématique

Informatique et

Génome

S. Robin

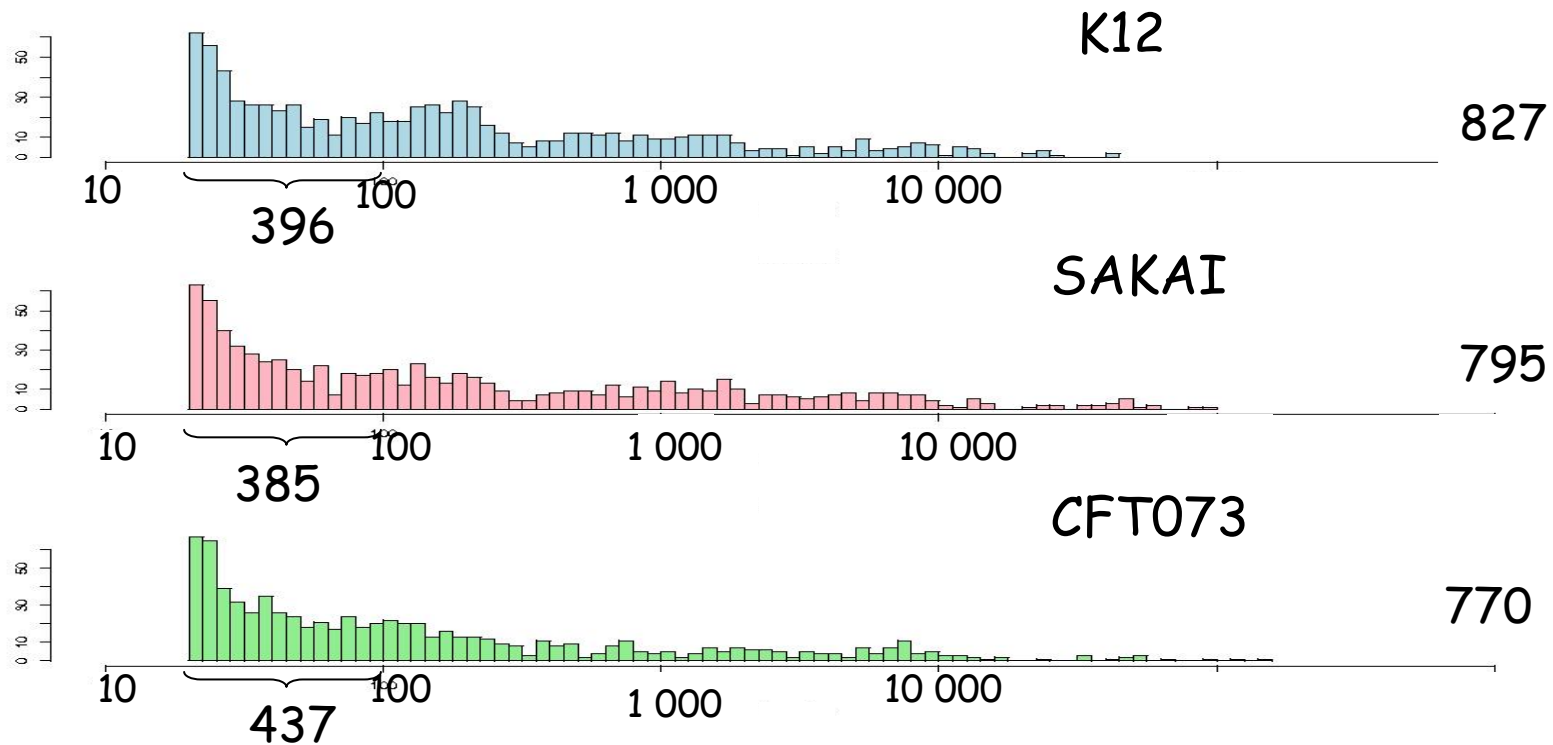


**Département Organisation et Modélisation
de l'Information et des Processus**

BOUCLES D'*Escherichia coli*

Taille du squelette : 3.73Mb (76%)

abscisse : taille de boucles
ordonnées : nombre de boucles



Axe B : Typologie et mécanismes d'acquisition des boucles

Croiser avec des annotations existantes
(banques publiques, collaborations)

Typologie : degré de redondance, composition, taille

Mécanismes : jonction squelette boucles