

# Rôle des transferts horizontaux dans l'acquisition de la virulence chez le bacille de la tuberculose

Jennifer Becq<sup>1</sup>, C. Gutierrez<sup>3</sup>, V. Rosas-Magallanes<sup>2</sup>, J. Rauzier<sup>2</sup>, B. Gicquel<sup>2</sup>, O. Neyrolles<sup>3,4</sup> & P. Deschavanne<sup>1</sup>

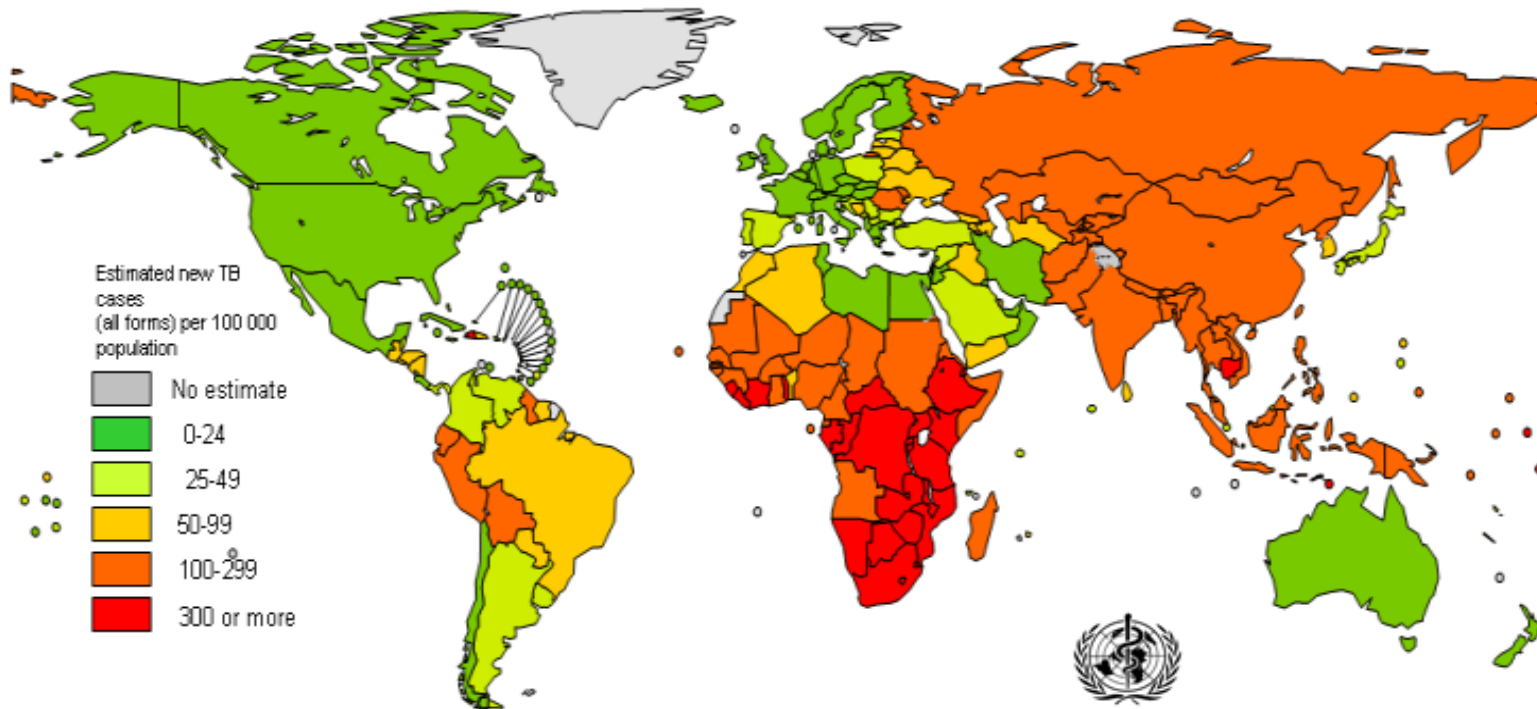
<sup>1</sup> Equipe de Bioinformatique Génomique et Moléculaire, Inserm U726, Paris 7

<sup>2</sup> Département Infection et Epidémiologie, Institut Pasteur, Paris

<sup>3</sup> Unité de Génétique Mycobactérienne, Institut Pasteur, Paris

<sup>4</sup> CNRS URA 2172, Institut Pasteur, Paris

# La tuberculose



Nombre de nouveaux cas de tuberculose par pays en 2005  
(source : OMS)

Maladie infectieuse  
(bacille de Koch)

1/3 de la population  
mondiale est infectée

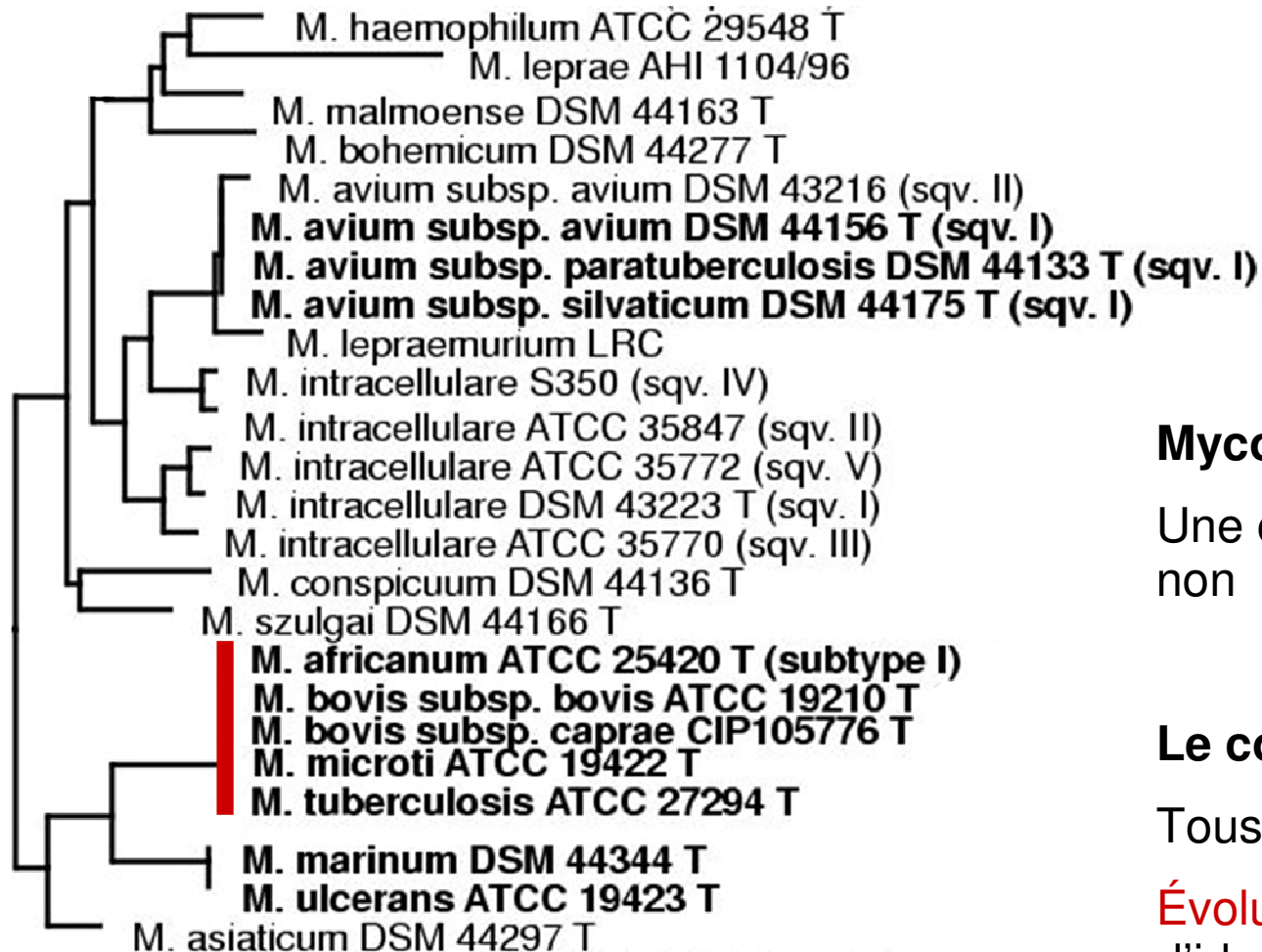
8,6 millions de  
nouveaux cas par an

1,5 millions de décès  
par an

dont 250 000  
associés au VIH

Émergence de  
souches multi-  
résistantes

# Le complexe *Mycobacterium tuberculosis*



## Mycobactéries

Une centaine d'espèces, pathogènes ou non

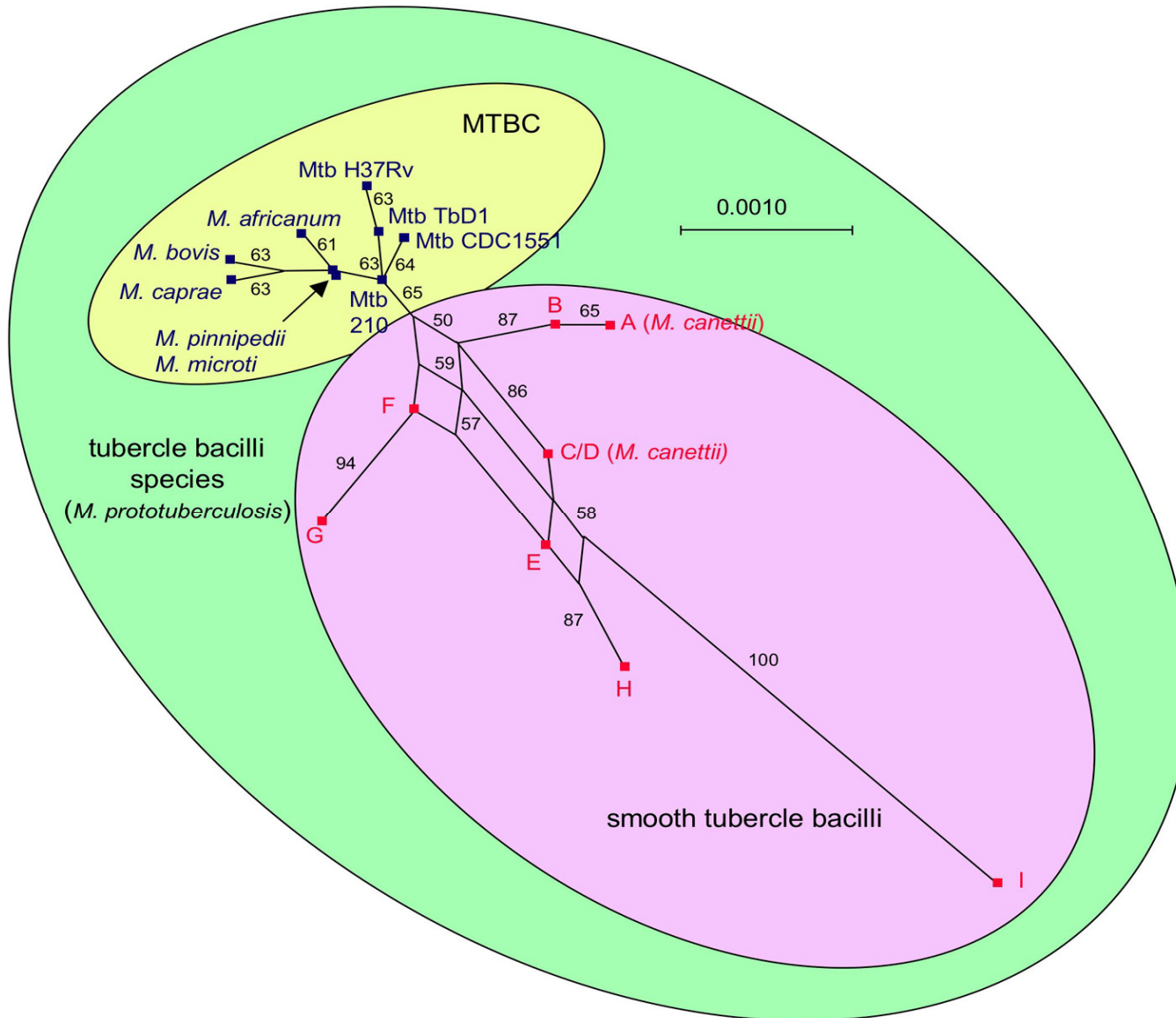
## Le complexe *M. tuberculosis*

Tous pathogènes de l'Homme

**Évolution clonale** « récente » (99% d'identité au niveau nucléotidique)

Diversité d'hôtes au sein des mammifères

# Le complexe *Mycobacterium tuberculosis*



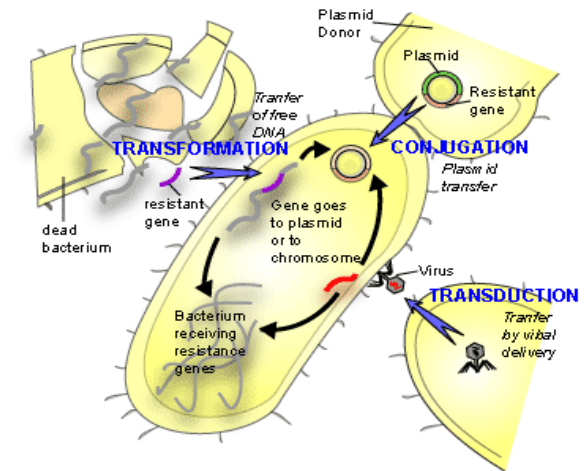
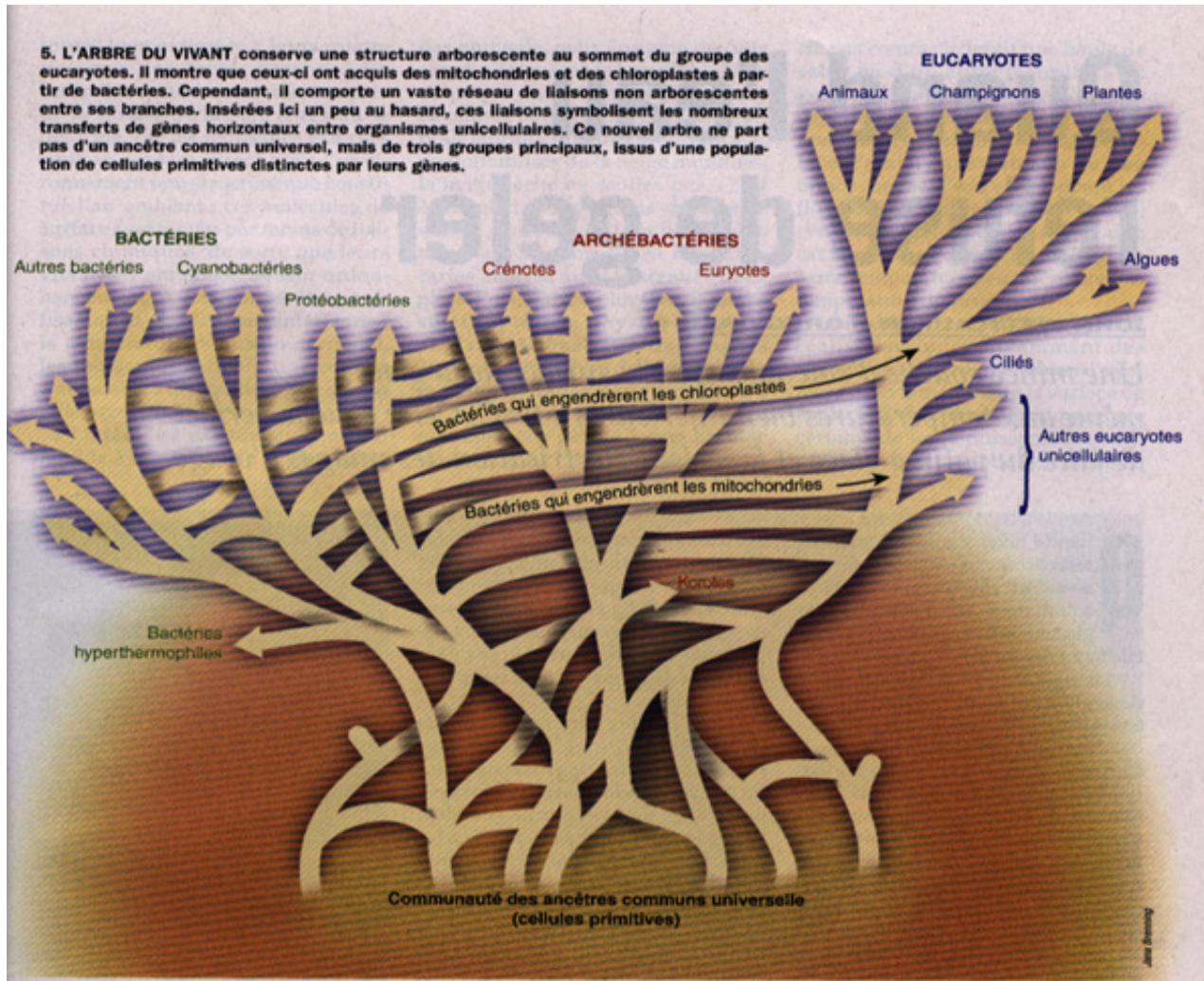
## Génomes séquencés

*M. tuberculosis* CDC1551

*M. tuberculosis* H37Rv

*M. bovis* AF2122/97

# Les transferts horizontaux



Facteurs d'évolution des génomes

« îlots génomiques » « îlots de pathogénicité »

Transfert de virulence, de résistance aux antibiotiques...

# Chez *M. tuberculosis*

Cole *et al.* 1998 *Nature*

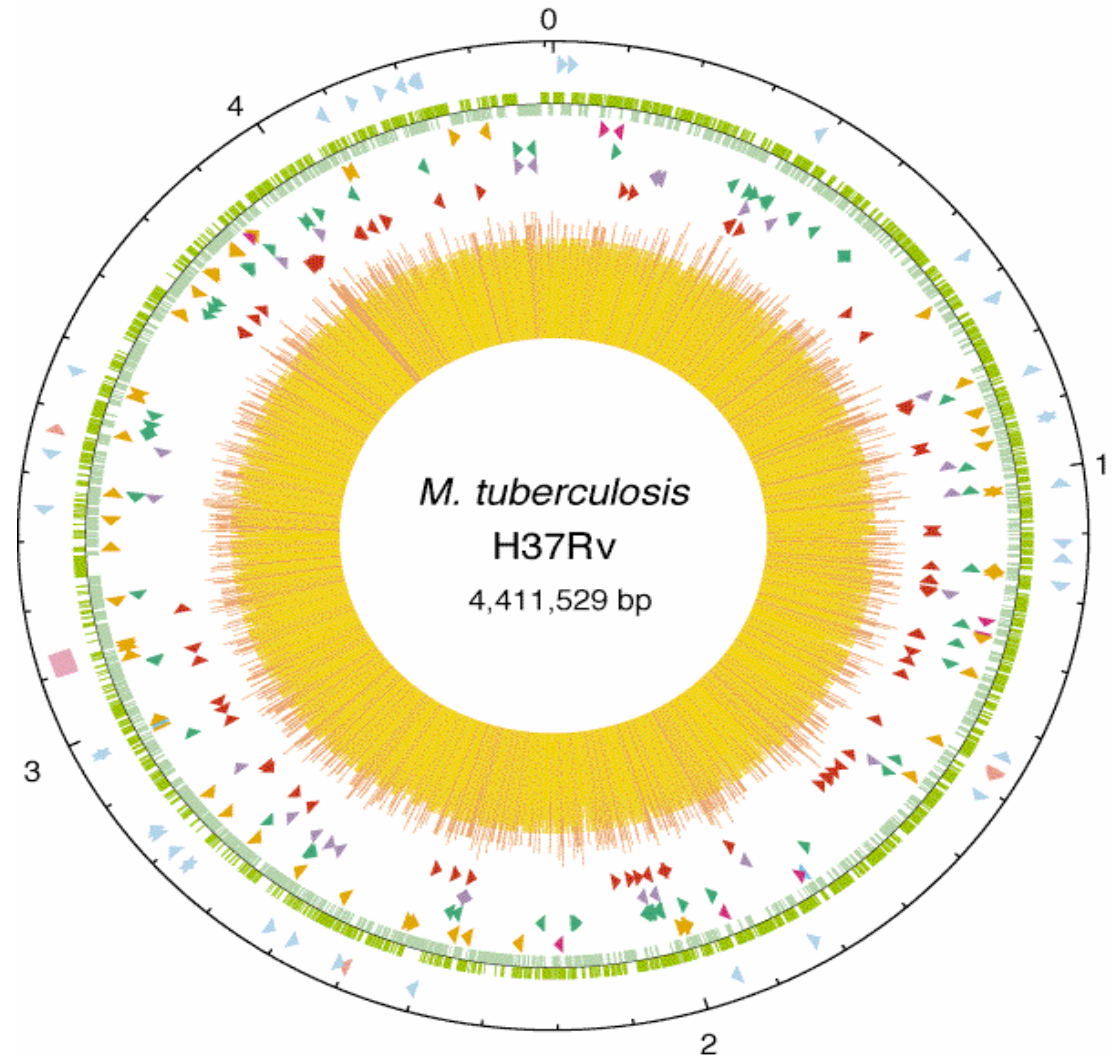
~ 4.4 Mb

~ 66% G+C

**Pas de transferts horizontaux décrits**

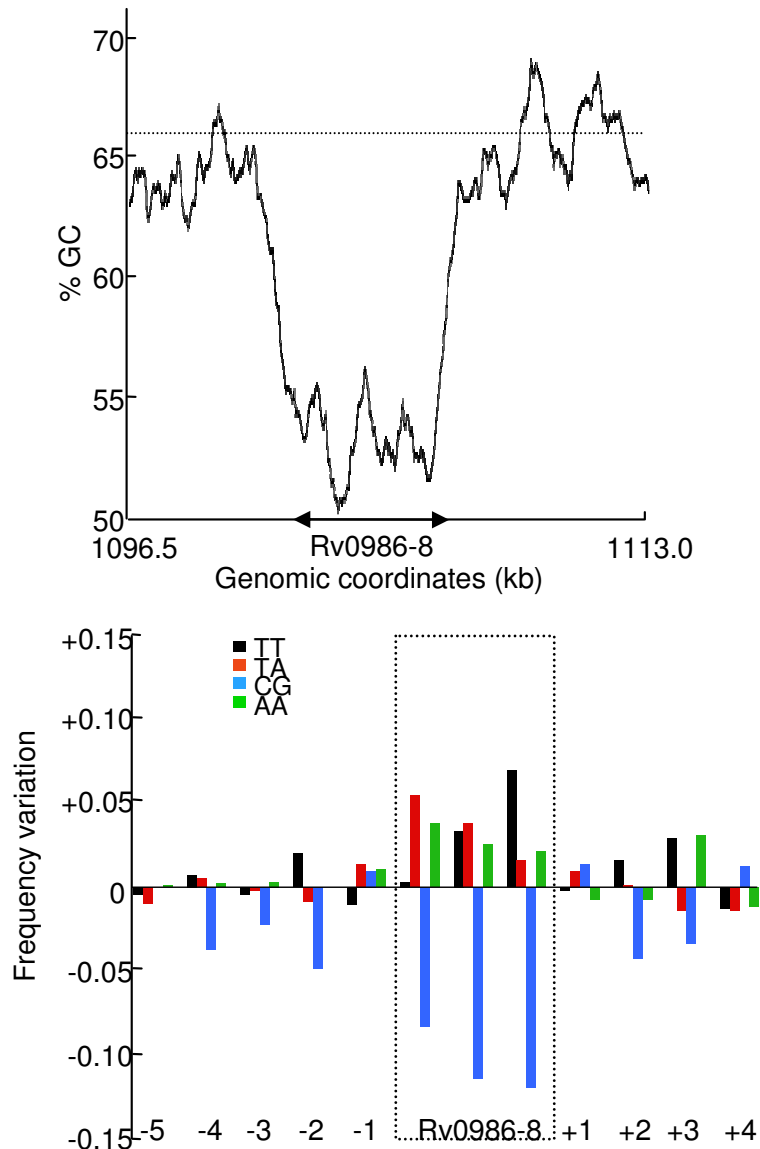
**Pas d'îlots de pathogénicité décrits**

Peu de gènes de virulence connus



# Rv0986-8 a été acquis par transfert horizontal

Des caractéristiques de séquence atypiques :

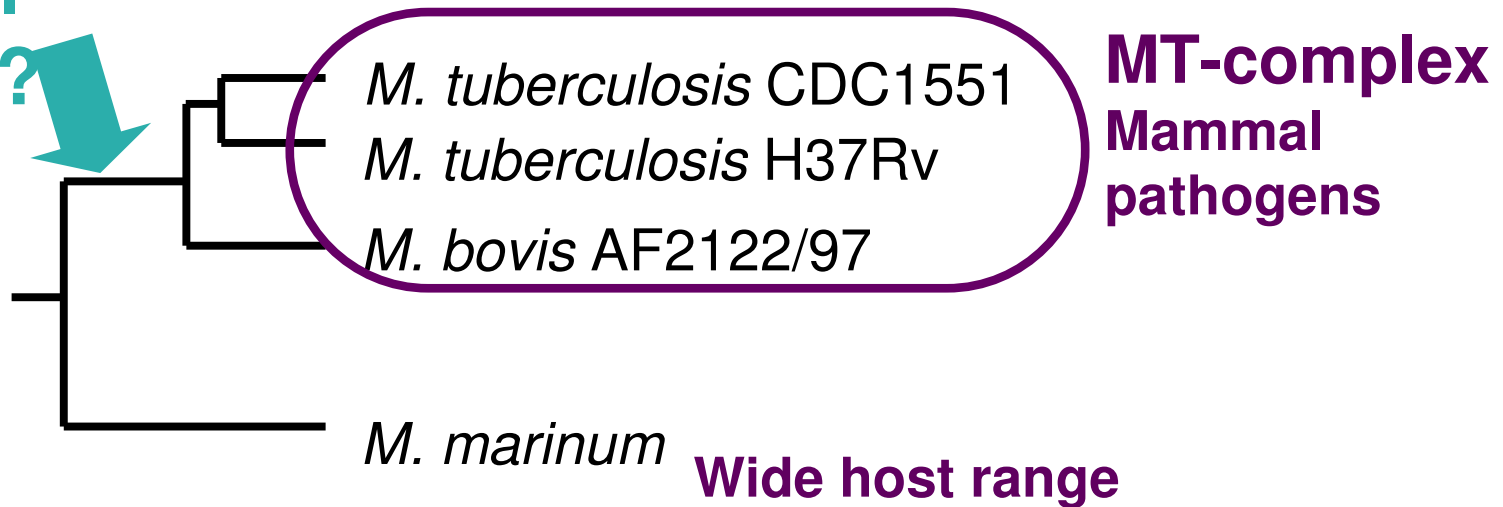


Un opéron de virulence spécifique au complexe *M. tuberculosis*

|              |                                            |             |        |        |              |         |
|--------------|--------------------------------------------|-------------|--------|--------|--------------|---------|
|              |                                            | Rv0986      | Rv0987 | Rv0988 |              |         |
|              |                                            | ←           | →      | →      | ←            | ←       |
|              |                                            | <i>mscL</i> |        |        | <i>grcC2</i> | Rv0990c |
|              |                                            | +           | +      | +      | +            | +       |
| Slow Growing | <i>M. tuberculosis</i><br>H37Rv<br>CDC1551 |             |        |        |              |         |
|              | TB complex                                 |             |        |        |              |         |
|              | <i>M. microti</i>                          | +           | +      | +      | +            | +       |
|              | <i>M. bovis</i> (incl BCG)                 | +           | +      | +      | +            | +       |
|              | <i>M. africanum</i>                        | +           | +      | +      | +            | +       |
| Fast Growing | <i>M. canetti</i>                          | +           | +      | +      | +            | +       |
|              | <i>M. leprae</i>                           | +           |        |        |              | +       |
|              | <i>M. avium</i>                            | +           |        |        |              | +       |
|              | <i>M. marinum</i>                          | +           |        |        |              | +       |
|              | <i>M. ulcerans</i>                         | +           |        |        |              | +       |
|              | <i>M. xenopi</i>                           | +           |        |        |              | +       |
|              | <i>M. smegmatis</i>                        | +           |        |        |              | +       |
|              | <i>M. fortuitum</i>                        | +           |        |        |              | +       |
|              | <i>M. vaccae</i>                           | +           |        |        |              | +       |
|              | <i>C. glutamicum</i>                       | +           |        |        |              | +       |

Dans quelle mesure les transferts horizontaux  
ont participé à l'évolution du complexe  
*Mycobacterium tuberculosis* ?

Horizontal  
transfers ?



# Recherche systématique de transferts horizontaux dans le complexe *Mycobacterium tuberculosis*

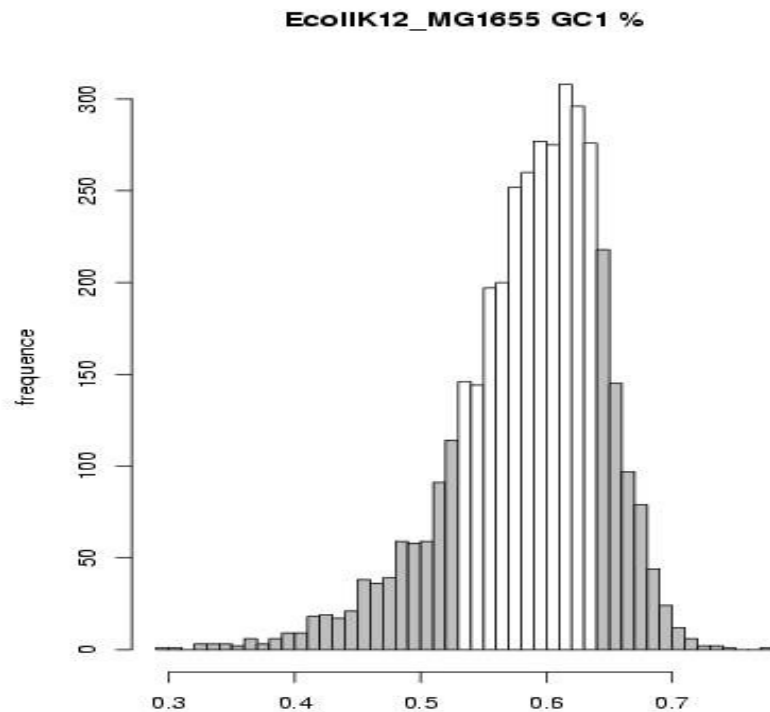
## Utilisation de méthodes paramétriques

« *Sequences that are new to a bacterial genome, i.e. those introduced through horizontal transfer, often bear unusual sequence characteristics and thus can be distinguished from ancestral DNA* »

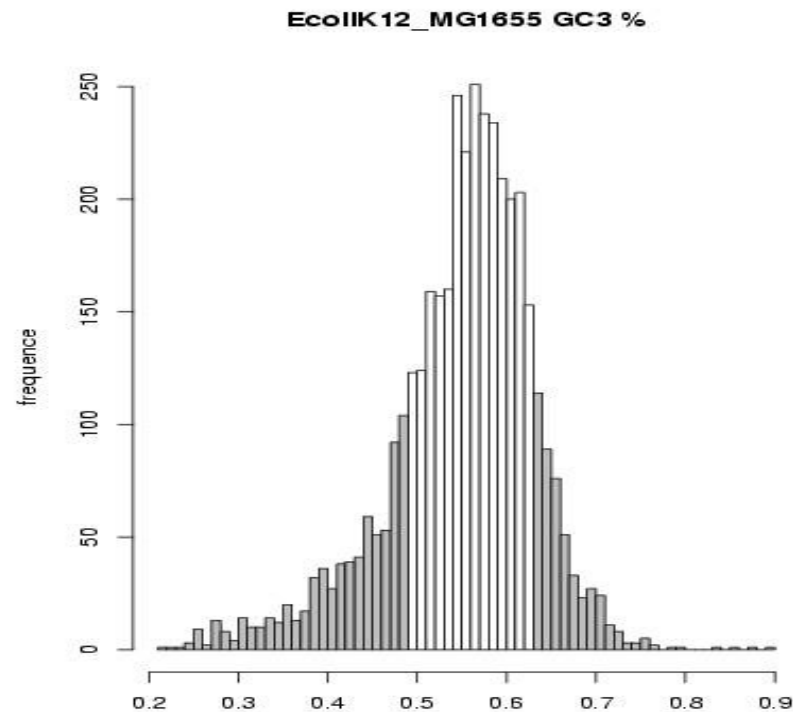
Lawrence & Ochman 1998 *PNAS*

# Le % GC en position 1 et 3 des codons

GC1



GC3



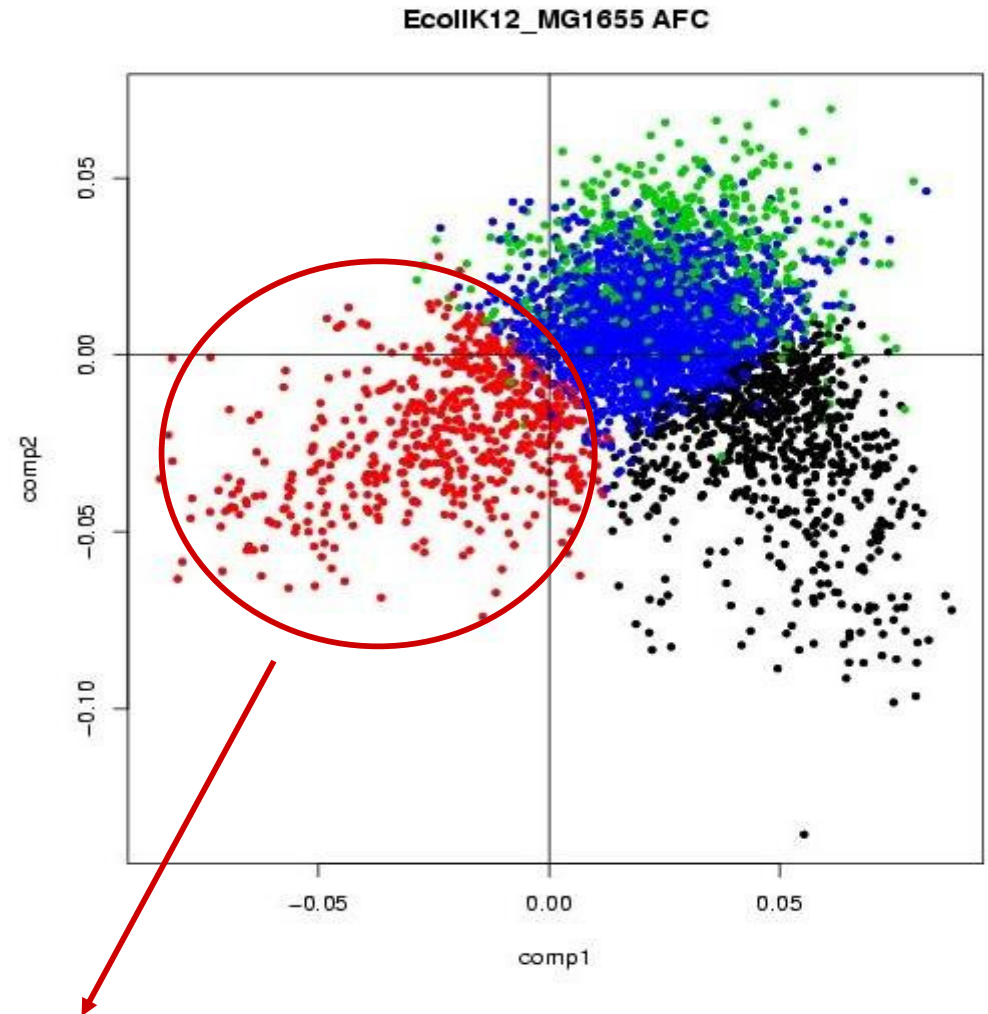
Seuil inférieur =  $\text{quartile}_{\text{inf}} - 0.25 \times \text{distance interquartile}$   
Seuil supérieur =  $\text{quartile}_{\text{sup}} + 0.25 \times \text{distance interquartile}$

**Gènes atypiques : GC1 % ET GC3% atypiques**

# L'usage des codons

|           | CCC   | ... | TTT   |
|-----------|-------|-----|-------|
| Gène 1    | 0.045 | ... | 0.134 |
| ...       | ...   | ... | ...   |
| Gène 4254 | 0.182 | ... | 0.267 |

K-means : 4 groupes

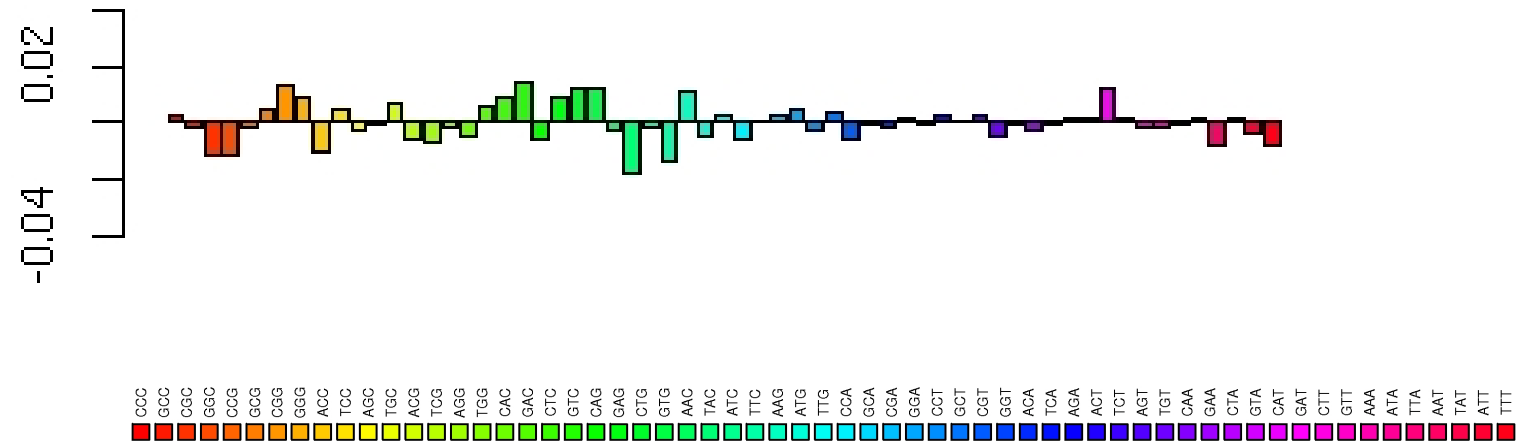


**Groupes de gènes à usage de codons atypique**

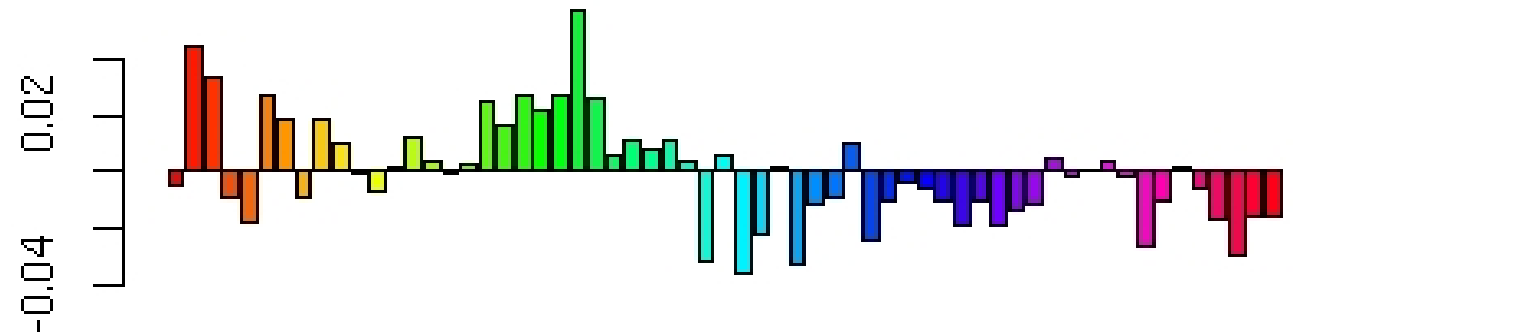
# L'usage des codons

## Déviations de la fréquence des codons

Gène Rv2045c (*lipT*)  
« typique »

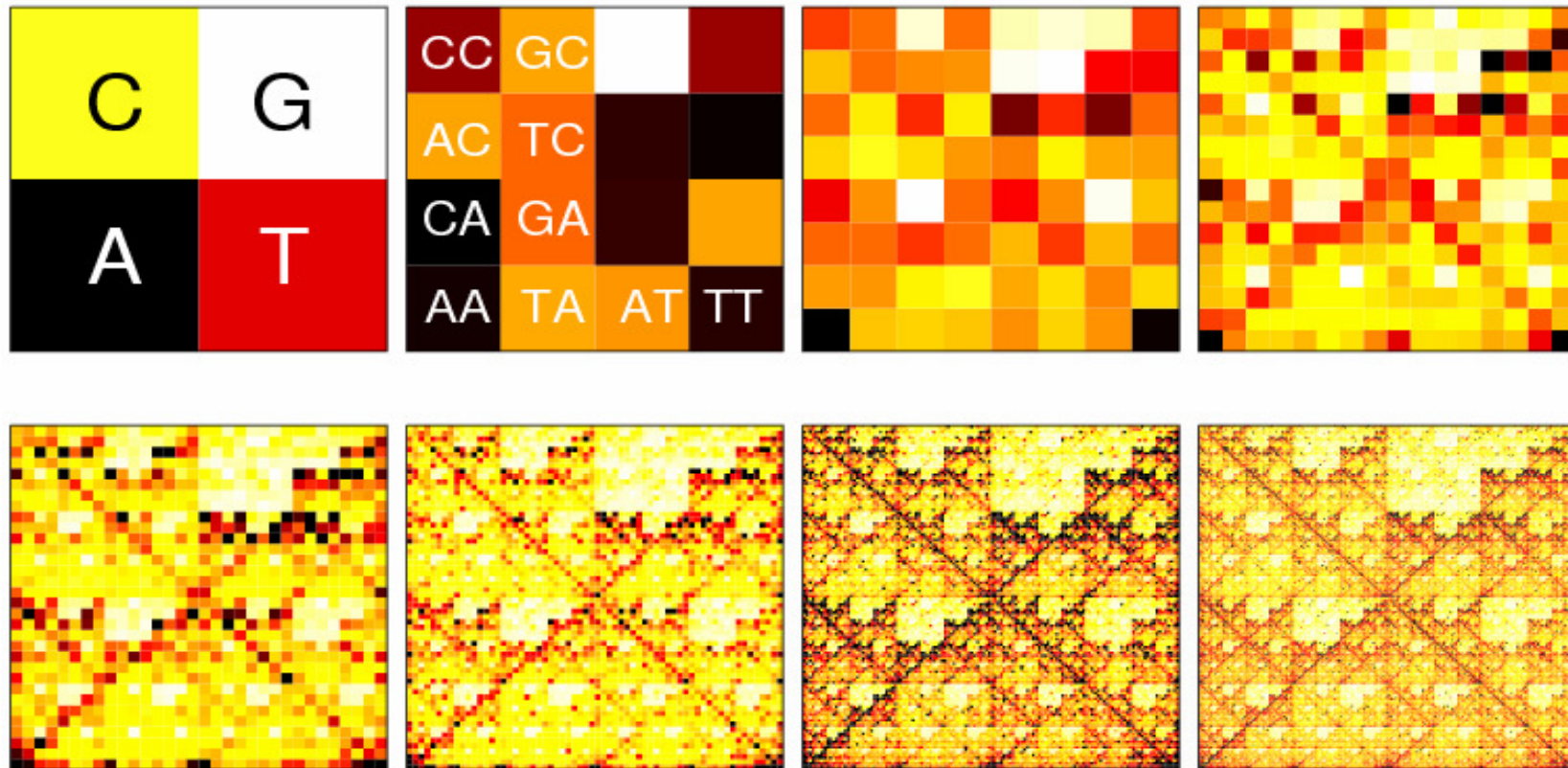


Gène Rv2307B  
« atypique »



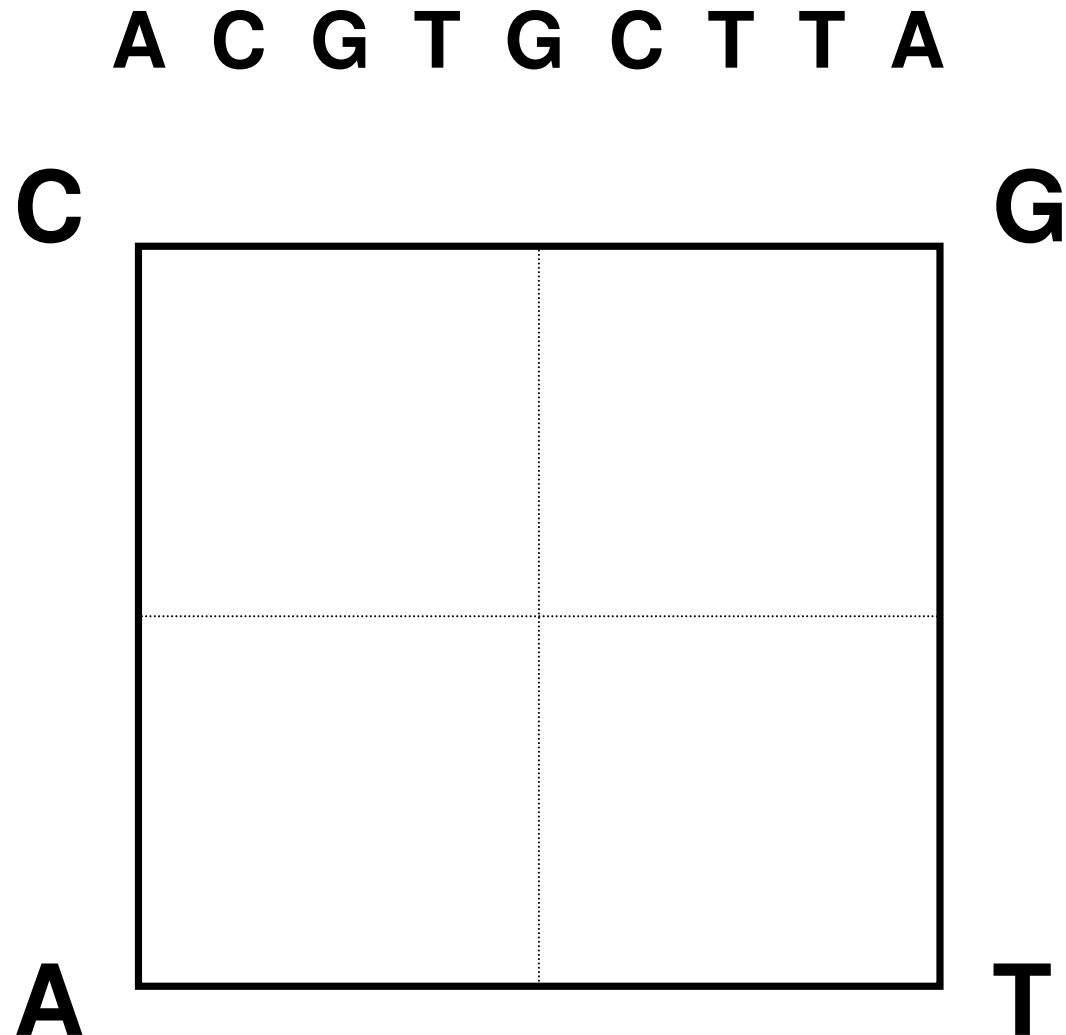
# La signature génomique

Ensemble des fréquences des mots d'une taille donnée dans une séquence



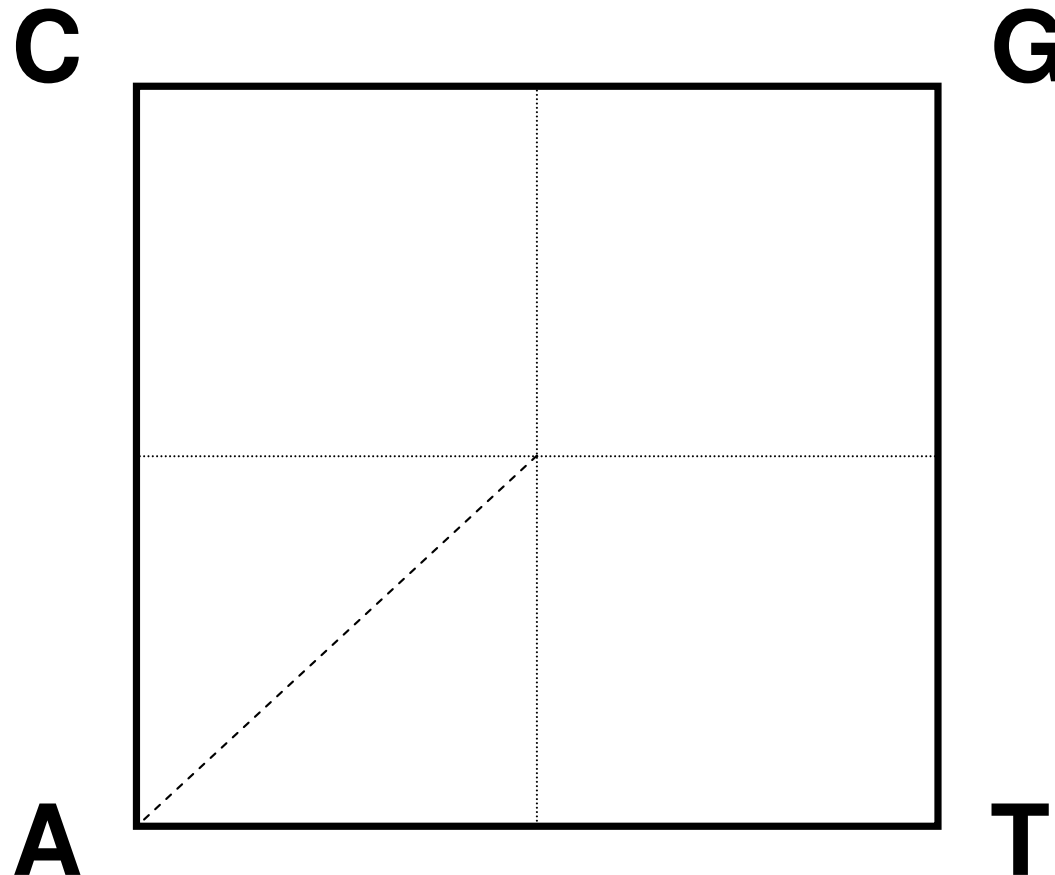
Signatures 1 à 8 du chromosome 22 d'*H. sapiens*

# La signature génomique - construction



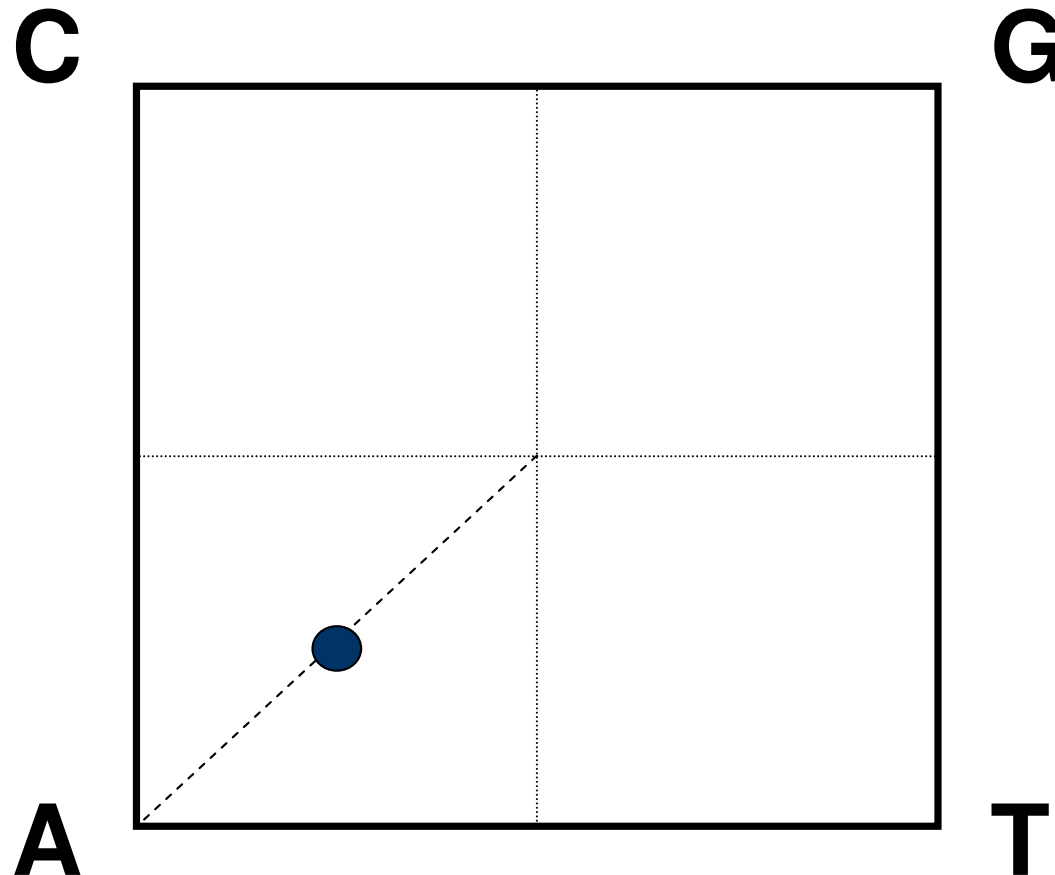
# La signature génomique - construction

**A** C G T G C T T A



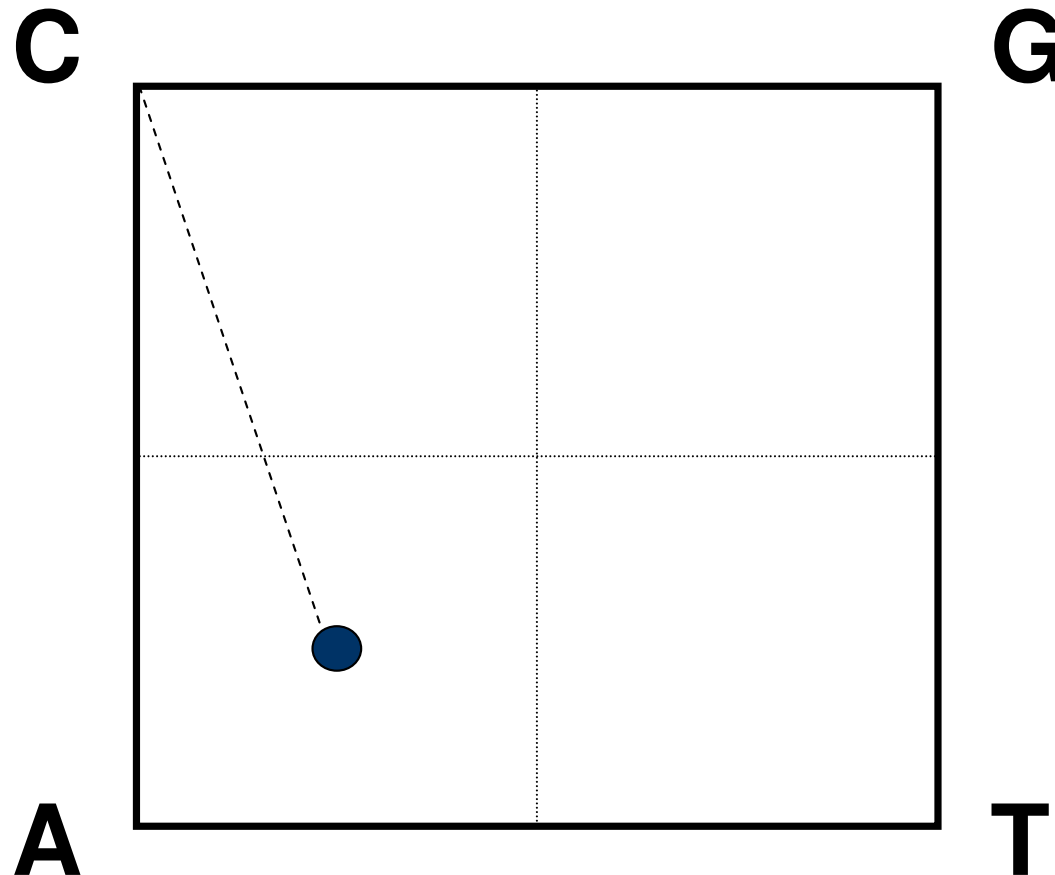
# La signature génomique - construction

**A** C G T G C T T A



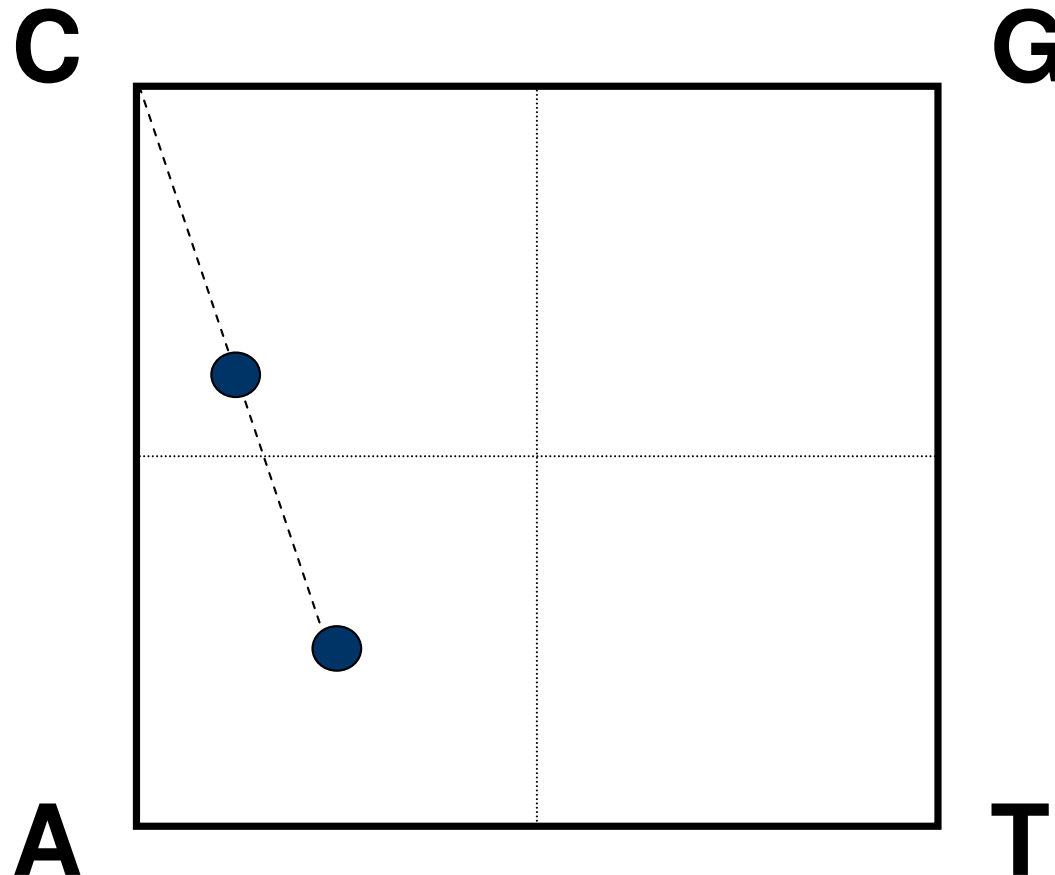
# La signature génomique - construction

A **C** G T G C T T A



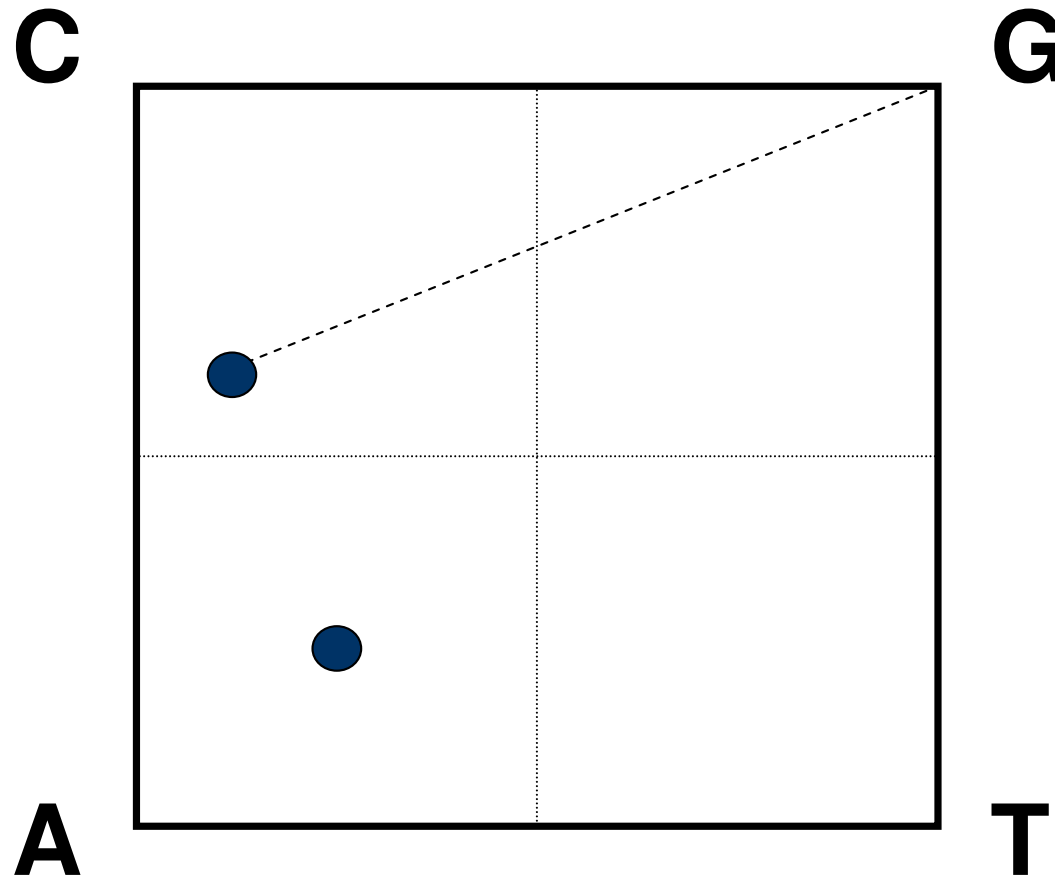
# La signature génomique - construction

A **C** G T G C T T A



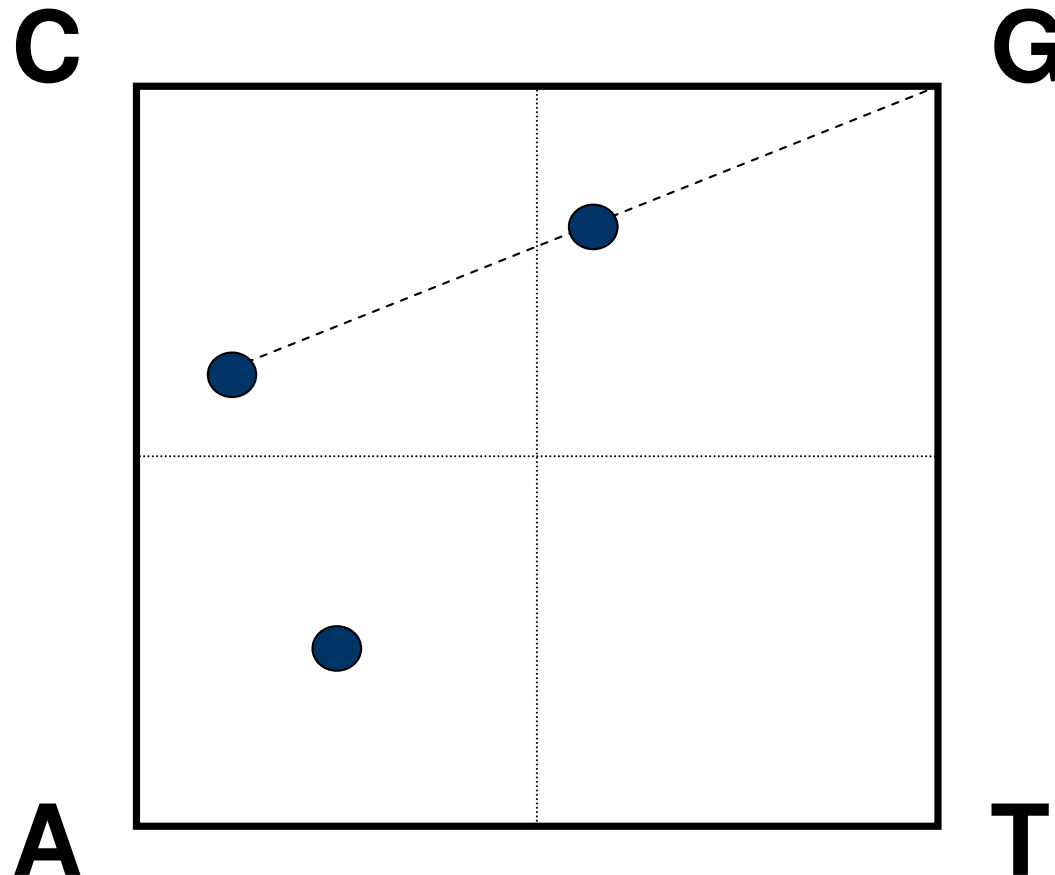
# La signature génomique - construction

A C **G** T G C T T A



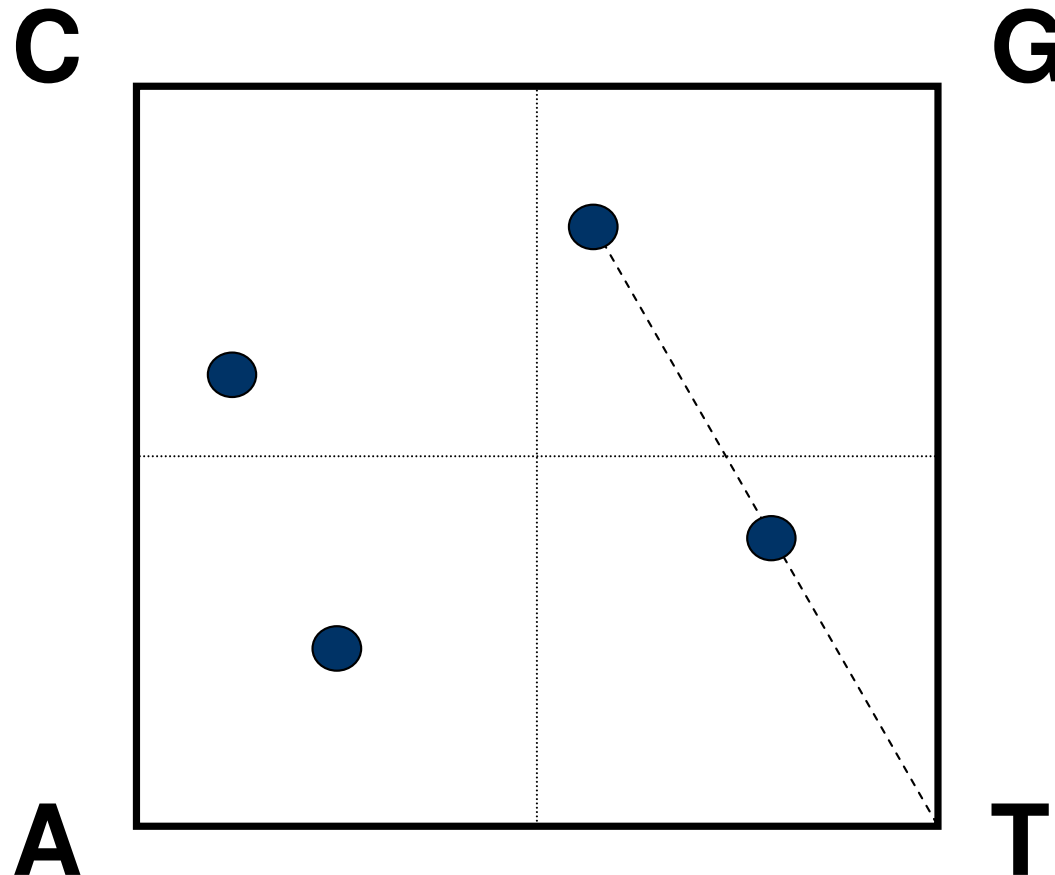
# La signature génomique - construction

A C **G** T G C T T A



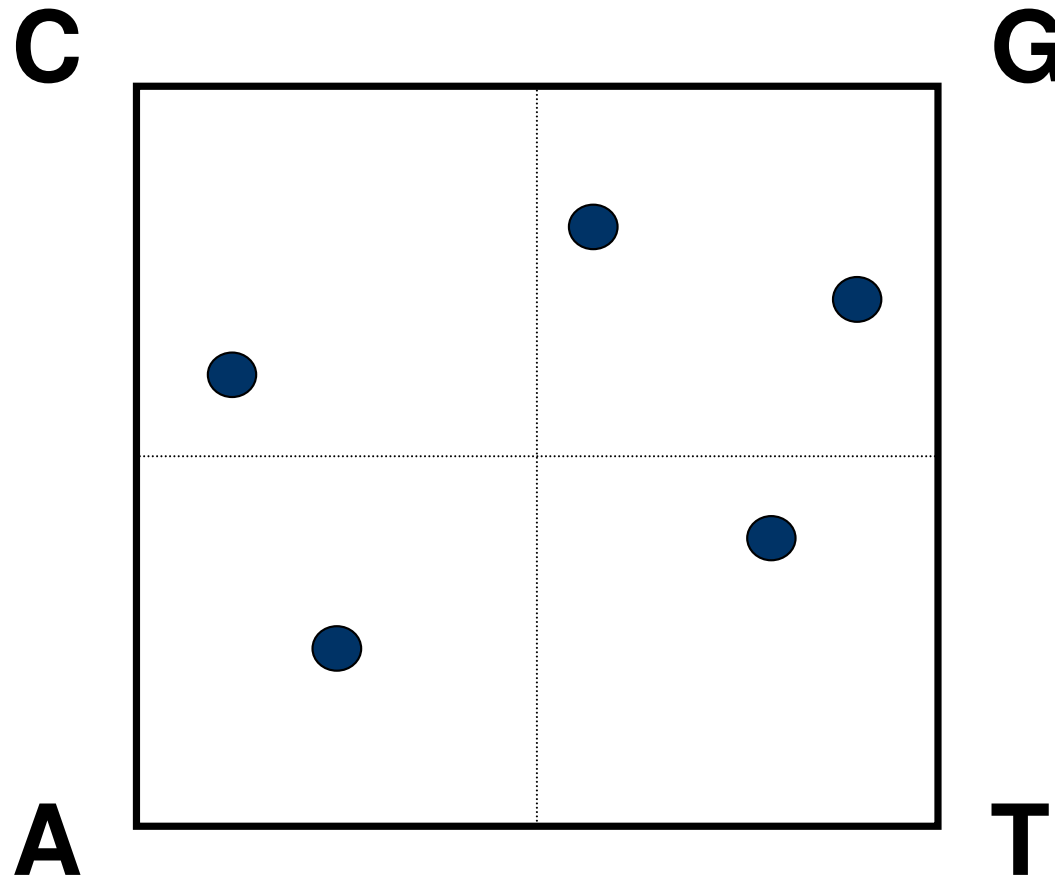
# La signature génomique - construction

A C G **T** G C T T A

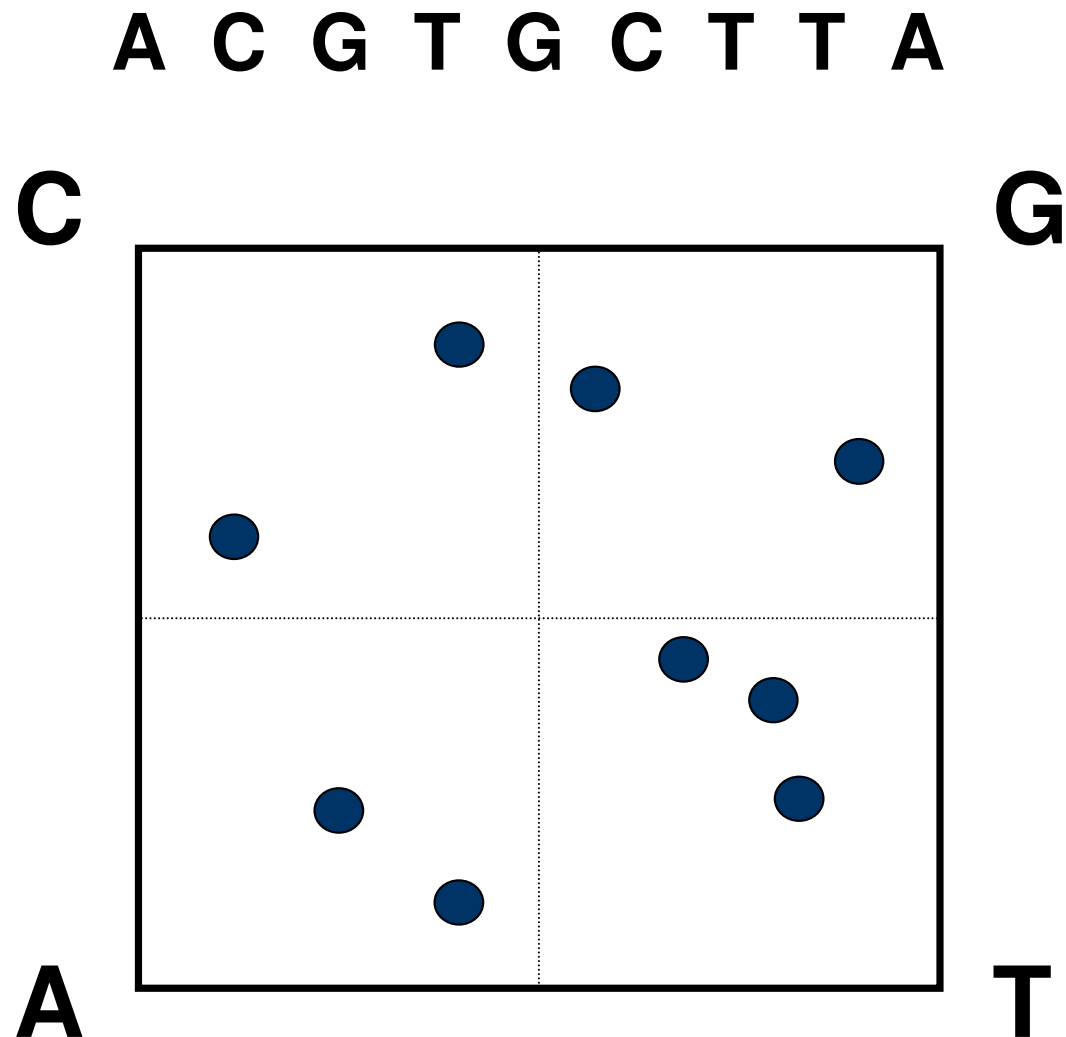


# La signature génomique - construction

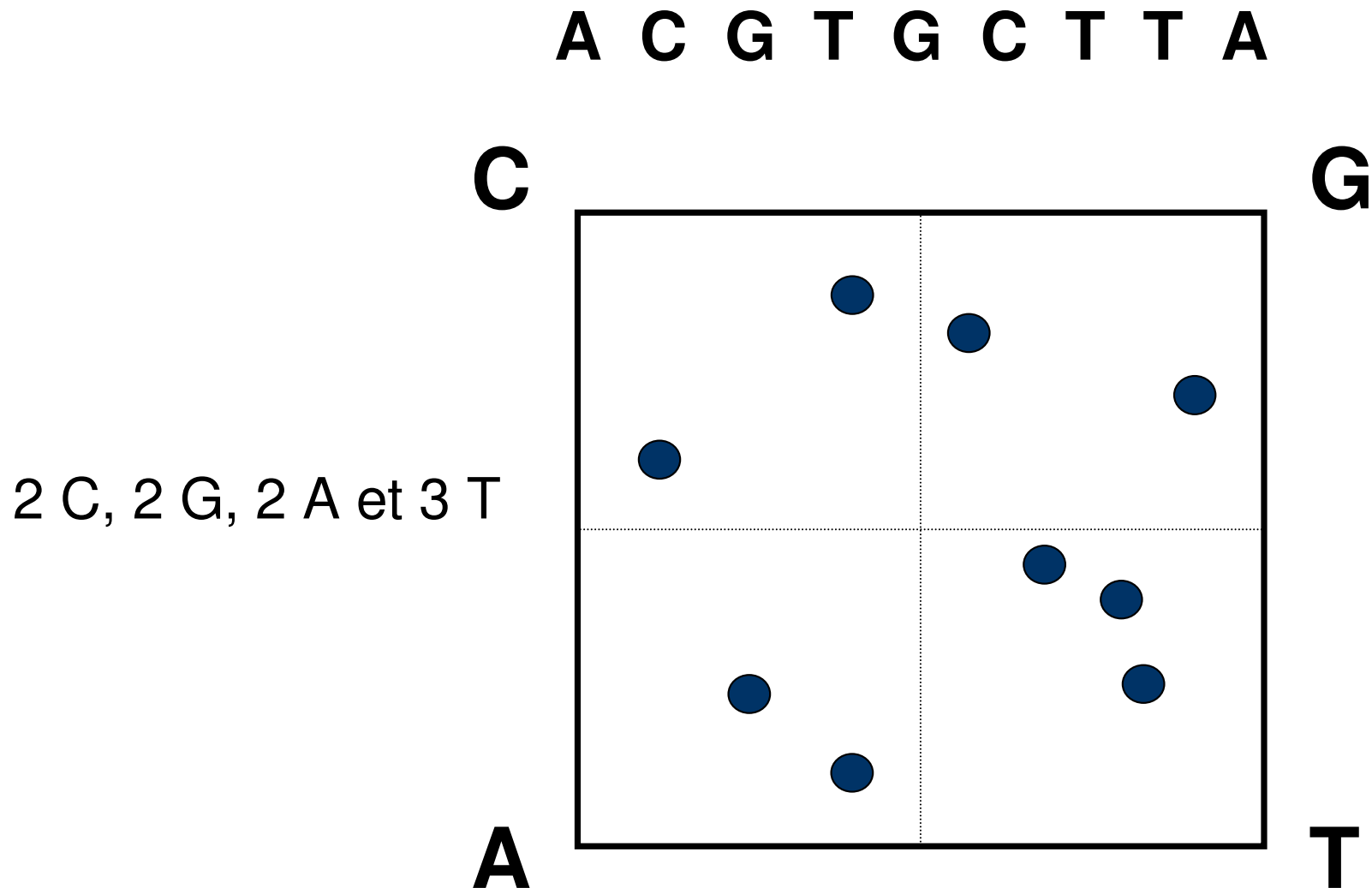
A C G T **G** C T T A



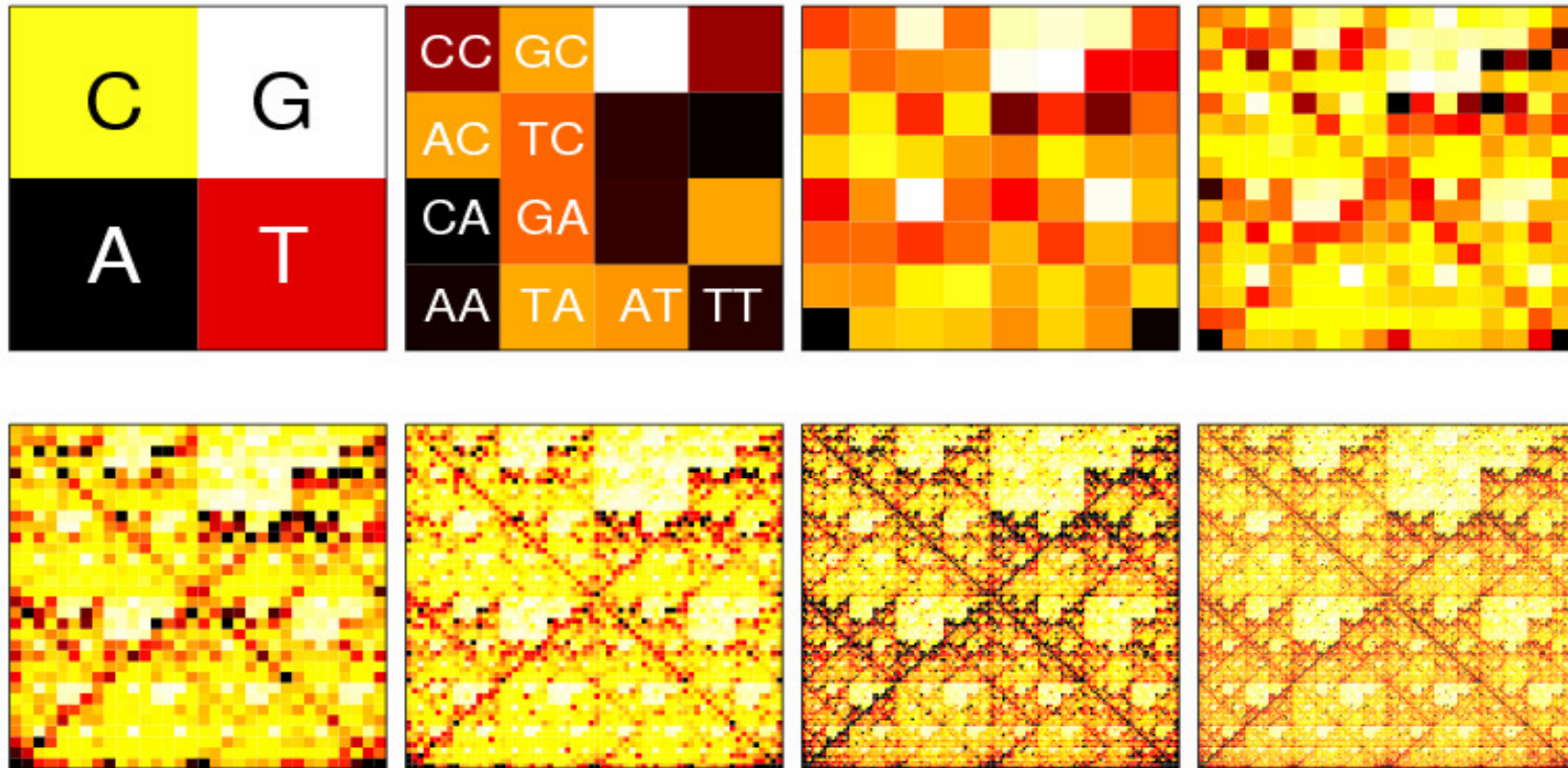
# La signature génomique - construction



# La signature génomique - construction



# La signature génomique - construction

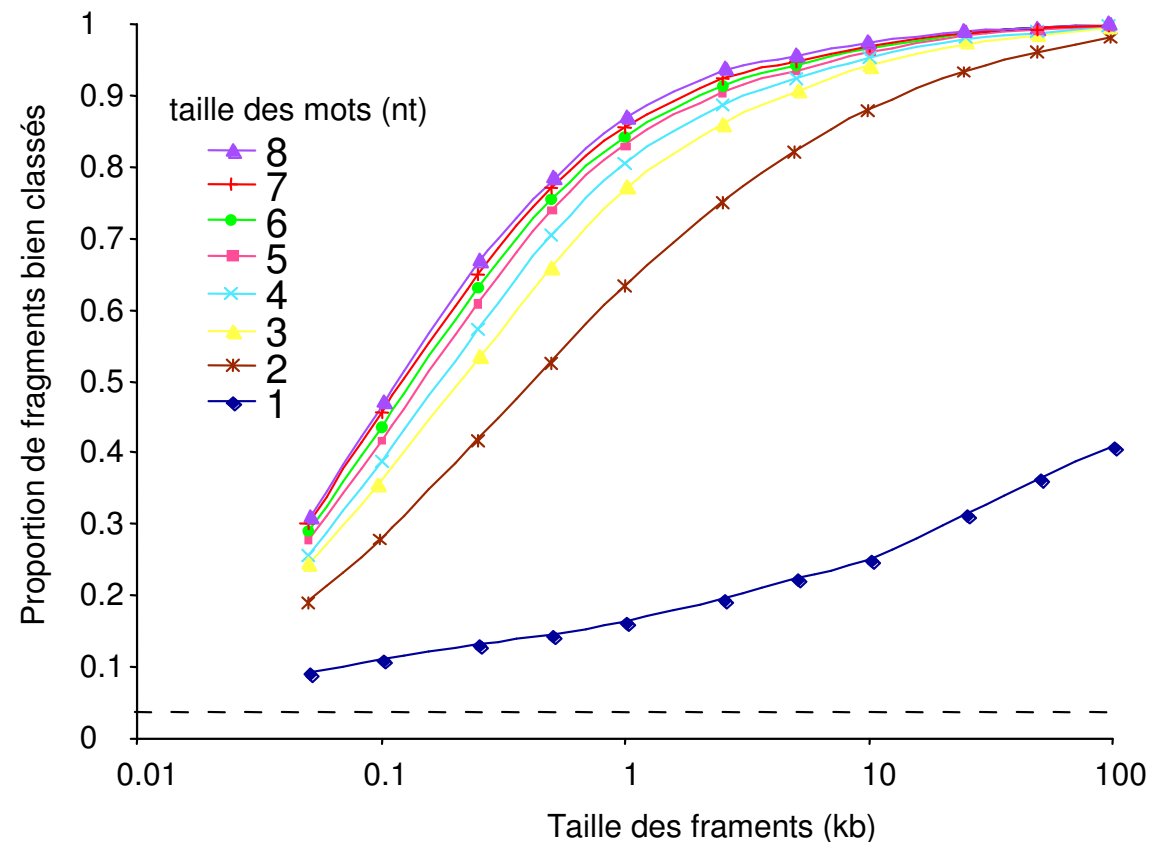
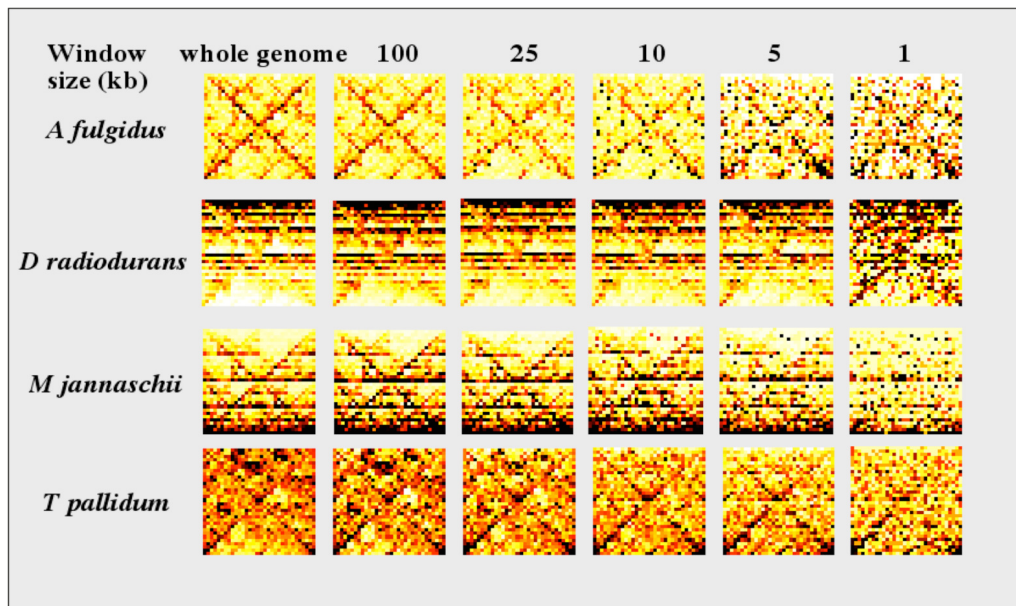


Fréquence des n-nucléotides ( $1 \leq n \leq 8$ ) du chromosome 22 de *H. sapiens*.

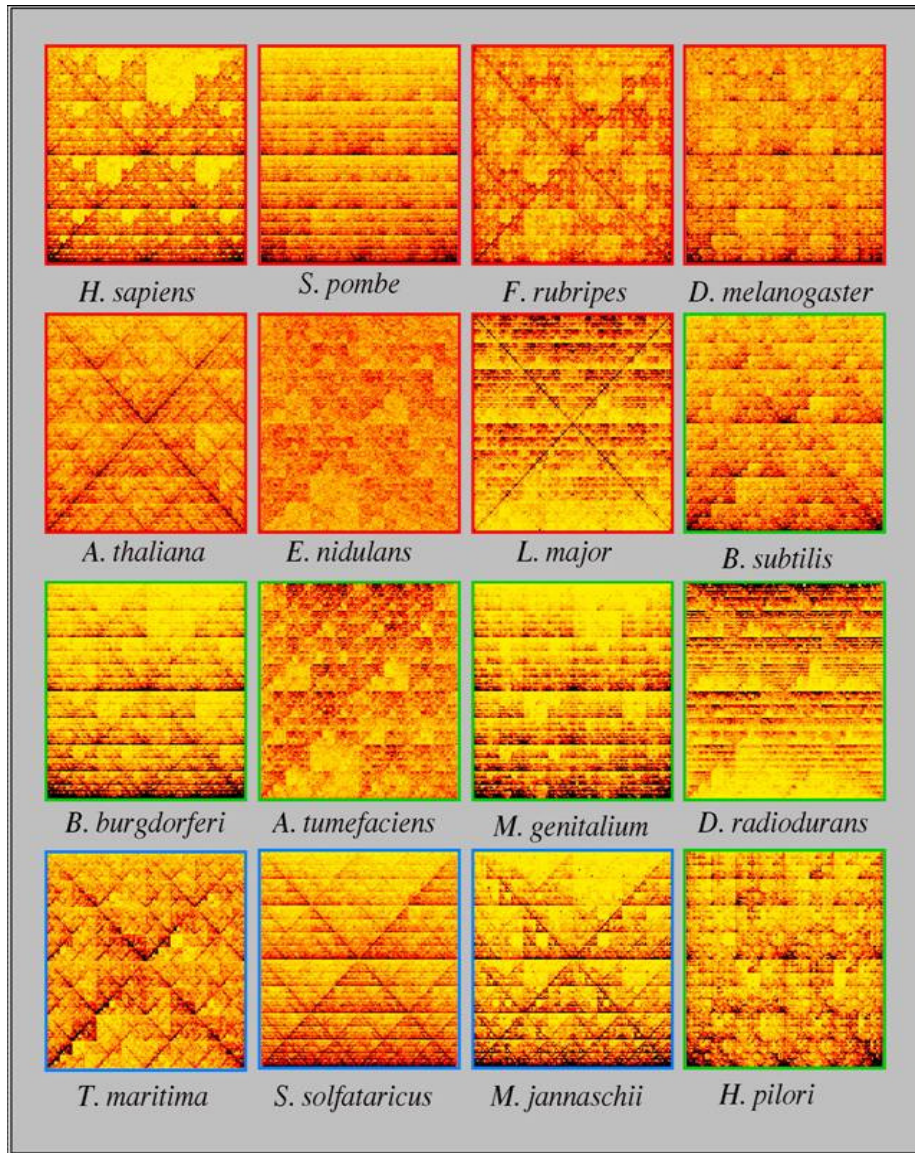
Temps < 1 s pour comptabiliser tous les trétanucléotides  
d'un génome de 5 Mb (processeur 2.6 GHz, RAM 2 Go)

# La signature génomique – propriétés

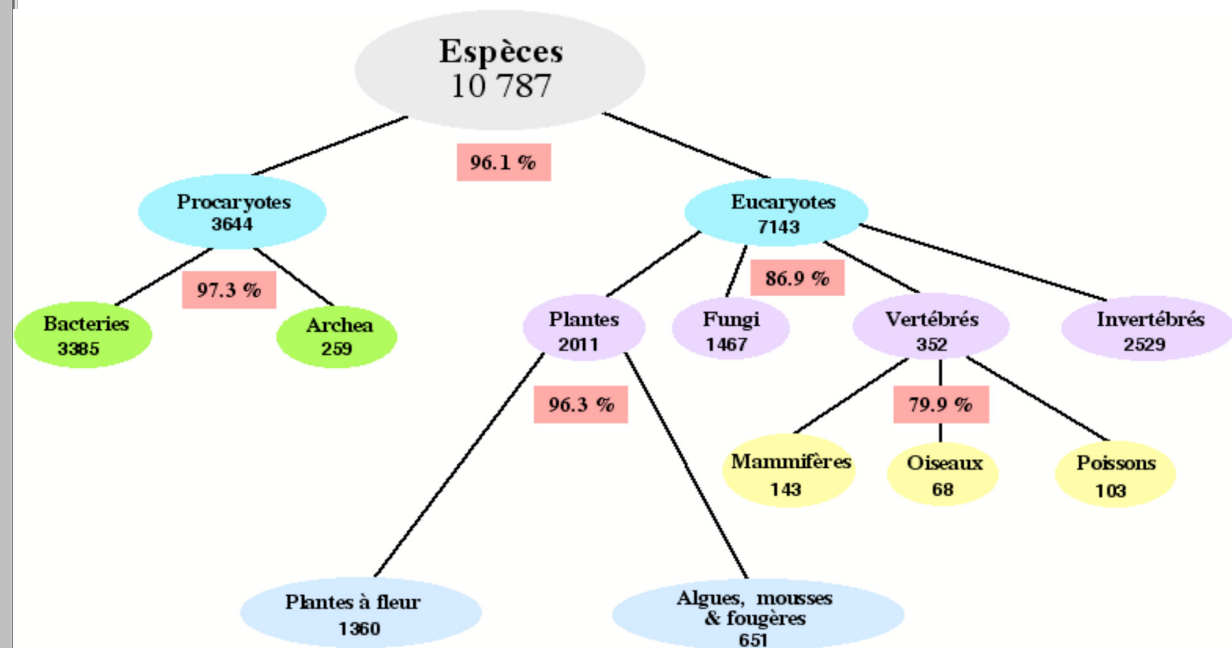
La signature est stable à partir de 1 kb, mots > 4nt



# La signature génomique – propriétés



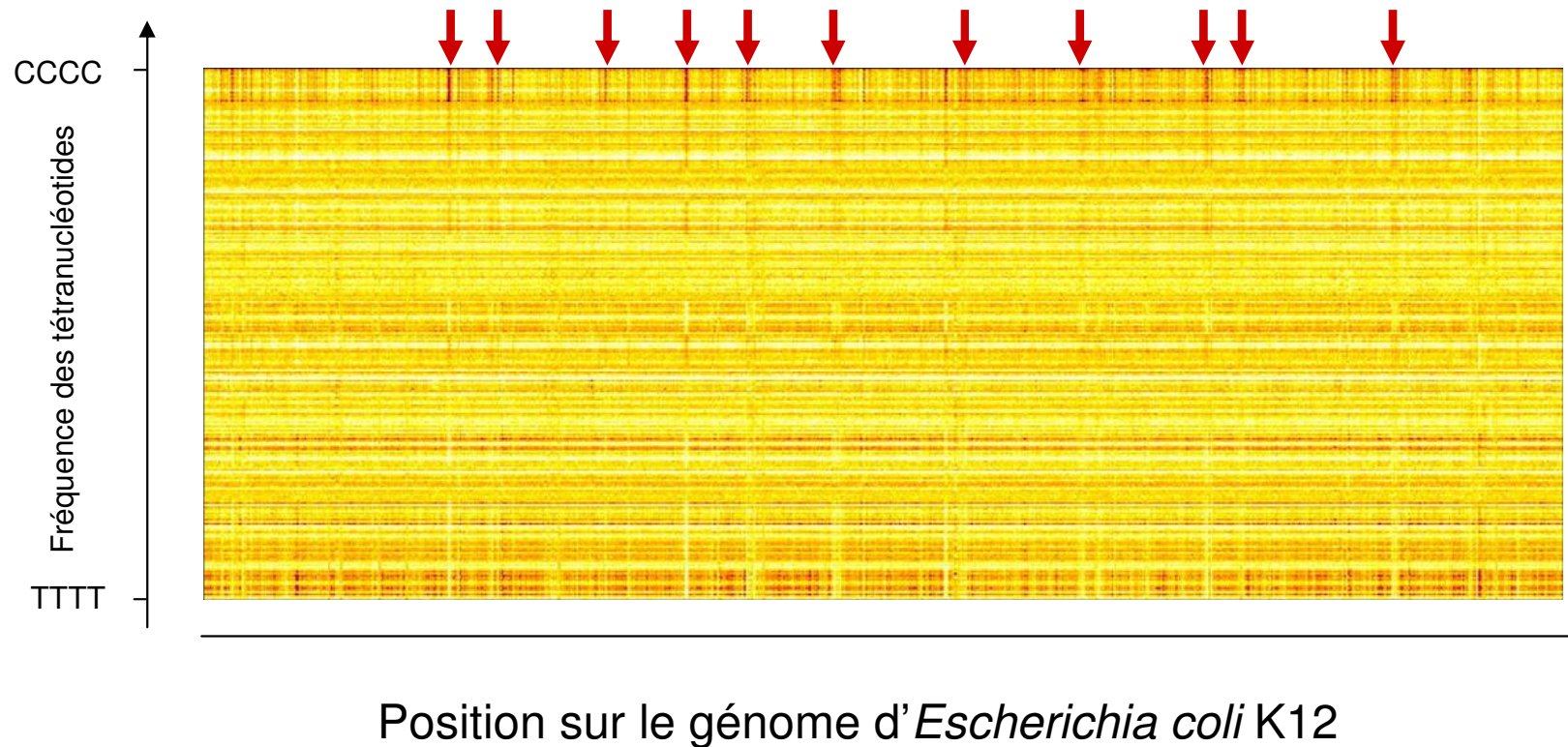
## La signature est espèce-spécifique



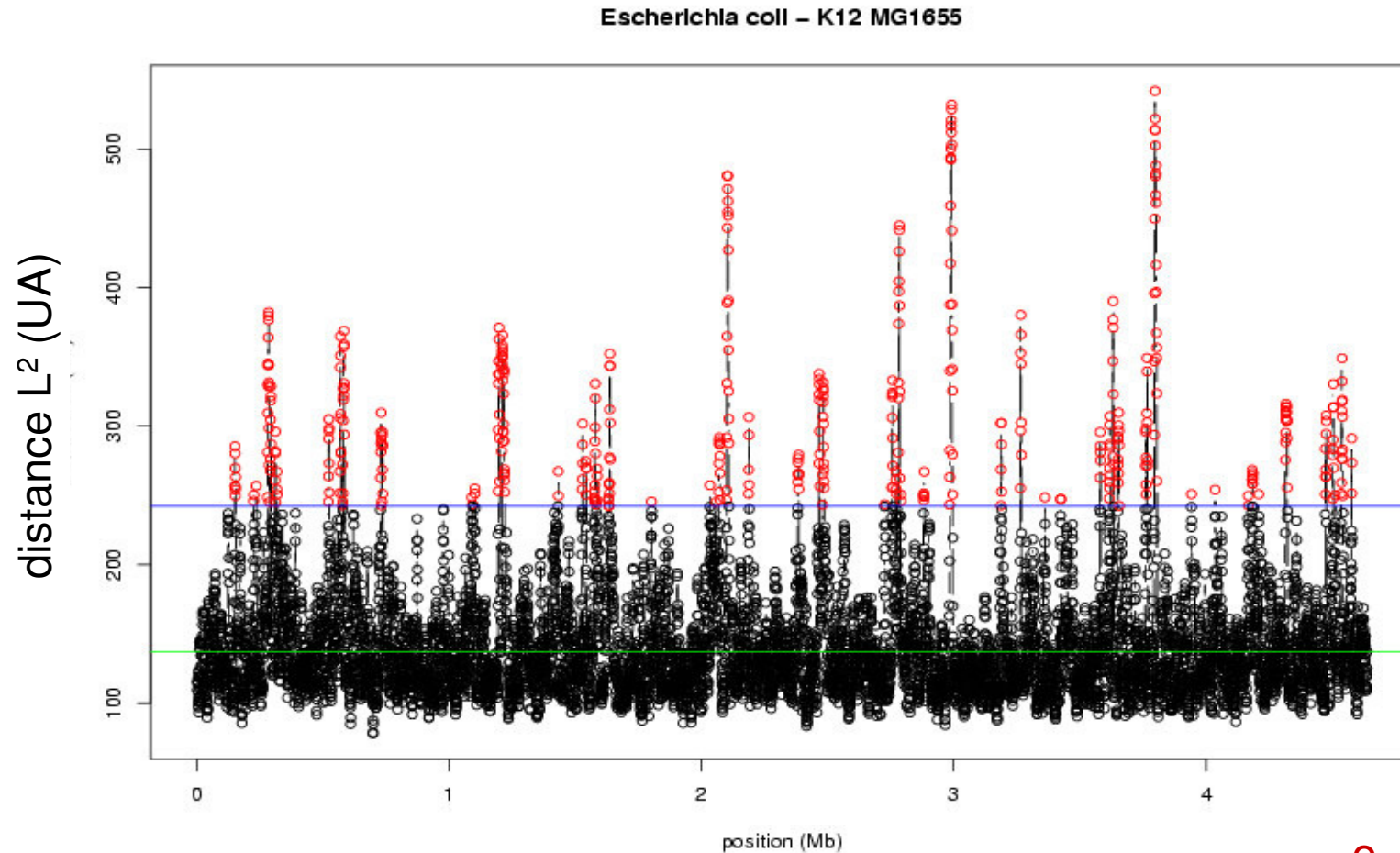
Classification de signatures par analyse discriminante

# Recherche de fragments « atypiques »

Calcul de la signature le long du génome sur des fenêtres de 5 kb, chevauchantes de 500 pb



# Recherche de fragments « atypiques »



○ = régions atypiques

$$l^2 = \sqrt{\sum_{i=words} (f_i^{window} - f_i^{mean})^2}$$

Moyenne 134 UA, seuil 237 UA

# Méthodes paramétriques - Résultats

## Application d'un consensus « 2 sur 3 »

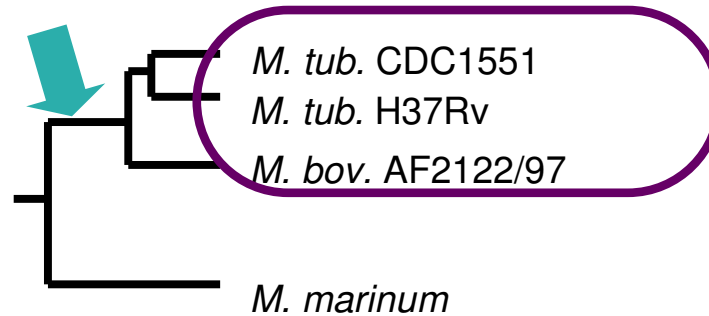
| atypical regions<br>(complete genome) | Number | Total size<br>(kb) | Size range<br>(kb) | Number of<br>genes | G+C %          |
|---------------------------------------|--------|--------------------|--------------------|--------------------|----------------|
| <i>M. bovis</i>                       | 165    | 220<br>(4.35 Mb)   | 0.3 – 6.9          | 240<br>(3953)      | 62.7<br>(65.6) |
| <i>M. tub.</i> CDC1551                | 207    | 259<br>(4.4 Mb)    | 0.3 – 10.1         | 279<br>(4189)      | 63.4<br>(65.6) |
| <i>M. tub.</i> H37Rv                  | 161    | 217<br>(4.41 Mb)   | 0.3 – 9.5          | 238<br>(3999)      | 62.5<br>(65.6) |



~ 5 % du  
génom

# Recherche des transferts **spécifiques** du complexe *M. tuberculosis*

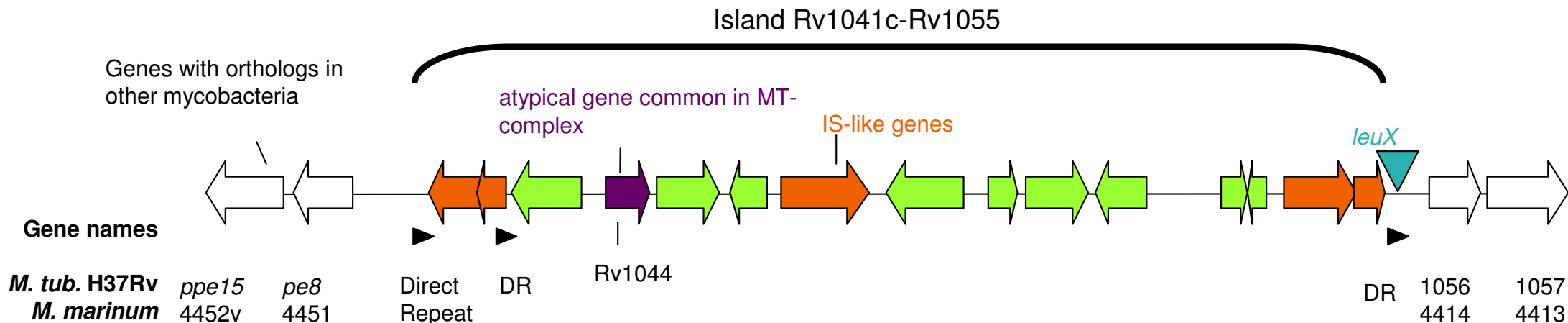
Horizontal transfers ?



1. Recherche des gènes atypiques **communs** aux 3 génomes du complexe
2. Sélection des gènes atypiques communs **absents** des autres mycobactéries
3. Extension de ces gènes en « îlots génomiques » en ajoutant les gènes adjacents **sans orthologue synténique** chez d'autres mycobactéries

# Résultats

- **48** îlots génomiques décrits, contenant **256** gènes
- Soit **4,5 %** du génome
- Présence de caractéristiques typiques d'îlots génomiques

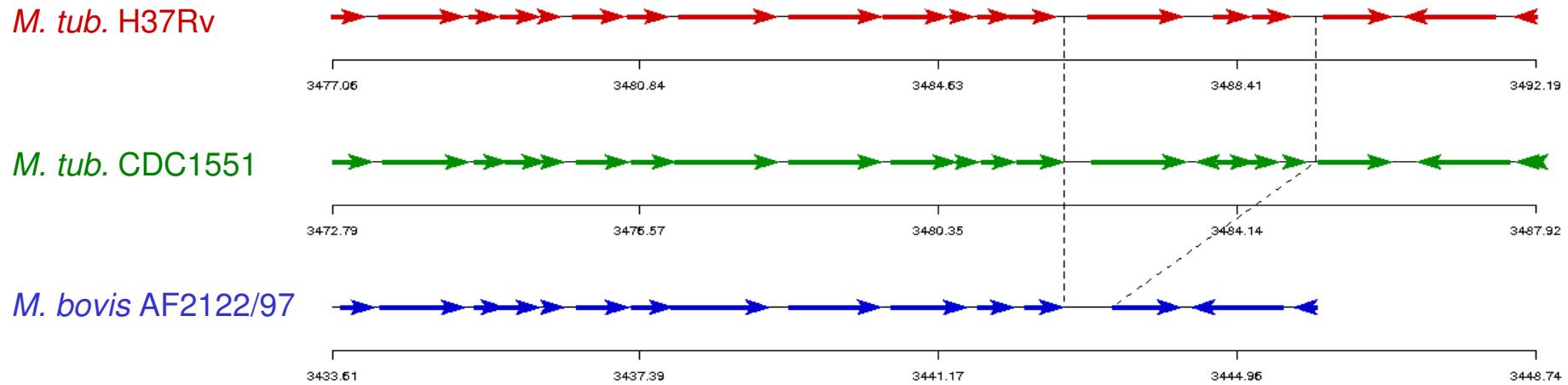


# Fonction des transferts horizontaux

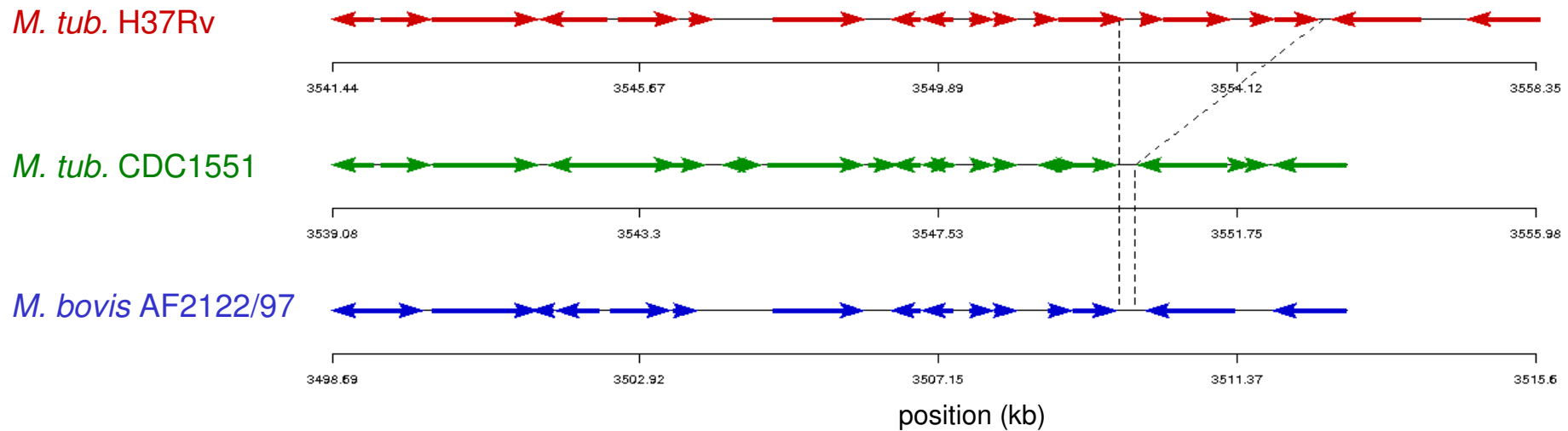
| Gene functional category                   | No. of genes | % in genomic islands | Average % in complete genomes |
|--------------------------------------------|--------------|----------------------|-------------------------------|
| IS elements, repeated sequences & phages   | 40           | 15.6                 | 3.7                           |
| Central intermediary metabolism            | 28           | 10.9                 | 27.8                          |
| <b>Virulence genes</b>                     | <b>21</b>    | <b>8.2</b>           | <b>6.5</b>                    |
| Regulatory functions                       | 13           | 5.1                  | 10.3                          |
| Cell envelope & possible membrane proteins | 12           | 4.7                  | 17.5                          |
| Others                                     | 10           | 3.9                  | NA                            |
| <i>Unknown</i>                             | <i>132</i>   | <i>51.6</i>          | <i>&lt; 48.0</i>              |
| <b>Total</b>                               | <b>256</b>   |                      |                               |

# Réarrangements des transferts horizontaux

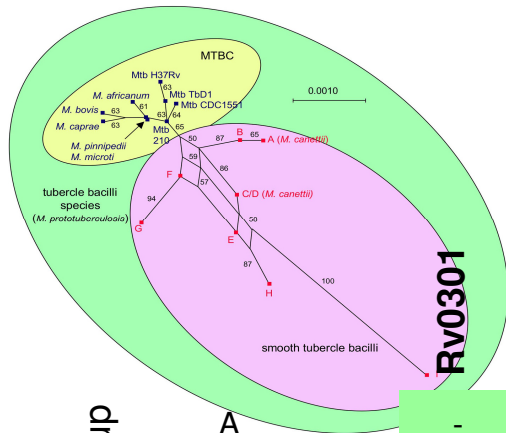
## îlot Rv3108-Rv3126c



## îlot Rv3173c-Rv3191c

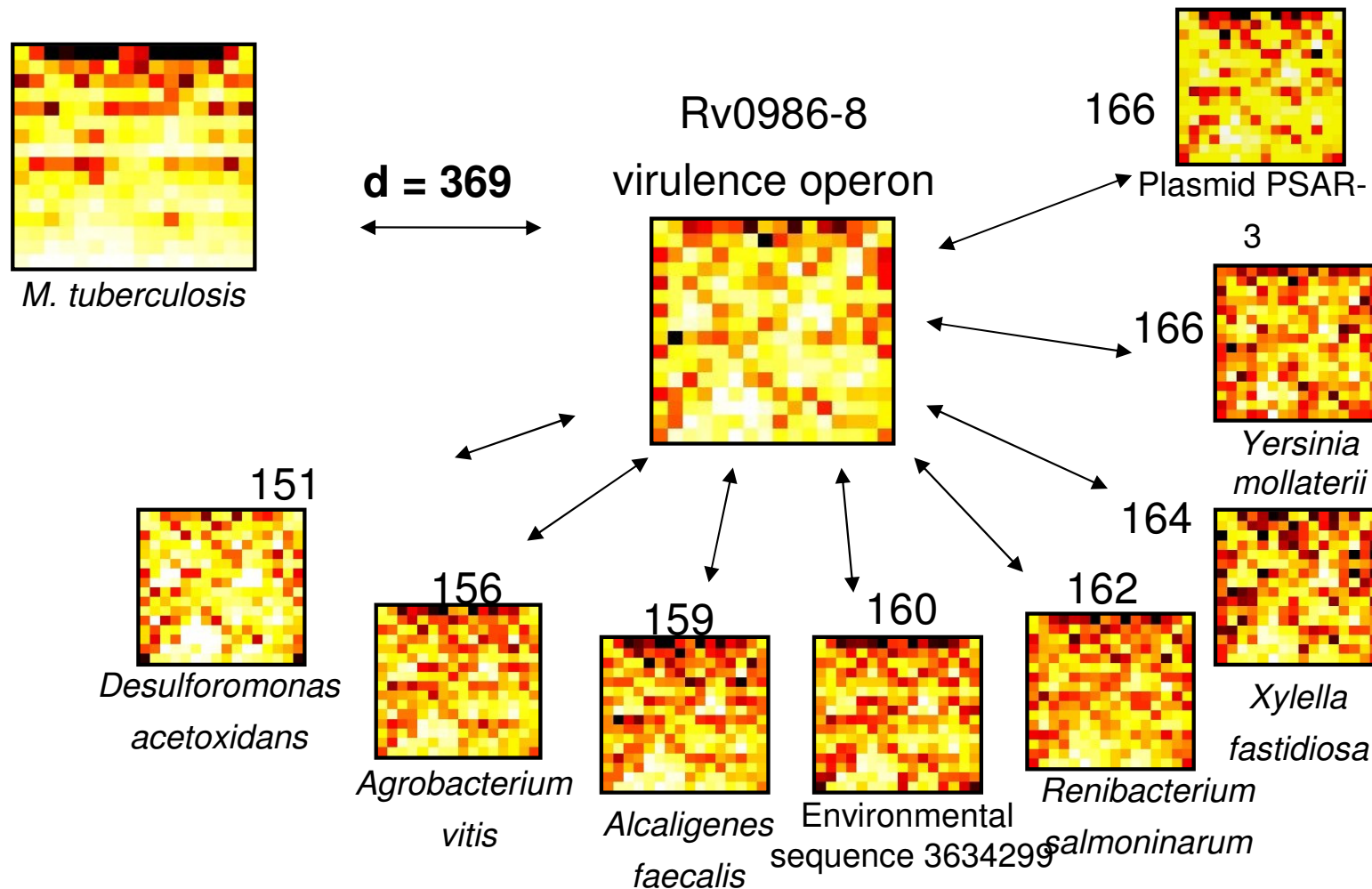


Présence / absence chez les « *prototuberculosis* »  
(Southern Blot)

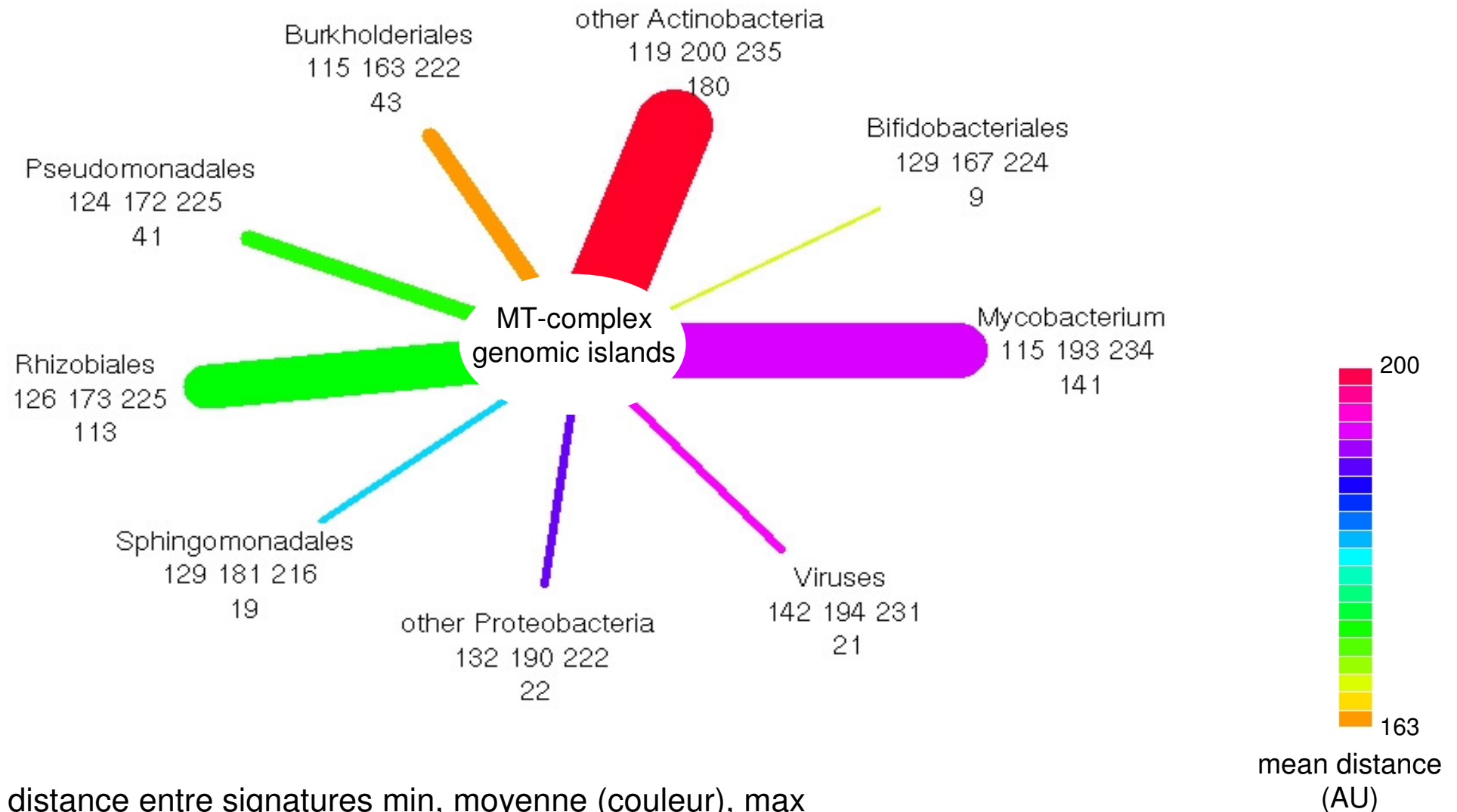
[illegible]

# Origine potentielle des transferts horizontaux

Utilisation d'une banque de données de signatures contenant 65 000 espèces, souches, organelles, virus et plasmides



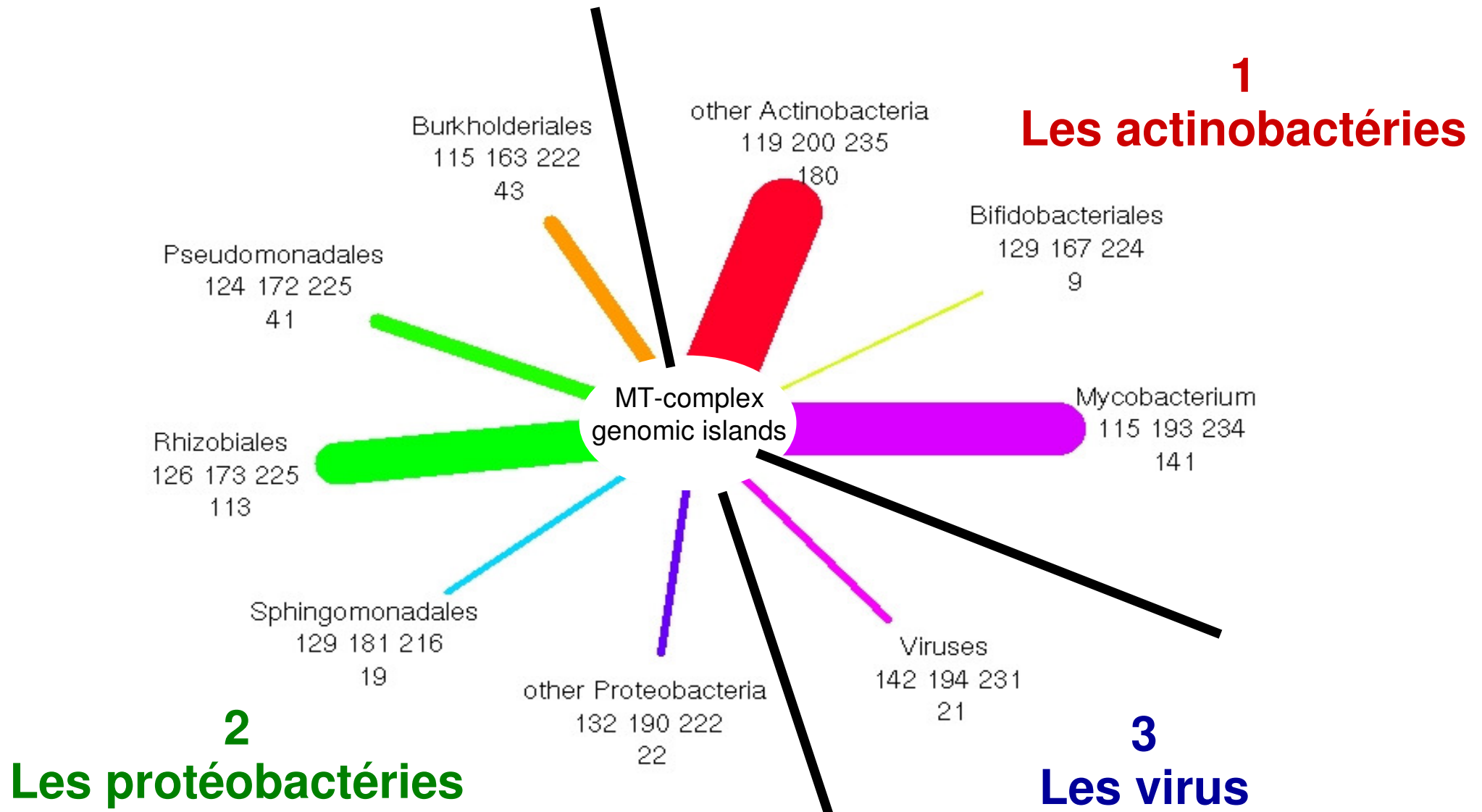
# Les groupes de donneurs potentiels



Rang sup : distance entre signatures min, moyenne (couleur), max

Rang inf : nombre d'espèces donneuses potentielles dans chaque groupe taxonomique

# 3 groupes majeurs de donneurs potentiels



# Conclusions

- Émergence il y a **1 Ma** du bacille de la tuberculose à partir d'une **souche ancestrale environnementale**
- Épisodes majeurs d'acquisition de **transferts horizontaux**
- Originaires de **bactéries du sol** (Rhizobiales) et de **bactéries commensales ou pathogènes** (Pseudomonadales, Burkholdériales)
- **Sur-abondance** de gènes de **virulence** dans les îlots génomiques et nombreux gènes « unknown »...
- Transferts horizontaux « jeunes », encore soumis à des **réarrangements**

# Perspectives

- Amélioration des **méthodes paramétriques** de détection des gènes atypiques
- Amélioration de la **banque de signatures**
- **Étude fonctionnelle** des gènes des îlots génomiques détectés
- Extension de l'étude aux **nouvelles souches** de *M. tuberculosis* séquencées, incluant des souches multi-résistantes aux antibiotiques