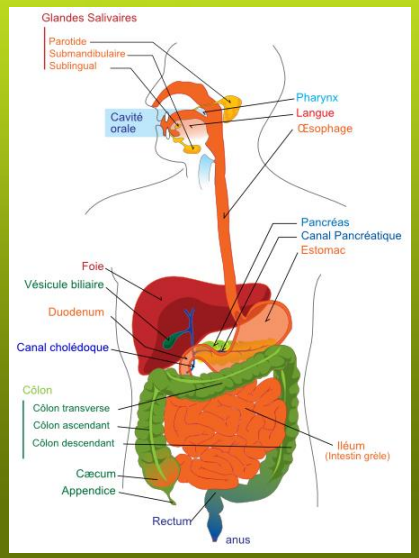
Université Paris Diderot
Ecole doctorale Gc2ID

Claire Hoede – INSERM et
Université Paris 7 UMR722
Erick Denamur
Olivier Tenaillon

Impact des processus de mutation et de recombinaison sur la diversité génomique au sein de l'espèce *Escherichia coli*

E. COLI : UNE BACTÉRIE VERSATILE



Habitat primaire : le tube digestif des vertébrés

Habitat secondaire : l'eau et les sédiments

Commensale

2 millions de morts par an

Infections intraintestinales et extraintestinales

Diarrhées
Infections urinaires
Méningites

Comprendre cette diversité à l’heure de la génomique

Evolution du génome de *E. coli*

La mutation

La recombinaison

La phylogénie

L'organisation générale du génome

LA MUTATION

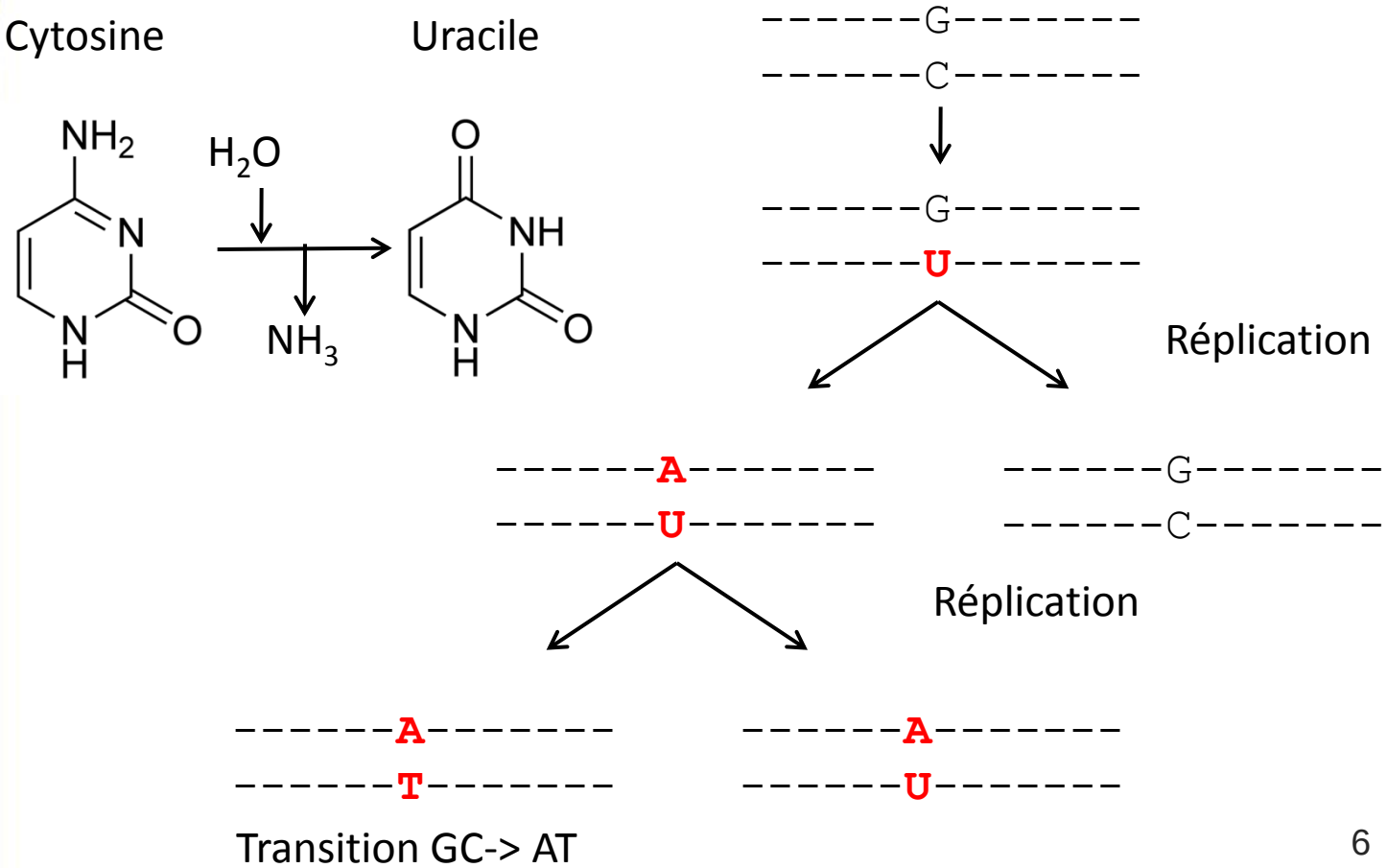
- Changement héréditaire dans la molécule d'ADN.
- Spontanée ou induite par l'action d'agents extérieurs.
- Génératrice de diversité génomique.

- Nombreux systèmes de réparation.

- Equilibre entre conservation de l'information génétique / diversité permettant l'adaptation.

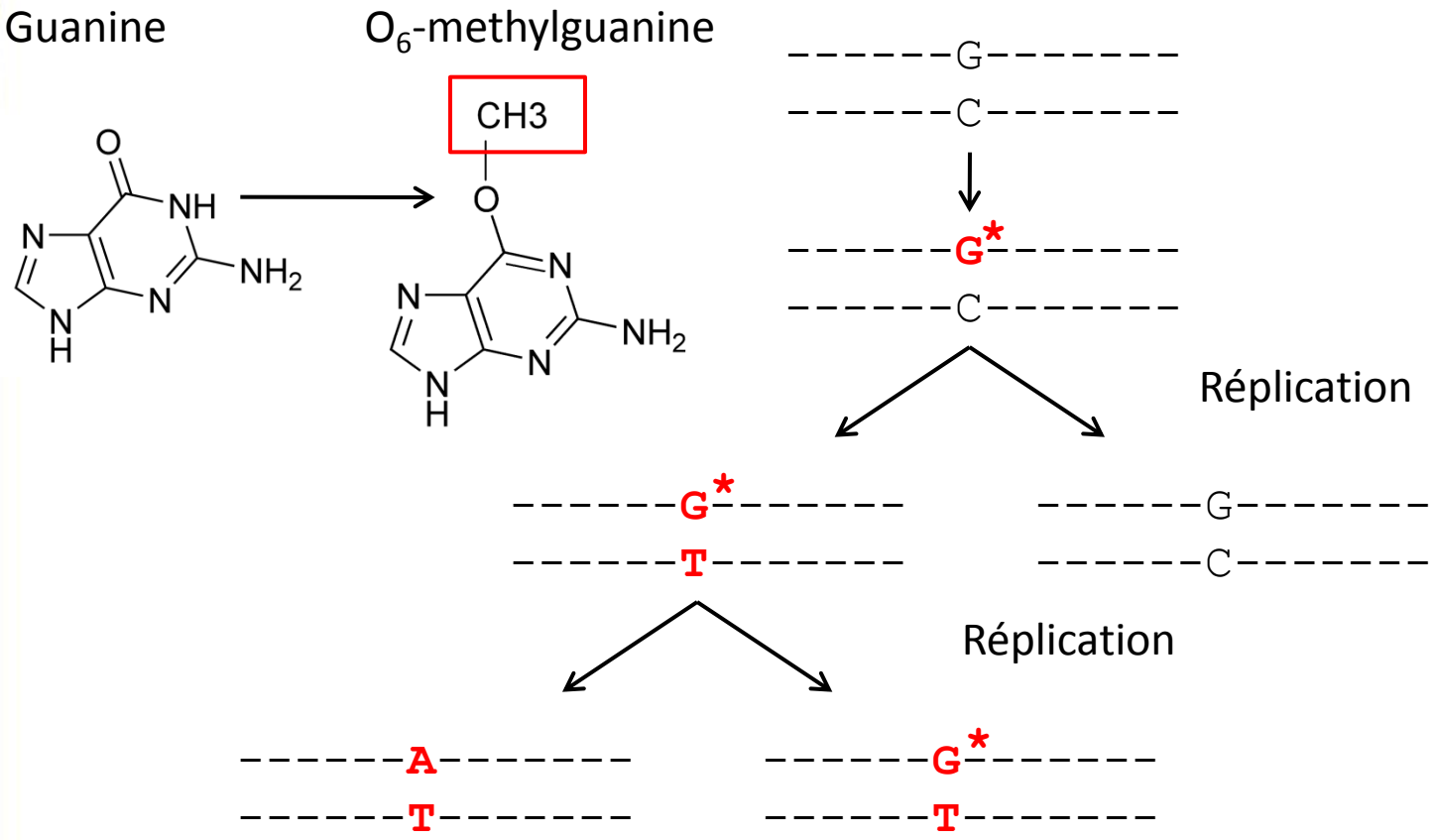
Les principales altérations chimiques spontanées

- La désamination oxydative.



Un exemple d'altération chimique induite

- L'alkylation.



Autres mutations spontanées

- Les erreurs réplicatives.
  - Mutations ponctuelles.
  - Dérapages de l'ADN polymérase.
  - Taux d'erreur de la machine réplicative par base et par réplication ≈ 10^{-7}.

Les mécanismes de réparation : un contrôle génétique du taux de mutation

- Nombreux mécanismes de réparation.
  - Pré-réplicatif.
    - BER (Réparation par excision de base)
  - Post-réplicatif.
    - SRM (Système de réparation des mésappariements).

Taux de mutation total

- Taux de mutation par base et par réplication de $5x10^{-10}$ (Miller 1996).
- Soit 0,003 mutation par génome et par réplication (Drake 1991).

- Lorsque une population bactérienne doit s’adapter à de nouvelles conditions, la sélection peut favoriser une augmentation du taux de mutation.

Quels sont les types de variations possibles du taux de mutation ?

Globale et permanente

- Souches mutatrices constitutives (Tenaillon *et al.* 1999).
- Les souches mutatrices (1 – 15% des populations naturelles) résultent de l'inactivation de certains mécanismes de réparation (dont le SRM).

➔Fort coût car des gènes essentiels peuvent être affectés et cela en dehors des phases d'adaptation.

Globale et transitoire

– Souches mutatrices inductibles (Bjedov *et al.* 2003) : système activé en condition de stress.

– Le système SOS semble impliqué : réinitialise la fourche réplicative lorsqu'un stress la bloque.

➔ Affecte tout le génome

Locale

- Des répétitions entrainent des décalages de cadre de lecture.
- Loci sous sélection diversifiante notamment pour échapper au système immunitaire de l'hôte (Moxon, 1994).

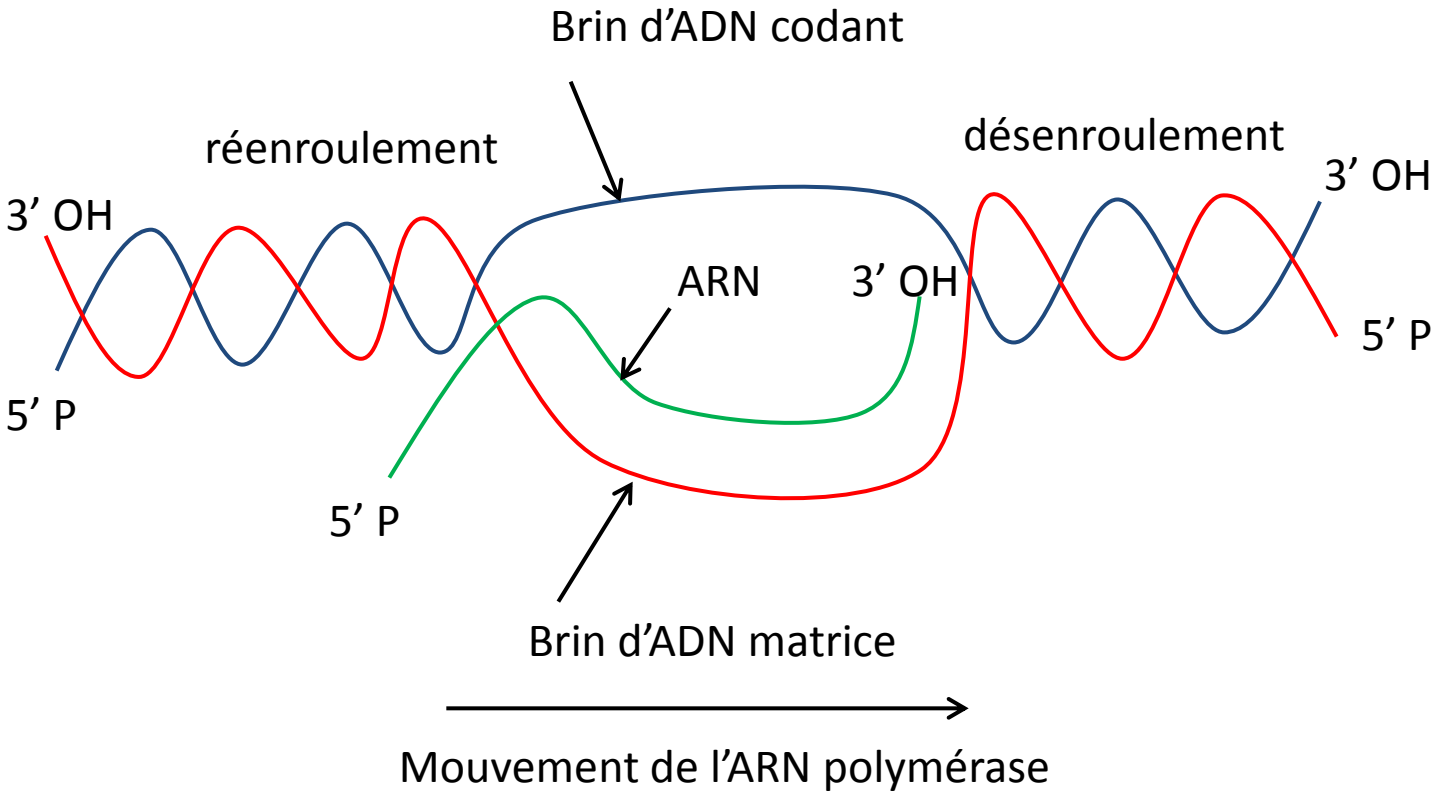
➔Augmentation ciblée du taux de mutation mais non inductible.

UN CONTRÔLE LOCAL ET INDUCTIBLE ?

- La mutagénèse transcriptionnelle ?

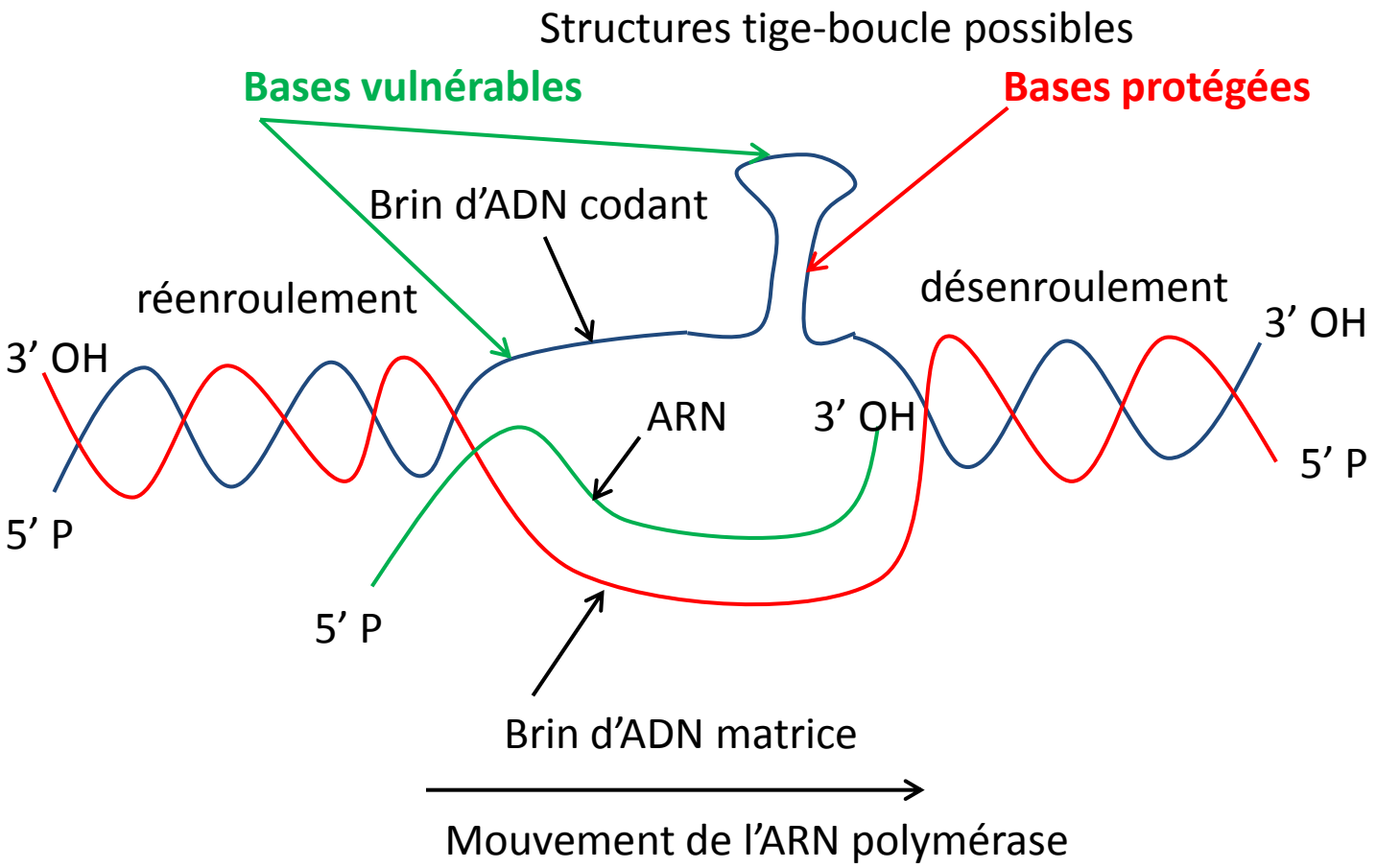
- Comme proposé par Barbara Wright en 2002 (Wright *et al.* 2002).

Pendant la transcription le brin codant est simple brin



Hoede *et al.*
2006
PLoS Genet.

Les structures secondaires protégeraient certaines bases



Hoede *et al.* 2006 PLoS Genet.

Etude de cette mutabilité à l'échelle du génome complet

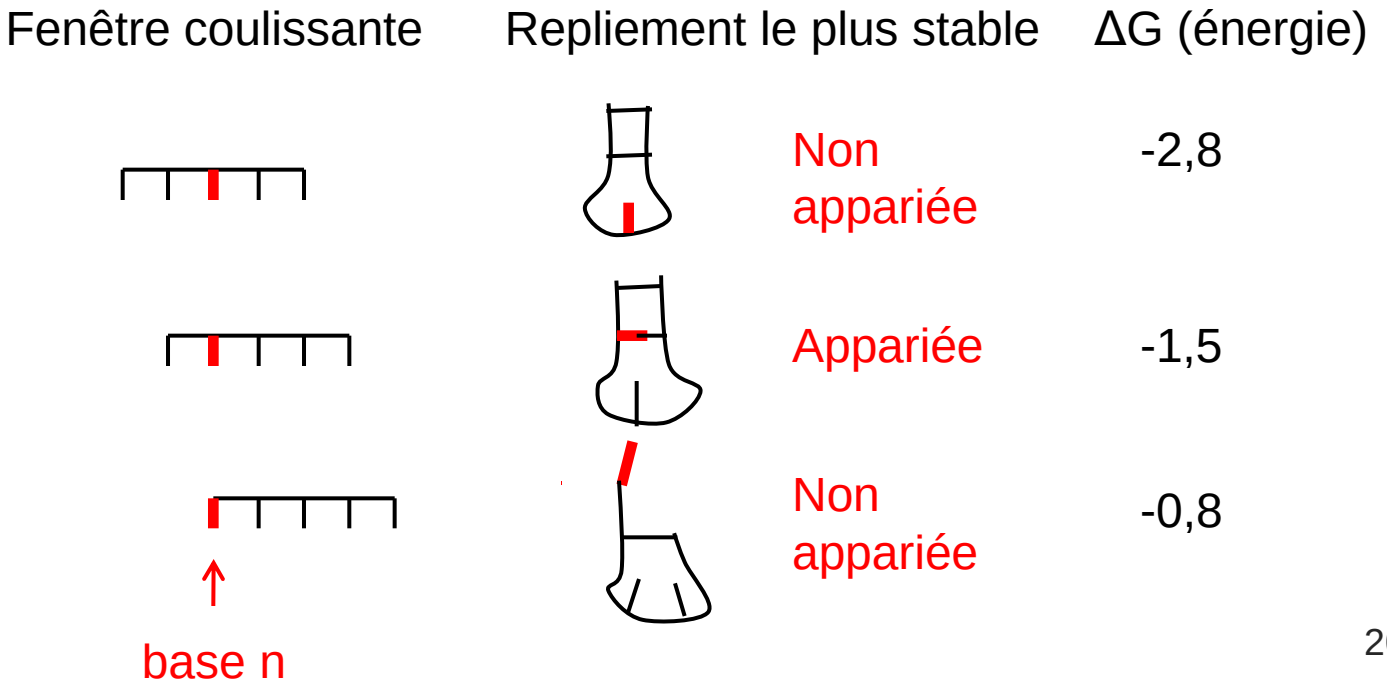
- Ce type de mutagenèse influence-t-il l'évolution des séquences à long terme ou est-il masqué par l'ensemble des facteurs pouvant influencer la mutagenèse?

- Existe-t-il des gènes plus ou moins robustes à cette forme de mutabilité?

Hoede *et al.*
2006
PLoS Genet.

Calcul des repliements

- Repliements les plus stables impliquant 30 nucléotides de tous les gènes (hormis les pseudogènes).



Un indice de mutabilité

- Exprime la proportion du temps qu'une base passe dans un état non-apparié.

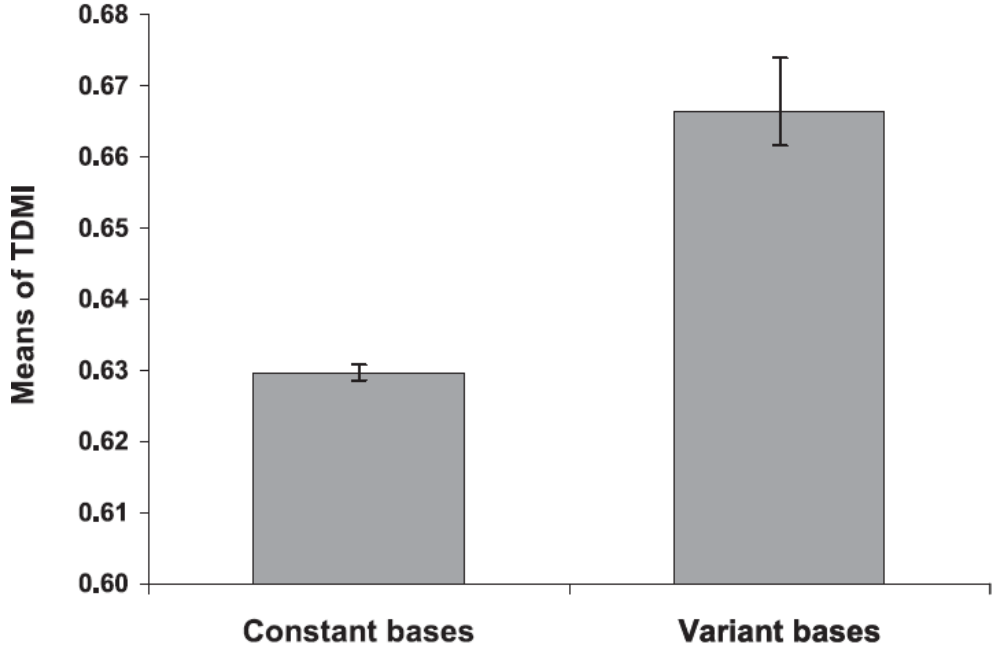
$$\text{TDMI (base n)} = \frac{\sum_{\text{formes non apparié}} \exp(-\Delta G/RT)}{\sum_{\text{toutes les formes}} \exp(-\Delta G/RT)}$$

« Transcriptional Driven Mutability Index »

Hoede *et al.*
2006
PLoS Genet.

Un indice corrélant avec la mutagénèse observée

- Les sites quatre fois dégénérés (L4) ayant changé entre *E. coli* K-12 et *E. coli* CFT073 ont un TDMI supérieur à ceux qui n'ont pas changé.

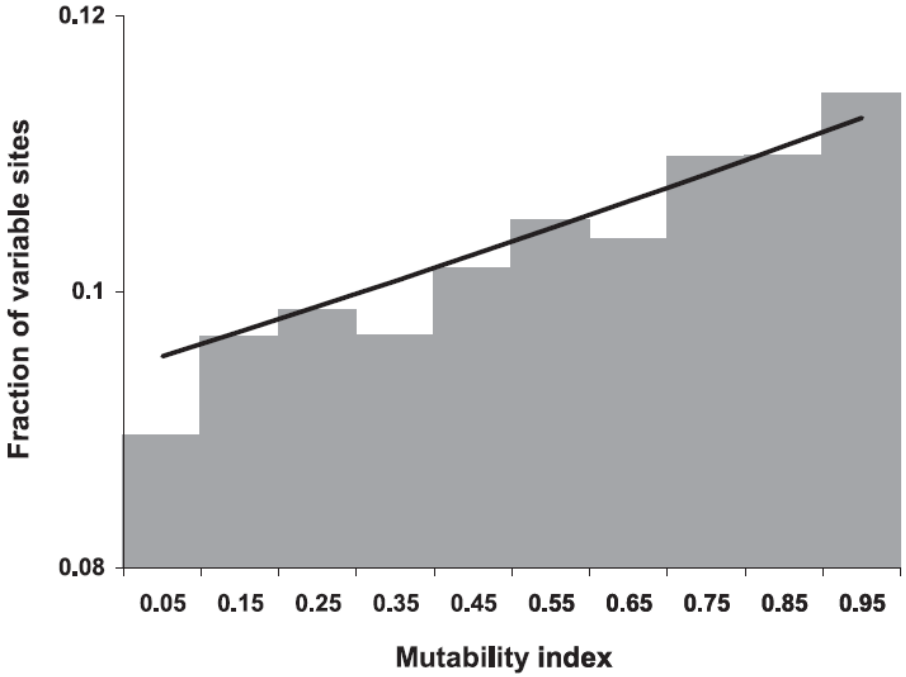


P-value (Mann-Whitney, bilatéral) < 2,2 x 10^{-16} ; N = 550575

Hoede *et al.* 2006 PLoS Genet.

Fraction de sites variables en fonction du TDMI

- Un site L4 avec un TDMI maximum a une probabilité de changer entre les deux souches augmentée de 20,7% par rapport à un site présentant un TDMI minimum.



Logistic regression

P-value < $2,2 \times 10^{-16}$

N = 550575

Hoede *et al.* 2006 PLoS Genet.

QUEL LIEN ENTRE TDMI ET EXPRESSION DES GÈNES ?

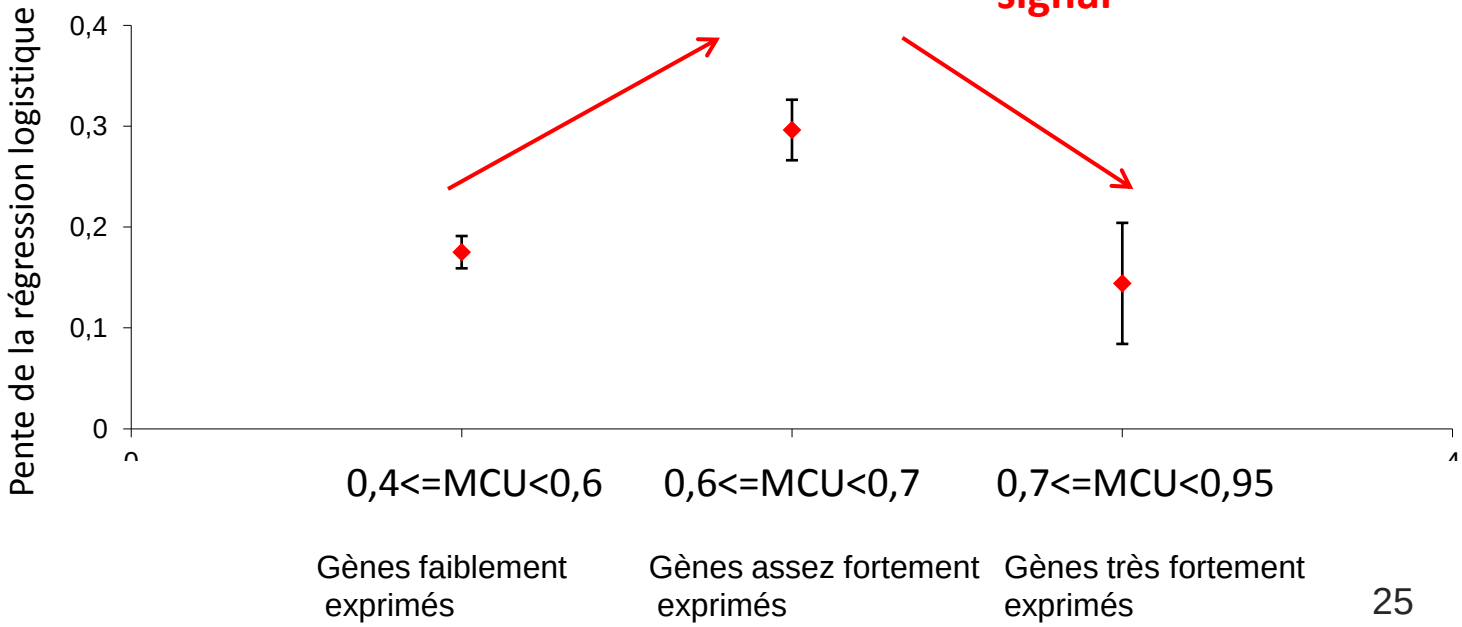
- Biais de codons mesuré par le Major Codon Usage (MCU).
- Bonne approximation de l’expression à long terme.
- Les gènes fortement transcrits sont-ils davantage affectés par la mutabilité transcriptionnelle ?

La balance entre TDMI et MCU

Impact du TDMI sur la variabilité des sites L4 entre K-12 et CFT073 par classe de MCU (Major Codon Usage)

Le TDMI semble avoir plus d'impact sur les gènes assez fortement exprimés

D'autres contraintes liées à un très fort biais de codon inverse le signal



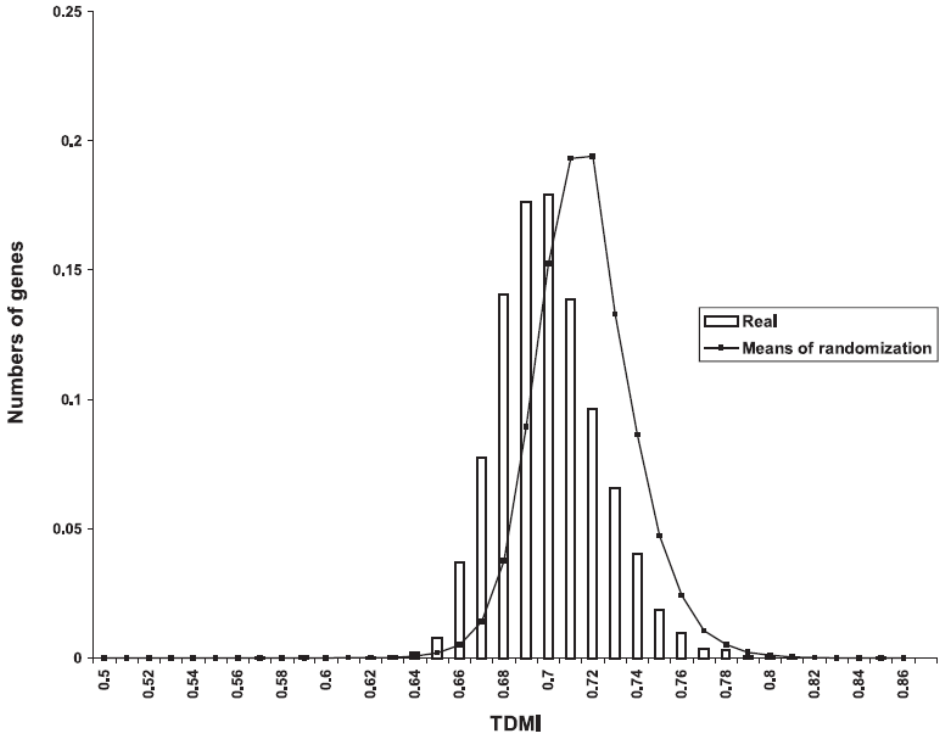
Hoede *et al.* 2006 PLoS Genet.

Une mutagénèse sous sélection ?

Une TDMI réduite

- Randomisation des nucléotides de chaque gène.

Les gènes réels sont en moyenne significativement plus stables que les gènes simulés du point de vue de la mutabilité transcriptionnelle



P-value (Wilcoxon paired, bilatéral) < 2,2 x 10^{-16} ; N = 4307

Hoede *et al.* 2006 PLoS Genet.

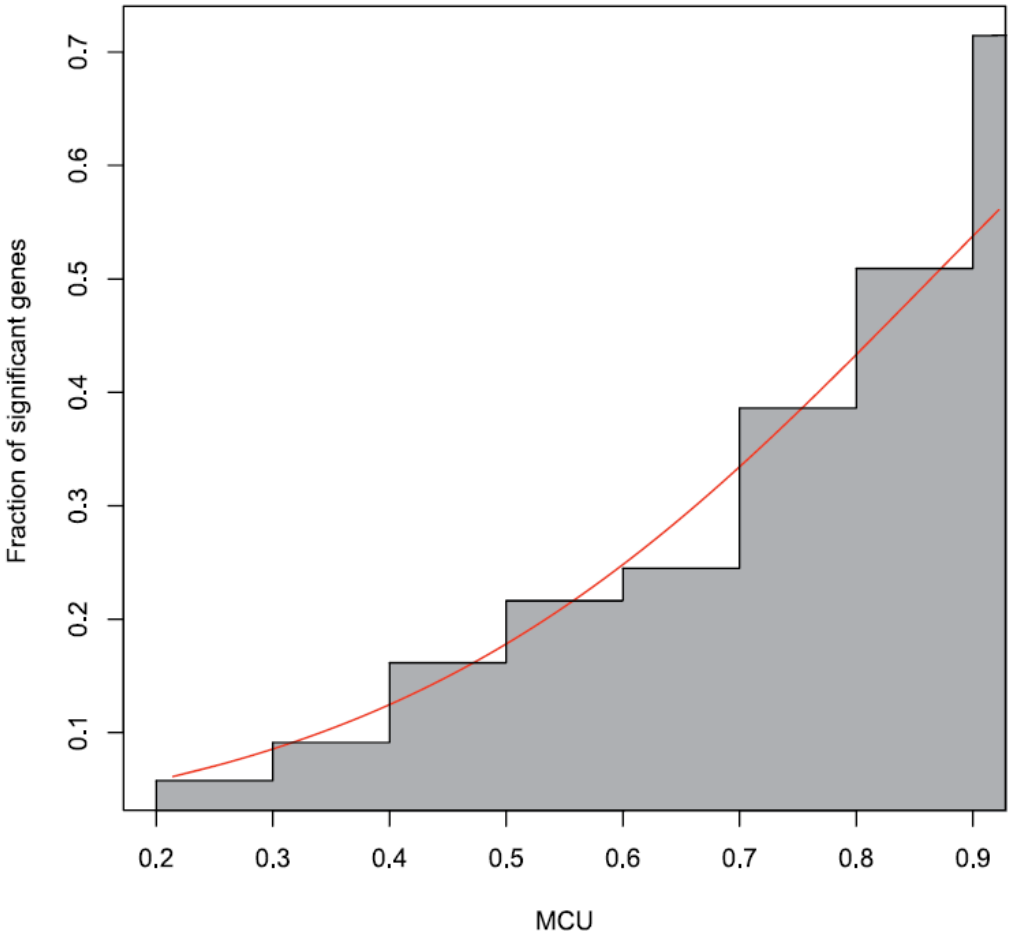
La fonction du gène impose une contrainte

- Randomisation en conservant la séquence codée et le biais de codons du gène.

	M	A	T	A	T	…
Séquence réelle	ATG	GCC	TAT	GCG	TAC	…
	M	**A**	**T**	**A**	**T**	**…**
Séquence randomisée	ATG	GCG	TAC	GCC	TAT	…

Hoede *et al.*
2006
PLoS Genet.

Plus de gènes robustes parmi les gènes à fort MCU



Fraction de gènes significativement robustes en fonction du MCU : en rouge la régression logistique en gris l'histogramme des valeurs réelles par classe d'intervalle 0,1

Hoede *et al.* 2006 PLoS Genet.

- Contrôle local et inductible du taux de mutation.

- La sélection naturelle semble favoriser l’existence de structure secondaire de l’ADN limitant la mutabilité transcriptionnelle.

LA RECOMBINAISON

- Secondaire à l'acquisition d'ADN par transfert horizontal.
- Recombinaison
  - –site spécifique
  - –illégitime
  - –homologue
- Recombinaison ➔ Hétérogénéisation des séquences?

La recombinaison site spécifique

- Echange entre des sites bien définis.
- Phage λ s'intégrant au locus *attB.*
- De l'ADN phagique est intégré au chromosome bactérien ➔ hétérogénéisation de la séquence.

La recombinaison illégitime

- Peut s'effectuer entre de courtes séquences homologues (<20 pbs).
- Ou sans aucune homologie.
- Lorsqu'elle implique de l'ADN étranger, il s'agit généralement d'ADN provenant d'une autre espèce ➔ hétérogénéisation de la séquence.

La recombinaison homologue

- Nécessite une homologie de séquence.
- Plusieurs mécanismes, plusieurs modèles.
- Un de ses produits est la conversion génique : transfert non réciproque de courts fragments d’ADN.

Homogénéisation ou hétérogénéisation ?

- La recombinaison homologue implique le plus souvent un fragment d'ADN intra-espèce.
  - A l'échelle du clone : hétérogénéisation de la séquence.
  - A l'échelle de la population ou de l'espèce : homogénéisation de la séquence.

La conversion génique

- En génétique des populations, désigne un ensemble de substitutions localisées qui aurait été acquis par transfert horizontal. Quelque soit le mécanisme sous-jacent.
- Proviendrait généralement de double crossing-over.

Hypothèse : la recombinaison homologue chez *E. coli* s'effectue essentiellement sous forme de conversion génique.

Quel est alors le taux de recombinaison dans le génome de *E. coli* ?

Un taux de recombinaison homologue très important

- Taux de mutation (θ) : 0,014.
- Ratio conversion génique / mutation : 2,47.
- Longueur des fragments : 50 pbs.
- Une base a une probabilité 100 fois supérieure d'être impliquée dans une conversion génique que de subir une mutation.

Touchon *et al.*
2009
PLoS Genet.

LA PHYLOGÉNIE

-Un tel taux de recombinaison est-il compatible avec l'établissement d'un arbre phylogénétique ?

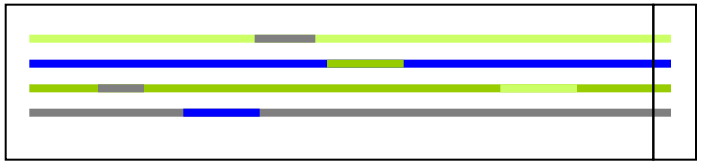
- Débat historique : *E. coli* définie comme clonale et recombinante à la fois.

- Approche de génomique comparative : 20 génomes complets disponibles provenant de souches représentatives de la diversité de l’espèce.
- Dont 7 que le consortium dans lequel nous étions impliqués a séquencés et annotés.

Touchon *et al.*
2009
PLoS Genet.

Simulations de divers taux de conversion génique

- 20 séquences de 25 kpbs (100x) longueur moyenne des fragments fixe (50 pbs) et taux de mutation fixe (0,014).



Arbre de coalescent **ms**

Évolution simulée de la séquence
Seq-gen

Arbre phylo
PhyML

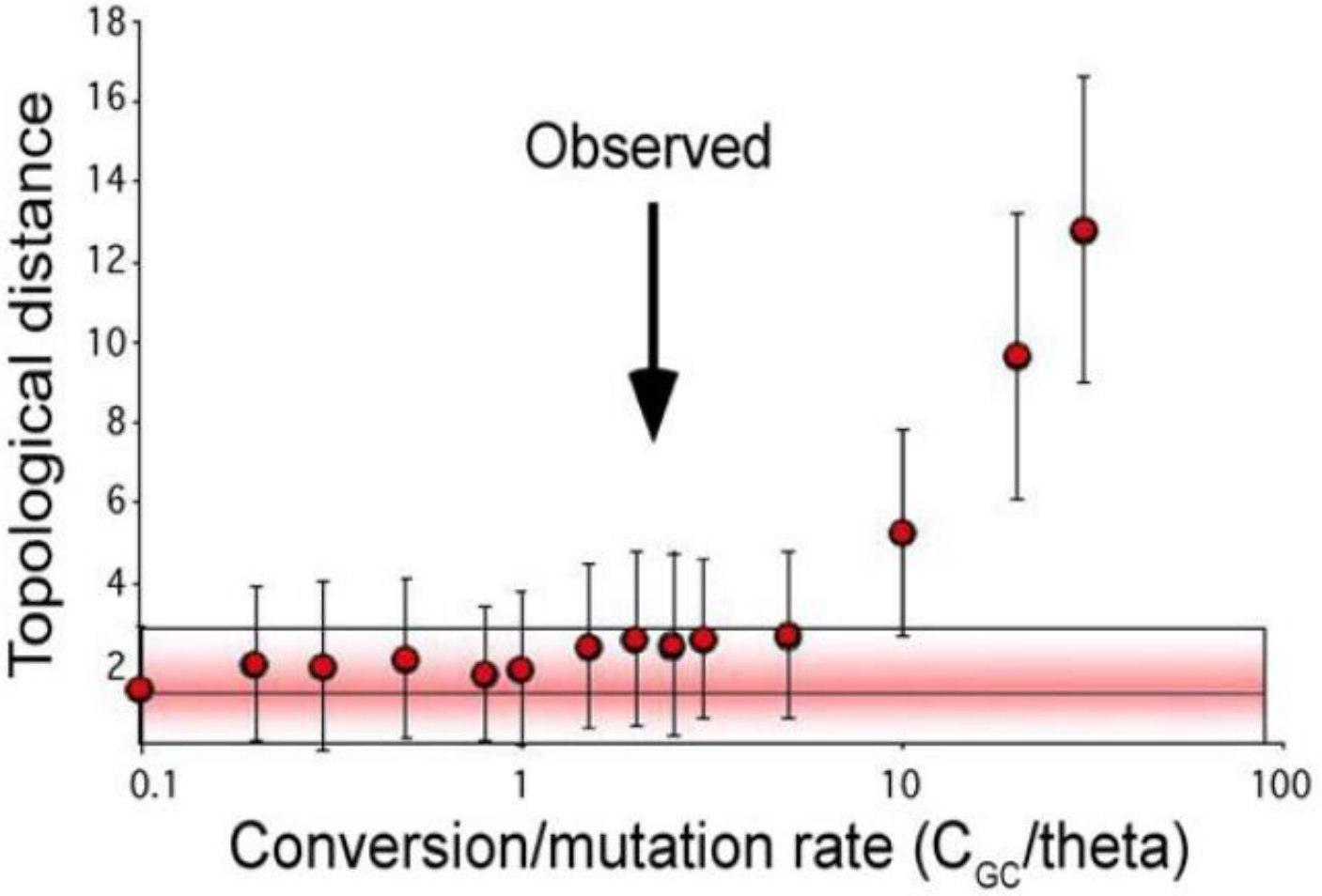
Pas de conversion génique sur le dernier nucléotide

Core genome
Arbre phylo

Test
R & F

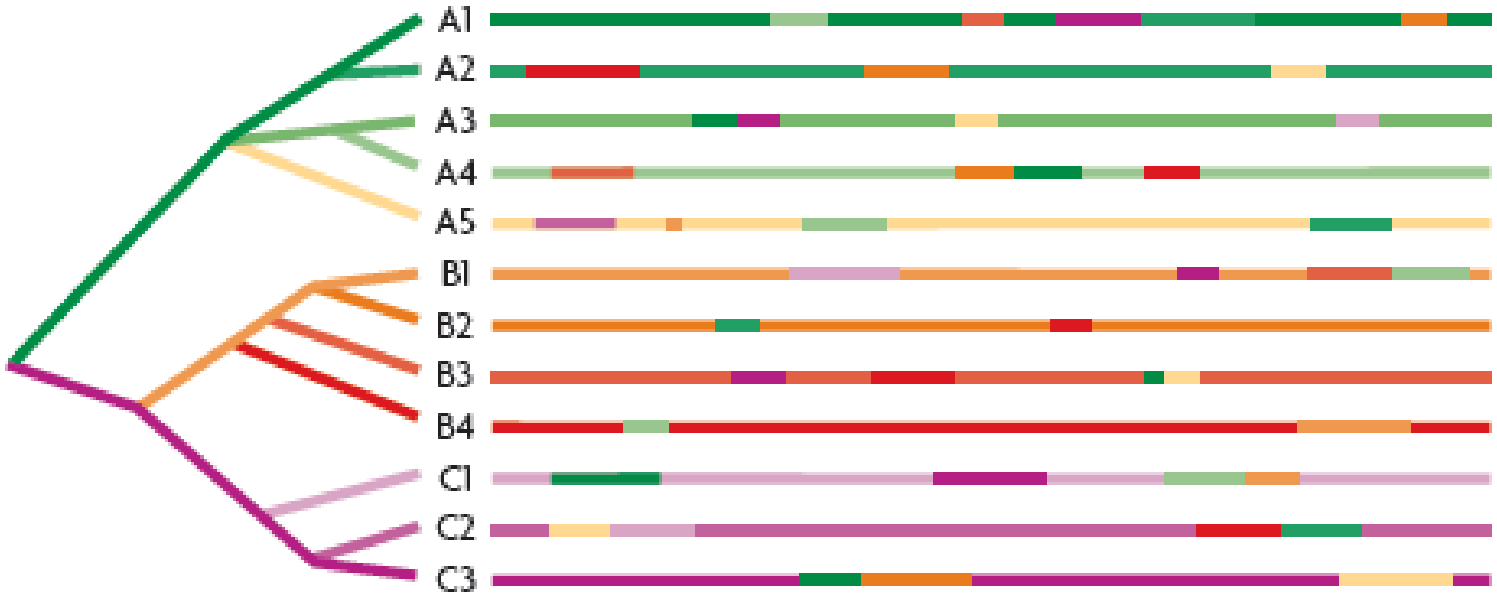
Touchon *et al.*
2009
PLoS Genet.

Le signal phylogénétique persiste



Touchon *et al.*
2009
PLoS Genet.

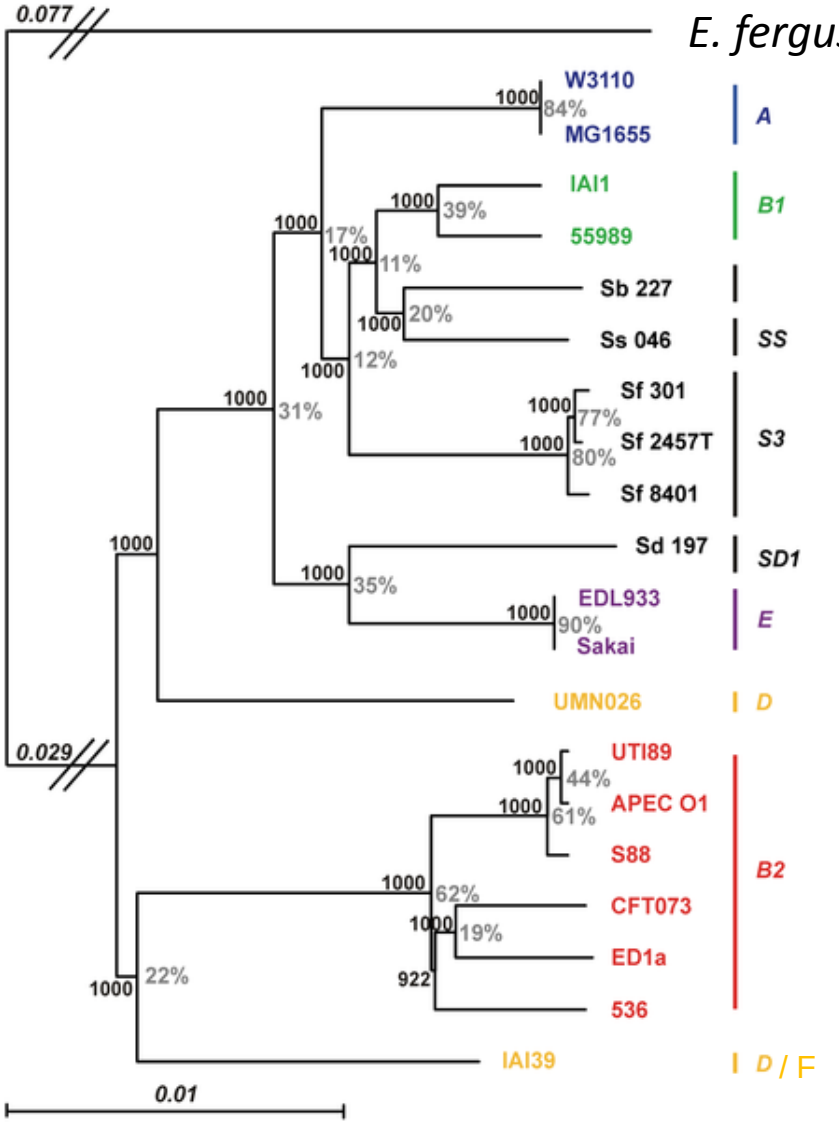
- La petite taille des fragments impliqués dans la recombinaison n'est pas suffisante pour obscurcir le signal phylogénétique global.



Touchon *et al.* 2009 PLoS Genet.

Tenaillon *et al*. 2010

Phylogénie de l'espèce *E. coli*

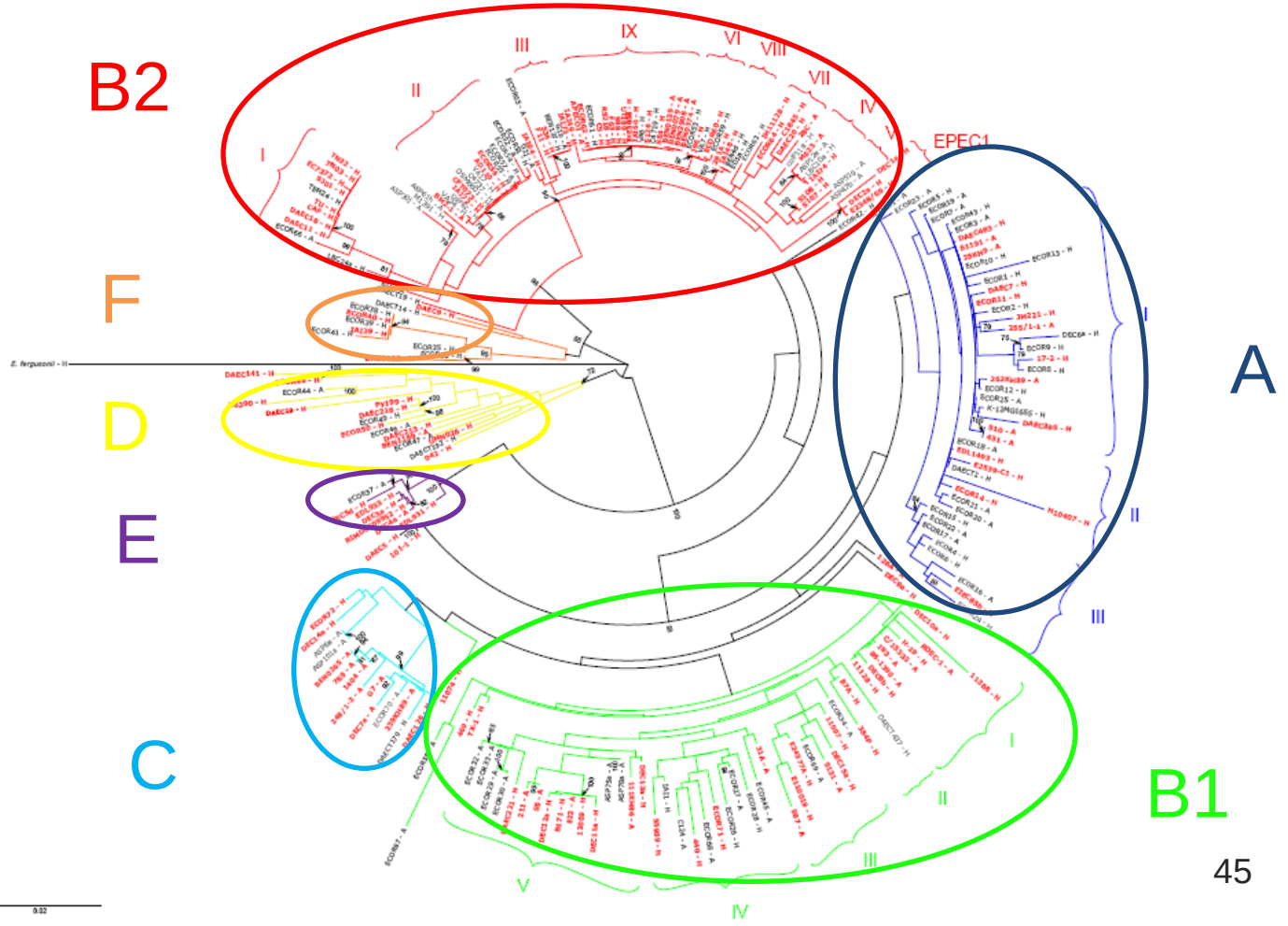


20 souches
Concaténât des gènes ayant un orthologue dans toutes les souches

Touchon *et al.* 2009 PLoS Genet.

234 souches diverses étudiées

- MLST (8 gènes) ➔ arbre similaire.



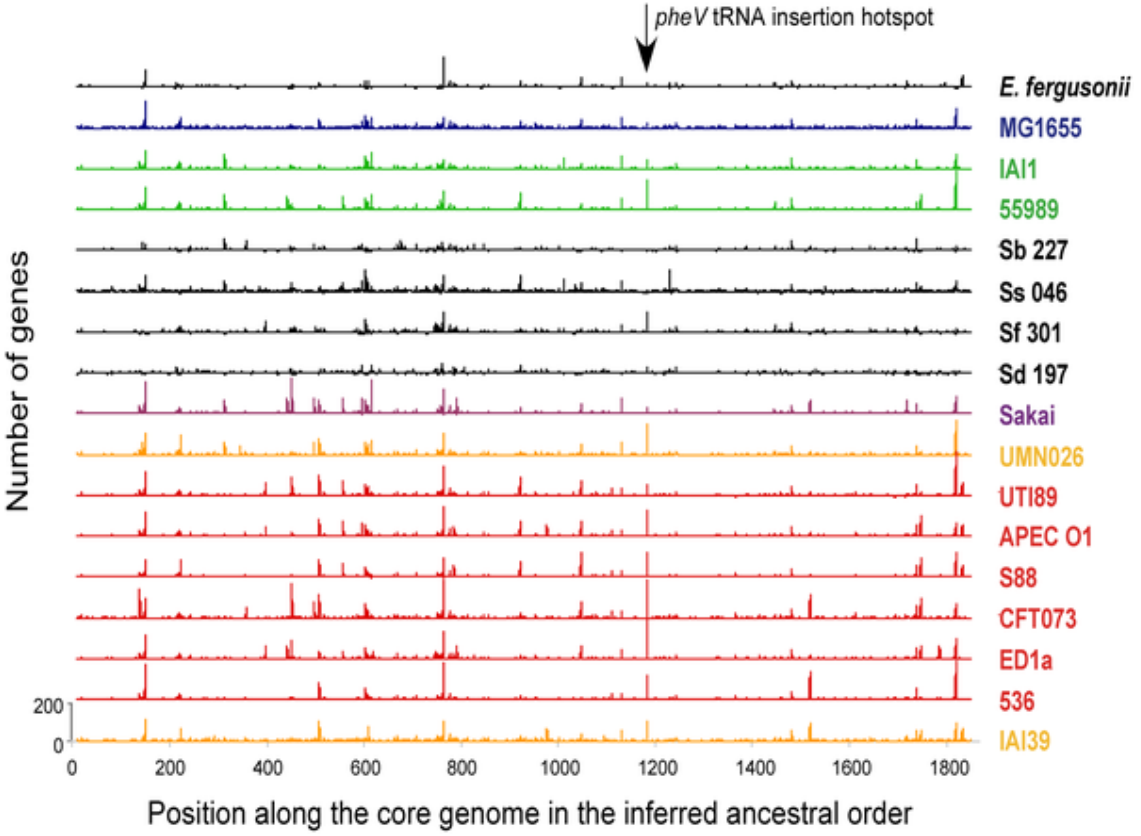
Clermont *et al.* 2011 Appl Environ Microbiol.

L’ORGANISATION DU GÉNOME

- Nous avons mis en évidence un très grand nombre de gènes variables (pas présents dans toutes les souches).
- Certains processus cellulaires nécessitent que le chromosome bactérien soit organisé.
- Comment cet important flux de gènes influe-t-il sur l’organisation du génome ?

Un génome organisé malgré un flux de gènes important

- 133 loci contiennent 71% des gènes variables.

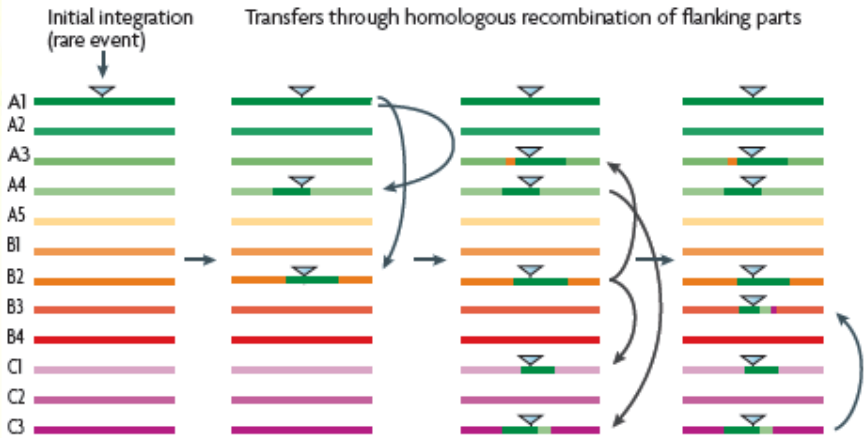


Touchon *et al.* 2009 PLoS Genet.

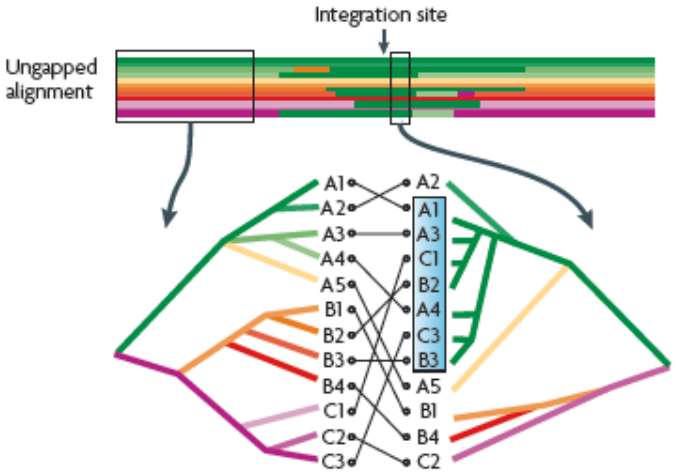
Neutralité des insertions multiples ?

- Si un premier fragment est intégré dans une région permissive, les nouvelles insertions au même endroit seraient alors neutres.

Dissémination des fragments insérés?



Dissémination du fragment intégré s'il procure un avantage sélectif par recombinaison homologue des séquences flanquantes

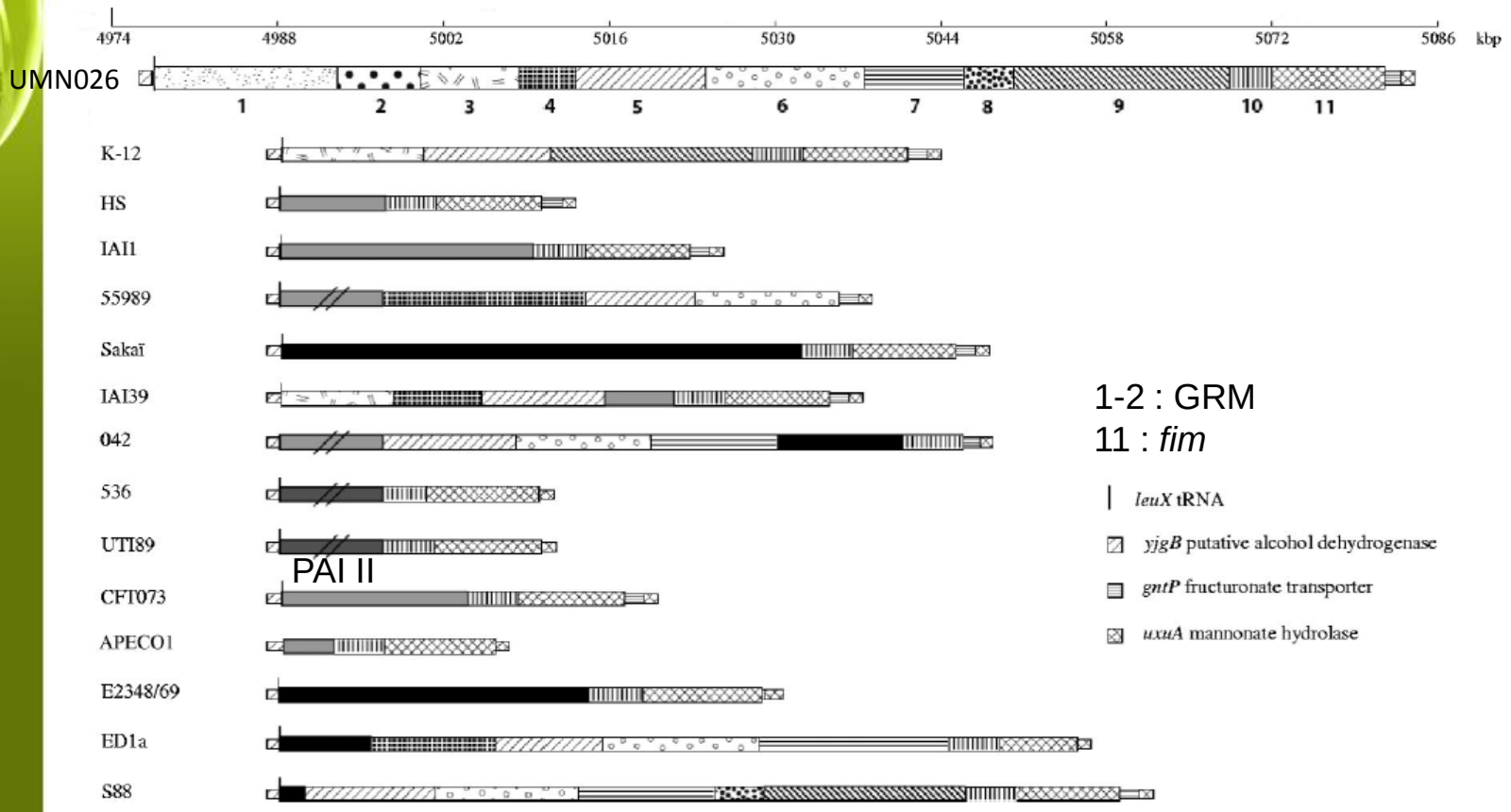


Il est alors observé un point chaud d'intégration et une incongruence phylogénétique entre l'arbre global et la phylogénie dérivée des séquences flanquant le site d'intégration (15 cas observés)

Touchon *et al.* 2009 PLoS Genet.

Tenaillon *et al.* 2010

Détails d'un point chaud d'insertion/délétion



Lescat *et al.* 2009 Antimicrob Agents Chemother.

CONCLUSIONS

- Apport de l'approche génomique
- La mutagénèse transcriptionnelle semble être un des facteurs sous sélection dans la séquence.
- Malgré un taux important de recombinaison, la population est plutôt clonale et la phylogénie robuste.
- Le génome d'*E. coli* est un désordre organisé.

PERSPECTIVES

- Groupes phylogénétiques affinés permettant une épidémiologie plus fine.
- Mettre en évidence des gènes présentant des profils atypiques (selon les groupes …)
- La multiplication des génomes disponibles permettra une meilleure estimation de la recombinaison et des processus mutationnels.

REMERCIEMENTS

- Je remercie les membres du jury.

- Merci à Olivier Tenaillon, Erick Denamur, Mathilde Lescat, Bertrand Picard.

- Et à tous les membres de l’équipe U722.

- Et aux collaborateurs : Marie Touchon, Eduardo Rocha, Allan Hance, Marie-Agnès Petit et Hélène Chiapello.