



**HAL**  
open science

# Recombinaison et évolution de la composition en base : démonstration expérimentale de la conversion génique biaisée vers GC chez les Bactéries

Samuel Barreto

► **To cite this version:**

Samuel Barreto. Recombinaison et évolution de la composition en base : démonstration expérimentale de la conversion génique biaisée vers GC chez les Bactéries. Bactériologie. Université de Lyon, 2020. Français. NNT : 2020LYSE1004 . tel-04531148

**HAL Id: tel-04531148**

**<https://theses.hal.science/tel-04531148>**

Submitted on 3 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THESE de DOCTORAT DE L'UNIVERSITE DE LYON**  
opérée au sein de  
**l'Université Claude Bernard Lyon 1**

**Ecole Doctorale N° 341**  
**Évolution, Écosystèmes, Microbiologie, Modélisation**

**Spécialité de doctorat :**  
**Discipline :** Biologie

Soutenue publiquement le 15/01/2020, par:  
**Samuel Barreto**

---

**RECOMBINAISON ET ÉVOLUTION DE LA  
COMPOSITION EN BASE:  
Démonstration expérimentale de la  
conversion génique biaisée vers GC  
chez les Bactéries**

---

Devant le jury composé de :

MOUCHIROUD Dominique, Professeure, Présidente  
Université Lyon 1

ELEZ Marina, Maîtresse de conférence, Rapportrice  
Sorbonne Université

ROCHA Eduardo, Directeur de Recherche, Rapporteur  
CNRS / Institut Pasteur

GUY Lionel, Chercheur, Rapporteur  
Université d'Uppsala

DURET Laurent, Directeur de Recherche, Directeur de thèse  
CNRS / Université Lyon 1

BERTOLLA Franck, Maître de Conférence, Co-directeur de thèse  
Université Lyon 1

UNIVERSITÉ CLAUDE BERNARD LYON 1

DIPLÔME DE DOCTORAT

**RECOMBINAISON ET  
ÉVOLUTION DE LA  
COMPOSITION EN BASE**

Démonstration expérimentale de la conversion  
génique biaisée vers GC chez les Bactéries

*Présentée par*

**Samuel BARRETO**

*Devant le jury composé de :*

<b>Laurent DURET</b>	Directeur de Thèse
<b>Franck BERTOLLA</b>	Co-directeur
<b>Marina ELEZ</b>	Rapportrice
<b>Lionel GUY</b>	Rapporteur
<b>Eduardo C. ROCHA</b>	Rapporteur
<b>Dominique MOUCHIROUD</b>	Examinatrice

*Pour l'obtention du diplôme de docteur en biologie*

15 janvier 2020

---

### **Université Claude Bernard - LYON 1**

Président de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	
Directeur Général des Services	M. Damien VERHAEGHE

### **COMPOSANTES SANTE**

Faculté de Médecine Lyon-Est – Claude Bernard	Doyen : M. Gilles RODE
Faculté de Médecine et Maïeutique Lyon Sud Charles. Mérieux	Doyenne : Mme Carole BURILLON
UFR d'Odontologie	Doyenne : Mme Dominique SEUX
Institut des Sciences Pharmaceutiques et Biologiques	Directrice : Mme Christine VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. Xavier PERROT
Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT

### **COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE**

UFR Biosciences	Directrice : Mme Kathrin GIESELER
Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
UFR - Faculté des Sciences	Administrateur provisoire : M. Bruno ANDRIOLETTI
UFR (STAPS)	Directeur : M. Yannick VANPOULLE
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Ecole Polytechnique Universitaire Lyon 1	Directeur : Emmanuel PERRIN
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Institut de Science Financière et d'Assurances	Directeur : M. Nicolas LEBOISNE
ESPE	Administrateur Provisoire : M. Pierre CHAREYRON



---

Cette thèse a été réalisée aux laboratoires de :

Biométrie et Biologie Évolutive UMR CNRS 5558  
Écologie Microbienne UMR CNRS 5557  
UMR INRA 1418

UCB Lyon 1 - Bât. Grégor Mendel  
Université Claude Bernard Lyon I  
43, Boulevard du 11 Novembre 1918  
69622 VILLEURBANNE CEDEX



## RÉSUMÉ

---

La composition en base des génomes varie très fortement entre espèce bactériennes, avec une gamme de taux de G+C génomique s'étalant de 13% jusqu'à 75%. Ce phénomène est connu depuis plus de 50 ans, mais sa signification biologique reste fortement débattue. Des expériences d'accumulation de mutations ont permis de démontrer que ces variations de G+C sont principalement dues à des différences de patron mutationnel. Cependant, ces expériences ont également montré que la grande majorité des génomes ont un taux de G+C supérieur à celui attendu à l'équilibre mutationnel. Deux forces évolutives peuvent expliquer cet écart : la sélection naturelle et la conversion génique biaisée vers GC (gBGC) – processus non-adaptatif lié à des biais de réparation des mésappariements au cours de la recombinaison homologue. Le phénomène de gBGC a été bien caractérisé chez plusieurs eucaryotes, et plusieurs indices suggèrent qu'il pourrait également affecter des bactéries.

Mon travail de thèse a été consacré à déterminer empiriquement si la recombinaison homologue provoque de la gBGC chez des bactéries. Pour cela nous avons réalisé des expériences de transformation chez *Acinetobacter baylyi* et *Bacillus subtilis* puis analysé les produits de recombinaison. Le séquençage des génomes de clones transformants a permis de quantifier les fréquences de conversion parmi 2023 régions recombinantes chez *A. baylyi*, et 549 régions recombinantes chez *B. subtilis*. Nous avons également ré-analysé les données issues d'expériences similaires publiées pour deux autres espèces : *Helicobacter pylori*, *Streptococcus pneumoniae*. Dans trois des quatre espèces analysées, nous avons observé un biais de réparation en faveur des bases GC, ce qui constitue une preuve directe de l'existence de la gBGC chez ces bactéries. L'intensité du biais que nous avons mesuré est compatible avec celui nécessaire pour expliquer l'écart entre la composition en base observée et celle attendue à l'équilibre mutationnel. Nos observations démontrent que ce processus non-adaptatif peut affecter fortement l'évolution des génomes de bactéries, qui, en raison des grandes tailles de population supposées, sont souvent perçus comme déterminés quasi-exclusivement par la sélection naturelle.

De façon intéressante, nous avons constaté que la gBGC se manifeste différemment selon les contextes génomiques et selon les espèces : chez *A. baylyi*, nous avons observé un fort gBGC spécifiquement dans les tracts de conversion impliquant un seul SNP — ce qui suggère un rôle possible de la machinerie BER (Base Excision Repair), comme cela a également été proposé chez les mammifères. A l'inverse, chez *H. pylori*,

---

et *S. pneumoniae*, la gBGC se manifeste uniquement aux extrémités de longs tracts de conversion — tout comme cela a été décrit chez la Levure. Cela suggère que différents mécanismes moléculaires sont à l'origine de la gBGC observée dans différentes lignées. Cette évolution convergente soulève la question des contraintes sélectives susceptibles de gouverner l'évolution de la gBGC.

Ce manuscrit sera consacré à la description de la question des origines de la variabilité en composition en base connue, à la question de la recombinaison chez les procaryotes, et à la façon dont les deux processus conjoints permettent d'apporter du poids à l'hypothèse de la conversion génique biaisée.

## REMERCIEMENTS

---

Après maintenant bientôt quatre ans passés dans deux laboratoires, il n'est que justice qu'il faille remercier une bien longue liste de personnes. Il m'a toujours semblé parfaitement contre-productif que la structuration en laboratoires contribue à segmenter artificiellement l'activité scientifique. Par conséquent la logique voudrait que je ne distingue pas, dans ce qui suit, celles et ceux que j'estime devoir remercier en fonction de leur laboratoire d'appartenance ; elles sauront s'y retrouver.

Je voudrais remercier :

- Toutes les personnes qui contribuent à ce que la vie d'un laboratoire soit possible, tout simplement, à commencer par celles de l'administration, de l'entretien, de l'hygiène et la sécurité, de l'informatique ou de l'aide technique en général ; et en particulier Danis Abrouk, Nathalie Arbasetti, Dominique Boulanger, Laetitia Catouaria, David Chapulliot, Audrey Dubost, Vincent Gaillard, Lionel Humblot, Simon Penel, Bruno Spataro.
- Toutes celles sans le travail desquelles le nôtre ne serait pas non plus tel qu'il est. Je pense à mes différents chefs ou cheffes d'équipe, notamment Laurent Duret, Céline Lavire, Ludovic Vial et Xavier Nesme. Je pense aussi à mes deux directeurs de laboratoire, Yvan Moëgne-Locez et Manolo Gouy.
- Les deux personnes sans lesquelles le travail de laboratoire n'aurait pas été possible non plus, notamment Cindy Garnier et plus encore Laurène Mirabel. J'espère que tu ne garderas pas à vie le souvenir cuisant de l'indocilité insolente avec laquelle les Archées se sont manifestées à toi. Je te suis sincèrement reconnaissant de toutes ces heures que tu as passées sous la hotte, en ma compagnie ou non. Certains voudraient, sinon te faire taire, au moins t'entendre moins ; pour ma part j'ai toujours su que les manip's à tes côtés se passeraient dans la bonne humeur. Je te fais confiance pour que jamais la corrosion ambiante ne vienne entamer ta joie de vivre.
- Tous les étudiants et les étudiantes ayant subit les enseignements que l'on m'a confiés, en particulier les enseignements d'écologie microbienne et de bioinformatique — ils et elles étaient pour moi la preuve qu'il est *possible* que l'enseignement soit émancipateur, et ont grandement contribué à ce que je n'ai plus la moindre espèce de considération pour les discours quasi-réactionnaires sur un quelconque déclin de la conscience scientifique ou politique des « étudiants de nos jours ».
- Toutes les étudiantes et tous les étudiants, en thèse ou non, avec

---

lesquels j'ai passé près de quatre ans mémorables ; en particulier Alexandre, Alexia, Aurélie, Benjamin, Djivan, Élodie, Guillaume, Julien, Jordan, Laura, Lucas, Marine, Maud, Morgane, Quentin, Rosa, Solène, Thibault L., Thibault M., Théo, Xavier, Yoann, William. (Une mention spéciale doit être faite aux compères d'SMBE 2019, sans lesquels la bière locale n'aurait pas eu tout à fait la même saveur.)

- Toutes les personnes sans lesquelles je n'aurais probablement pas songé à une carrière scientifique, notamment Séverine Allégra, Florence Girardot, et plus encore Laurence Mouton et Fabrice Vavre.
- Toutes les personnes qui se sont intéressées de plus ou moins près à mes travaux ou à ce que j'avais à en dire : notamment Nicolas Lartillot ; Daniel Kahn ; les membres des clubs [RS]AGE ; les membres de l'ANR DaSiRe ; les membres de mon comité de pilotage, Tristan Lefébure, Sophie Abby, Ivan Matic, avec une mention particulière pour Xavier Charpentier, qui tous ont largement contribué à ce que je manque un peu moins de confiance en mes travaux.
- Ce travail de thèse se serait probablement déroulé très différemment s'il ne m'avait pas fourni l'occasion de rencontrer deux très belles personnes : celle qui n'a été pour moi que certitudes quand tout le reste n'était que doutes, et ce petit cristal d'avenir, ce petit bout d'homme qui a tout rendu tellement limpide.
- Bien évidemment, je remercie vraiment très sincèrement les trois personnes qui m'ont supervisé pendant ces trois ans. Je commencerai par Vincent, dont l'activité cérébrale irrémédiablement ralentie a probablement contribué grandement à ce qu'il ne sache pas déjà tout le mal que je pense de lui. J'espère bien que j'aurai l'occasion de le lui faire comprendre au cours des trois ans qui vont suivre.

Merci ensuite à Franck, avec lequel je n'ai jamais eu besoin de lutter pour me sentir considéré d'égal à égal ; pas un moment je n'ai senti que je n'avais pas sa confiance pour conduire le projet de la façon dont je le jugeais bon. J'ai déjà vu ce genre de remerciements masquer en fait une forme de ressentiment pour l'absence d'encadrement ; ça n'a jamais été mon cas, et je suis quasiment certain que toute autre forme d'encadrement aurait été contre-productive au bout du compte. Je lui serai toujours reconnaissant de m'avoir laissé l'opportunité d'avoir raison parfois, et tort souvent.

---

Merci enfin à toi Laurent. J'ai mis longtemps à réfléchir à la façon la plus efficace de te remercier, et avais considéré un temps de recourir à des extraits du Monsieur Teste de Valéry. Sauf que je ne suis déjà plus tout à fait sûr que ce manuscrit passe le crible de l'anti-plagiat au vu du nombre de citations dont je n'ai pas réussi à me passer ; et que je ne suis pas certain qu'il soit tout à fait aussi flatteur de te comparer à M. Teste que je le voudrais. Bref, il reste que je garderai à jamais l'empreinte de la façon dont tu as élevé la rigueur et l'ascétisme de la lucidité à une forme d'idéal de conduite, si bien qu'il suffit de t'avoir fréquenté suffisamment pour connaître le mieux quelles vertus éthiques la science correctement menée peut engendrer. Pas de meilleure défense pour moi contre les relativistes qui poussent des hauts cris lorsqu'on invoque des valeurs comme la rigueur et la vérité, que de te montrer comme exemple du genre de personne qui fleurirait dans une démocratie pensée comme un espace de raison. « Vous les reconnaîtrez à leurs fruits », lit-on quelque part. Que le champ scientifique veuille bien de moi ou non, l'avenir le dira. Quoi qu'il en soit, tu auras tellement contribué à façonner ma vision du monde qu'il m'arrive parfois dans le même temps de me demander si je pense ce que je pense parce que tu l'as déjà pensé tellement fort, et de douter d'avoir raison de penser ce que je pense parce que tu l'as déjà pensé. En tout les cas, je n'ai jamais été aussi heureux qu'à cette limite quasi-schizophrénique où conduit la conduite du doute que tu mènes véritablement à la perfection.

Dédiée à mon frère Miguel,  
à ma sœur Lisa,  
et à toutes celles et ceux qui  
*cherchent.*



## SOMMAIRE

---

<b>I</b>	<b>Introduction</b>	<b>1</b>
<b>1</b>	<b>Prologue</b>	<b>3</b>
<b>2</b>	<b>Aspects historiques du GC%</b>	<b>9</b>
2.1	Découvertes des acides nucléiques . . . . .	11
2.2	Expliquer la variabilité du GC%? . . . . .	20
<b>3</b>	<b>Distribution des taux de GC entre génomes</b>	<b>27</b>
3.1	Diversité inter-génomique . . . . .	27
3.2	Diversité intra-génomique . . . . .	29
<b>4</b>	<b>Processus mutationnels</b>	<b>35</b>
4.1	Balance mutationnelle . . . . .	36
4.2	Déterminants de la balance mutationnelle . . . . .	42
4.3	Distribution phylogénétique des gènes impliqués . . . . .	58
<b>5</b>	<b>Sélection sur la composition en base (?)</b>	<b>63</b>
5.1	Corrélations GC% et traits environnementaux . . . . .	63
5.2	Indices expérimentaux de sélection sur le GC% . . . . .	70
<b>6</b>	<b>Conversion Génique Biaisée : quelle importance?</b>	<b>73</b>
6.1	Sélection, dérive & recombinaison . . . . .	76
6.2	Indices de la gBGC chez les Procaryotes . . . . .	86
<b>II</b>	<b>Quantification de la gBGC chez les Bactéries</b>	<b>95</b>
<b>7</b>	<b>Mesure des fréquences de conversion</b>	<b>97</b>
7.1	Comment faire? . . . . .	97
7.2	Transformation génomique d' <i>Acinetobacter baylyi</i> . . . . .	99
7.3	Transformation génomique de <i>Bacillus subtilis</i> . . . . .	105
7.4	Mesure des fréquences de conversion . . . . .	109
7.5	Ré-analyse des données publiques . . . . .	113
<b>8</b>	<b>Démonstration de la gBGC chez des Bactéries</b>	<b>121</b>
8.1	Répartition des régions converties sur le génome . . . . .	123
8.2	<i>A. baylyi</i> : Preuves de gBGC dans les conversions isolées .	128
8.3	Preuves de gBGC dans deux autres espèces de Bactéries .	133
8.4	Intensité populationnelle de la gBGC . . . . .	135

---

<b>9 Discussions</b>	<b>143</b>
9.1 Est-ce que la sélection <i>cause</i> le biais en faveur de GC? . .	143
9.2 Les mécanismes responsables de la gBGC . . . . .	145
9.3 Détecter la gBGC : une tâche délicate . . . . .	154
9.4 La nature convergente de la gBGC . . . . .	155
<b>Perspectives et conclusions générales</b>	<b>159</b>
<b>Appendices</b>	<b>I</b>
<b>Taux de croissance et taux de GC%</b>	<b>III</b>
<b>Bibliographie</b>	<b>XI</b>
<b>Table des matières</b>	<b>XXXI</b>

Je vis également Sisyphe, en proie à ses tourments.  
Il soulevait de ses deux bras un rocher gigantesque ;  
Arc-bouté des pieds et des mains, il poussait ce grand bloc  
Vers le sommet d'une hauteur ; mais à peine allait-il  
Le faire basculer, qu'il retombait de tout son poids,  
Et le bloc sans pitié roulait de nouveau vers la plaine.  
Mais lui recommençait, bandant ses muscles ; la sueur  
Ruisselait sur son corps, et la poussière le nimbait.

Homère, *L'Odyssée* (rhapsodie XI),  
traduit par Frédéric Mugler, Arles France : Actes Sud, 1999.



# **Première partie**

## **Introduction**



## PROLOGUE

---

Dans le monde inorganique elle semble faire défaut, dans le monde organique commence la ruse [...].

Friedrich Nietzsche, *La Volonté de Puissance*, §262

Règle première : *Il ne faut admettre de causes, que celles qui sont nécessaires pour expliquer les Phénomènes.*

La nature ne fait rien en vain, & ce seroit faire des choses inutiles que d'opérer par un plus grand nombre de causes ce qui peut se faire par un plus petit.

Isaac Newton, *Principes mathématiques de la philosophie naturelle*, tome 2, « Règles qu'il faut suivre dans l'étude de la physique. », traduction du latin par Mme du Chastelêt.

## De quoi la composition en base est-elle le signe ?

Quel peut bien être l'intérêt porté à la composition en base ? À quoi sert-il de s'y intéresser ? En quoi l'étude de la composition en base permet-elle de rendre le vivant plus intelligible ? Le permet-elle d'ailleurs ?

En effet, comme le dit PEIRCE [1992a, pages 322–324], « à quoi sert de tenter de former une conception de l'univers sinon de rendre les choses intelligibles ? »

Historiquement, on peut définir trois courants qui se sont intéressés à la composition en base. Le premier part de la découverte des acides nucléiques à la confirmation qu'ils constituent l'un des matériaux essentiels de l'hérédité, et au premier chef l'ADN. Le fait que la composition en base des acides nucléiques varie d'une espèce à l'autre a largement contribué à cette confirmation. Le second courant a vu dans la composition en base un caractère pratique permettant de discriminer empiriquement les espèces les unes des autres, du fait de sa variabilité entre espèces et de sa stabilité au sein d'une espèce, et par conséquent d'améliorer la résolution de la taxonomie des microorganismes. Le dernier courant s'inscrit dans un débat relativement ancien entre ceux que l'on en est venu à appeler « neutralistes » et « sélectionnistes ». C'est dans ce débat que ce travail de thèse s'inscrit également.

## 1. PROLOGUE

---

Depuis Darwin, et probablement longtemps avant lui, les biologistes cherchent à comprendre ce qui, dans le vivant, résulte de pures contingences historiques et ce qui contribue à l'*adaptation* d'un organisme à son milieu. Pour cela, certains d'entre eux s'intéressent à l'évolution des organismes, probablement parce qu'en biologie *rien* — ou en tout cas *pas grand chose* — n'a de sens sauf à la lumière de l'évolution. Le génome des organismes est l'un des composants majeurs de l'hérédité, c'est-à-dire de ce qui est transmis d'un individu biologique à sa descendance, et de ce que cet individu hérite de son ascendance. Par conséquent les génomes portent des marques de l'évolution des organismes; aussi un moyen fécond pour comprendre cette évolution consiste à comparer leur génome.

Or, jusqu'à preuve du contraire, le génome de tous les êtres vivants que l'on connaisse est composé d'acides nucléiques — souvent ADN, parfois ARN pour certains virus —, eux-mêmes composés de cinq bases nucléotidiques : adénines et guanines (bases puriques), et cytosine, thymine, ou uracile pour les ARN, (bases pyrimidiques). L'une des avancées majeures réalisées notamment par Erwin Chargaff a permis de comprendre qu'au sein d'une molécule d'ADN, la molarité de cytosine et de guanine d'une part, d'adénine et de thymine d'autre part, sont respectivement identiques, à de rares exceptions près. Pour *résumer* la composition en base d'un ADN de taille  $L$ , on peut donc la réduire à son taux de guanine et cytosine ou taux de GC, noté  $GC$ , tel que

$$GC = \frac{n_G + n_C}{n_A + n_T + n_G + n_C} = \frac{n_G + n_C}{L}$$

Ce qui correspond à la proportion de bases G et C dans un génome de taille  $L = n_A + n_T + n_G + n_C$ .

Cette proportion de G et C est assimilable à une fonction d'état, qui dépend des variables d'état que sont les probabilités de *substitutions* de A ou T vers G ou C, ou de G ou C vers A ou T. L'on notera ces deux variables d'états  $v$  et  $u$ , telles que

$$G : C \xrightleftharpoons[v]{u} A : T, \text{ où } \begin{cases} u = \Pr(G : C \rightarrow A : T) \\ v = \Pr(A : T \rightarrow G : C) \end{cases}$$



Ces deux variables d'états  $u$  et  $v$  résument elles-mêmes les probabilités des différentes substitutions représentées par :

$$\begin{bmatrix} & A & T & C & G \\ A & \ddots & \bullet & \bullet & \bullet \\ T & \bullet & \ddots & \bullet & \bullet \\ C & \bullet & \bullet & \ddots & \bullet \\ G & \bullet & \bullet & \bullet & \ddots \end{bmatrix}$$

Les variables  $u$  et  $v$  sont des variables d'états parce que l'état de la composition en base à l'équilibre ne dépend que d'elles. On peut ainsi définir la composition en base à l'équilibre par

$$GC_{eq} = \frac{v}{v + u}$$

Or les probabilités de *substitutions* de  $AT \rightarrow GC$  et de  $GC \rightarrow AT$  sont affectées, d'une part par les probabilités de *mutation*, notées  $\mu$ , de  $AT \rightarrow GC$  et de  $GC \rightarrow AT$ , et d'autre part par les probabilités que ces mutations ont d'être *fixées* dans la population, notées  $\phi$ .

$$\begin{aligned} u &= \mu_{GC \rightarrow AT} \cdot \phi_{GC \rightarrow AT} \\ v &= \mu_{AT \rightarrow GC} \cdot \phi_{AT \rightarrow GC} \end{aligned}$$

De sorte que la composition en base à l'équilibre peut s'écrire

$$GC_{eq} = \frac{v}{v + u} \quad (1.1)$$

$$= \frac{\mu_{AT \rightarrow GC} \cdot \phi_{AT \rightarrow GC}}{\mu_{AT \rightarrow GC} \cdot \phi_{AT \rightarrow GC} + \mu_{GC \rightarrow AT} \cdot \phi_{GC \rightarrow AT}} \quad (1.2)$$

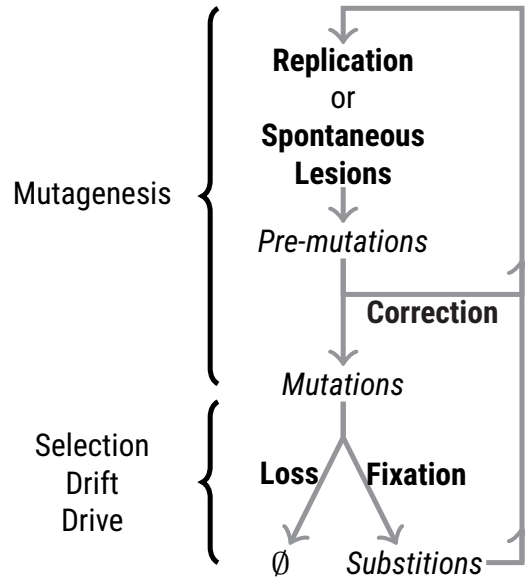
$$= \frac{\mu_{AT \rightarrow GC}}{\mu_{AT \rightarrow GC} + \mu_{GC \rightarrow AT} \cdot \frac{\phi_{GC \rightarrow AT}}{\phi_{AT \rightarrow GC}}} \quad (1.3)$$

Quel est alors le point de départ entre sélectionnistes et neutralistes sur cette question ? Avant de l'explicitier, j'introduit ici la terminologie que j'emploierai dans le reste de ce manuscrit.

Le processus qui conduit à une mutation est la *mutagénèse* (figure 1.1). La mutagénèse dépend des erreurs qui sont commises cours de la répliation de l'ADN, ou plus simplement de dommages occasionnés à la double hélice par des agents mutagènes comme le rayonnement ultraviolet ou le stress oxydatif. Ces erreurs, ou *pré-mutations* sont susceptibles d'être corrigées avec plus ou moins d'efficacité par une machinerie spécifique de correction des lésions de l'ADN. Les pré-mutations qui ne

## 1. PROLOGUE

FIGURE 1.1 –  
Terminologie  
employée dans  
ce manuscrit



sont pas corrigées donne lieu à des *mutations*, à proprement parler. Le destin évolutif de ces mutations dépend alors de processus par nature stochastiques, comme la dérive génétique, et de processus directionnels, tels que la sélection naturelle et certains biais moléculaires (ou *molecular drive*), desquels je serai amené à discuter plus en détail. Une mutation donnée, dans une population de taille infinie, est soit perdue, soit fixée, ce qui donne lieu à une *substitution*. On voit alors que le taux de GC d'un génome dépend des probabilités différentielles des pré-mutations et de l'efficacité avec lesquelles elles sont corrigées ou non, ainsi que des probabilités différentielles que ces pré-mutations non-corrigées, ou mutations, ont d'être fixées dans la population.

Cela posé, l'on peut définir à quel niveau se situe le débat entre neutralistes et sélectionnistes concernant la composition en base. Les neutralistes, depuis les travaux de SUEOKA [1962] et FREESE [1962], défendent que la composition en base d'un génome résulte de processus qui ne contribuent pas par eux-mêmes à la fitness de l'organisme. Autrement dit, la différence de composition en base entre deux génomes devrait ne résulter que des probabilités différentielles de mutations  $\mu_{GC \rightarrow AT}$  et  $\mu_{AT \rightarrow GC}$  auxquelles ils sont soumis, et pas de leurs probabilités respectives d'être fixées dans la population. Les neutralistes sont donc partisans de dire que  $\phi_{AT \rightarrow GC} = \phi_{GC \rightarrow AT}$ , de sorte que, dans ce cas

$$GC_{eq} = \frac{\mu_{AT \rightarrow GC}}{\mu_{AT \rightarrow GC} + \mu_{GC \rightarrow AT} \cdot 1}$$

---

En clair, pour les neutralistes, la mutation est le déterminant majeur de la variabilité de composition en base entre les génomes.

Les sélectionnistes quant à eux proposent que les probabilités de fixation d'une mutation  $AT \rightarrow GC$  et  $GC \rightarrow AT$  soient différentes, parce que ces deux mutations sont susceptibles d'être plus ou moins avantageuses pour un organisme donné, dans un environnement donné. Si, pour une quelconque raison, une mutation  $AT \rightarrow GC$  est plus favorable à l'organisme qu'une mutation  $GC \rightarrow AT$ , le terme  $\frac{\phi_{GC \rightarrow AT}}{\phi_{AT \rightarrow GC}}$  est inférieur à 1, et alors le taux de GC à l'équilibre des substitutions augmente. Dans son principe, la sélection n'opère que sur les traits qui ont une quelconque *fonction* pour l'organisme. Si donc elle doit contribuer à déterminer le taux de GC du génome d'un organisme, c'est soit parce qu'elle opère directement sur les régions fonctionnelles du génome, et au premier chef les gènes, soit parce que globalement la composition en base se trouve être elle-même *fonctionnelle*.

Mais chez les mammifères, des travaux de la fin des années 1990 ont montré que les bases G et C ont une probabilité plus forte d'être fixée dans la population que les bases A et T. Dix ans plus tard, plusieurs travaux montrent le même patron chez les Procaryotes [LONG et al. 2018]. Les bases G et C ont donc une probabilité de fixation supérieure aux bases A et T dans de nombreuses espèces très diverses du point de vue phylogénétique. Dit autrement il existe un *biais de fixation* en faveur des bases G et C. Cette observation pourrait s'expliquer d'un point de vue sélectionniste : puisque  $\phi_{AT \rightarrow GC} > \phi_{GC \rightarrow AT}$ , c'est qu'il doit y avoir un avantage sélectif à l'enrichissement d'un génome en G et C.

Néanmoins on connaît au moins un autre phénomène qui n'est pas à première vue adaptatif, et qui contribue à accroître les probabilités de fixation des bases G et C : la conversion génique biaisée vers GC, un mécanisme lié à la correction des mésappariements dans les intermédiaires de recombinaison homologue. Ce mécanisme est dit non adaptatif parce qu'il ne contribue pas, du moins pas à première vue, à la valeur sélective de l'organisme (voire même parfois la diminue).

Si l'on dispose jusqu'alors d'indices quant à la présence éventuelle du mécanisme de conversion génique biaisée vers GC chez les Procaryotes, cela n'a jamais été testé directement. Mon travail de thèse a consisté à déterminer expérimentalement si la conversion génique est également biaisée vers G et C chez les Bactéries.

De quoi la composition en base est-elle le signe ? C'est cette question que j'ai cherché à explorer dans ce travail. Pour moi, la question de la composition en base n'a réellement d'intérêt qu'en tant qu'elle est le signe des forces évolutives à l'œuvre. Et précisément, c'est tout

## 1. PROLOGUE

---

l'enjeu de la génomique moderne que d'essayer de déterminer dans quelle mesure le monde biologique est le fruit de la nécessité et dans quelle mesure il l'est de contingences. Or nous avons dans le taux de GC un trait partagé par l'ensemble des êtres vivants, fonction d'état d'une matrice de substitution, elle-même reflet de patrons de mutation et de probabilités de fixation. Si, à l'égard du taux de GC, nécessité fait loi, alors c'est sur les probabilités de fixation que la sélection naturelle devrait peser. Si cependant c'est au hasard des mutations qu'il évolue, alors la composition en base n'est vraiment signe que de l'histoire évolutive contingente des génomes. Si enfin une autre force contribue à la variabilité de ce trait, il s'agit de la comprendre le mieux possible ; pour cela rien de mieux que d'en étudier les effets.

Pour ma part, j'ai tenté dans cette thèse de contribuer à cette question en m'attardant particulièrement sur la composition en base des Procaryotes. Pour introduire la problématique, je n'ai pas trouvé de meilleure façon que d'en retracer sommairement l'histoire. Cela permettra d'en dire plus sur les deux premiers courants qui ont traversé l'histoire de la question, courants que je n'ai fait que mentionner ci-dessus.

## « NAISSANCES » D'UNE PROBLÉMATIQUE

---

In the old times, the knowledge of biology was perhaps similar to what could be made out in a very large, very dark house. Many objects could be more felt than seen with equal dimness, once the eyes got used to the darkness; and scientists were conscious of the limiting conditions under which they worked. In our time, however, a few very powerful and very narrow beams of light have been thrown into a few corners of this dark house, and several things can be seen in a clarity and illumination that almost distort their significance. But at the same time we have lost our dark-adaptation; and since we all have a tendency to follow the light, we have moved into these cosy corners, to the detriment of the rest, which still is, by far, the major part of nature. In pointing this out one runs the risk of being accused of trying to spread darkness.

Erwin CHARGAFF [1963]. *Essays on Nucleic Acids*. Amsterdam, New York, Elsevier Pub. Co. 236 p., chapitre 3, page 39-40

### *Introduction*

La pleine compréhension d'un sujet de recherche ne peut faire l'économie d'une enquête historique, même aussi sommaire que celle que l'on s'apprête à exposer, sur les questionnements originels, si l'on veut pouvoir appréhender pleinement les buts qu'il se fixe. Jean Gayon disait en effet que « des cadres de pensée qui se donnent émancipés de toute histoire passée apparaissent alors comme le résultat de choix; de manière ingénue, ils conservent néanmoins la trace du sol originaire, préscientifique dans lequel ils demeurent ancrés, ainsi que du cheminement par lequel ils se sont arrachés à leur propre histoire. » [GAYON 2018, p. 77]

Cette étude est d'autant plus difficile que, disait Duhem, « en la genèse d'une doctrine scientifique, il n'est pas de commencement absolu; si haut que l'on remonte la lignée des pensées qui ont préparé, suggéré, annoncé cette doctrine, on parvient toujours à des opinions qui, à leur tour, ont été préparées, suggérées et annoncées; et si l'on cesse de suivre cet enchaînement d'idées qui ont procédé les unes des autres, ce n'est pas qu'on ait mis la main sur le maillon initial, mais c'est que la chaîne s'enfonce et disparaît dans les profondeurs d'un insondable

## 2. ASPECTS HISTORIQUES DU GC%

---

passé. » [DUHEM 1913-1959] Aussi nous fixerons-nous une limite historique arbitraire, beaucoup moins ambitieuse que celle que s'était fixée Duhem pour son *Système du Monde* (la cosmologie des présocratiques!), à la découverte des acides nucléiques à Tübingen il y a maintenant cent cinquante et un ans.

Conscient du manque d'exhaustivité des paragraphes qui suivent, je n'avais pas pour eux l'ambition d'un travail de recherche historique abouti (s'il existe même une telle chose), mais plutôt d'une présentation succincte des différents « jalons » qui ont marqué la question de la composition en base des génomes. Ce genre de démarche s'expose nécessairement à l'incomplétude liée au choix de quelques grands noms, et masque d'autant tous les travaux dits mineurs, toutes les étapes successives qui ont permis de poser lesdits jalons. Il faut néanmoins lui reconnaître des vertus heuristiques, qui nous permettront d'aller plus directement au cœur des questions qui nous ont intéressées.<sup>1</sup>

L'intrigue qui me semble se dessiner dans l'histoire de cette question est la suivante. Les acides nucléiques sont découverts à l'université de Tübingen en 1869 par Friedrich Miescher, et les monomères qui les composent par Albrecht Kossel et son équipe au tournant du XX<sup>e</sup> siècle. Pendant une trentaine d'années environ, l'on croit des acides nucléiques, avec Phoebus Levene, qu'ils se composent de tétramères régulièrement agencés de ces quatre bases nucléotidiques. À ce titre, ils sont proprement incapables de supporter quelque « information » héréditaire que ce soit, et sont plus ou moins mis au ban de la recherche du matériau de l'hérédité au profit des protéines, et ce jusqu'à ce qu'Oswald T. Avery montre astucieusement que c'est probablement la fraction nucléique d'une fragmentation d'un extrait cellulaire de pneumocoques qui est responsable de la transmission de la pathogénie entre différentes souches. Ces travaux, que d'aucuns disent être restés relativement confidentiels, suscitent quasiment immédiatement l'intérêt de deux grands noms de la biologie d'alors, le biochimiste Erwin Chargaff d'une part et le généticien Theodosius Dobzhansky d'autre part. Le second, dans les dernières pages du chapitre « Heredity and Mutation » de la deuxième édition de son ouvrage phare *Genetics and the Origin of Species*, détaille les découvertes d'Avery *et al* et discute du caractère « induit » (Lamarckien ?) de

1. Je voudrais souligner ici que je réfute la vision de sens commun, volontiers empiriciste d'une « vie des idées », d'un découlement *logique* des idées de celles qui les précèdent, dans une vision quasi-évolutionniste qui ne tient naïvement pas compte des processus

sociaux dans lesquels sont englués leurs émetteurs. ABIR-AM [1985] est l'autrice d'un excellent travail de déconstruction des mythes de la biologie moléculaire, et son article constituerait un crucial complément critique des paragraphes qui vont suivre.

ces mutations. Erwin Chargaff quant à lui réoriente tout bonnement l'ensemble des activités de son équipe de recherche vers la caractérisation des acides nucléiques, ce qui l'a conduit à de nombreuses découvertes fondamentales. Au nombre d'entre elles, on compte la découverte de la variabilité en composition en base des acides nucléiques.

Dans un paragraphe spéculatif d'un article de revue, Chargaff écrit explicitement qu'il y a probablement à chercher dans l'ADN quelque chose comme le cristal aperiodique que proposait un autre Erwin, Schrödinger celui-là, dans *What is life?*. Il contribue par là à replacer l'ADN sous les projecteurs et autres diffractomètres, notamment ceux des physiciennes et physiciens convertis à la biologie désormais célèbres pour leur découverte de la structure de l'ADN : Rosalind Franklin, Maurice Wilkins et Francis Crick, ainsi que du généticien James Watson. (Chargaff dira sarcastiquement de ces deux derniers qu'ils furent à la quête de la structure en double-hélice ce que Perceval était au Saint Graal — écarté de la récompense par le comité Nobel, certains jugent qu'il a regardé l'essor de la biologie moléculaire avec l'amertume de la jalousie ; il me semble qu'il s'agissait plutôt d'une forme de lucidité, une fois tempérées les ardeurs des candides promesses d'une jeune discipline.)

La prise de pouvoir symbolique et institutionnelle de la biologie moléculaire dans les années 1970 a conduit au réinvestissement de la question de la variabilité en composition en base. Le développement des techniques de séquençage notamment dévoile une variabilité dont on ne faisait que soupçonner l'existence. Sont notamment découvertes les variabilités en taux de GC le long de la molécule d'ADN, entre brins de la double-hélice, entre positions des codons, entre séquences codantes et non-codantes . . .

Survient alors la question des origines de cette variabilité, question à laquelle les chapitres suivants seront consacrés.

### 2.1 XIX<sup>e</sup> siècle : Découvertes des acides nucléiques

L'éclat retentissant des avancées scientifiques des années 1950-60 qui ont marqué l'essor de la biologie moléculaire<sup>2</sup>, notamment la proposition de la structure de l'ADN [J. D. WATSON et CRICK 1953], la proposition et la confirmation des acides nucléiques comme l'un des agents de l'hérédité [AVERY et al. 1944; HERSHEY et CHASE 1952] et l'élucidation du code génétique, ont tendance à masquer le fait que les acides nucléiques ont été découverts quelques soixante-quinze ans plus

2. Au grand dam de Chargaff, qui mieux que personne a su y voir les effets d'autorité exercés par une dynamique plus sociale que proprement scientifique.

Erwin CHARGAFF [1963]. *Essays on Nucleic Acids*. Amsterdam, New York, Elsevier Pub. Co. 236 p., chapitre 11, *Amphisbaena*

## 2. ASPECTS HISTORIQUES DU GC%

---

tôt en Allemagne par les fondateurs de la biochimie. Je me pencherai ici brièvement sur leurs travaux d'une part parce qu'ils sont les premiers à s'être intéressés à la question de la composition en base, et d'autre part parce qu'on y verra que les questions qu'ils se posaient ont traversé le temps scientifique peu ou prou inchangées. La première partie de ce chapitre retrace en ce sens les apports majeurs de Friedrich Miescher et Albrecht Kossel, ainsi que la première théorie concernant le taux de GC des organismes formulée par LEVENE et W. A. JACOBS [1909].

C'est à Erwin Chargaff que l'on doit d'avoir rectifié l'erreur de Levene, qui pensait que les acides nucléiques étaient composés à équimolarité des quatre bases ; il contribua ainsi à ce changement de perspective majeur que fut le glissement d'une conception essentiellement protéique à une conception essentiellement nucléique du matériau de l'hérédité. À sa suite, c'est la science nouvelle de la biologie moléculaire qui s'empara de la question de la composition en base, non sans qu'il y voit, avec une ironie certaine, les piètres biochimistes d'hier devenir les grands molécularo-biologistes d'alors.

Parmi les contributeurs majeurs de cette science émergente, on compte Noboru Sueoka, dont la contribution théorique et expérimentale à la question du taux de GC fait date ; c'est si vrai que l'on peut raisonnablement dire qu'essentiellement toute contribution à cette question s'est construite en accord ou en opposition avec ses positions d'alors. Assurément notre problématique de thèse a cherché à répondre aux questions qu'il a posé il y a maintenant soixante ans, aussi terminerons-nous ce survol de la genèse historique par une description de ses conceptions.

### *Haeckel, Hoppe–Seyler et la chimie physiologique*

On doit la découverte des acides nucléiques à Friedrich Miescher, alors étudiant dans le laboratoire de Felix Hoppe–Seyler à l'université de Tübingen, au sud de Stuttgart. Les années 1860 marquent l'essor de ce qu'on appelait alors en France la « chimie physiologique », devenue depuis la biochimie, sous l'impulsion notoire de l'école allemande de Felix Hoppe–Seyler. Parmi ses étudiants célèbres, on compte entre autres, Eugen Baumann (qui découvrit le polychlorure de vinyle, PVC), Friedrich Miescher et Albrecht Kossel, deux figures majeures de la chimie des acides nucléiques.<sup>3</sup>

En 1866, Ernst Haeckel, parfois qualifié de « Darwin allemand », avait suggéré que le noyau était l'organelle susceptible de contenir les « facteurs » responsables de la transmission des traits héréditaires [HAECKEL

3. Les travaux datant de cette époque étant publiés essentiellement en Allemand et pour la plupart non-traduits, je dois les paragraphes suivants aux travaux

d'historiens des sciences, notamment DAHM [2008], JONES [1953], KENNAWAY [1952] et OLBY [2012]



1866; OLBY 1969]. Ernst Haeckel contribua en tout cas à populariser une vision matérialiste de l'évolution, se félicitant de ce que la biologie d'alors soit parvenue à réduire certains phénomènes vitaux à leurs propriétés physiques :

L'explication générale de la vie n'est donc pas plus difficile pour nous maintenant que celle des propriétés physiques des corps inorganiques. Tous les phénomènes vitaux, tous les faits de l'évolution des organismes, dépendent étroitement de la constitution chimique et des forces de la matière organique comme les phénomènes vitaux des cristaux inorganiques, c'est-à-dire leur croissance, leurs formes spécifiques, *dépendent de leur composition chimique* et de leur état physique. [p. 293 HAECKEL 1877, Je souligne. On voit bien dans son enthousiasme, certainement exagéré a posteriori, tout ce qu'il attendait de la biochimie.]

Plus généralement, cette époque marque l'avènement de la théorie cellulaire, c'est-à-dire de la définition de la cellule comme unité organique primaire, que résume bien Haeckel : « En effet, au début de sa vie individuelle, tout animal et toute plante sont représentés par une simple cellule. L'homme, aussi bien que tout autre animal, n'est d'abord qu'une simple cellule contenant un noyau. » [HAECKEL 1877, p. 305] Aussi peut-on comprendre l'essor de la chimie physiologique, c'est-à-dire la caractérisation et la classification des matériaux organiques, comme répondant à la volonté de réduire les phénomènes vitaux à leurs propriétés physiques matérielles (sans que cette réduction ne soit à entendre dans le sens péjoratif auquel on a tendance, parfois abusivement, à l'associer de nos jours). Point n'est besoin de rappeler que c'est encore, sous une forme quelque peu remaniée, l'un des buts de la biologie moléculaire.

C'est dans ce courant que s'inscrivent les travaux de Friedrich Miescher, figure majeure dans la découverte des acides nucléiques. (Bien qu'indirectement liée à la question de la composition en base, j'ai la sensation de contribuer à rectifier le genre d'injustice qui caractérise si souvent l'histoire des sciences en ce qu'elle a tendance à procéder d'une mémoire sélective, selon le mot d'Olivier Perru : qui d'autre en effet qu'une poignée d'historiennes de la biologie connaît le nom de Miescher?)

### *Friedrich Miescher, la découverte de la nucléine et des acides nucléiques*

Friedrich Miescher était étudiant en médecine de l'université de Bâle ; il préféra une carrière scientifique à celle de praticien et choisit la célèbre université de Tübingen pour se former, poussé par son oncle His [DAHM 2008; OLBY 1969]. Ce dernier, médecin et physiologiste

## 2. ASPECTS HISTORIQUES DU GC%

---

aussi, pressentait qu'il fallait chercher à expliquer les désordres physiologiques par les mécanismes chimiques ayant lieu à l'échelle cellulaire. Miescher rejoignit le laboratoire de Felix Hoppe–Seyler, et s'assigna la tâche d'élucider « la composition chimique des cellules lymphoïdes. » Pour diverses raisons, il finit par se concentrer sur les leucocytes, qu'il pouvait obtenir aisément depuis les pus fournis par l'hôpital voisin. Il entreprit d'en déterminer la composition protéique et lipidique. Il obtint alors (« par hasard, direz-vous peut-être, mais souvenez-vous que dans les champs de l'observation le hasard ne favorise que les esprits préparés »<sup>4</sup>) un précipité dans des conditions qui suggéraient qu'il ne s'agissait d'aucune substance chimique connue alors. Il fut amené à conclure que cette substance provenait du noyau des leucocytes, et la baptisa *nucléine*. Miescher y vu l'occasion d'approfondir la connaissance de la chimie et des fonctions du noyau. L'analyse des éléments chimiques composant la nucléine lui indiquait une quantité importante de phosphore organique, à la différence de la plupart des protéines et autres molécules organiques connues alors. Il écrivit, dans le style précautionneux et largement hésitant caractéristique de ces pionniers de la biochimie :

Je crois que les présentes analyses, si incomplètes soient-elles, permettent de conclure que nous ne sommes pas face à une mixture aléatoire, mais, en dehors de contaminations sous forme de traces au mieux, face à un individu chimique ou une mixture d'entités très liées. [...] Nous avons plutôt ici des entités *sui generis* [de leur propre genre] incomparables à des groupes connus jusqu'alors.<sup>5</sup>

Ayant depuis quitté le laboratoire de Hoppe–Seyler, il lui communiqua ses résultats en 1869, en lui soumettant un article à publier dans la revue que Hoppe–Seyler éditait. Ce dernier trouva les propositions de l'article si peu crédibles qu'il décida de répéter les expériences par lui-même et avec l'aide de deux de ses étudiants [JONES 1953]. C'est seulement deux ans après qu'il fut convaincu de leur vraisemblance, et l'article parut en 1871.<sup>6</sup> La conclusion de l'article montre toute l'étendue de la prescience dont Miescher a fait preuve :

Je crois que les présents résultats, pour fragmentaires qu'ils soient, sont d'importance suffisante pour inviter d'autres, particulièrement les chimistes, à se pencher sur la question. *La connaissance des relations entre les substances nucléaires, les protéines et leurs produits*

4. PASTEUR 1922, page 131.

5. MIESCHER [1871]. Il remarqua également que la quantité de nucléine augmentait pendant la « progression générative » dans les tumeurs, ce qui correspond respectivement à la phase S du cycle

cellulaire pendant laquelle les chromosomes sont répliqués.

6. À notre époque marquée d'une accélération du temps, social et scientifique [ROSA 2013] il faut méditer sur ce sain scepticisme et la longue durée d'une science consciente d'elle-même.

*de conversions vont aider graduellement à dévoiler les processus internes du développement cellulaire.*<sup>7</sup>

Il proposa plus tard une théorie sur la façon dont l'information héréditaire pouvait être « encodée » dans les cellules étonnamment proche de ce qui est désormais admis aujourd'hui. DAHM [2008] rapporte que Miescher avait perçu qu'il était possible d'encoder ce que l'on appellerait aujourd'hui « l'information héréditaire » dans une large molécule organique composée d'une diversité réduite de carbones asymétriques ; de plus, il semble que Miescher ait envisagé la possibilité de ce que l'on appellerait désormais les mécanismes de correction des erreurs de réplication :

Pour empêcher les erreurs, causées par des facteurs environnementaux susceptible de changer l'état stéréochimique des atomes, de se manifester au cours du développement de l'embryon, Miescher a supposé que ces erreurs devaient être corrigées par la fusion de l'information entre deux cellules germinales au cours de la fertilisation. [DAHM 2008]

Si ça n'est pas ainsi que cela fonctionne vraiment, Miescher avait au moins entr'aperçu que la fusion des gamètes pouvait jouer un rôle génétique. S'il y a bien un « pionnier » du concept de recombinaison homologue, c'est certainement Miescher.

### *Albrecht Kossel, caractérisation des cinq bases nucléotidiques*

Ce sont toutes questions dont Albrecht Kossel, un autre étudiant de Felix Hoppe–Seyler, s'empara ; il est resté célèbre pour le prix Nobel qu'il obtint en 1910 « en reconnaissance pour la contribution que ses travaux sur les albumines, ainsi que les nucléines, ont apportée à notre connaissance de la cellule. » [JONES 1953 ; KENNAWAY 1952].

Il passa ses premières années d'études de médecine à l'université de Strasbourg en 1872, à suivre, entre autres, les cours de botanique dispensés par Heinrich Anton de Bary — le célèbre botaniste à l'origine du concept de « symbiose » et l'un des fondateurs de l'écologie des microorganismes —, ainsi que ceux de Hoppe–Seyler, qui dirigeait à Strasbourg l'unique département de biochimie d'Allemagne de l'époque. Il termina ses études de médecine à l'université de Rostock en 1877 et rejoignit le laboratoire de Hoppe–Seyler sept ans après sa publication des travaux de Friedrich Miescher. Albrecht Kossel hérita donc de ces questions qui l'occupèrent pour le reste de sa carrière.

7. J'ai souligné ici pour montrer qu'une certaine forme de réductionnisme *tout-génétique* des années 1970 s'est inscrit en faux par rapport à la tradition relationnelle et intégrative des premiers

temps de la chimie biologique. C'est ce réductionnisme précis que Woese qualifia rétrospectivement d'erreur conceptuelle. WOESE 2004

## 2. ASPECTS HISTORIQUES DU GC%

---

C'est, en effet, dans son laboratoire que furent isolées les cinq bases nucléotidiques composant les acides nucléiques : c'est en 1871 qu'il annonce avoir isolé la guanine et l'adénine de la nucléine, ainsi qu'un carbohydrate (probablement le désoxyribose); deux ans plus tard la thymine et la cytosine sont isolées par lui et son étudiant Neumann; la découverte, par lui et Ascoli, de l'uracile dut attendre 1901 [JONES 1953]. Pour Kossel, il était clair que les acides nucléiques n'avaient pas de fonction de stockage ou énergétique, mais étaient bien plutôt liées à la synthèse d'une nouvelle cellule au cours de la croissance [DAHM 2008].

### *Phoebus Levene, l'hypothèse tétranucléotidique*

L'intérêt pour les acides nucléiques est allé décroissant, en partie en raison de l'hypothèse tétranucléotidique proposée par P.A Levene et H. Steudel. [DAHM 2008; OLBY 2012, chapitre 6] Ceux-ci tenaient que les acides nucléiques étaient composés d'unités identiques composés des quatre nucléotides. Cette hypothèse avait deux conséquents évidents : (1) que les quatre bases devaient se trouver en équimolarité et (2) que le matériau de l'hérédité était à chercher ailleurs, une telle composition étant proprement inapte à encoder quelque information génétique que ce soit, puisque dépourvue de tout caractère « séquentiel ». Cette hypothèse s'est donc inscrite en faux par rapport à celle de Miescher, qui avait bien perçu le potentiel des acides nucléiques comme candidats au matériau support de l'hérédité. Tout à la fois en raison du prestige scientifique de Levene, de sa mainmise quasi-impérialiste sur le champ de la chimie cytologique, et sans doute de la simplicité relative de l'hypothèse nucléotidique, elle est restée l'hypothèse dominante pendant 20 à 30 ans, pendant lesquels la recherche du matériau de l'hérédité s'est essentiellement concentrée sur les protéines.

Il ne faut pas oublier que, dans le même temps, cette période est marquée de l'essor de la génétique mendélienne, récompensée d'un prix Nobel en 1934 pour Thomas Hunt Morgan. Ce dernier pose, dans son discours de réception du prix la question suivante : « Que sont les gènes? Maintenant qu'on les localise dans les chromosomes, sommes-nous fondés à les voir comme des unités matérielles; comme des corps chimiques d'un ordre supérieur aux molécules? [...] Au niveau où se place l'analyse génétique, cela ne fait pas la moindre différence que le gène soit une unité hypothétique, ou que le gène soit une particule matérielle [MORGAN 1934] ». Cela pour rappeler que la quête du support de l'hérédité était marquée d'une ambivalence entre (certains) des généticiens qui pouvaient, pour leurs fins propres, se passer d'une telle

« Le savant ne voit pas que l'ignorance est un tissu d'erreurs positives, tenaces, solidaires. » — Bachelard, *La Philosophie du Non*, p. 8.

entité matérielle, et les chimistes de la cellule occupés à comprendre le processus matériel du développement cellulaire.

C'est en partie grâce aux travaux d'Oswald AVERY et al. [1944] que les acides nucléiques revinrent sur scène, notamment grâce aux travaux du groupe mené par Erwin Chargaff à l'Université de Columbia.<sup>8 9</sup> À la découverte des travaux d'Avery, Chargaff réoriente toutes les activités de son laboratoire vers l'étude des acides nucléiques.<sup>10</sup> (Il faudrait aussi

8. Baptiste Bedessem, chercheur en philosophie des sciences, nous a indiqué au cours d'une présentation au laboratoire que les travaux d'Avery sont restés quasiment ignorés du reste de la « communauté » scientifique, jusqu'à leur redécouverte quelques années plus tard. Il me semble que ça ne correspond pas tout à fait à ce qu'il s'est passé. Olby rapporte que Theodosius Dobzhansky, l'un des acteurs majeurs de la synthèse moderne, a visité le laboratoire d'Avery en 1941, et a même jugé nécessaire d'ajouter à l'un des ses chapitres de la seconde édition de son *Genetics and the Origin of Species* : « Si cette transformation était décrite comme une mutation génétique — et il est difficile de ne pas le faire —, nous avons affaire à un cas authentique d'induction de mutations spécifiques par des traitements spécifiques — un exploit que les généticiens ont vainement tenté d'accomplir dans des organismes supérieurs. » DOBZHANSKY 1941, page 49. Dobzhansky a quelque peu édulcoré son propos dans l'édition de 1951 suivante, mais il continue d'y insister sur l'importance des travaux d'Avery. De plus, comme les paragraphes suivants l'indiquent, ils ont occasionné un changement d'orientation majeur pour le laboratoire de Chargaff, dont on sait toute la contribution à la découverte de l'ADN.

9. S'il faut souligner les vertus épistémologiques de la microbiologie, c'est certainement avec des exemples de ce genre ; pour notre part, il nous est assez agréable que l'une des avancées majeures en génétique moléculaire ait été accomplie par la transformation naturelle, encore qu'évidemment, nous n'ayons pas beaucoup d'illusions sur l'importance de nos travaux relative à ceux d'Avery.

10. Je voudrais m'attarder ici sur le personnage d'Erwin Chargaff, qui est un personnage historique de la discipline qu'il a contribué à fonder malgré lui et qui fut appelée « biologie moléculaire », terme qu'il aurait qualifié de *buzzword* dans un dialecte actuel.

Sa vie durant, il n'a considéré comme son professeur que le personnage de Karl Kraus, un fameux satiriste autrichien dont l'activité intellectuelle quasi-entière fut consacré à la critique satirique de l'activité intellectuelle sociale et de l'emprise des médias. Il est tentant de rapprocher le style littéraire et les orientations critiques de Chargaff de celui et celles de Kraus. ABIR-AM [1980] s'est appliquée à la socio-analyse d'Erwin Chargaff et à montrer la genèse sociale de ce critique scientifique dont l'« anti-héroïsme » s'est construit contre les tendances et les modes de la biologie moléculaire.

Il faudrait expliciter en détail son style d'écriture pour montrer à quel point cette appellation de « règle » se trouve ironiquement accolée à son nom, lui qui mettait en garde contre les dangers des généralisations et des dogmes scientifiques dans les mêmes articles d'où ses fameuses « règles » ont été extrapolées ; je crois qu'il aurait préféré léguer des « régularités » que des « règles » ... Il fut en tout cas, à tout point de vue, un scientifique en dehors du commun, de par sa formation dans la Vienne de l'entre-deux guerres, son attachement à la langue que lui ont légué Kraus, dont il ne manquait aucun récital, et Wittgenstein, son admiration pour Shakespeare, son style d'écriture rigoureux en science, mordant pour les affaires du monde, son mépris des faux-semblants et des effets de manche particulièrement courant à cette époque de prise de pouvoir symbolique des biologistes moléculaires sur les anciens érudits qu'il représente. Il s'est fait le héraut, dans son autobiographie *Heraclitean Fire*, d'une science désabusée, pleinement consciente de ce dont elle était capable, du meilleur mais surtout du pire. Voici par exemple ce qu'il dit de sa réception de la nouvelle des bombes atomiques larguées sur Hiroshima et Nagasaki : « Jamais loin d'une vision apocalyptique du monde, j'ai vu la fin de l'essence de l'humanité ; une fin rendue plus proche, voire même rendue possible,

## 2. ASPECTS HISTORIQUES DU GC%

---

mentionner les travaux de HERSHEY et CHASE [1952] sur les phages, qui démontrèrent très astucieusement que la réplication des phages dépendait de l'ADN encapsidé, et pas des protéines qui les constituent.)

### *Erwin Chargaff, vers la quantification du GC%*

Erwin Chargaff est peut-être l'un de ceux qui a le plus marqué le champ en formulant ce qui a été appelé plus tard les « règles de Chargaff » [Erwin CHARGAFF 1951]. Ce serait pourtant réducteur de restreindre l'apport de Chargaff à ces deux règles, tant il a contribué à l'étude des acides nucléiques. Ce biochimiste né en Autriche a considérablement œuvré au perfectionnement des méthodes de purification des acides nucléiques, et à leur appliquer la méthode de quantification par chromatographie de Vischer, l'un de ses étudiants.<sup>11</sup> En les appliquant à une large gamme d'organismes [ZAMENHOF et al. 1952] et de tissus [E. CHARGAFF, LIPSHITZ et al. 1951 ; E. CHARGAFF et E. VISCHER 1949], il en est venu à comprendre que les concentrations d'un ADN en thymine et adénosine d'une part, en guanine et cytosine d'autre part sont quasiment égales — c'est la règle de Chargaff la plus connue (la règle de complémentarité).<sup>12</sup> C'est lui qui peut-être a formulé le premier l'hypothèse d'un appariement entre les bases AT et les bases GC.<sup>13</sup>

Ses travaux ont permis de renforcer l'hypothèse d'AVERY et al. [1944] selon laquelle les acides désoxypentoses nucléiques « pourraient servir comme un agent, potentiellement l'agent, impliqué dans la transmission des propriétés héréditaires. » (Erwin CHARGAFF 1950)

par la profession dont j'étais. À mon sens, toutes les sciences naturelles ne sont qu'une ; et si l'une ne peut plus plaider l'innocence, aucune ne le peut. » Erwin CHARGAFF [1978]. *Heraclitean Fire : Sketches from a Life Before Nature*. Rockefeller Univ. Press. 270 p., page 3. On l'aura sans doute compris, mais c'est un personnage pour lequel j'ai la plus grande admiration.

11. La lecture rétrospective de ses travaux donne un vague aperçu de la difficulté de la tâche, à l'heure des extractions d'ADN en routine et en une matinée
12. La seconde « règle de parité de Chargaff » a été formulée plus tard par KARKAS et al.

[1968] et RUDNER et al. [1968a,b]. Elle tient au fait qu'au sein d'un même brin, les règles de parité *entre* brins sont également vraies : il y a autant de bases A que de bases T globalement sur l'ensemble d'un brin, et de bases G que de bases C.

13. Encore que ça ne soit pas très clair à la consultation des articles qu'il y soit vraiment dit explicitement quoi que ce soit qui puisse permettre de penser que Chargaff avait réellement en tête une règle d'appariement ; en tout cas certainement pas de la structure en double hélice. Erwin CHARGAFF [1950, 1963]



## Découvertes des acides nucléiques

Table V  
Molar proportions of purines and pyrimidines in desoxyribose nucleic acids from different species.

Species	Number of different organs	Number of different preparations	Adenine/Guanine			Thymine/Cytosine		
			Number of hydrolyses <sup>3</sup>	Mean ratio	Standard error	Number of hydrolyses <sup>3</sup>	Mean ratio	Standard error
Ox <sup>1</sup> . . . . .	3	7	20	1.29	0.013	6	1.43	0.03
Man <sup>2</sup> . . . . .	2	3	6	1.56	0.008	5	1.75	0.03
Yeast. . . . .	1	2	3	1.72	0.02	2	1.9	
Avian tubercles bacillus	1	1	2	0.4		1	0.4	

FIGURE 2.1 – Première réalisation que les molarités en base varient d'une espèce à l'autre. Tiré de Erwin CHARGAFF [1950]. (Ox = Bœuf.)

La seconde observation que formule Erwin CHARGAFF [1950] est probablement des deux celle qui nous concerne de plus près. Elle propose que la composition en base soit liée à l'espèce, donc globalement constante entre les individus d'une même espèce. Elle s'appuie sur l'observation que la proportion des acides nucléiques occupée par les différents nucléotides varie d'une espèce à l'autre. L'un des premiers articles à l'avoir montré est celui de Ernst VISCHER et al. [1949] : ils montrent que, pour aller vite, le taux de guanine et cytosine est plus élevé chez *Mycobacterium tuberculosis* (une Actinobactérie avec un taux de GC proche de 65%) que chez la levure (38%). La figure 2.1 montre les ratios A/G (purines) et T/C (pyrimidines) des ADN extraits de différents organismes ; on y voit clairement que ces ratios sont plus faibles chez *Mycobacterium tuberculosis*. Chargaff se félicitera plus tard de son choix particulièrement heureux d'organismes tels que la levure et le bacille tuberculeux, car ils « différaient l'un de l'autre dramatiquement en composition en base. [...] Aurais-je, d'un autre côté, décidé de comparer l'ADN du thymus de veau avec l'ADN du pneumocoque, que j'en eusse probablement conclu qu'ils étaient indistinguables chimiquement » [Erwin CHARGAFF 1978, p. 92].

Il résume ainsi les éléments que son équipe a démontrés dans son article de revue [Erwin CHARGAFF 1951], dont on peut raisonnablement dire qu'ils constituèrent une étape majeure dans notre compréhension du rôle des acides nucléiques :

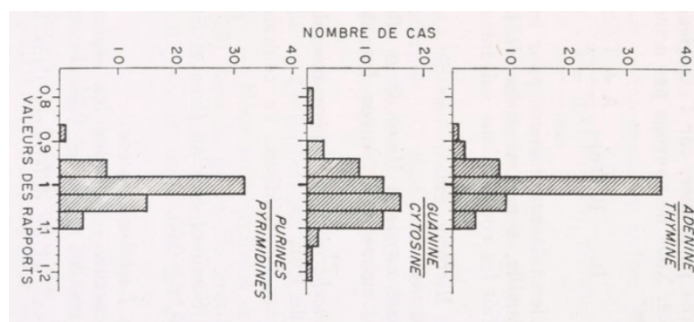
1. Les acides nucléiques désoxyribose sont dans leur composition caractéristiques de l'espèce dont ils proviennent.
2. Différents tissus de la même espèce produisent les mêmes acides nucléiques désoxyribose.
3. L'hypothèse tétranucléotidique est incorrecte.

Certains français s'emparèrent de la question : par exemple K. Y. LEE et al. [1956], qui ont appliqué les protocoles de Ernst VISCHER et al. [1949] à 60 espèces de bactéries des collections de l'Institut Pasteur (voir figure 2.2).

Point n'est besoin ici de rappeler l'essor de la biologie moléculaire, la découverte de la structure en double hélice par Watson et Crick,

## 2. ASPECTS HISTORIQUES DU GC%

FIGURE 2.2 – Tiré de K. Y. LEE et al. [1956]



grandement aidés en cela d’injustement oubliés par le comité Nobel en 1962, quand ils n’en ont pas été franchement écartés par les velléités misogynes d’un Jim Watson déjà détestable. Je pense bien sûr à Rosalind Franklin, mais il est vraisemblablement tout aussi injuste que Chargaff n’ait pas été récompensé, lui sans qui les « deux Perceval » n’auraient probablement pas atteint le Saint Graal. [Erwin CHARGAFF 1976]

*Les généralisations en science sont autant nécessaires qu’hasardeuses; elles véhiculent un semblant d’achèvement qui masque leur caractère essentiellement provisoire; elles font avancer autant qu’elles retardent; elles donnent tout autant qu’elles prennent. — Erwin Chargaff, 1950*

En 1962, quand le Prix Nobel pour la découverte de l’ADN fut remis, Chargaff en fut écarté malgré son rôle de pionnier dans l’établissement de la crédibilité de l’ADN comme le matériau génétique. Watson et Crick, qui jamais n’ont manipulé d’ADN, reçurent la récompense. Aussi, l’on doit conclure qu’en une décennie les perceptions influentes dans la communauté scientifique en sont venues à dissocier l’ADN de sa tradition chimique. En outre, les « découvertes » théoriques, particulièrement celle de nature prédictive, en sont venues à être évaluées en elle-même, jusqu’à prendre le dessus sur les découvertes empiriques. [ABIR-AM 1980]

## 2.2 Expliquer la variabilité du GC% ?

### GC% et taxonomie moléculaire

Assez généralement, c’est l’époque où le critère du taux de GC s’est popularisé grâce à l’une de ses applications pratiques évidentes, la taxonomie. Les taxonomistes bactériens des années 1960 étaient à la recherche de critères stables permettant de classer les bactéries en genres et espèces [SAPP 2005, chapitre 1, *The Bacterium’s Place in Nature*]; devant l’impossibilité d’utiliser un critère morphologique ou phénotypique unique, les taxonomistes mettaient en pratique une classification dite *numérique* ou *adansonienne*, qui attribuait un poids égal à un nombre important de caractères phénotypiques et tentaient d’utiliser les matrices



obtenues pour classer les souches en espèces et genres. Ces méthodes laissaient cependant la place à de nombreux biais subjectif.<sup>14</sup>

La taxonomie bactérienne était donc à la recherche d'un critère stable et universel, et qui permettrait de réduire l'intensité de la lutte pour la « vérité taxonomique » en diminuant la part d'interprétation subjective des données expérimentales; critère qu'elle a cru trouver dans l'étude de la composition en base de l'ADN :

La composition en nucléotide a été étudiée de façon relativement intensive chez les bactéries, en raison de son utilité dans la classification. [...] Par les méthodes phénotypiques, il était quasiment impossible de déterminer le vrai degré d'apparentement (*relatedness*) des bactéries. Il en résulte une classification des bactéries confuse et biaisée par les opinions personnelles des chercheurs. Il est probable que les méthodes quantitatives plus récemment développées, la composition en nucléotide de l'ADN et l'homologie de l'ADN remédient à cet état de chose. [DE LEY 1968b]

D'autres encore, à l'Institut Pasteur, disent :

Depuis les travaux de Chargaff, on sait en effet que la teneur GC% de l'ADN des bactéries est susceptible de larges variations d'une espèce à l'autre (28 à 75). Un GC% identique constitue un critère nécessaire mais non suffisant de parenté entre deux souches bactériennes dont les phénotypes sont identiques. À l'inverse, chaque fois que le GC% d'une souche ou d'une espèce diffère des autres représentants du même groupe, on peut douter de la validité de la position systématique qui lui a été attribuée [SEBALD et al. 1965].

Les premières méthodes pour déterminer la composition en base reposaient sur des gradients de densité en chlorure de césium [SCHILDKRAUT et al. 1962], sur des chromatographies en couche mince [BELOZERSKY et SPIRIN 1958] ou sur la température de dénaturation de l'ADN [MARMUR et DOTY 1962]; elles ont montré que le taux de GC de l'ADN de bactéries précédemment affiliées aux mêmes genres était relativement stables au sein d'un genre. Elles ont également révélé l'extrême diversité du taux de GC des êtres vivants, notamment celui des bactéries.<sup>15</sup>

14. Je suis à peu près certain qu'il y aurait là pour le sociologue un terrain d'enquête très intéressant. Il faudrait montrer dans quelle mesure la structure taxonomique bactérienne de l'époque reflète la structure de la lutte des microbiologistes pour définir la *vérité* taxonomique. Une grande part de ces luttes subsistent encore aujourd'hui, malgré l'utilisation d'outils de classification prétendant à l'impartialité; l'introduction de nouveaux

outils peut nonobstant servir les intérêts sociaux de ceux qui les développent ...  
15. L'excellente revue de DE LEY [1968b] dans DOBZHANSKY et al. [1968] montre la première diversité observée et les conséquences de la mesure de taux de GC aberrants au sein du genre *Pseudomonas* qui a dû conduire à la réaffectation taxonomique d'une partie des espèces du genre.

## 2. ASPECTS HISTORIQUES DU GC%

---

### *Freese & Sueoka : formulation de la théorie neutraliste de l'évolution de la composition en base*

C'est précisément cette variabilité que certains théoriciens ont tenté d'expliquer, par des travaux fondateurs au nombre desquels on peut compter les travaux des époux Freese [FREESE 1962], notamment Ernst Freese, l'un des premiers à avoir travaillé sur la mutagénèse et les conséquences des agents mutagènes sur l'ADN. C'est lui qui proposera les concepts de *transitions* et *transversions* pour qualifier les différents types de mutations que les bases nucléiques sont susceptibles de subir [DRAKE 1991].

On lui doit d'avoir su formuler explicitement les questions théoriques auxquelles ce travail de thèse à chercher à contribuer, aussi les traduirai-je ici en entier :

Certains changements de paire de base sont léthaux, d'autres altèrent la valeur sélective d'un organisme, d'autres encore peuvent avoir lieu dans une région de l'ADN génétiquement inerte ou redondante, d'autres enfin ne font guère que changer la séquence de paire de base d'une région sans en altérer les propriétés fonctionnelles. Dépendant de la fréquence relative de ces changements, l'on peut concevoir deux causes bien différentes pour l'évolution de la composition en base de l'ADN :

1. *Sélection* : Les organismes actuels ont des ratios en base différents de leurs précurseurs présumés parce que l'essentiel des changements de paire de bases ont doté les organismes altérés d'un avantage sélectif, dans les conditions génétiques et environnementales particulières auxquels ils étaient exposés.
2. *Dérive non-sélective du ratio en base* : L'essentiel des paires de bases de l'ADN peut changer sans avantage sélectif, ou insignifiant, pour l'organisme altéré. L'évolution des propriétés fonctionnelles, par la sélection naturelle, n'est guère que superposée à l'évolution de la composition en base moyenne mais ne la contrôle pas [FREESE 1962].

(Il faut souligner qu'une bonne partie du reste de l'article est consacré à une spéculation traitant de l'influence de la dérive sur la composition en base, et ce six ans avant l'article séminal de Motoo Kimura qui posera les bases de la théorie neutre de l'évolution moléculaire [KIMURA 1968].<sup>16</sup> De là à y voir une théorie neutre de la composition en base, il n'y a qu'un pas.)

16. Encore que, pour être tout à fait honnête, Kimura ait déjà écrit sur des sujets

relativement voisins plus tôt, KIMURA 1957, 1961

L'autre théoricien de la mutation et de ses conséquences sur le taux de GC est Noboru Sueoka, dont il serait malvenu de cantonner l'importance à ses résultats théoriques étant donné l'influence de ses différents travaux expérimentaux.

Ce dernier est l'un des acteurs majeurs de la problématique de la composition en base. Il a notamment travaillé sur l'influence du biais mutationnel sur la composition en base, en cherchant à constituer une théorie à même d'expliquer les découvertes expérimentales de la diversité de la composition en base. Ses premiers travaux datent des années 1960 [SUEOKA 1962], à l'époque des premières descriptions de la variabilité de la composition en acides nucléiques des bactéries, notamment par l'un des fondateurs de la taxonomie moléculaire des bactéries, Jozef de Ley [KERSTERS 1998].<sup>17</sup>

En 1961 et 1962, il publie deux articles qui constituent le fondement historique des controverses sur le taux de GC. Le premier [SUEOKA 1961] compile les différentes données disponibles sur la composition en base des bactéries et postule explicitement une relation entre la composition en base et la taxonomie ; le second [SUEOKA 1962] pose les fondements théoriques de l'importance de la balance mutationnelle dans la fixation du contenu en base de l'ADN. J'essaierai ici d'en exposer les principaux tenants.

### *L'hypothèse de la balance mutationnelle*

En guise d'introduction à la théorie de Sueoka, il me paraît pertinent de traduire les différentes observations connues à l'époque qu'il pose comme postulat de départ.

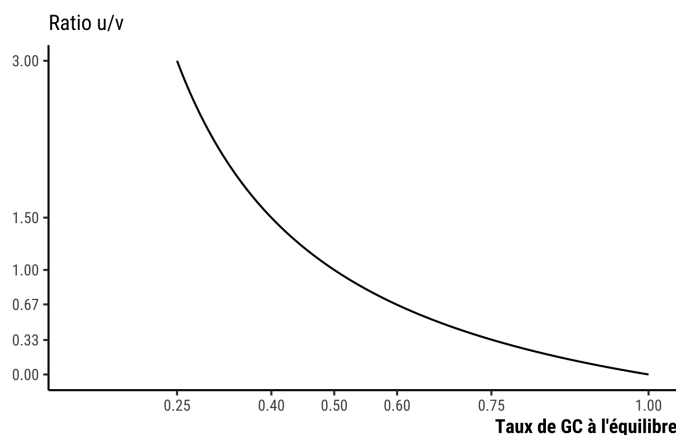
1. Parmi les bactéries, le taux de guanine et cytosine de l'ADN varie de 15 % à 75 %.
2. Le taux de GC moyen reflète les relations phylogénétiques. Les organismes phylogénétiquement proches montrent des compositions en base similaires.
3. La distribution du taux de GC entre les molécules d'ADN d'un organisme est unimodale et l'hétérogénéité relative faible. La conséquence en est que deux organismes dont le taux de GC moyen diffère de plus de 10 % ne devraient avoir que peu de molécules d'ADN de taux de GC similaires en commun.

Sueoka expose ensuite sa théorie qui repose sur des postulats purement logiques. (Je me contente ici de traduire le premier paragraphe

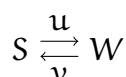
17. BELOZERSKY et SPIRIN [1958], DE LEY [1964, 1968a,b], DE LEY et MUYLEM [1963], L. R. HILL [1966] et SUEOKA [1961]

## 2. ASPECTS HISTORIQUES DU GC%

FIGURE 2.3 – Relation entre le taux de GC moyen  $\hat{p}$  et la balance mutationnelle entre  $u$  (GC  $\rightarrow$  AT) et  $v$  (AT  $\rightarrow$  GC)



d'exposition.) « La cause d'un changement du taux de GC d'une molécule d'ADN doit se trouver dans la conversion des appariements AT en GC ou vice-versa. On peut ici considérer indifféremment les appariements AT et TA ou GC et CG. Les paires AT et TA seront donc dénotés  $\alpha$ , GC et CG dénotés  $\gamma$ . » Par soucis de cohérence avec la suite du manuscrit, je me permettrai ici de remplacer  $\alpha$  et  $\gamma$  par les code IUPAC  $W$  et  $S$ . « Les taux de conversions par génération,  $u$  pour  $S \rightarrow W$  et  $v$  pour  $W \rightarrow S$  sont supposés plutôt uniformes pour tous les paires de bases des molécules d'ADN d'un organisme. Il faut noter que pour la présente théorie, c'est la seule supposition majeure. » Il s'agit en effet du seul postulat de départ supposé et qui n'est ni logiquement déduit ni expérimentalement prouvé à l'époque et à ma connaissance, et ce sera l'un des points de discussion de ce manuscrit. « Schématiquement, nous avons donc : »



On peut déduire de cette relation le taux de GC moyen  $\hat{p}$  à l'équilibre mutationnel :

$$\hat{p} = \frac{v}{u + v}$$

Sueoka déduit ensuite de ces équations qu'il faudrait  $2 \times 10^7$  générations à un organisme pour passer d'un taux de GC  $\bar{p}_0 = 0.30$  à un taux de GC de 0.40 dans des conditions physiologiques où  $u/v = 0.43$  correspondant au taux de GC à l'équilibre mutationnel  $\hat{p} = 0.70$ , avec un taux de substitution  $u$  ou  $v$  de l'ordre de  $10^{-8} \text{ bp}^{-1} \text{ gen}^{-1}$ . Pour une bactérie de temps de génération 1 h, cela correspond à deux mille ans. Étant donné que le temps de génération d'une bactérie n'est d'une heure que dans des conditions optimales de croissance et qu'une fraction seulement des mutations n'est pas susceptible d'être contre-sélectionnée, « dans ce modèle le taux de GC d'un ADN est une caractéristique extrêmement

stable, en accord avec la cohérence entre la composition en base et la taxonomie. » (Il faut bien reconnaître ici que Sueoka est assez loin de la réalité; le taux de substitution estimé par exemple dans l'expérience à long terme de Richard Lenski chez *Escherichia coli* est plutôt de l'ordre de  $10^{-11}$  bp<sup>-1</sup> gen<sup>-1</sup>, soit trois ordres de grandeurs d'écart, d'après WIELGOSS et al. 2011. D'ailleurs 2000 ans à l'échelle évolutive, ça n'est pas très long.)

Le reste de l'article expose des prédictions découlant de l'application logique de cette théorie; elles seront assez largement confirmées expérimentalement par la suite.

### Séquençage et composition en base

En partie en raison des effets d'autorité exercé par la biologie moléculaire, en partie en raison de ses avantages évidents déjà soulignés pour la taxonomie, l'analyse moléculaire des acides nucléiques a pris un essor considérable dans les années 1970, et tout particulièrement au début des années 1980, probablement grâce au développement des techniques de séquençage (voir figure 2.4 B.) C'est l'application du programme que résume DE LEY [1968b], qui consistait à chercher dans la description des acides nucléiques des caractères classificatoires stables (conscient néanmoins qu'ils ont une valeur de classification toute relative, c'est-à-dire par exemple que deux espèces distantes peuvent avoir des taux de GC très proches.)

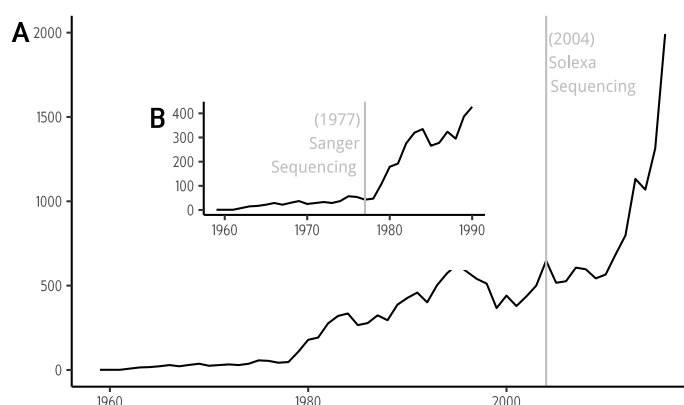


FIGURE 2.4 – Nombre de résultats Pubmed pour « DNA base composition » par an, de 1959 à 2017

Qui consulterait la littérature sur la composition en base des années 1970 à 1990 y trouverait essentiellement des articles descriptifs, ayant trait par exemple à la description d'une nouvelle espèce isolée du sol ou d'autres environnements, qui s'accompagnait systématiquement d'une description du taux de GC de l'espèce, et parfois d'un degré d'hybridation moléculaire avec des souches de référence. Ça n'est pas

## 2. ASPECTS HISTORIQUES DU GC%

---

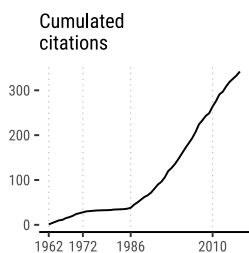


FIGURE 2.5 – Nombre de citations de l'article SUEOKA 1962 dans le Web Of Science. S'il n'a guère été cité dans les années 1970, il faut attendre le milieu des années 1980 pour qu'il suscite un véritable intérêt.

dire que fut abandonnée toute tentative d'explication des origines de la variabilité en taux de GC des génomes; seulement qu'un relatif déclin dans l'effort théorique à ce sujet est notable.

Pour preuve, une analyse du nombre de citations de l'article SUEOKA [1962], dont on peut raisonnablement penser qu'il n'était pas censément possible de faire l'économie de le citer dans quelque travail sur l'évolution de la composition en base que ce soit, recensées dans la base de données Web of Science (voir figure 2.5), montre que ça n'est qu'après les années 1980 qu'il suscita un véritable intérêt.

## Conclusion

Les années 1980 voient le développement de technologies majeures qui viendront bouleverser assez drastiquement la façon dont la biologie moléculaire se pratiquait. Les techniques de PCR et de séquençage de l'ADN sont développées à la toute fin des années 1970, permettant de constituer les premières bases de données de séquence, marquant du même coup la naissance de la bioinformatique. La croissance exponentielle du volume de donnée de séquence qui a suivi a permis d'aborder la biologie de l'évolution moléculaire d'une toute autre façon.

En parallèle, la génétique des populations a commencé dans les années 1980 à confronter ses modèles d'évolution aux données moléculaires, notamment par le travail pionnier que KREITMAN [1983], au laboratoire de Richard Lewontin, a réalisé sur le locus *Adh* de la *Drosophila*. La fusion de ces courants théoriques que sont la bioinformatique, la biologie moléculaire et la génétique des populations a donné lieu à l'affirmation progressive du champ de la biologie évolutive moléculaire.

C'est ce champ qui s'est emparé de la question de l'origine de la variabilité en composition en base, et c'est dans ce champ que ces travaux s'inscrivent également. J'arrêterai donc ici ma description de l'histoire de la problématique. Je commencerai, dans le chapitre suivant, à en dire ce que l'on admet désormais, en montrant la diversité des taux de GC qu'il nous faut expliquer.

## DISTRIBUTION DES TAUX DE GC ENTRE GÉNOMES

---

On peut appliquer aux mélodies des différents compositeurs le principe théorique suivant : Toute espèce d'arbre est « arbre » en un *sens* différent. En d'autres termes : Ne te laisse pas égarer par le fait que l'on dise que toutes sont des mélodies. Ce sont des degrés sur un chemin qui conduit d'un point de départ, que tu refuserais de nommer mélodie, à un point d'arrivée que tu ne nommerais pas non plus de ce nom. Si l'on se contente de considérer la suite des sons et la variation de la tonalité, toutes ces figures apparaissent certes en coordination. Mais si tu regardes le champ dans lequel elles se trouvent (je veux dire leur signification), alors tu inclineras à dire : la mélodie est toute autre ici et là (elle a ici une autre origine, elle joue un autre rôle, etc.)

Ludwig WITTGENSTEIN, *Remarques Mêlées* — 1946, page 110, traduit par Gérard Granel; Flammarion

Avant toute chose, il s'agit de montrer de quelle diversité du taux de GC il est question, et dans quelle mesure elle varie selon les positions que l'on considère dans les génomes. Nous procéderons pour cela, lorsque cela s'avérera pertinent, à quelque chose comme une « réactualisation » d'éléments que l'on connaît depuis plus de trente ans, en utilisant les données de la base version 7 de Hogenom.<sup>18</sup> La diversité des taux de GC est susceptible d'être décrite à deux échelles : soit du point de vue de sa variabilité entre génomes (partie 3.1 ci-dessous), soit de celui de sa variabilité à l'intérieur d'un génome (partie 3.2 page 29).

### 3.1 Diversité inter-génomique

La variabilité du taux de GC entre génomes est quelque chose de bien connu, comme décrit plus haut (voir figure 2.2 page 20 notamment). La distribution du taux de GC des génomes complets de la base Hogenom est décrite dans la figure 3.1.

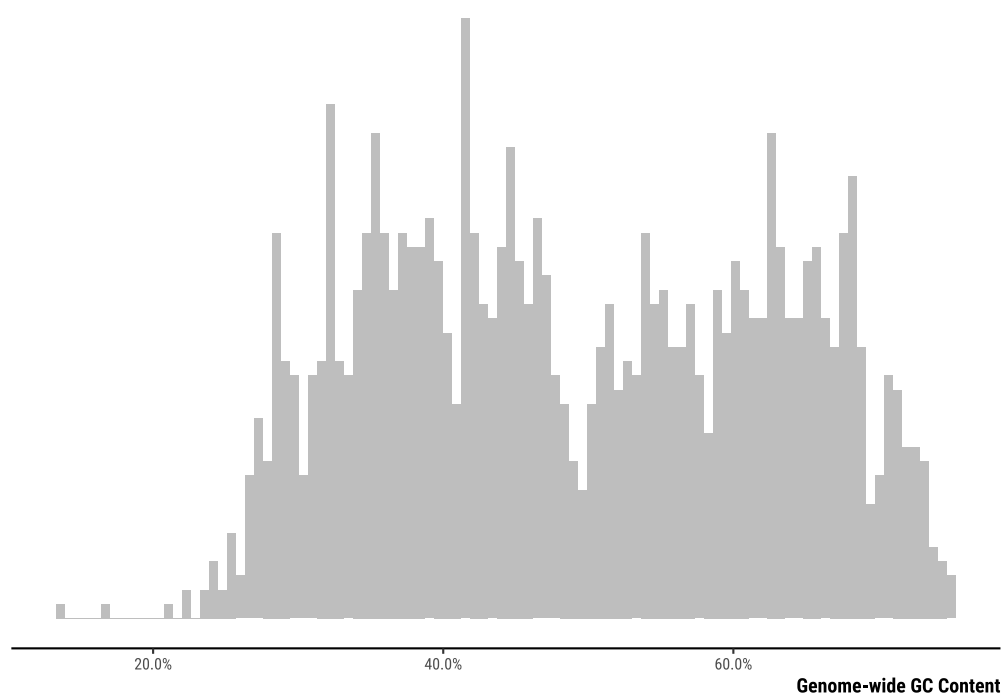
Il se dessine nettement une distribution bimodale caractéristique, avec un premier mode autour d'un taux de GC de 40%, et un second autour de 60%. À l'évidence une bonne part de cette bimodalité a trait

18. PENEL et al. 2009, Hogenom est une base de données comprenant les séquences génomiques de 12 326 bactéries et 224 archées, et qui rassemble leur gènes en

familles d'homologie, « homologie » est définie par des critères de similarité de séquence.

### 3. DISTRIBUTION DES TAUX DE GC ENTRE GÉNOMES

**FIGURE 3.1 –** Distribution du taux de GC génomique parmi les génomes complets de Hogenom7. (Pour cette figure et toutes les autres de ce chapitre provenant de la base de donnée Hogenom, N = 1636, dont 1472 bactéries et 146 archées. Ces génomes correspondent à l'ensemble des génomes complets, non fragmentés, et à un échantillonnage aléatoire d'un génome par espèce, dans le cas des espèces où plusieurs génomes sont disponibles.)



à l'histoire évolutive des génomes; il est de notoriété commune que certains *phyla* entiers regroupent des génomes qui tendent à un GC% élevé — que l'on pense aux Actinobactéries par exemple.

#### *Distribution phylogénétique du GC%*

Quelle est alors la répartition phylogénétique de ce trait? C'est ce que nous avons tenté de représenter dans la figure 3.2. Se distinguent nettement des taxons entiers où la variabilité du taux de GC est relativement restreinte, à l'exemple des  $\gamma$ -Protéobactéries (à 40'); d'autres où ce trait varie davantage, à l'exemple des  $\alpha$ -Protéobactéries (à 35'). Peut-être que l'exemple le plus marquant est celui des Terrabactéries (celles que l'on dit Gram positives), qui contiennent à la fois des bactéries à taux de GC élevé (les Actinobactéries, par exemple le genre *Streptomyces*), d'autres à taux de GC bas (les Firmicutes, dont le genre *Bacillus*); certains auteurs y font référence comme aux « Bactéries Gram+ à haut et bas GC » [DAUBIN et PERRIÈRE 2003].

La figure 3.3 décrit cette variabilité au sein d'un phylum donné, pour les huit groupes les plus représentés dans la base de donnée (les autres groupes sont rassemblés dans la catégorie « other »).

Là encore la dispersion du taux de GC à l'intérieur d'un phylum varie selon les taxons : les Tenericutes (qui contiennent le genre *Mycoplasma* par exemple) regroupent exclusivement des organismes à bas taux de GC (jusqu'à preuve du contraire). La variabilité du taux de GC que l'on



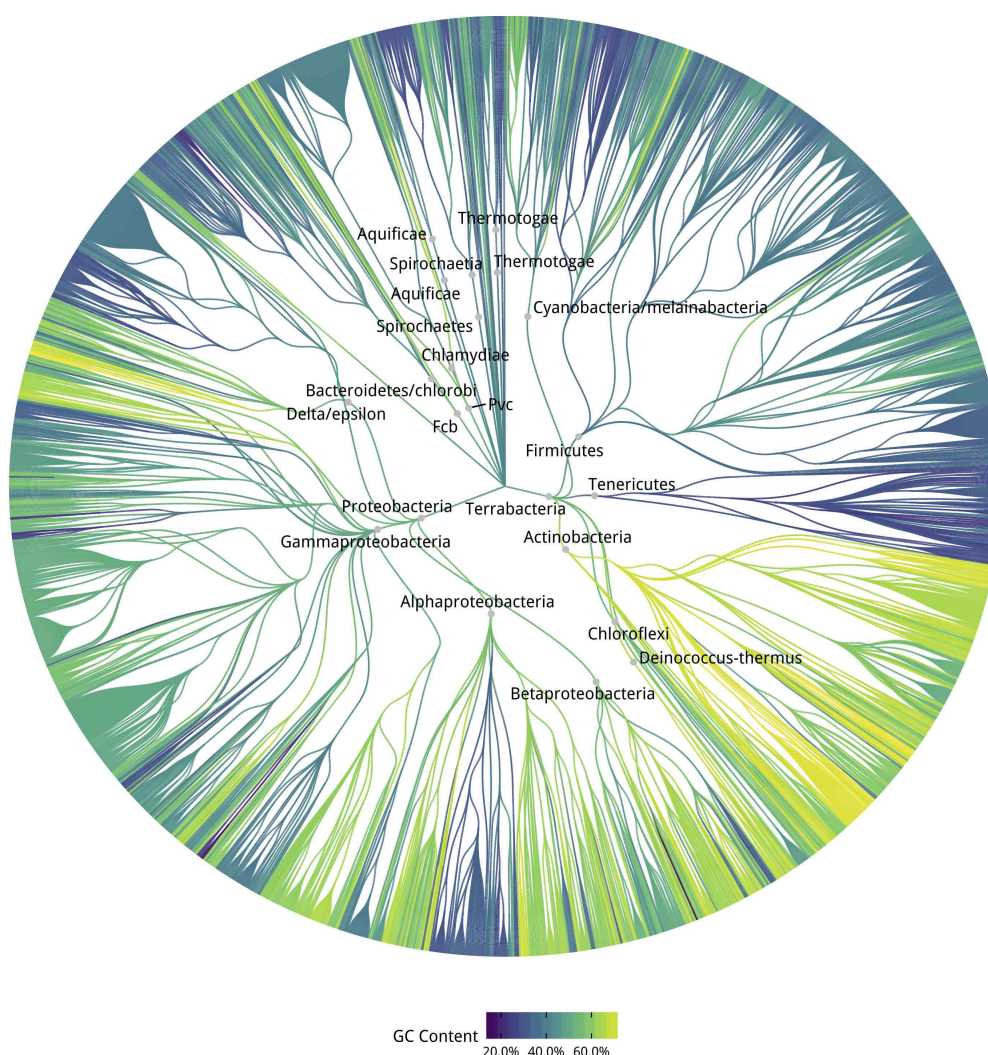


FIGURE 3.2 – Distribution phylogénétique du taux de GC génomique. La couleur d’une branche représente le taux de GC médian aux différentes feuilles de l’arbre qui lui sont rattachées (la couleur de la branche de la racine au noeud des Terrabactéries représente le taux de GC médian des Terrabactéries.) Les *phyla* contenant le plus de représentants ont été indiqués à la base de leur branche. L’arbre repose sur la taxonomie annotée dans la Taxonomy Database du NCBI; les longueurs de branche n’ont pas de sens.

observe dans le clade le plus représenté, celui des Protéobactéries, se retrouve également dans d’autres clades.

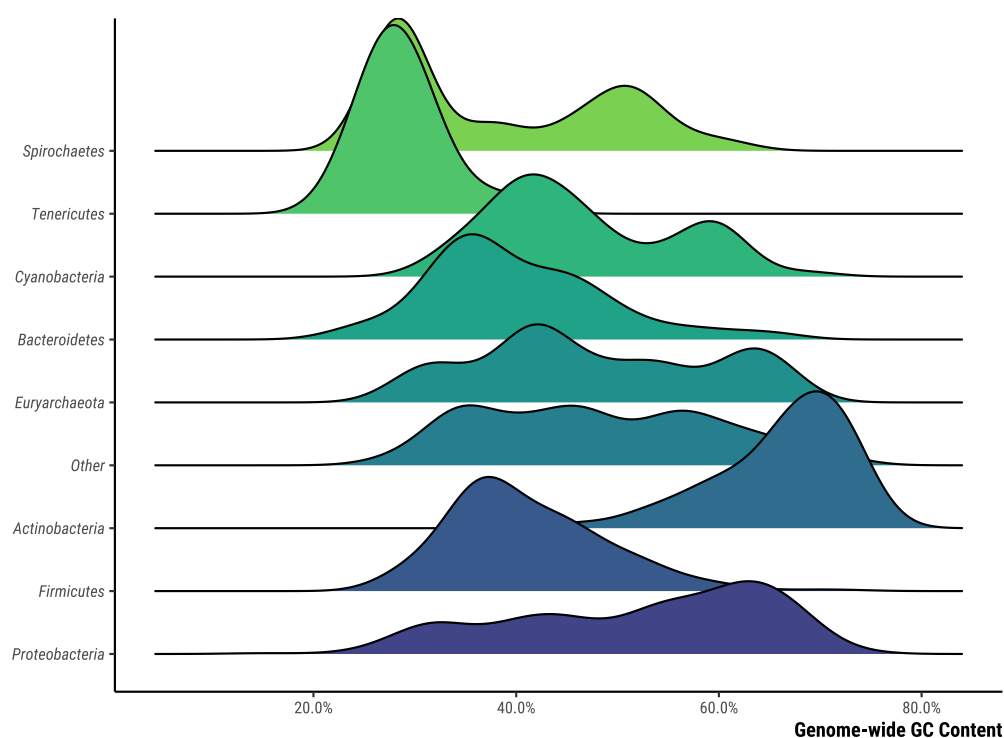
De sorte que l’on voit que la variabilité du taux de GC, si elle dépend pour une part de contingences historiques, ne s’y réduit pas : on ne peut pas déterminer l’ensemble des déterminants de la variabilité du taux de GC par la seule connaissance de la taxonomie.

## 3.2 Diversité intra-génomique

S’il est intéressant de quantifier globalement la composition en base d’un génome, la division du génome en ce qu’on pourrait appeler des « catégories fonctionnelles » permet d’identifier des structures en termes

### 3. DISTRIBUTION DES TAUX DE GC ENTRE GÉNOMES

FIGURE 3.3 – Variabilité intra-phylum en taux de GC génomique



de composition en base plus fines. C'est, sauf erreur de notre part, l'article de MUTO et OSAWA [1987] qui a popularisé cette approche pour ce qui concerne les bactéries, alors que les bases de données de séquence commençaient tout juste à sortir du berceau (la première version de GenBank date de 1982). Il est probablement l'un des premiers à avoir démontré cette caractéristique désormais bien connue de la composition en base : elle n'est pas tout à fait homogène selon que l'on considère des positions fortement contraintes par la sélection (les ARN de transfert, ou les premières et deuxième positions des codons), ou que l'on considère des positions *a priori* neutres ou presque neutres d'un point de vue fonctionnel (la troisième position des codons, les régions dites intergéniques, etc...).

C'est bien décrit par la figure 3.4 : si l'on corrèle (en A) le taux de GC d'un génome avec celui des positions codantes (Protein, tRNA et rRNA) ou non (Spacer), l'on voit que les positions les moins contraintes semblent présenter un taux de GC inférieur à celui du génome pour les génomes riches en AT, supérieur pour les génomes riches en GC ; c'est vrai aussi de la troisième position des codons (en B). Cela fait sens pour les ARN ribosomiaux : ils font probablement partie des gènes sur lesquels la sélection pèse le plus fort, aussi leur composition en base ne peut être prédite par celle du génome. En fait l'apport principal de ces figures est de montrer qu'il existe une corrélation linéaire positive forte

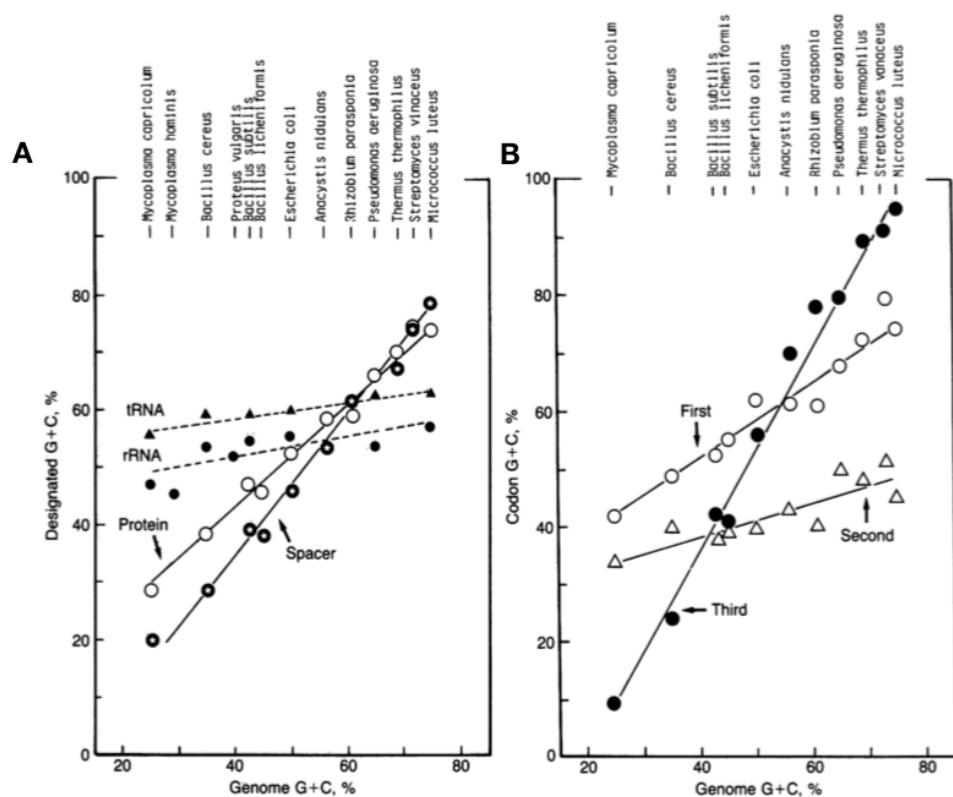


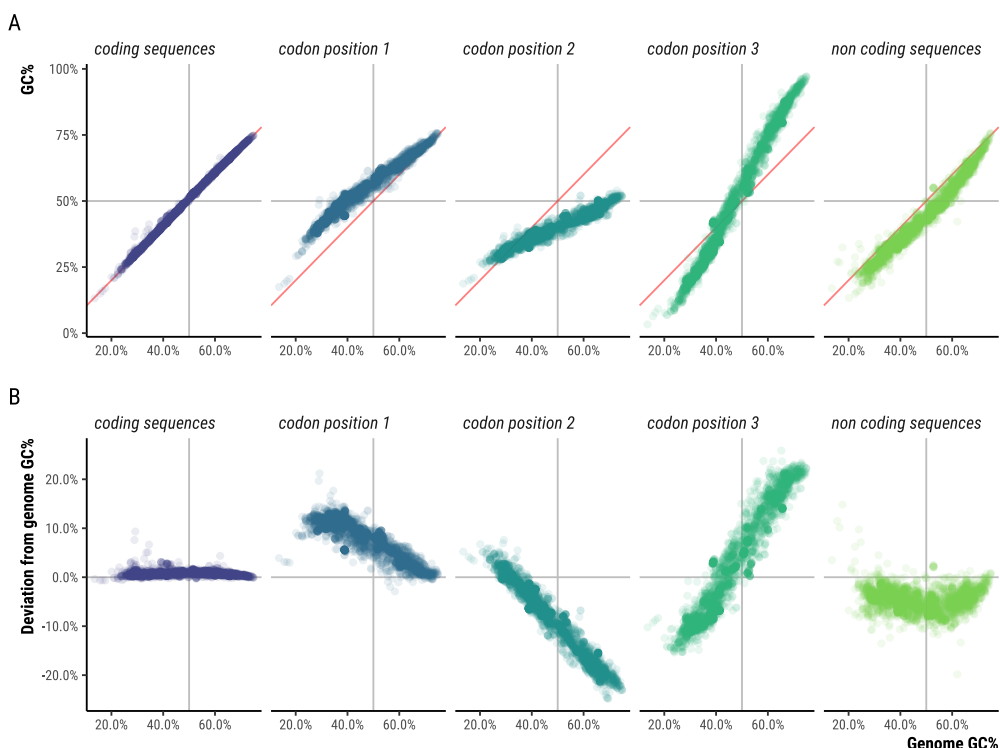
FIGURE 3.4 – Extrait de MUTO et OSAWA [1987]

entre d'une part le taux de GC des séquences codant les protéines (CDS), qui composent entre 70 et 80% d'un génome bactérien, et le taux de GC génomique ; et d'autre part le taux de GC des séquences intergéniques (Spacer) et le taux de GC génomique. « Ainsi, pour une bactérie donnée, les taux de GC des régions intergéniques, des gènes de protéines, [...] sont tous biaisés dans la même direction que le taux de GC du génome entier. » Et d'ajouter plus loin : « Ces colinéarités montrées dans la figure [3.4] soutiennent fortement l'idée que la pression AT/GC a largement influencé le taux de GC de l'ADN entier au cours de l'évolution. » En clair, l'idée force dégagée de ce genre d'analyse était que le taux de GC des génomes est pour l'essentiel le fait de patrons de substitution globalement biaisés vers A·T ou G·C, de sorte que les génomes riches en GC ont tendance à résulter de patrons de substitution biaisés vers G·C, au contraire des génomes riches en AT. Il s'agissait donc d'une confirmation directe des prédictions de la théorie de Sueoka, que ce dernier s'empressa de rappeler au goût du jour dans son article de 1988 [SUEOKA 1988].

Nous avons simplement reproduit ce genre d'analyse en utilisant les génomes complets de Hogenom ; c'est ce que nous avons représenté dans la figure 3.5. Nous avons indiqué en A la relation entre le taux

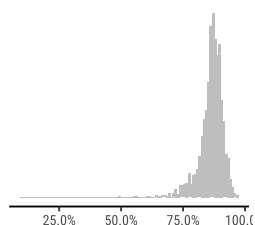
### 3. DISTRIBUTION DES TAUX DE GC ENTRE GÉNOMES

**FIGURE 3.5 –**  
Distribution de la composition en base à différentes positions des génomes de bactéries et d'archées en fonction du taux de GC global.



de GC d'une catégorie de site génomique donnée et le taux de GC génomique (ce genre de graphe est parfois appelé « graphe de Sueoka », abusivement visiblement, puisque ce sont Muto et Osawa qui les ont introduit). (La droite rouge est la projection en y du taux de GC génomique en x.) En B a été représenté l'écart entre le taux de GC de ces mêmes catégories et le taux de GC génomique (ce qui correspond géométriquement à l'écart entre les points et la droite rouge en A). Il est évident que :

1. le taux de GC des séquences codantes est étroitement corrélé avec le taux de GC génomique, puisque l'essentiel du génome d'une bactérie est constitué de séquences codantes (voir figure 3.6);
2. le taux de GC des séquences non codantes l'est également; or les séquences non-codantes étant moins contraintes par la sélection que les séquences codantes, on est fondé à penser que ce sont, et pour les séquences codantes et pour les séquences non-codantes, les mêmes processus qui déterminent l'essentiel de la variabilité observée;
3. la composition en base des trois positions des codons est également corrélée avec celle du génome;
4. cependant, les trois positions ne sont pas influées par les processus qui déterminent la composition en base de la même façon : les



**FIGURE 3.6 –**  
Distribution de la proportion de séquences codantes.

positions 1, et *a fortiori* les positions 2 sont les plus contraintes par la sélection, puisque c'est à ces positions que la probabilité qu'une mutation soit non-synonyme est la plus élevée;

5. la troisième position des codons est la moins sujette à l'action de la sélection, de par la redondance du code génétique. On observe ainsi que le taux de GC de la troisième position des codons tend à être plus riche en AT dans les génomes riches en AT que le génome lui-même, et tend inversement à être plus riche en GC dans les génomes riches en GC que le génome lui-même.

Selon SUEOKA [1988] et MUTO et OSAWA [1987], ces observations suggèrent que les différences de composition en base entre génomes sont le fait de biais *mutationnels*. En effet, on observe que les processus à l'origine des variations de composition en base entre génomes opèrent sur l'ensemble du génome (codants ou non-codants, synonymes ou non); or le biais mutationnel n'est pas censé distinguer entre séquences fonctionnelles ou non, et donc devrait influencer l'ensemble du génome. Sur les positions contraintes par la sélection (positions 1 et 2 des codons), l'influence du biais mutationnel est probablement contrebalancée par la sélection. Sur les positions les moins contraintes (positions 3 des codons, intergénique), les différences de composition en base entre génomes sont très élevées. MUTO et OSAWA écrivent ainsi que « ces faits [...] suggèrent fortement qu'une pression mutationnelle biaisée s'est exercée uniformément sur l'ensemble de l'ADN ».

Si ces travaux confirment clairement que les processus qui causent les différences de taux de GC entre génomes ne sont pas liés au codage des protéines, ils ne permettent pas d'exclure l'hypothèse sélectionniste d'une pression de sélection sur la composition en base globale des génomes. En effet, une hypothèse qui serait tout aussi compatible avec ces observations serait que les patrons mutationnels sont identiques d'une espèce à l'autre, mais que les probabilités de fixation varient en raison de différences de pression de sélection pesant sur la composition en base du génome, de manière globale.

Néanmoins, ces travaux contribuent à conforter de façon majeure l'hypothèse de *pression de mutation directionnelle*.<sup>19</sup> C'est près de vingt ans après cette quasi-confirmation que les contours de l'hypothèse ont commencé à bouger, puisque les données de polymorphisme, qu'il devenait de plus en plus commode de rassembler avec la « démocratisation » du séquençage de génome complet, ont révélé que la composition en

19. C'est sans doute parce qu'il n'était pas très clair de quelle façon l'hypothèse alternative sélectionniste pouvait être

testée, pas plus que n'était claire la forme que cette pression de sélection pouvait prendre.

### 3. DISTRIBUTION DES TAUX DE GC ENTRE GÉNOMES

---

base des procaryotes était tout compte fait moins déterminée par l'équilibre mutationnel qu'escompté.



## PROCESSUS MUTATIONNELS

---

Car, au niveau le plus fondamental de la vie, les jeux du code et du décodage laissent place à un aléa qui, avant d'être maladie, déficit ou monstruosité, est quelque chose comme une perturbation dans le système informatif, quelque chose comme une « méprise ». À la limite, la vie — de là son caractère radical — c'est ce qui est capable d'erreur. Et c'est peut-être à cette donnée ou plutôt à cette éventualité fondamentale qu'il faut demander compte du fait que la question de l'anomalie traverse de part en part toute la biologie. À elle aussi qu'il faut demander compte des mutations et des processus évolutifs qu'elles induisent. Elle également qu'il faut interroger sur cette erreur singulière, mais héréditaire, qui fait que la vie a abouti avec l'homme à un vivant qui ne se trouve jamais tout à fait à sa place, à un vivant qui est voué à « errer » et à « se tromper ».

Michel Foucault, Dits & Ecrits, tome IV texte n°361

L'article de MUTO et OSAWA [1987] précédemment décrit a contribué à ce que l'hypothèse dominante pour expliquer la variabilité des taux de GC entre génomes et au sein d'un génome soit de nature neutraliste. L'hypothèse en question peut être formulée de la façon suivante : « La composition en base est un trait qui évolue de façon neutre sous l'influence de spectres mutationnels différents entre génomes et au sein d'un génome. » Deux types de travaux sont venus à la fois conforter cette hypothèse et en appeler à son amendement : 1) les travaux consacrés à l'analyse du polymorphisme, rendus possibles par l'augmentation du nombre de génomes complets dans les bases de données (paragraphe page 36); et 2) les travaux qui ont analysé directement les patrons de mutations par des expériences d'accumulation de mutations ou assimilées (paragraphe page 38). C'est dans cet ordre que je les présenterai dans la partie 4.1.

Je passerai ensuite en revue les différents éléments moléculaires que l'on soupçonne ou que l'on sait conditionner la variabilité du taux de GC (section 4.2). Ils comprennent la diversité des mécanismes de réparation de l'ADN ou sa polymérisation, comme l'indiquent deux types de preuves différentes : les données expérimentales sur la mutagenèse (paragraphe page 46), et les données comparatives sur la répartition phylogénétique des gènes impliqués dans lesdits mécanismes (paragraphe page 58). Je mentionnerai ici le fait que la composition en base

n'est pas tout à fait homogène entre les deux brins de la double-hélice, en raison de l'asymétrie liée à leur réplication (paragraphe page 53).

### 4.1 Balance mutationnelle : confirmations et mise en balance

Cette partie se fixe deux buts :

1. montrer que la pression mutationnelle est le déterminant majeur de la variabilité en taux de GC entre les génomes ;
2. montrer que cette pression seule ne suffit pas à expliquer *toute* la variabilité de la composition en base, parce qu'il existe un biais de fixation en faveur des bases G et C.

Deux types de preuves ont contribué à ce que l'on prenne conscience du fait que toutes les positions du génome ne sont pas affectées de la même façon par la pression mutationnelle, et que d'autres forces évolutives sont requises pour expliquer l'écart entre la composition en base attendue à l'équilibre mutationnel et la composition en base observée. Par ordre chronologique, ce sont les données de polymorphisme (intra-spécifique) qui l'ont d'abord suggéré (paragraphe page 36), avant que des travaux empiriques, aussi rigoureux que fastidieux, ne viennent les confirmer directement (paragraphe page 38).

#### *Inférence d'un biais de fixation en faveur des bases GC par analyse du polymorphisme*

En 2006, l'article de Eduardo P. C. ROCHA, TOUCHON et al. [2006] montre une première fois chez des Bactéries que la composition en base des génomes des sept genres analysés n'est pas celle attendue à l'équilibre mutationnel. La méthodologie employée est à peu de choses près la même que celle des deux articles suivants, c'est pourquoi elle sera détaillée ici. Elle repose sur le regroupement des gènes de plusieurs génomes complets d'un même genre en famille d'*orthologues*, c'est-à-dire que les gènes d'une famille partagent un ancêtre commun — ce sont des gènes homologues —, et leurs divergences sont le fait d'évènements de spéciation (non de duplication). À partir de l'alignement des différents gènes, le taux de GC des codons aux positions dites quatre fois dégénérées (*fourfold degenerate*) est calculé (noté GC4). Les substitutions qui ont lieu à ces positions sont également détectées en les polarisant par rapport à l'orthologue le plus fréquent : si l'ensemble des espèces aux branches terminales de l'arbre possèdent à une position une base A, quand tous les autres possèdent une base T, il s'agit plus probablement



d'une substitution  $T \rightarrow A$  que  $A \rightarrow T$ . De l'ensemble de ces positions peut-être déduit un patron de substitution aux positions dites neutres. On peut alors comparer la composition en base observée à celle attendue d'après le patron de substitution. Eduardo P. C. ROCHA, TOUCHON et al. [2006] montrent que pour les quelques genres bactériens analysables alors, la composition en base n'est pas celle attendue à l'équilibre déduit du patron de mutation.

Deux articles majeurs sont venus conforter cette observation. Le premier est celui de HERSHBERG et PETROV [2010], dans lequel les auteurs se sont intéressés à calculer le patron de substitution dans les génomes de cinq espèces pathogènes à expansion clonale, ayant été séquencés à la suite d'épidémies. Les génomes en question proviennent d'évènements récents d'expansion d'une souche ancestrale, essentiellement clonalement, de sorte que 1) l'action de la sélection naturelle est supposée s'être restreinte aux locus impliqués dans la pathogénie et 2) les substitutions observées sont plus probablement le fait de mutations ayant eu lieu au cours de la radiation que d'évènements de recombinaison entre souches. Il est alors possible d'estimer le spectre mutationnel, en comptant les différentes substitutions, puisqu'elles sont supposées neutres. Ils montrent, comme attendu, que les génomes d'espèces à taux de GC bas, comme *Bacillus anthracis*, tendent à accumuler plus de transitions de GC vers AT que l'inverse. Mais, et c'est en cela qu'ils viennent s'opposer à l'hypothèse de Sueoka, les génomes à taux de GC intermédiaires — comme *Yersinia pestis* ou *Salmonella typhi* — ou haut — comme *Burkholderia mallei*, accumulent également plus de transitions de GC vers AT.<sup>20</sup> Ils montrent ainsi, pour les quatre espèces étudiées, que le taux de GC attendu à l'équilibre mutationnel est inférieur au taux de GC actuel. C'est bien la preuve qu'aucun de ces génomes n'est à l'équilibre du point de vue de leur composition en base. Puisque la pression mutationnelle est biaisée vers les bases A et T, il doit y avoir un biais de fixation des bases G et C.

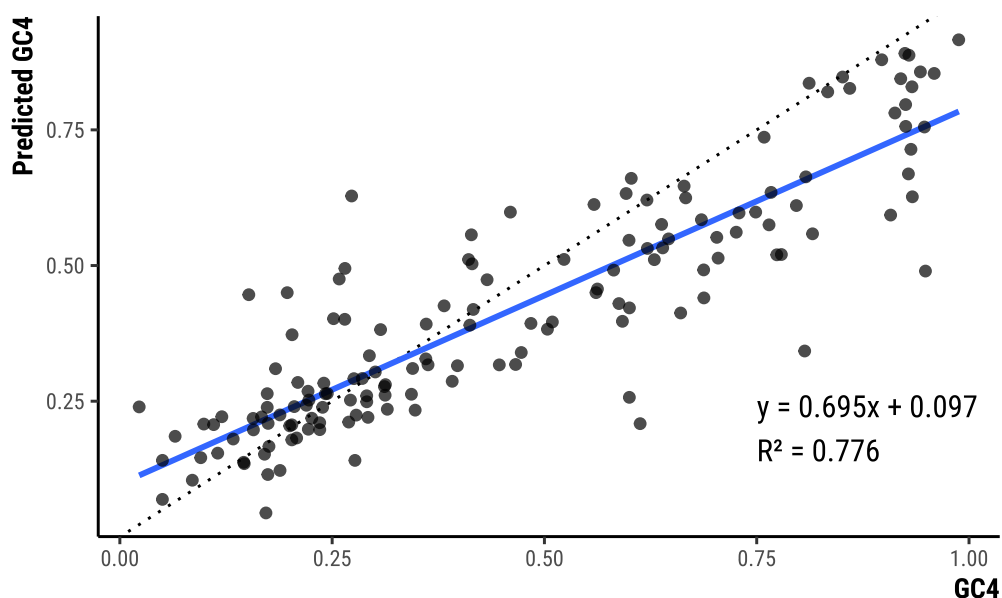
C'est à la même conclusion que parviennent HILDEBRAND et al. [2010]. À la différence de HERSHBERG et PETROV, les auteurs s'appuient sur un jeu de données plus grand, comprenant 149 bactéries, pour lesquelles au moins huit séquences étaient disponibles, et avec lesquelles il était possible de calculer un GC4 incluant suffisamment de positions. De façon similaire aux deux articles déjà mentionnés, ils estiment le patron de substitution aux positions les moins contraintes par la sélection, et comparent le taux de GC auquel il devrait conduire à celui observé (figure 4.1). On voit clairement que ces deux taux sont très corrélés

20. Évidemment, les probabilités pour un génome riche en AT et un génome riche en GC de muter de GC vers AT ne sont pas les mêmes. Aussi faudrait-il, en toute

rigueur, rapporter plutôt le taux de substitutions de GC vers AT conditionnellement au taux de GC des sites étudiés.

#### 4. PROCESSUS MUTATIONNELS

FIGURE 4.1 –  
Corrélation entre le  
taux de GC aux po-  
sitions quatre fois  
dégénérées (GC4) at-  
tendu à l'équilibre et  
GC4 observé. (n = 149,  
données tirées de HIL-  
DEBRAND et al. 2010)



( $R^2 = 0,776$ ), ce qui suggère que la pression mutationnelle détermine l'essentiel de la variabilité du taux de GC de ces positions supposées les moins soumises à la sélection. Néanmoins, 62,4 % des points se situent sous la droite de correspondance  $y = x$ . Ceci démontre, pour plus des deux tiers des génomes étudiés, que le taux de GC observé est supérieur à celui attendu à l'équilibre [Eduardo P. C. ROCHA et FEIL 2010].

D'après ces analyses du polymorphisme, on peut en conclure qu'il existe un biais de fixation en faveur des bases G et C dans les génomes de presque toutes les bactéries. Néanmoins, il s'agit d'une observation indirecte du patron de mutation, puisque les substitutions étudiées au travers du polymorphisme ont déjà subit l'action de forces comme la sélection naturelle. La meilleure façon d'obtenir une estimation non-biaisée (la moins biaisée possible en tout cas) du patron de mutation est encore de l'étudier le plus directement qu'il soit possible le faire, notamment à travers des expériences d'accumulations de mutations.

##### *Démonstration d'un biais de fixation par expériences d'accumulation de mutations*

C'est vraiment avec la compilation, par LONG et al. [2018], des expériences d'accumulation de mutation réalisées, entre autres, par l'équipe de Michael Lynch, que l'on obtient la confirmation directe de l'existence d'un biais de fixation en faveur des bases G et C. Le principe d'une expérience d'accumulation de mutations a été développé dans les années 1960 par BATEMAN [1959] puis MUKAI [1964], c'est pourquoi elle est parfois qualifiée de méthode Bateman-Mukai [DRAKE 2012]. Cette

méthode repose sur des cultures successives des organismes d'intérêt entrecoupées de goulot d'étranglement drastiques, de façon à ce que la probabilité de fixation des mutations ne dépendent presque pas de leur effet sur la fitness. Dans le cas des bactéries, ces cultures successives se font en isolant et repiquant régulièrement une unité formant colonie (UFC), et ce pendant un grand nombre de cycles de culture — 217 pour 75 lignées indépendantes chez *Vibrio fischeri* par exemple [DILLON, SUNG, LYNCH et al. 2015]. À la suite de quoi le génome de souches isolées est séquencé, et comparé au génome de la souche initiale. L'absence de compétitions au cours de l'expérience implique que les seules mutations léthales sont contre-sélectionnées.<sup>21</sup> On peut donc obtenir une estimation directe du taux et du patron de mutation. L'accumulation de mutation (MA) est une approche très puissante pour déterminer les effets d'un gène sur le spectre mutationnel, aussi en reparlerons-nous au paragraphe sur la mutagenèse page 46.

La contribution de LONG et al. [2018] consiste à rassembler et compiler dans un article de synthèse les apports d'expériences d'accumulation de mutation réalisées aussi bien chez des eucaryotes (pas seulement unicellulaires) que chez des bactéries. Leur article décrit 22 bactéries différentes.

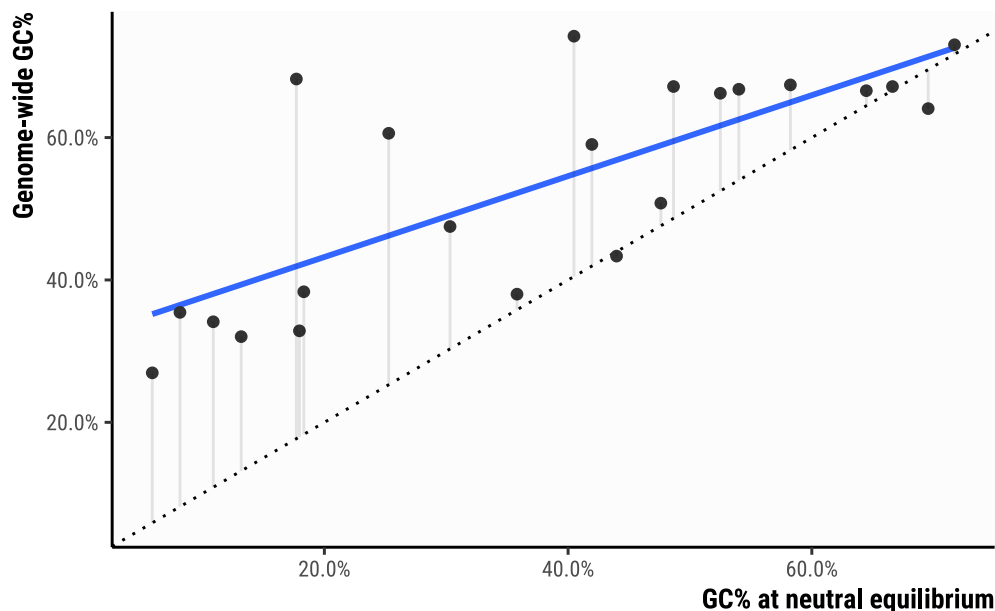


FIGURE 4.2 – Relation entre taux de GC observé et taux de GC à l'équilibre mutationnel. La droite bleue est la droite de régression linéaire; la courbe en pointillés est la droite d'équation  $y = x$ . (Données tirées de LONG et al. 2018, en se limitant aux Procaryotes.)

La figure 4.2 est une adaptation par nous d'une partie de la figure 1 de leur article, en se limitant aux espèces de Procaryotes. On y voit très clairement d'abord qu'il y a bien une corrélation linéaire positive

21. Mais SUNG, ACKERMAN, GOUT et al. [2015] montrent qu'il y a probablement

une forme de sélection aux effets mineurs même dans ce type d'expérience.

#### 4. PROCESSUS MUTATIONNELS

---

Pour rappel, le taux de GC à l'équilibre « substitutionnel », à ne pas confondre avec le taux de GC attendu à l'équilibre mutationnel peut s'écrire :

$$GC_{eq} = \frac{v}{v+u}, \text{ où}$$

$$v = \mu_{AT \rightarrow GC} \cdot \phi_{AT \rightarrow GC}$$

$$u = \mu_{GC \rightarrow AT} \cdot \phi_{GC \rightarrow AT}$$

Voir équation 1.1 page 5.

( $R^2 = 0.568$ ) entre le taux de GC attendu à l'équilibre mutationnel et le taux de GC observé ( $p = 3 \cdot 10^{-5}$ ); ensuite que presque tous les génomes ont un taux de GC supérieur à celui attendu à l'équilibre mutationnel, comme prédit par les données de polymorphisme. C'est donc la confirmation que :

1. le patron de mutation explique 56,8% de la variance du GC génomique, et donc que la mutation est un déterminant fort, sinon le plus fort, du taux de GC des génomes;
2. presque aucun génome étudié n'est à l'équilibre du point de vue de sa composition en base, et donc qu'il y a un biais de fixation en faveur des bases G et C.

L'une des avancées majeures de ces travaux tient à ce qu'ils démontrent que le spectre mutationnel contribue à diminuer le taux de GC du génome, et ce de façon quasi-universelle. C'est donc aussi bien une confirmation de la théorie de Sueoka concernant la balance mutationnelle, qu'un appel à son amendement, puisqu'elle ne permet pas d'expliquer toute la variabilité observée.

*Composition en base à l'équilibre :  
biais de fixation en faveur des bases G et C*

*Il faut donc recourir à une autre force évolutive pour expliquer l'écart que l'on observe entre le taux de GC attendu à l'équilibre mutationnel et le taux de GC observé.*

C'est l'une des prémisses essentielles de nos travaux de thèse. Pour reprendre la notation indiquée dans le prologue :

$$\phi_{GC \rightarrow AT} < \phi_{AT \rightarrow GC}$$

c'est-à-dire, la probabilité de fixation d'une mutation de GC vers AT est inférieure à celle d'une mutation de AT vers GC.

Pour en donner un exemple, supposons, pour un génome donné, que le taux de mutation GC→AT soit m fois plus élevée que le taux de mutation AT→GC, donc que  $\mu_{GC \rightarrow AT} = m \cdot \mu_{AT \rightarrow GC}$  [LONG et al. 2018]. On peut en déduire  $GC_n$ , la « composition en GC attendue à l'équilibre neutre (où la mutation est la seule force évolutive directionnelle) » [LONG et al. 2018]. Notons maintenant  $\kappa$  le ratio des probabilités de fixation d'une mutation de GC→AT et de AT→GC. Par exemple,  $\kappa$  vaut 1 si  $\phi_{GC \rightarrow AT} = \phi_{AT \rightarrow GC}$ , et  $\kappa = 5/6$  si  $\phi_{GC \rightarrow AT} = 5/6 \cdot \phi_{AT \rightarrow GC}$ .

Dans ce cas, on peut écrire  $GC_n$  et  $GC_{eq}$  :

$$\begin{aligned}
 GC_n &= \frac{\mu_{AT \rightarrow GC}}{\mu_{GC \rightarrow AT} + \mu_{AT \rightarrow GC}} & GC_{eq} &= \frac{\mu_{AT \rightarrow GC}}{\mu_{AT \rightarrow GC} + \mu_{GC \rightarrow AT} \cdot \kappa} \\
 &= \frac{\frac{1}{m} \mu_{GC \rightarrow AT}}{\mu_{GC \rightarrow AT} + \frac{1}{m} \mu_{GC \rightarrow AT}} & &= \frac{\frac{1}{m} \cdot \mu_{GC \rightarrow AT}}{\mu_{GC \rightarrow AT} + \frac{1}{m} \cdot \mu_{GC \rightarrow AT} \kappa} \quad (4.1) \\
 &= \frac{1}{1 + m} & &= \frac{1}{1 + m\kappa}
 \end{aligned}$$

Par exemple, si  $m = 1/2$ ,  $GC_n = 1/(1/2)+1 \approx 0,67$  ; si le taux de mutation  $AT \rightarrow GC$  est deux fois supérieur au taux de mutation  $GC \rightarrow AT$ , le taux de GC attendu à l'équilibre mutationnel est de 67 %. Si maintenant  $\kappa = 5/6$ , alors  $GC_{eq} = 1/(1+(1/2)(5/6)) \approx 0,71$  ; si la probabilité de fixation d'une mutation  $GC \rightarrow AT$  vaut  $5/6$  de celle d'une mutation  $AT \rightarrow GC$ , le taux de GC attendu à l'équilibre substitutionnel est de 71 %.

Quelles sont les forces que l'on peut envisager pour expliquer que  $\kappa$  soit inférieure à 1, c'est-à-dire pour que la probabilité de fixation d'une mutation  $GC \rightarrow AT$  soit plus inférieure à celle d'une mutation  $AT \rightarrow GC$  ? À ce jour, on peut en proposer deux, comme l'ont fait LONG et al., HILDEBRAND et al. ou encore Eduardo P. C. ROCHA et FEIL : la sélection naturelle ou la conversion génique biaisée. Ces deux forces font l'objet des deux chapitres qui suivent.

Pour donner une idée de l'intensité des forces en jeu, je reviens maintenant sur le taux de GC à l'équilibre « substitutionnel ». En utilisant les travaux théoriques de BULMER [1991], LONG et al. reformulent l'équation 4.1 en

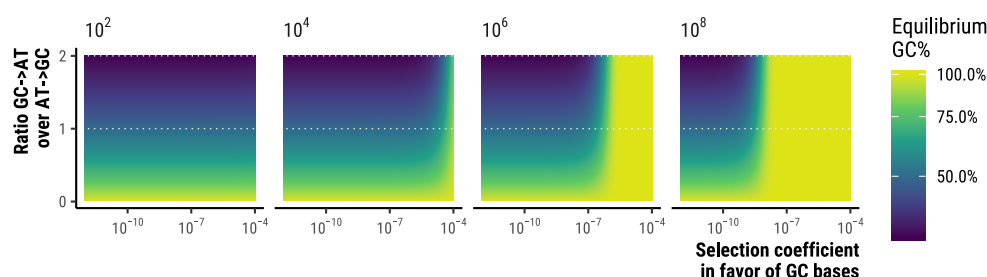
$$GC_{eq} = \frac{1}{1 + m e^{-S}}$$

où  $S = 2\psi N_e s$ ,  $N_e$  étant la taille de population efficace,  $\psi$  le degré de ploïdie (1 pour les haploïdes, 2 pour les diploïdes), et  $s$  l'avantage sélectif des bases G et C par rapport aux bases A et T.

La figure 4.3 montre la variation du taux de GC attendu à l'équilibre  $GC_{eq}$  en fonction de  $m$ , de  $s$  et de  $N_e$ . On y voit que pour des tailles de population très petites ( $N_e \in [10^2, 10^4]$ ), l'avantage sélectif lié aux bases G et C n'a pas d'influence sur le taux de GC à l'équilibre, et seul  $m$  le ratio des probabilités de mutation  $AT \rightarrow GC$  et  $GC \rightarrow AT$  influe. Néanmoins, dans de grandes populations, où  $N_e \geq 10^8$  — ce qui est réaliste dans des populations bactériennes [BATUT et al. 2014 ; BOBAY et OCHMAN 2018] —, même de très faibles coefficients de sélection tels que  $s \approx 10^{-7}$  prennent le pas sur le patron mutationnel et conduisent à une composition en base de 100 % de G et C. (Évidemment ça n'est pas très réaliste pour l'ensemble du génome, mais ça l'est pour les

FIGURE 4.3 –  
Composition en base attendue à l'équilibre en fonction du coefficient de sélection  $s$  en faveur des bases G et C, de la probabilité de mutation GC→AT par rapport à la probabilité de mutation de AT→GC, notée  $m$ , et de la taille efficace  $N_e$ ; avec  

$$GC_{eq} = \frac{1}{1+m e^{-s}}$$
et  $S = 2N_e s$ .



sites dits « silencieux » comme les positions des codons quatre fois dégénérées, d'après les observations de LONG et al.) Or il se trouve précisément que l'avantage que procure un biais tel que la conversion génique biaisée vers GC à une base, à l'égard de sa fixation, se comporte exactement comme un avantage sélectif [NAGYLAKI 1983b]. De sorte qu'on pourrait remplacer dans tout ce qui précède  $S$  par  $B$  l'intensité de la gBGC à l'échelle populationnelle et aboutir aux mêmes conclusions. (J'y reviendrai plus loin, au chapitre 6.)

Mais avant de détailler plus avant les deux forces de la sélection et des biais moléculaires, il me faut revenir sur les mécanismes moléculaires qui contribuent à la mutagenèse. Comme on l'a vu, la mutation est le déterminant le plus fort de la variabilité en composition en base, aussi je tenterai dans la partie qui suit d'explicitier les différents mécanismes que l'on sait contribuer à la variabilité de la balance mutationnelle.

## 4.2 Déterminants moléculaires de la balance mutationnelle

On l'a vu, la balance mutationnelle, c'est-à-dire l'équilibre qui tient entre le taux de mutations de AT vers GC d'une part, de GC vers AT d'autre part, est le principal déterminant de la variabilité en composition en base. Afin de comprendre ce qui contribue à l'équilibre de cette balance (ou à l'écart à cet équilibre), l'on mettra à profit deux types de preuves : 1) celles fournies par la littérature sur la mutagenèse (paragraphe page 46); et 2) celles que semble indiquer la distribution phylogénétiques de certains gènes et la relation avec le taux de GC des génomes (paragraphe page 58).

### Bref rappels sur la correction des mésappariements chez les Bactéries

Avant de décrire ce que les expériences de mutagenèse nous ont appris sur le patron de mutation, je souhaiterais simplement faire un bref tour d’horizon sur ce que l’on sait de la correction des mésappariements chez les Procaryotes. En effet, les mutations qui nous intéressent dans le cadre de l’évolution de la composition en base sont essentiellement des mutations ponctuelles, donnant lieu à l’échelle de la population à des polymorphismes portant sur une seule base, ou SNP (*single nucleotide polymorphism*).<sup>22</sup>

RADMAN et WAGNER [1986] définissent un mésappariement comme « n’importe quelle paire de base non-complémentaire dans un ADN double-brin ». Un mésappariement peut survenir à cause d’une erreur de réplication de l’ADN, au sein d’un intermédiaire de recombinaison homologue entre deux molécules différentes (*hétéroduplexe*), ou à cause de l’altération chimique de la double-hélice par des agents hydrolytiques, oxydants ou alkylants [BHAGWAT, HAO et al. 2016]. Ces mésappariements, qui ne sont à ce stade que des *pré-mutations* sont susceptibles d’être corrigées par une machinerie adaptée. On peut distinguer, comme l’ont fait JOSEPH et al. [2006], deux voies de réparation des mésappariements : 1) celles qui impliquent la réparation de longues portions d’ADN (*long-patch repair*), comme le MMR (*MisMatch Repair*), qui dépend généralement d’homologues des protéines MutS et MutL; et 2) celles qui impliquent de courtes portions (*short patch repair*), comme les voies du NER et BER — pour respectivement *nucleotide* et *base excision repair* (table 4.1).

22. Évidemment, c’est passer sous silence le rôle de réarrangements génomiques plus grands, au premier chef ceux occasionnés par l’acquisition d’un gène d’une espèce distante. C’est-à-dire, passer sous silence le rôle des transferts de gènes. Une bonne raison de le faire tient à ce que lesdits transferts à « grande distance » évolutive

sont relativement rares [BONHAM et al. 2017; POPA et al. 2011]; or ce sont ceux qui sont les plus susceptibles de changer drastiquement la composition en base. Il n’est pour autant pas exclu que des transferts récurrents ou par vagues puisse contribuer à la variation du taux de GC.

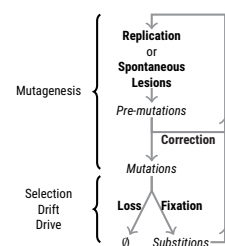


FIGURE 4.4 – Terminologie employée. (Voir page 6 pour la figure taille réelle.)



#### 4. PROCESSUS MUTATIONNELS

TABLE 4.1 – Sources, dommages et réparation des différentes lésions de l'ADN chez les Pro-caryotes. (Adaptée de JOSEPH et al. 2006.)

Sources	Dommages	Réparation
agents alkylants radicaux libres de l'oxygène rayons X	sites abasiques oxydations déaminations alkylations cassures simple-brin	directe
réactions spontanées radicaux libres de l'oxygène	O <sup>6</sup> -méthyl guanine 8-oxo guanine	Base Excision Repair (BER)
Rayonnements UV	dimères de pyrimidine	Nucleotide Excision Repair (NER)
Rayons X	Cassures double-brins (DBS)	Recombinaison homologue Jonction d'extrémités non-homologues (NHEJ)
Erreurs de réplication	Mésappariements Insertions Délétions	Mismatch Repair (MMR)

**Short Patch Repair** Parmi les voies du *Short Patch Repair*, l'une des plus importantes est liée au système de réparation par excision de base, (*base excision repair*, BER), dont la fonction première est, d'après KROKAN et BJØRÅS [2013] de « corriger les lésions de l'ADN résultant d'oxydations, de déaminations ou d'alkylations ». Ces dommages occasionnent de légères distortions de la structure tridimensionnelle de l'ADN, qui sont reconnues et réparées par des DNA-glycosylases spécifiques.<sup>23</sup> Ces glycosylases excisent l'une des bases du mésappariement, laissant un site abasique (une lésion simple brin, ou *nick*), qui est réparé par polymérisation via des DNA polymérases et ligaturé par des ligases généralistes (LigA). Ces mécanismes de type BER opèrent donc sur une courte portion d'ADN, et affectent uniquement la base excisée.

**Long Patch Repair** Une autre voie canonique de correction des mésappariements est celle qui implique de plus grandes portions de l'ADN, au premier chef le MMR. Les mésappariements dans la double-hélice qui résultent d'erreurs de polymérisation de l'ADN au cours de sa réplication, ou de lésions spontanées sont reconnues par le MMR. Chez *Escherichia coli*, probablement l'une des espèces pour lesquelles le MMR a été le plus étudié, cette machinerie repose essentiellement sur l'action des trois protéines MutS, MutL et MutH, mais en implique beaucoup d'autres. La protéine MutS est une protéine dimérique [LAMERS et al. 2000] qui se comporte comme une sorte de « pince coulissante » (*sliding clamp*) parcourant la double-hélice. Le changement de structure tridimensionnelle

23. Il devient de plus en plus prégnant que les DNA-glycosylases impliquées dans le BER sont impliquées dans d'autres

systèmes de « maintenance » du chromosome [A. L. JACOBS et SCHÄR 2012; MULLINS et al. 2019]



qu'occasionne un mésappariement à la double-hélice est reconnu par MutS, qui recrute alors la protéine MutL [GROOTHUIZEN et al. 2015], libère le site du mésappariement pour permettre son excision [SPIES et FISHEL 2015] et forme avec MutL un autre complexe de type *sliding clamp* [J. LIU et al. 2016], qui joue probablement un rôle stabilisant les hétéroduplexes [ELEZ et al. 2007]. MutL recrute ensuite l'hélicase UvrD et les exonucléases ExoX ou RecJ, qui dégradent l'un des deux brins<sup>24</sup> sur une portion de 3000 nt en moyenne chez *E. coli* [WAGNER et M. MESELSON 1976]; un nouveau brin d'ADN corrigé est ensuite synthétisé par les polymérases de type III, suivi pour terminer d'une ligature [KUNKEL et ERIE 2005].

La fonction première du MMR est la reconnaissance et la réparation des mésappariements liés aux erreurs de réplication; néanmoins il joue également un rôle majeur dans la réparation des mésappariements qui surviennent dans les hétéroduplexes de *recombinaison homologue* [THAM, KANAAR et al. 2016]. Je décrirai ici le modèle classique de réparation des cassures double-brin par recombinaison homologue. L'on verra cette réparation jouer un rôle clé pour ce qui concerne la conversion génique biaisée.

Les cassures double-brins sont des lésions importantes de l'ADN, qui sont léthales pour la cellule et empêchent la réplication de progresser. L'un des moyens utilisés pour les réparer est la recombinaison homologue, une autre est la voie NHEJ [KOWALCZYKOWSKI 2015]. La recombinaison homologue repose sur la recherche d'une séquence homologue à la séquence bordant la cassure et susceptible de servir de matrice pour sa réparation. Chez les Protéobactéries, les extrémités d'une cassure double-brins sont prises en charge par le complexe RecBCD, un complexe de nucléases et d'hélicases générant par résection 3'-5' des extrémités 3' simple-brin libres, qui sont ensuite liées à des protéines de la famille de RecA pour générer un filament nucléo-protéique [SYMINGTON 2014]. Chaque unité de RecA est liée à trois bases de l'extrémité simple-brin [Ja Yil LEE et al. 2015]. Le filament nucléo-protéique ainsi formé « parcourt » pour ainsi dire le génome avec un coefficient de diffusion estimé à  $7700 \text{ bp}^2 \text{ s}^{-1}$  [RAGUNATHAN et al. 2012], à la recherche de régions homologues. RAGUNATHAN et al. [2012] rapportent que la reconnaissance d'homologie peut nécessiter seulement 6 nucléotides. Cette reconnaissance repose sur l'appariement Watson-Crick entre le double-brin parcouru par le

24. Chez *E. coli*, le « choix » du brin dégradé dépend de l'état de méthylation des sites

GATC environnants, et est guidé par la courte incision pratiquée par MutH.

filament et l'extrémité sortante de la cassure [Ja Yil LEE et al. 2015]. Un degré d'appariement suffisant déclenche un changement de conformation synergique entre les différentes sous-unités de RecA, ce qui catalyse l'échange de brin et l'invasion des extrémités de la cassure dans la double-hélice de la région homologue, donnant lieu à un *hétéroduplexe*. L'invasion de brin peut, sous certaines conditions, s'ensuivre d'une re-synthèse du brin complémentaire et par conséquent de la réparation de la cassure.

La formation de l'hétéroduplex peut entraîner la formation de mésappariements, si les deux régions homologues ne sont pas identiques<sup>25</sup> Ces mésappariements peuvent être réparés notamment *via* le MMR.

Il faut ici mentionner que, au cours de la réparation des cassures par recombinaison homologue, la probabilité d'acceptation ou de rejet d'un brin dépend du niveau d'homologie entre les deux séquences pour deux raisons : la première, dont j'ai déjà parlé, repose sur l'appariement Watson-Crick catalysé par RecA ; la seconde tient à ce que les protéines du MMR sont susceptibles de provoquer le rejet du brin en déstabilisant l'hétéroduplexe et en bloquant la migration de branche, lorsqu'il présente trop de mésappariements [THAM, HERMANS et al. 2013].

Les gènes que j'ai décrit ici et ceux de la réplication sont vraiment ceux desquels dépendent les taux de mutation et l'équilibre de la balance mutationnelle. Chez *E. coli*, la fréquence de mésappariements de base au cours de la synthèse de l'ADN est de  $10^{-1} \text{ bp}^{-1} - 10^{-2} \text{ bp}^{-1}$  [JOSEPH et al. 2006]. L'activité de correction intrinsèque des polymérases le réduit à  $10^{-5} \text{ bp}^{-1} - 10^{-7} \text{ bp}^{-1}$ , et les systèmes de correction des erreurs post-réplicatifs le réduisent encore d'un facteur  $10^3$ , de sorte qu'il est généralement de l'ordre de  $10^{-10} \text{ bp}^{-1}$ .

Un bon moyen de connaître la fonction de ces gènes et leurs conséquences sur le spectre mutationnel consiste à étudier le phénotype de souches mutantes pour ces gènes. C'est l'étude de la mutagenèse, à laquelle la partie suivante sera consacrée.

#### *Mutagenèse et mécanismes mutateurs*

La *mutagenèse* est le nom donné aux *processus* qui génèrent la mutation. La mutation est l'une des forces essentielles à l'évolution. « En essence, écrivent BLOUNT et al. [2018], l'évolution est le fait de quatre processus fondamentaux, la mutation, la recombinaison, la sélection

25. Certains utilisent le terme d'*homéologues* [THAM, KANAAR et al. 2016] — je me

contendrai pour ne pas multiplier les termes de la polysémie d'*homologue*.

naturelle et la dérive génétique. [...] Trois de ces processus — à l'exception de la sélection naturelle — sont de nature stochastique, en ce sens que les variants spécifiques produits ou perdus à une génération résultent (ou semblent le faire) du hasard. Le hasard est un concept délicat, ceci dit. » L'étude des probabilités de mutation et des effets qu'elles ont sur la fitness de l'individu a montré que, si la mutation est bien un processus par essence indéterminé (ou contingent), certaines mutations sont plus probables que d'autres. Un argument fort concernant la nature stochastique de la mutation a été apporté par ROBERT et al. [2018]. En utilisant une technologie microfluidique, couplée à de la microscopie fluorescente permettant de visualiser à l'échelle de la cellule individuelle la survenue d'une pré-mutation au cours de la réplication, cette étude a démontré que les temps écoulés entre deux pré-mutations (traduites par un *focus* de fluorescence associé au recrutement de la protéine MutL) « sont exponentiellement distribués et dé-corrélés, ce qui est caractéristique d'un processus Poissonien ». Les auteurs étant en mesure d'observer les pré-mutations indifféremment de leur éventuel effet sur la fitness de la souche, ils montrent que les pré-mutations adviennent d'une façon typiquement aléatoire. Seulement, le devenir évolutif d'une pré-mutation, à court et long termes, est conditionné par l'efficacité avec laquelle elle est prise en charge par les systèmes de correction des mésappariements, la probabilité qu'elle a d'être impliquée dans un évènement de recombinaison, et les coefficients de sélection qui pèsent sur elle. Or l'efficacité avec laquelle une pré-mutation est prise en charge est largement dépendante du type de mutation et du contexte dans lequel elle advient, lequel est soit génomique (paragraphes suivants) ou environnemental (paragraphe page 51). Les paragraphes qui suivent décrivent ce que l'on a appris des probabilités d'occurrence des mutations selon les contextes et selon les espèces.

### *Approches pour l'étude du spectre mutationnel*

On peut envisager deux approches pour l'étude du spectre mutationnel. L'une consiste à étudier, par le biais du polymorphisme, les fréquences des mutations que l'on suppose exemptées de l'action de la sélection, par exemple les mutations aux positions des codons quatre fois dégénérées, comme l'ont fait HILDEBRAND et al. L'autre consiste à étudier le taux de mutation dans des populations pour lesquelles un ou plusieurs gènes de la réparation de l'ADN ont été délétés, et de le comparer au spectre mutationnel de souches sauvages.

Une voie médiane empruntée par WIELGOSS et al. [2011] a consisté

à étudier le taux de substitution dans les lignées d'*Escherichia coli* non-mutatrices de la LTEE (*long term evolution experiment*), l'expérience d'évolution à long terme lancée il y a plus de trente ans par Richard Lenski. Le séquençage du génome de 19 clones répartis sur 8 lignées a montré 25 substitutions synonymes indépendantes, après  $3 \cdot 10^5$  générations ; sur  $941 \cdot 10^3$  sites synonymes potentiels, cela représente un taux de substitution observé de  $5,7 \cdot 10^{-11} \text{ bp}^{-1} \text{ gen}^{-1} - 1,3 \cdot 10^{-10} \text{ bp}^{-1} \text{ gen}^{-1}$ . (C'est un taux de substitution moindre que celui déjà observé sur des populations naturelles d'*E. coli*, ce que DRAKE [2012] estime être dû à la sélection sur l'usage du code dans la LTEE.) Ils montrent que le nombre de substitutions qui ont eu lieu aux positions synonymes est dominé par le nombre de transitions de G·C vers A·T, comme déjà montré par HERSHBERG et PETROV [2010] sur d'autres modèles. Cependant, comme le dit DRAKE, ces substitutions sont affectées par la sélection, et ne représentent donc pas le taux de mutation sous-jacent. C'est pourquoi l'approche par accumulation de mutations, qui s'affranchit des coefficients de sélection, est riche d'enseignement concernant le spectre mutationnel.

##### *Spectre mutationnel de mutants MMR*

La voie MMR est vraiment l'une des voies majeures de correction des mésappariements, comme en témoigne l'augmentation drastique du taux de mutation bien décrite chez divers organismes. Par exemple, H. LEE et al. [2012] ont observé que le taux de mutation d'un mutant de l'espèce *E. coli* inactivé pour le gène *mutL* est 173 fois plus élevé que le taux de mutation de la souche sauvage (respectivement  $326 \cdot 10^{-10} \text{ bp}^{-1}$  et  $1,88 \cdot 10^{-10} \text{ bp}^{-1}$ ). Dans quelle mesure l'inactivation du MMR est-elle susceptible d'influer sur la composition en base ?

ELEZ et al. [2007] ont montré une augmentation de la proportion de mutations qui sont des *transitions* dans un mutant d'*E. coli*  $\Delta$ -MutL. La proportion de transitions AT→GC passe de 28 % dans la souche sauvage à 88 % dans la souche mutante ; celle des transitions GC→AT de 49 % à 11 %. Toujours chez *E. coli*, les travaux de H. LEE et al. [2012] ont permis d'obtenir un taux et un patron de mutation, en analysant 233 mutations dérivées d'expériences d'accumulation de mutation pour la souche sauvage, et 1625 pour un mutant inactivé pour le gène *mutL*. Ces auteurs observent également un excès de transitions de G·C vers A·T (82 contre 49 dans le sens inverse) dans la souche sauvage, alors que la souche MutL<sup>-</sup> montre un excès de transitions AT→GC (1141 contre 447). L'inactivation de MutL conduit à une multiplication par 465 du nombre de transitions AT→GC et à une multiplication par 108 du nombre de transitions GC→AT. Cette inactivation affecte moins les transversions, dont le nombre est multiplié par 7. La protéine MutL, et

donc le MMR, est par conséquent impliquée dans une voie qui répare préférentiellement les transitions chez cette bactérie ; son inactivation contribue à biaiser le spectre de mutations vers GC.

FOSTER, NICCUM et al. [2018] ont ensuite répété les expériences d'accumulation de mutation avec d'autres mutants, délétés pour chacun des gènes essentiels de la voie du MMR, ou une combinaison de délétions doubles. Le phénotype des souches délétées pour MutS, MutL, MutH, MutLS ou MutSHL n'étant pas significativement différent, les auteurs ont choisi de les analyser conjointement, ce qui représente un total de 30 061 BPS (*base pair substitutions*), réparties sur 334 lignées après  $250 \cdot 10^3$  générations. Ils montrent à nouveau que le spectre de BPS de souches  $\Delta$ -MMR est dominé par les transitions aux sites A·T, qui sont trois fois plus fréquentes qu'aux sites G·C. Grâce au grand nombre de BPS analysés, ils montrent également que cet excès est « presque entièrement dû aux mutations aux sites 5'NAC3' + 5'GTN3' ». L'une des originalités de l'article est d'avoir montré que le taux de mutations de souches  $\Delta$ -MMR, croissant en milieu minimal, est divisé par deux, et que le taux de transitions en pareilles conditions est le même aux sites A·T et aux sites G·C. De façon similaire, la réduction du taux de croissance par incubation à des températures plus basses réduit également le taux de BPS de 60 %, mais réduit davantage le taux de transitions aux sites A·T qu'aux sites G·C. (FOSTER, NICCUM et al. proposent que cette réduction soit due à une meilleure efficacité de l'activité d'auto-correction des polymérase, peut-être parce qu'elles auraient plus de temps pour cette correction.) Ces deux résultats indiquent que le milieu et les conditions de croissance influent sur la probabilité d'occurrence d'une mutation.

Des résultats similaires ont été observés dans d'autres modèles distants de *E. coli*, qui, bien qu'il soit souvent le modèle de référence d'étude des mécanismes de réparation de l'ADN, possède un système de correction des mésappariements loin d'être canonique en ce qu'il est dirigé par la méthylation du brin matrice. Ça n'est pas le cas de la majorité du reste du vivant. Cependant, les effets sur le spectre mutationnel de la déletion de tout ou partie de la voie MMR dans d'autres modèles sont similaires. Chez *Bacillus subtilis*, l'inactivation de gènes du MMR conduit aussi à une multiplication du taux de mutation par un facteur qui va de 43 à 76 selon les gènes délétés [SCHROEDER et al. 2016]. Là encore, ce sont essentiellement les transitions qui sont affectés, *a fortiori* les transitions de A·T vers G·C. Chez *Pseudomonas aeruginosa* aussi, l'inactivation du gène MutS conduit à une multiplication par 919 du taux de transitions de A·T vers G·C, contre 264.7 de G·C vers A·T [DETTMAN et al. 2016]. Néanmoins, pour les espèces *Vibrio fischeri* et *cholerae* étudiées par DILLON, SUNG, SEBRA et al. [2016], l'inactivation de MutS

n'affecte pas significativement différemment le taux de transitions de A·T vers G·C et celui de G·C vers A·T, même s'ils augmentent tout deux par rapport aux lignées sauvages.

Dans tous ces modèles, la probabilité d'une mutation dépend fortement du contexte nucléotidique immédiat, phénomène d'autant plus visible dans les souches dites mutatrices. Le contexte nucléotidique comprend en fait d'une part les nucléotides flanquant la base focale de la substitution, et d'autre part l'orientation de celle-ci par rapport à la réplication. (Je reparlerai de l'influence de la polarité de la réplication sur la probabilité d'une mutation et la composition en base aux paragraphes page 53.) Sachant les  $4^3 = 64$  triplets de base possibles, l'on peut comparer les propensions des bases focales (en position 2 du triplet) à muter. Dans le modèle *Mesoplasma florum*, SUNG, ACKERMAN, GOUT et al. [2015] ont montré qu'il y a une corrélation positive forte ( $r^2 = 0.90$ ) entre la composition du génome en triplet de base et celle attendue à l'équilibre des mutations contexte-dépendantes. Cette relation ne tient pas chez *Bacillus subtilis* ou *Escherichia coli*, un effet que les auteurs attribuent à la plus faible taille efficace estimée de l'espèce *Mesoplasma florum* ( $\approx 10^6$ ) par rapport à *B. subtilis* ( $\approx 10^7$ ) et *E. coli* ( $\approx 10^8$ ). Cependant, SCHROEDER et al. [2016] montrent qu'il y a bien une relation forte entre le taux de *transitions* contexte-dépendantes observé expérimentalement et le taux attendu uniquement d'après le contexte nucléotidique chez *B. subtilis*. Ceci « révèle que le contexte nucléotidique local est le facteur le plus important influant sur le taux de transition due aux erreurs de l'ADN-polymérase répliquative ».

#### *Spectre mutationnel de mutants BER*

FOSTER, H. LEE et al. [2015] ont également caractérisé le spectre mutationnel de différentes souches d'*E. coli* pour lesquelles les mécanismes de correction des mésappariements par *base excision repair* [BER] ont été inactivés. Les protéines du BER inactivées sont MutY et MutM, respectivement adénine et formamidopyrimidine DNA glycosylase, toutes deux impliquées dans la correction des mésappariements liés à l'oxydation des guanines. La diphosphatase MutT n'est pas directement liée au BER, mais catalyse l'élimination des guanines oxydées du pool de dNTP. À ce titre, son inactivation est susceptible de conduire à davantage d'erreurs de réplication liées à l'oxydation des guanines, qui sont pris en charge en conditions normales par le BER.

Trois types de souches mutatrices ont donc été générées : des souches pour lesquelles MutT, MutY et une combinaison MutM/MutY ont été inactivés. L'inactivation de MutT conduit à une multiplication par 900 du taux de transversions de A·T vers C·G, ce qui concorde avec le rôle



de MutT dans le « nettoyage » (*sanitization*) du pool de 8-oxo-dGTP, dont l'incorporation face à une adénine conduit à une augmentation du taux de transversions de A·T vers C·G [FOWLER et al. 2003]. L'inactivation de MutY conduit au phénotype inverse : à une multiplication par 80 du nombre de transversions de C·G vers A·T.<sup>26</sup> COUCE, CAUDWELL et al. [2017] montrent enfin que deux des populations de la LTEE ayant évolué un phénotype mutateur par perte de la fonction du gène MutT ont un taux de transversions de A·T vers C·G plus élevé que les populations non-mutatrices.

Il semble ainsi que les protéines liées au BER sont susceptibles d'affecter le taux de transversions A·T→C·G, notamment le gène MutY.

### *La mutation chez Burkholderia est biaisée vers GC%*

Dans tous les modèles dont il a été question ici, à l'état sauvage, le spectre mutationnel est biaisé vers l'introduction de base A et T, biais qui s'inverse notamment lorsque le MMR est inactivé. La seule exception que l'on connaissait à ce tableau est la bactérie riche en GC *Burkholderia cenocepacia*, avec laquelle des expériences d'accumulation ont été réalisées [DILLON, SUNG, LYNCH et al. 2015]. Elles ont montré, en tenant compte des transitions et des transversions simultanément, que le taux de mutations de AT vers GC est supérieur à celui des mutations de GC vers AT. C'était donc en 2015 la seule espèce que l'on connaissait pour laquelle le patron de mutation était biaisé vers G et C (depuis, avec la parution de l'article de LONG et al., on connaît 6 autres bactéries pour lesquels  $m < 1$ ). Néanmoins, avec un  $m = 0.81$ , son taux de GC à l'équilibre mutationnel est de  $GC_n = 1/1+0.81 = 55\%$ , et son taux de GC observé est de 67%. Donc même pour *Burkholderia* (et les six autres bactéries avec  $m < 1$ ), le taux de GC à l'équilibre mutationnel, bien que supérieur à 50%, est inférieur au taux de GC observé. S'il n'est donc pas tout à fait exact de dire que la « mutation est universellement biaisée vers AT », il reste que le biais de fixation vers G et C se trouve aussi dans les organismes où  $m < 1$ .

### *Environnement, stress et mutation*

Comme déjà indiqué dans l'article de FOSTER, NICCUM et al. [2018], les conditions de culture et de croissance jouent probablement un rôle dans le déterminisme du taux de mutation. C'est cette question que

26. À ce propos, COUCE et TENAILLON [2019] ont montré par simulation que les effets idiosyncratiques de la perte de l'un de ces deux gènes aux effets antagonistes

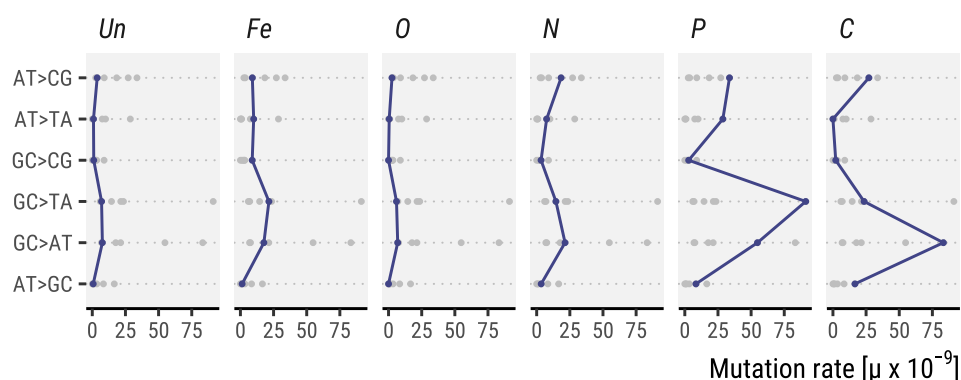
conduisent à une augmentation de la fitness lié à la perte du mutateur inverse, ou antimutateur.

#### 4. PROCESSUS MUTATIONNELS

MAHARJAN et FERENCI [2017] ont entrepris d'explorer. Leur article s'inscrit dans la lignée des travaux sur la mutation induite par un stress (*stress-induced mutation*, SIM). Leur étude a consisté à étudier la fréquence de chaque mutation possible chez *E. coli* en cultivant une souche initiale sensible à la cyclosérine dans un milieu carencé en l'un des cinq éléments nutritifs suivants : fer, oxygène, azote, phosphore et carbone. L'avantage de leur système expérimental est qu'il n'est pas biaisé du point de vue de la sélection qui peut opérer dans le système, comme peuvent l'être les expériences d'accumulation de mutation. L'acquisition de la résistance à la cyclosérine passe en effet par un large spectre de mutations causant une perte de fonction au locus *cycA*, « incluant les six types de BPS possibles, différents types d'insertions courtes (SI) et longues (LI), et des transpositions d'IS à travers toute la longueur du gène *cycA*. » De plus, la perte de fonction au locus *cycA* a un effet en fitness négligeable, de sorte qu'on peut obtenir une estimation des influences environnementales sur le patron de mutation non-biaisée par la sélection. La culture en conditions homéostatiques (en *batch*) dans un milieu carencé, pendant trois jours, suivie d'un criblage des souches résistantes à la cyclosérine, d'une amplification et du séquençage du gène *cycA* sur toute sa longueur pour 1399 mutants Cyc<sup>R</sup> leur a fourni une idée précise de l'influence de stress environnementaux sur le taux de mutation.

La figure 4.5 est une adaptation de la figure 5 de l'article, en se restreignant aux seules substitutions. Elle représente le taux de mutation, pour différentes catégories de mutations, obtenu quand le milieu est carencé en un élément nutritif donné.

FIGURE 4.5 – Influence des carences nutritives sur le taux de mutation. Données extraites de MAHARJAN et FERENCI 2017.



Baptisé « paysage mutationnel » par MAHARJAN et FERENCI, ce genre de figure montre clairement que la carence en un élément nutritif influe sur la probabilité de mutation. Par exemple, le taux de transversions de G·C vers T·A est multiplié par 13 par rapport au témoin quand le milieu est carencé en phosphore. De même, le taux de transitions de G·C vers A·T est multiplié par 11 par rapport au témoin quand le



milieu est carencé en carbone. Ils montrent aussi, en figure 6, que la protéine MutY est significativement réprimée dans toutes les conditions de carences à l'exception des carences en carbone, alors que la protéine MutS ne l'est qu'en condition de carence en phosphore, en azote ou en carbone.

Cet article souligne ainsi que la composition en base et les processus mutationnels qui la sous-tendent dépendent pour partie des conditions environnementales. Ceci pourrait expliquer pourquoi certains auteurs ont détecté des corrélations entre composition en base et certains traits environnementaux, mais cela reste à tester.

### Asymétrie entre brins liée à la réplication

Au nombre des déterminants clés de la composition en base et de sa variabilité le long du génome, il faut enfin mentionner la découverte de l'asymétrie liée à la réplication. Celle-ci dut attendre que l'on soit en mesure d'étudier des fragments de génomes suffisamment longs, ce qui fut rendu possible par les laborieuses tentatives de séquençage des génomes de bactéries modèles dans les années 1995 (notamment *H. influenzae*). J. R. LOBRY [1996b] le premier montra que le taux de GC tend à être structuré par le sens de la réplication du chromosome bactérien.<sup>27</sup> Il s'appuyait pour cela sur les travaux théoriques de Sueoka (encore lui), et les siens propres [J. R. LOBRY 1995; SUEOKA 1995]. En étudiant les quelques contigs alors disponibles pour les génomes de *Escherichia coli*, *Haemophilus influenzae* et *Bacillus subtilis*, il a montré que l'on pouvait diviser leur génomes en deux segments (baptisés « *chirochores* », probablement par analogie avec les « *isochores* » des génomes de mammifères — nous adopterons plutôt le terme de *réplichores*, que l'usage semble avoir consacré depuis) polarisés par l'origine et la terminaison de la réplication. Ces deux segments montrent des patrons de substitutions asymétriques entre le brin sens et le brin anti-sens, de sorte que l'on trouve généralement une proportion plus élevée de bases C que de bases G par fenêtre de 1kb le long du brin sens (voir figure 4.6 pour une illustration de ce phénomène dans le génome d'*Acinetobacter baylyi*).

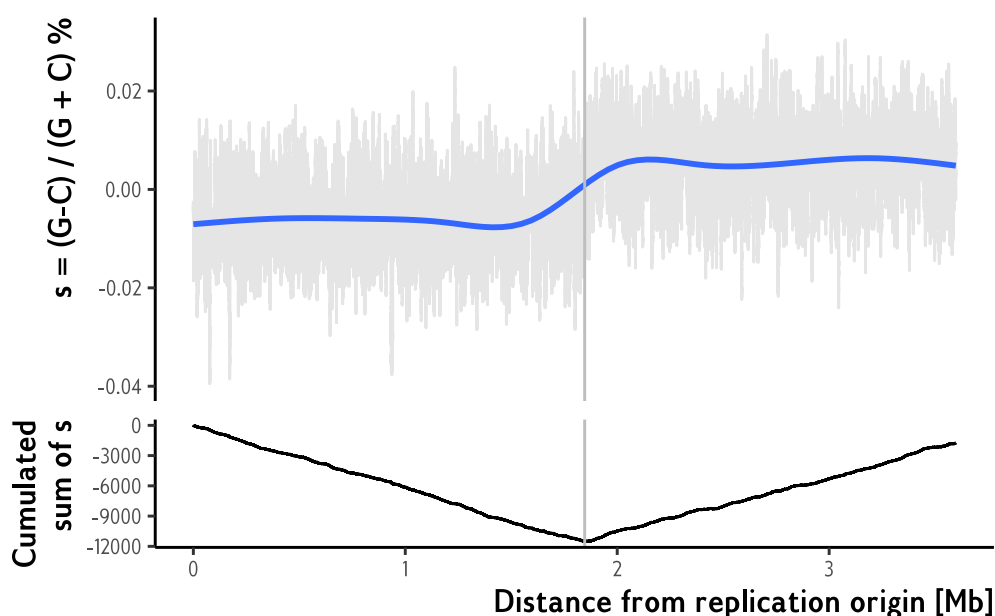
Il utilisa pour cela la métrique dite du *GC-skew* (proposée, à ma connaissance, par C.-I. WU et MAEDA [1987]), soit l'inclination différentielle des deux brins à accumuler d'avantage de bases C que de bases G,

27. Encore que l'idée ne soit pas sortie du front de Zeus, d'autres évidemment avaient approché ce résultat d'assez près, par exemple [FILIPSKI 1990; C.-I. WU et MAEDA 1987]; en science il est toujours malaisé d'identifier rigoureusement ceux ou celles à l'origine d'une « découverte »,

quand précisément cette origine est le produit d'un champ social en activité. Tout l'enjeu est d'identifier les « ruptures épistémologiques » chères à Bachelard, celles par quoi la perception théorique du monde s'est trouvée brusquement bousculée.

#### 4. PROCESSUS MUTATIONNELS

FIGURE 4.6 –  
Structuration en  
GC% le long du  
brin sens du génome  
d'*Acinetobacter baylyi*



traduite par :

$$\sigma = \frac{G - C}{G + C}$$

Elle correspond à l'écart avec  $[C] = [G]$ , c'est-à-dire aux concentrations égales en cytosine et guanine sur un brin donné en l'absence d'un biais lié au brin. L'inversion du signe de la somme cumulée de  $\sigma$  (voir figure 4.6) correspond ainsi au terminus de la réplication ou à son origine, ce qui permet leur détection dans les génomes [A. C. FRANK et J. R. LOBRY 2000; J. R. LOBRY 1996a].

À sa suite, FREEMAN et al. [1998] ont démontré d'autres asymétries entre répliques. La somme cumulée du nombre de purine par rapport à la somme cumulée du nombre de pyrimidine de l'origine de réplication à la position  $i$

$$\sum_1^i A_i + G_i - \sum_1^i T_i + C_i$$

atteint son maximum dans la région du terminus de réplication. La même fonction appliquée aux bases cétones (GT) et aminées (AC) (de par leur conformation stéréochimique et leurs origines métaboliques) conduit à la même observation que la dérivée de cette somme cumulée change de signe autour de l'origine et du terminus de réplication. Enfin, le même raisonnement appliqué aux bases elles-mêmes situées dans

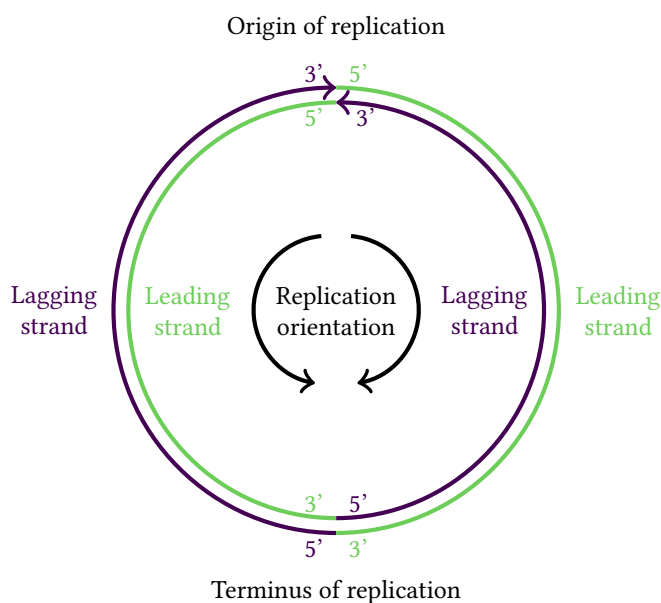


FIGURE 4.7 – Représentation schématique de l'orientation des brins par rapport à l'origine de réplication d'un chromosome bactérien

une région codante  $C$ , ou dont le complément l'est  $\bar{C}$ ,

$$\sum_1^i C_i - \sum_1^i \bar{C}_i$$

montre que la propension de certaines bases à se trouver dans les régions codantes dépend du réplichoire en question. Enfin, FRANCINO, CHAO et al. [1996] ont montré que le taux de transitions de C vers T ou de A vers G dépend d'un processus « distinguant entre les brins codants et non-codants d'un gène, et non le sens de la réplication ».

Ces résultats ont contribué à ce que l'on prenne conscience que les processus mutationnels sont moins aléatoires, en tout cas peut-être moins uniformément répartis sur les génomes, qu'on ne le pensait alors.<sup>28</sup> Cela s'est traduit par une cristallisation progressive de l'idée que l'on trouve formulée par DAUBIN et PERRIÈRE [2003], selon laquelle « la composition en codon d'un gène résulte de forces évolutives concurrentes, dont l'équilibre varie d'un gène à l'autre et dépend de sa fonction, du niveau d'expression et de la localisation sur le chromosome ». À quoi l'on pourrait ajouter que *la composition en base d'un génome résulte de forces évolutives concurrentes, agissant à différents niveaux, et dont l'équilibre varie*

28. FRANCINO, CHAO et al. commencent d'ailleurs leur article par dire que « l'un des postulats fondamentaux de

l'évolution moléculaire tient que les mutations sont également probables à quelque site du génome que ce soit. »

#### 4. PROCESSUS MUTATIONNELS

---

*d'une espèce à l'autre.*<sup>29</sup> Autrement dit, la littérature de l'époque commence à poser clairement que, pour comprendre la sélection au niveau des gènes, ce qui était jusqu'alors l'un des intérêts fondamentaux de la génétique, l'on ne peut se passer de la compréhension des processus aux niveaux d'organisation plus élevés dans lesquels ils sont impliqués.

Un bon exemple de cette tendance est justement l'article qui vient d'être cité [DAUBIN et PERRIÈRE 2003]. Il montre que bon nombre de traits génomiques alors utilisés pour détecter les gènes provenant d'espèces plus ou moins lointaines et potentiellement acquis horizontalement [RAVENHALL et al. 2015], dont le taux de GC en troisième position des codons, GC3, (et d'autres comme l'indice d'adaptation des codons), ne sont pas répartis uniformément le long du génome. On est donc susceptible de confondre un gène dont le taux de GC diverge du taux de GC attendu par ailleurs, avec un gène d'acquisition « étrangère ». (C'est vrai d'autres traits que la composition en base. En fait les indices en questions sont rarement indépendants les uns des autres, comme le montre l'article : « la structuration du taux de GC3 va souvent de pair avec celle de l'indice d'adaptation des codons » DAUBIN et PERRIÈRE 2003.)

La revue de E. P. C. ROCHA [2004] résume l'ensemble des avancées permises par le séquençage des génomes bactériens et leur analyse bio-informatique. Il pose explicitement l'enjeu épistémique lié à la compréhension de la structuration des génomes : « Puisque certains aspects de la structure du chromosome semblent se trouver sous une forte pression de sélection, il importe d'appréhender pareils biais pour comprendre l'organisation et l'adaptation du chromosome. » Je me contenterai ici de rappeler les apports, résumés dans cette revue, concernant la composition en base, et dans quelle mesure elle est structurée par la réplication du chromosome. D'abord, le sens de la réplication influe sur la composition en base des deux brins, puisque l'un des brins (5'→3') est répliqué directement, quand l'autre (3'→5') implique une phase simple-brin avec appariement des fragments d'Okazaki. Comme le dit Rocha, « certains types de substitutions sont plus fréquents [dans l'ADN simple-brin], et donc la réplication différentielle entre les deux brins entraîne des différences de compositions entre eux » [FRANCINO et OCHMAN 1997; A. C FRANK et J. R LOBRY 1999]. Ensuite, comme un cycle de réplication se déroule sur un temps relativement long (1 H chez *E. coli*), les conditions physiologiques dans lesquelles se trouve la cellule au moment

29. Je sais bien toute la difficulté que certains accordent parfois probablement un peu gratuitement à la notion d'espèce chez les bactéries; j'esquiverais volontiers le débat

pour ma part en invoquant Wittgenstein et la dernière phrase de son *Tractatus* : « Sur ce dont on ne peut parler, il faut garder le silence ».

de l'initiation de la réplication et de sa terminaison sont probablement sensiblement différentes [DAUBIN et PERRIÈRE 2003]. Ce qui pourrait entraîner des biais mutationnels différents entre les régions proches de l'origine de réplifications et celles proches du terminus. Enfin, il rappelle après DAUBIN et PERRIÈRE que la composition en base en elle même n'est pas tout à fait homogène le long du génome, sans que l'on ait pour l'instant d'explication mécanistique à ce phénomène (voir figure 4.8).

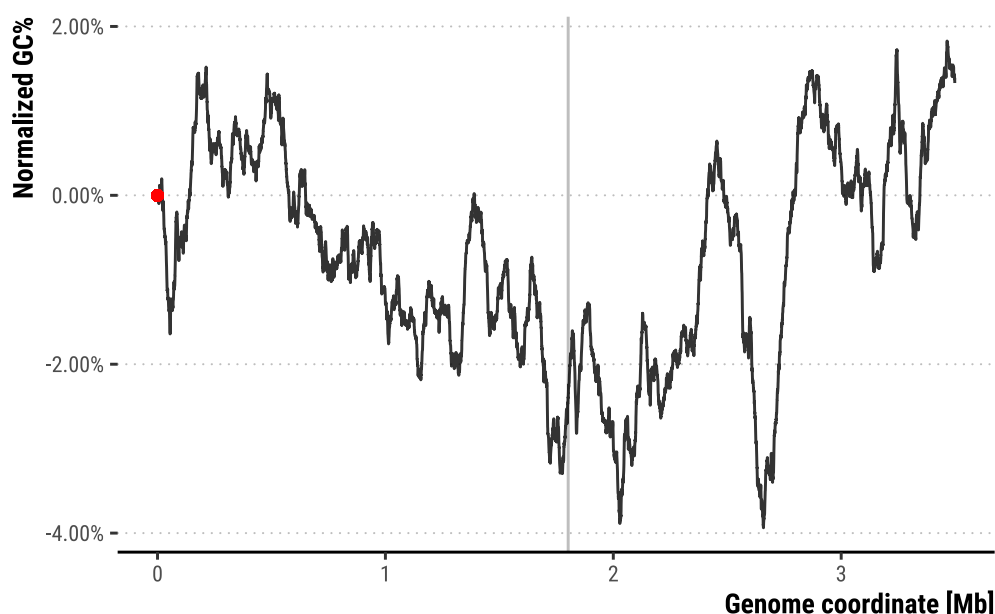


FIGURE 4.8 – Distribution de la variabilité du GC% le long du génome d'*A. baylyi*. Le taux de GC est calculé par fenêtre glissante de  $1 \cdot 10^5$  bp, et est normalisé en lui retranchant le taux de GC de la première fenêtre. Est représenté la fonction cumulative de la première dérivée de ce taux de GC, pour visualiser la variation en GC. La barre verticale représente la région *ter*. Le taux  $\tau$  normalisé en  $i$  est calculé par  $\tau_i = -GC_0 + \sum_{n=1}^i GC_n - GC_{n-1}$

La différence de composition en base entre les deux brins d'un réplique était alors pensée comme résultant de la déamination spontanée des cytosines, plus fréquente lorsque la base en jeu se trouve dans un contexte simple-brin [A. C FRANK et J. R LOBRY 1999; E. P. C. ROCHA 2004]. La déamination d'une cytosine consiste en la perte de son groupement amine, ce qui donne, quand elle n'est pas méthylée, une base uracile, généralement excisée par une uracile DNA-glycosylase; quand la cytosine est méthylée, la déamination produit une base thymine. La synthèse du brin 5'→3' étant discontinue, ce brin est plus fréquemment que le brin complémentaire sous forme simple-brin. Il est donc en théorie plus susceptible d'occasionner des déaminations spontanées des cytosines, d'où vient leur accumulation dans le brin *lagging*. À quelques réserves prêt, c'était l'hypothèse privilégiée pour expliquer la quasi-universalité du biais de composition en base des deux brins chez les Bactéries. Néanmoins, l'article de Eduardo P. C. ROCHA, TOUCHON et al. [2006] a montré qu'en réalité ce sont probablement différents mécanismes à l'œuvre chez différentes espèces. « Comme un défi au rasoir d'Ockham, concluent-ils, même de simple biais de

composition quasiment ubiquitaires, causés par le processus essentiel et hautement conservé de la réplication, peuvent être sous-tendus par une vaste complexité de phénomènes biologiques. »

Les expériences d'accumulation de mutation sont venues corroborer ces observations. Chez *E. coli*, FOSTER, H. LEE et al. [2015] décrivent que les transitions de G·C vers A·T sont deux fois plus probables lorsque que les bases C et G sont respectivement sur le brin antisens (*lagging*) et sens (*leading*) que dans le cas inverse. Par contre, les transversions A·T vers C·G dans le mutant  $\Delta$ -*mutT* ne sont pas biaisées du point de vue du brin, et se produisent aussi fréquemment que la base A soit sur le brin sens ou antisens. Chez *P. aeruginosa*, DETTMAN et al. [2016] observent le même phénomène que chez *E. coli* : les transitions de G·C vers A·T sont 3,53 plus fréquentes quand la base G est sur le brin sens et C sur le brin antisens que l'inverse. Réciproquement, le taux de transitions de A·T vers G·C est 0,43 fois moins élevé quand la base A est sur le brin sens que lorsqu'elle est sur le brin antisens. En revanche, chez *Salmonella typhimurium*, LIND et ANDERSSON [2008] n'ont pas observé d'asymétrie entre les deux brins dans la probabilité de mutations de C→T ou de G→T. On voit donc que l'origine de l'asymétrie en composition en base entre les deux brins polarisés par la réplication varie d'une espèce à l'autre, comme le suggéraient Eduardo P. C. ROCHA, TOUCHON et al. [2006].

### 4.3 Distribution phylogénétique des gènes impliqués

Si donc la balance mutationnelle est le principal déterminant de la variabilité en composition en base, et que cette balance est elle-même le fruit d'un équilibre entre la propension des polymérase à commettre certaines erreurs et celle des systèmes de réparation desdites erreurs, on devrait s'attendre à trouver dans l'arbre du vivant des relations entre la composition en base des génomes et la présence, ou l'absence, des gènes impliqués dans la réparation et la polymérisation.

ACOSTA et al. [2015] ont quantifié, chez les Bactéries et les Archées, l'écart de taux de GC entre les génomes pour lesquels un gène de réparation est présent à ceux pour lesquels il est absent.

La figure 4.9 représente l'écart entre le taux de GC des génomes possédant un gène de réparation donné et le taux de GC de ceux qui ne le possèdent pas, pour les toutes les bactéries et les bactéries dites non-intracellulaires (adapté de la table 1 de l'article cité). L'absence du gène *mutY* par exemple est associé à un taux de GC significativement plus bas, tout comme celle du gène *vsr*, impliqué dans la réparation des mésappariements T·G en C·G. (Il faut noter que les auteurs ne

## Distribution phylogénétique des gènes impliqués

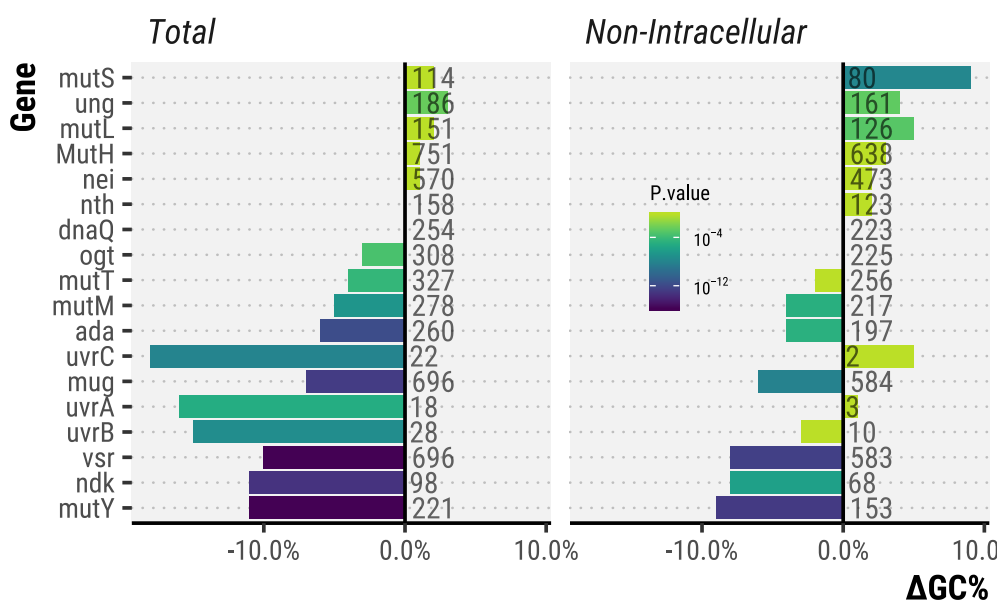


FIGURE 4.9 – Écart de taux de GC entre génomes possédant un gène de la réparation et ceux qui ne le possèdent pas. Les valeurs obtenus avec les génomes de bactéries sont représentés ici. Ont été distingués les génomes des bactéries non-intracellulaires de l'ensemble. La couleur représente la valeur du test de comparaison effectif. Le nombre de génomes analysés par catégorie est représenté à droite de la ligne verticale de coordonnée  $x = 1$ . (Données extraites de ACOSTA et al. 2015)

contrastent pas les résultats pour l'inertie phylogénétique, mentionnant qu'il est difficile de le faire quand il est question de traits évoluant lentement comme le taux de GC, et donc plus sujets à l'inertie que d'autres.) GARCIA-GONZALEZ et al. [2012] ont analysé spécifiquement la distribution des gènes de réparation, et montrent aussi l'effet de la présence des gènes du BER (MutT et MutY) sur la composition en base.

En amont de la réparation des mésappariements, les différentes DNA-polymérase diffèrent dans leur propension à commettre des erreurs. H. WU et al. [2012] ont ainsi proposé que la « sous-unité  $\alpha$  de la DNA-polymérase III et ses isoformes [...] jouent un rôle déterminant dans la variabilité du taux de GC ». Là encore, il est possible qu'une bonne part du signal observé soit dû à l'inertie phylogénétique, dont on peut raisonnablement croire qu'elle est la plus marquée lorsqu'elle concerne des protéines aussi essentielles que les DNA-polymérase. Néanmoins, il semble que le recours à différentes DNA-polymérase ait quelque influence sur le taux de GC : les groupes qui recourent à la polymérase dnaE1 et dnaE2 ont un taux de GC significativement plus élevé que ceux qui recourent à la polymérase dnaE3 et polV.

Une étude récente a également observé une relation forte entre la présence du gène *Ku*, impliqué dans la réparation de l'ADN par la voie des jonctions d'extrémités non-homologue (NHEJ) et le taux de GC [WEISSMAN et al. 2019]. (La voie NHEJ est une voie possible pour la réparation des cassures double-brin [DSB], et consiste en la ligation, catalysée par les homologues de *Ku* des extrémités de la cassure; elle est dite non-conservative en ce qu'elle peut causer des pertes de matériel



#### 4. PROCESSUS MUTATIONNELS

---

génétique.) Plus précisément, il semble que les bactéries qui possèdent ce gène aient un taux de GC génomique plus élevé que les bactéries qui ne le possèdent pas, et ce dans différents embranchements de l'arbre des bactéries.<sup>30</sup>

#### *Conclusion*

Il apparaît ainsi que la variabilité de la composition en base entre espèces est bien le fait d'abord de différences dans leurs patrons de mutations. Ce patron de mutation, c'est-à-dire l'inclination qu'ont certaines bases ou certains triplets de bases à muter préférentiellement dans une direction que d'autres, est essentiellement déterminé par 1) le taux d'erreur des DNA-polymérase, lui-même sans doute fonction des DNA-polymérase elle-même; 2) l'efficacité avec laquelle les erreurs commises sont prises en charge par les systèmes de réparation, elle-même fonction des différents systèmes de réparation en jeu, et éventuellement du contexte génomique (notamment lié à la réplication); 3) la nature des stress environnementaux auxquels sont soumises les souches. En ce sens, on observe bien des différences en termes de composition en base entre bactéries qui possèdent ou non un système de réparation donné, par exemple les DNA-glycosylases impliquées dans la voie du BER. Il ne faut pas oublier que des contraintes liées à l'inertie phylogénétique pèsent lourdement sur ce trait, comme l'a montré la distribution du taux de GC dans l'arbre des bactéries. Elles peuvent ainsi affecter la qualité des inférences que l'on tire d'observations à grands traits, comme celle de la relation entre composition en base et distribution des enzymes de réparation ou de polymérisation de l'ADN.

Néanmoins, toute la variabilité observée ne peut-être expliquée par les seuls patrons de mutations, puisque les probabilités de *fixation* des bases G et C sont plus élevées que celles des bases A et T, et ce dans (presque) tous les génomes étudiés. Comme le soulignent Eduardo P. C. ROCHA et FEIL [2010], « les preuves d'une pression mutationnelle commune en faveur d'un taux de GC plus bas sont claires. [...] Deux

30. Les auteurs l'interprètent comme un signal de sélection en faveur d'un taux de GC plus élevé dans les génomes enclins aux DSB, en faisant l'hypothèse que la présence de *Ku* est un indicateur de cette inclination... Si le résultat est intéressant, j'ai, à l'égard de l'interprétation qui en est faite, les réserves les plus grandes : d'abord parce que l'enrichissement en GC des régions recombinantes est précisément la signature que laisse la gBGC; ensuite parce qu'il n'y a pas de raisons de penser que la possession de *Ku*

soit un indicateur de la disposition du génome aux DSB. POPA et al. [2011] montrent ainsi que la voie NHEJ pourrait être un moyen pour des éléments génétiques mobiles [MGE] de franchir les barrières d'homologie requises pour l'intégration hétérologue *via* la recombinaison homologue aux extrémités. Il se pourrait donc bien que la possession de *Ku* ne soit au bout du compte qu'un indicateur de l'invasion du génome par les MGE.



## Distribution phylogénétique des gènes impliqués

processus distincts peuvent opérer : la conversion génique biaisée (BGC) et la sélection naturelle. »

Dans la partie suivante, je décrirai certaines hypothèses sélectives proposées pour la détermination du taux de GC, avant de décrire la conversion génique biaisée vers GC et les indices dont on dispose de son existence chez les Bactéries.



## SÉLECTION SUR LA COMPOSITION EN BASE (?)

---

- *Le Théorème de Tante Jobiska (d'Edward Lear)* :  
« C'est un fait que tout le monde sait. »
- *Le Théorème de l'Homme à la Cloche (de Lewis Carroll)* :  
« Si je te le dis trois fois, c'est vrai. »
- *Le Théorème de Pangloss (de Voltaire)* :  
« Tout est pour le mieux dans le meilleur des mondes possibles. »

Les trois « théorèmes » du sophisme scientifique, par J.B.S. Haldane, cité par J.M. Smith, cité par G. C. WILLIAMS, *Adaptation and Natural Selection*, page xi.

« L'accoutumance à des principes intellectuels sans raisons, on appelle cela croyance. »

F. Nietzsche, *Humain, Trop Humain*, cité par J. Bouveresse, « L'éthique de La Croyance et La Question Du « Poids de l'autorité » ».

L'explication la plus évidente, à première vue, pour expliquer l'écart entre composition en base observée et composition en base attendue à l'équilibre mutationnel est celle de la sélection naturelle. Autrement dit, il se pourrait qu'il y ait un avantage sélectif, pour une bactérie donnée dans un environnement donné, à ce que le génome soit enrichi en G et C. Il faut alors que la composition en base soit envisagée comme un trait, qui contribue en lui-même à la fitness de l'individu. Ainsi, Giorgio BERNARDI et Giacomo BERNARDI [1986] indiquent que « les contraintes de composition [en base] pointent vers une nouvelle composante du phénotype de l'organisme, que l'on pourrait appeler le "phénotype du génome" ». Différents éléments inclinent en ce sens. C'est l'objet de cette partie que de les passer en revue.

### 5.1 Corrélations GC% et traits environnementaux

*Tante Jobiska : Température optimale de croissance.*

Une hypothèse sélectionniste formulée qui indiquerait que la sélection opère bien sur la composition en base globale tient à l'observation

## 5. SÉLECTION SUR LA COMPOSITION EN BASE (?)

---

que la composition en base de certains *gènes* corrèle avec la température optimale de croissance ( $T_{opt}$ ) de l'organisme.<sup>31</sup>

Cette hypothèse repose sur les deux observations suivantes : d'une part qu'une molécule d'acides nucléiques est composée d'une même molarité d'adénines que de thymines et de guanines que de cytosines, et cette relation est vraie que la molécule soit composée de deux ou d'un seul brin (les fameuses règles de parité de Chargaff) ; d'autre part que les paires de G et C sont liées par trois liaisons hydrogène, contre deux pour les paires de bases A et T, et par conséquent que la molécule devrait être stabilisée dans sa structure secondaire par une composition élevée en paires de bases G et C [NAYA et al. 2002]. Par exemple, Giorgio BERNARDI et Giacomo BERNARDI [1986], en contrastant les taux de GC de différentes espèces isolées à différentes températures, proposent que « ces découvertes fournissent des exemples précis d'un facteur environnemental, la température, comme responsable de contraintes sur la composition en base du génome. » Plus haut, ils proposent, sur la base des 340 séquences qu'ils avaient à disposition, ainsi que sur la tendance qu'ont les bactéries thermophiles à remplacer certains acides aminés par d'autres (*e.g.* Ser→Ala), et du fait que ces remplacements correspondent en moyenne à des augmentations en bases GC au niveau nucléotidique, que l'augmentation en taux de GC des séquences nucléotidique de bactéries thermophiles correspond à une adaptation à des températures supérieures. Ce serait une adaptation pour deux raisons, d'abord parce que l'accroissement en G·C augmente la thermostabilité de l'ADN, ensuite parce qu'il se traduit par une augmentation de la thermostabilité des protéines encodées.

Cependant, GALTIER et J. R. LOBRY [1997] ont montré que la corrélation entre  $T_{opt}$  et le taux de GC n'est significative que lorsqu'on considère les acides nucléiques structuraux, au nombre desquels on compte les sous-unités ribosomiques 16 et 23S. Il n'y a pas de corrélation entre le taux de GC génomique et  $T_{opt}$ , pour les 224 genres procaryotes disponibles en 1997.

À leur suite, HURST et MERCHANT [2001] ont ajouté que cette absence de corrélation était robuste à la prise en compte de l'inertie phylogénétique ; que la troisième position des codons variait comme le taux de GC du génome plutôt que comme le taux de GC des ARN structuraux ; donc que le taux de GC aux positions évoluant les plus librement comme le GC3 ne peut être considéré comme une adaptation en réponse à des pressions thermiques exercées par l'environnement. De plus, comme

31. Cette théorie introduite dans les années 1990, en dépit du fait qu'elle ait été démontrée invalide dès les années 2000 s'est, inexplicablement à mon sens, répandue dans le champ de la

microbiologie, si bien qu'il m'a moi-même souvent fallu lutter contre cette préconception. Après tout, « c'est un fait que tout le monde sait ».

souligné par AGASHE et SHANKAR [2014], la culture de *Pasteurella multocida* pendant 14 000 générations à température croissante a causé une accumulation de significativement plus de A·T que de G·C [AGASHE et SHANKAR 2014; XIA et al. 2002].

Après quoi MUSTO et al. [2004] ont rétorqué qu'en analysant dans un cadre phylogénétique mieux défini, c'est-à-dire au niveau de la famille taxonomique, parmi les 20 familles pour lesquelles les auteurs disposaient de plus de 10 espèces, il y a bien une corrélation entre température optimale de croissance et composition en base pour 15 d'entre elles. Mais H.-C. WANG et al. [2006] reprennent leur données de température optimale de croissance, elles-mêmes provenant de GALTIER et J. R. LOBRY [1997], elles-mêmes du Bergey's Manual, et les confrontent aux valeurs plus récemment calculées par d'autres chez les *Halobacteriaceae*, pour montrer qu'elles sont largement discordantes. En corrigeant les valeurs utilisées par MUSTO et al. [2004] par les leurs, la relation entre température optimale et taux de GC ne tient plus. De plus, chez les *Enterobacteriaceae*, si la taille du génome est intégrée dans une régression multiple, le facteur température n'est plus significatif non plus [H.-C. WANG et al. 2006].

Si donc l'adaptation à la vie en milieux chauds implique vraiment une augmentation du taux de GC global, il n'y en a pas de traces dans les données disponibles, ou à tout le moins sont-elles difficiles à détecter.

### *Environnements différents et taux de GC différents ?*

Si le taux de GC devait être considéré comme contribuant à la fitness, on devrait s'attendre à ce que des espèces, ou même des populations d'une même espèce, vivant dans deux milieux différents présentent des compositions en bases différentes. C'est précisément à cette question que tentent de répondre FOERSTNER et al. [2005] et REICHENBERGER et al. [2015].

FOERSTNER et al. [2005] inaugure en quelque sorte l'analyse de métagénomés. Ils se sont appuyés sur les données fournies par les premiers séquençages « en masse »<sup>32</sup> d'environnements comme la surface de la mer des Sargasses, le sol d'une ferme du Minnesota, le biofilm d'eaux percolées de mines ou d'une carcasse de baleine. Ils observent que le taux de GC des lectures de séquençage diffère significativement de celui attendu d'après la composition taxonomique de l'environnement en question, connue par ailleurs. Par exemple, le taux de GC obtenu à partir du métagénome de la mer des Sargasses est en moyenne inférieur à celui de la ferme du Minnesota, et ce d'autant plus aux positions trois des codons. Ils concluent en disant que cela « suggère que des facteurs

32. En « *shotgun* », encore que la métaphore relève d'un bien mauvais choix de mot.

## 5. SÉLECTION SUR LA COMPOSITION EN BASE (?)

---

essentiellement environnementaux et externes influencent la composition nucléotidique en GC d'une communauté, soit sélectivement, soit en causant un biais mutationnel, mécanistique et dirigé ».

Cette approche a néanmoins trois inconvénients majeurs : 1) sont comparées des communautés qui n'ont probablement pas beaucoup de similitudes en terme de composition taxonomique ; 2) les protocoles d'obtention des métagénomés sont différents d'une étude à l'autre, et l'on sait bien tous les biais inhérents à la collecte de métagénomés [LOZUPONE et al. 2013 ; WESOLOWSKA-ANDERSEN et al. 2014] ; 3) seulement quatre communautés sont contrastées.

REICHENBERGER et al. [2015] pallient au moins aux points 1 et 3, en analysant le taux de GC de Bactéries de même genres dans des 183 métagénomés issus de 14 environnements différents, dont 111 du tube digestif humain. Là aussi, le taux de GC moyen est significativement corrélé à l'environnement duquel le métagénome a été isolé. Ça n'est pas dû à une composition taxonomique spécifique à chaque environnement, puisqu'une comparaison deux à deux de chacun des métagénomés révèle que la plus grande différence qui ait été observée dans leur composition en genre est de 3 %, ce qui veut dire que tous les métagénomés ont en commun quasiment tous les genres analysés. De plus, le taux de GC moyen d'un phylum donné au sein d'un environnement donné tend à être corrélé à celui des autres phyla ; autrement dit, l'augmentation de GC moyen lié à un environnement donné tend à se traduire à travers tous les phyla étudiés (à l'exception des Ténéricutes). Enfin, les métagénomés issus du tube digestif humain peuvent être classés en deux catégories : ceux qui sont enrichies en Actinobactéries et les autres. Le taux de GC moyen des phyla dans les métagénomés riches en Actinobactéries (Actinobactéries exclues) est significativement supérieur à celui des autres métagénomés. Ces résultats, d'après les auteurs, suggèrent que « les environnements dominés par les *Actinobacteria* tendent aussi à être ceux pour lesquels il y existe une sélection en faveur des taux de GC supérieur, ce à travers de nombreux phyla, sinon tous ».

Il me semble que l'on peut proposer trois hypothèses pour expliquer ces différences de taux de GC observées entre échantillons :

1. Elles sont effectivement dues à la sélection naturelle, mais il est difficile de concevoir que la sélection naturelle agisse uniformément à travers les phyla et les environnements.
2. Elles sont dues aux patrons de mutations différents inhérents aux différents environnements.
3. Elles sont dues à l'absence d'uniformité des protocoles d'échantillonnage, de séquençage et de traitement des séquences. Certains

protocoles pourraient être plus enclins que d'autres à enrichir l'extraction en acides nucléiques riches en GC. C'est, il me semble, l'hypothèse la plus probable pour expliquer qu'il y ait une corrélation entre les variations de composition en base de différents phyla : il n'est pas non plus attendu que les patrons de mutations se comportent uniformément à travers les phyla. C'est-à-dire qu'il n'y a à priori aucune raison de penser que les patrons de mutations des *Bacillus* s'ajustent à un nouveau milieu de la même façon que ceux des *Helicobacter* par exemple.

Ces travaux suggèrent donc qu'il pourrait y avoir des relations entre des traits environnementaux et la composition en base ; en l'état, si cela s'avérait vrai, ç'aurait été affirmé sur la base de mauvaises données.

### Réquisits en oxygène

NAYA et al. [2002] ont observé que le taux de GC des procaryotes aérobies est supérieur à celui des procaryotes anaérobies, et ce à travers différents grands clades. ROMERO et al. [2009] montrent ensuite qu'il existe une corrélation entre le taux de GC et la quantité d'oxygène disponible dans une colonne d'eau. En effet, à 500 m et 4000 m de profondeur, qui correspondent d'après les auteurs à des strates de la colonne d'eau en conditions oxiques, la distribution du taux de GC montre un pic qui pourrait correspondre à la catégorie des procaryotes aérobies. Dans le même sens, la région à 770 m de profondeur, correspondant à la région la moins riche en oxygène, montre un pic de taux de GC bas, ce qui pourrait correspondre à la catégorie des anaérobies.

Quel serait alors l'avantage sélectif associé à un taux de GC plus élevé ? Sachant que le stress oxydatif est généralement associé à des transversions de G·C vers T·A, certains ont défendu que les guanines en position synonymes pourraient jouer un rôle « sacrificiel », protégeant le reste du génome du stress oxydatif [KANVAH et SCHUSTER 2005]. Même avec beaucoup d'efforts, j'éprouve les plus grandes difficultés à ne pas y voir l'une de ces « "just-so" stories » que Gould et Lewontin raillent dans « The Spandrels of San Marco and the Panglossian Paradigm : A Critique of the Adaptationist Programme ».

En tout les cas, ASLAM et al. [2019] ont récemment montré, en utilisant les taux de GC de 2154 aérobies et 1758 anaérobies, ainsi que des méthodes statistiques rigoureuses contrastant pour l'inertie phylogénétique, que la relation précédemment observée n'est pas statistiquement soutenue, à l'exception des aérobies obligatoires. Les aérobies obligatoires ont un taux de GC significativement supérieur à celui des aérobies et des anaérobies facultatifs ou obligatoires. Ils corroborent ainsi les résultats

## 5. SÉLECTION SUR LA COMPOSITION EN BASE (?)

---

obtenus par d'autres équipes indépendamment [BOHLIN et al. 2010; VIEIRA-SILVA et Eduardo P. C. ROCHA 2008].

### *Ressources métaboliques et fitness*

Eduardo P. C. ROCHA et DANCHIN [2002] ont proposé, en comparant le taux de GC de bactéries associées à des hôtes (typiquement des bactéries pathogènes) et celui de bactéries à style de vie « libre », que l'explication du taux de GC inférieur des bactéries « non-libres » pourrait être dû aux plus faibles coûts métaboliques associés à la synthèse de A et T, par rapport à celui des G et C. Il se pourrait que le « contenu en AT élevé des bactéries non-libres soit le résultat de la sélection par compétition pour des ressources limitées ». En ce sens, on retrouve le même type de « phénotype » vis à vis de la composition en base pour différents éléments qui se comportent comme des pathogènes à l'égard d'un hôte : les plasmides, les phages et les transposons à l'égard du génome ; et certaines bactéries pathogènes à l'égard de leur hôte. Tous présentent un taux de GC inférieur à celui de leur hôte. Dans le contexte métabolique de l'hôte, la plus faible disponibilité des différents acides nucléiques pourrait ainsi contribuer à ce qu'il y ait un avantage, pour un pathogène, à voir le coût métabolique de sa réplication diminuer.

Cependant, comme Eduardo P. C. ROCHA et DANCHIN le disent explicitement, cette hypothèse ne peut pas expliquer l'entière du spectre de variation du taux de GC, et au premier chef, l'existence de génomes riches en GC.

Une approche intéressante développée récemment par HELLWEGGER et al. [2018] a consisté à modéliser l'évolution d'une population de bactéries en paramétrant explicitement le coût métabolique en carbone et azote, associé à la synthèse d'une molécule d'acide nucléique donnée et des acides aminés qu'elle encode. C'est une approche qu'on pourrait qualifier de « stœchiométrie évolutive ». En faisant varier la disponibilité du milieu en carbone ou azote, les coûts respectifs des acides nucléiques, des acides aminés et donc des codons changent. Par exemple, la synthèse d'une paire de base A·T requiert un atome d'azote en moins qu'une paire de base G·C (7 contre 8) ; la synthèse de la phénylalanine requiert 9C et 1N, celle de l'arginine 6C et 4N. Le modèle implémente par ailleurs d'autres fonctions plus classiques des modèles d'évolution, tels que le coût associé à une mutation non-synonyme, le taux de recombinaison et de mutation, la taille efficace, le patron de mutation, etc ...

Dans deux modèles appliqués au cas de *Roseobacter*, un genre d'Alphaprotéobactérie marine riche en GC (55 à 66%), sont paramétrés simultanément :



1. le biais lié au fait que la probabilité pour une base d'un génome riche en GC de muter de AT vers GC est inférieure à la probabilité de la mutation inverse;
2. le patron mutationnel spécifique à ce modèle;
3. une pénalité liée aux mutations non-synonymes;
4. les réquisits en ressources métaboliques liés à la synthèse de l'ADN;
5. ceux liés à la synthèse des protéines.

En faisant évoluer *in-silico* des populations de *Roseobacter* ainsi modélisées, dans les deux cas de carences en azote ou en carbone, les auteurs constatent une augmentation du taux de croissance déduit après 4000 ans, augmentation plus forte dans le cas de la carence en azote. « Cela suggère que la carence en C et en N représentent de fortes forces sélectives qui peuvent surmonter les biais mutationnels intrinsèques et contraindre l'évolution Darwinienne de bactéries marines en ajustant la composition en acides aminés. » L'un des résultats étonnants est que le modèle prédit, à partir d'une population qui présenterait les traits génomiques reconstitués de l'ancêtre commun le plus récent (MRCA) aux *Roseobacter* (GC  $\approx$  50%), et qui évoluerait dans un environnement limité en carbone, un taux de GC de 64,5[62,7 – 66,4]%. Or celui de l'espèce DSS-3, isolé dans un milieu limité en carbone, est de 64,1 %.

Si ce modèle reste ce qu'il est — un modèle —, il suggère qu'il est *possible*, dans certaines conditions, que la sélection opère sur la composition en base. La question de la manière dont la sélection s'y prend pour opérer directement sur le taux de GC reste cependant entièrement ouverte.

### Résumé

Dans cette partie j'ai voulu présenter quelques unes des hypothèses sélectives les plus fréquemment avancées pour rejeter l'hypothèse nulle selon laquelle l'évolution de la composition en base est guidée par des forces neutres ou quasi-neutres. Toutes n'ont pas été présentées (voir pour d'autres hypothèses les revues de Eduardo P. C. ROCHA et FEIL [2010] et AGASHE et SHANKAR [2014]). Pour certains, il est clair que si sélection il y a, elle est conçue comme s'appliquant moins à l'échelle des substitutions individuelles qu'à l'échelle des mécanismes qui génèrent ces mutations [Eduardo P. C. ROCHA et DANCHIN 2002; SUEOKA 1993]. Pour d'autres, il n'est pas très clair de quelle façon la sélection opère effectivement sur la composition en base. De quelle façon une mutation portant sur un nucléotide isolé dans un génome de plusieurs mégabases

## 5. SÉLECTION SUR LA COMPOSITION EN BASE (?)

---

peut-elle affecter suffisamment les coefficients de fitness [PIGANEAU et al. 2001]?

S'il est évidemment pertinent de chercher à déterminer s'il existe des relations entre traits génomiques et traits environnementaux, les approches à grands traits employées ici font face au problème qu'elles ignorent une bonne partie des détails environnementaux confondants. Les approches expérimentales détaillées dans la partie qui suit tentent de s'en affranchir.

### 5.2 Indices expérimentaux de sélection sur le GC%

*The Bellman : plasmides, GFP et taux de croissance*

RAGHAVAN et al. [2012] ont publié en 2012 un article dans lequel ils rapportent avoir observé une corrélation linéaire positive significative entre la quantité de bases G et C d'un gène codant pour la protéine GFP et le taux de croissance d'une souche d'*E. coli*. Pour cela, différents clones porteurs de plasmides différant par l'allèle du gène qu'ils comportent et codant la GFP, allant de 40,4 à 53,7% de GC%, ont été cultivés en milieu LB [KUDLA et al. 2009]. Une autre version de l'expérience a utilisé différentes versions de la DNA-polymérase du phage  $\Phi_{29}$  de *Bacillus* au lieu de la GFP. Les plasmides utilisés sont des plasmides pET, dérivés du pBR322, dont le nombre de copies par cellule est compris entre 15 et 20 [BALBÁS et BOLÍVAR 2004]. Dans les deux cas, le taux de croissance des porteurs de l'allèle riche en GC est supérieur à celui des porteurs de l'allèle riche en AT. (Cela reste vrai quand l'indice d'adaptation des codons est pris en compte.) Les auteurs l'interprètent donc comme un indice de sélection en faveur d'un taux de GC plus élevé.

*Et les autres modèles ?*

Le même laboratoire a ensuite reproduit l'expérience avec d'autres modèles, incluant *Caulobacter crescentus*, *Pseudomonas aeruginosa* et *Escherichia coli* à nouveau [KELKAR et al. 2015]. Là encore est observée une relation positive entre le taux de croissance et la composition en base du gène GFP, chez deux des trois espèces : *C. crescentus* et *E. coli*. Cependant, dans un *corrigendum* post-publication, les auteurs rapportent qu'une fois ôtées les valeurs aberrantes chez *C. crescentus*, la relation n'est plus significative. De sorte qu'il ne reste vraiment que *E. coli* chez qui cette relation ait vraiment été démontrée.

Cela n'a pas empêché les auteurs d'étendre la portée de leurs résultats à l'ensemble du règne des Procaryotes, et d'en conclure que c'est

bien vers la sélection naturelle qu'il faut se tourner pour expliquer l'écart entre composition en base à l'équilibre et composition en base observée. Le moins qu'on puisse dire est qu'il s'agit probablement d'une exagération de l'importance du résultat : d'abord parce qu'il n'est visiblement vrai que chez *E. coli*, *a fortiori* quand le gène est inséré dans un plasmide multi-copies ; ensuite parce que je ne suis pas certain qu'il soit tout à fait sensé de réduire la fitness au taux de croissance (même si évidemment c'en est une composante importante, sinon la plus importante).

Avant même tout cela, il n'est pour le moins pas évident que la façon dont est conçue l'expérience permette même de répondre à la question posée. Quel peut bien être en effet l'avantage sélectif lié à une composition en base plus élevée sur un gène, la GFP, qui, telle qu'est conçue l'expérience, n'est d'aucun intérêt pour la cellule ?<sup>33</sup>

### *Pangloss et Clifford*

J'ai tenté dans cette partie de décrire certaines des hypothèses avancées pour expliquer par la sélection naturelle l'écart entre la composition en base attendue à l'équilibre mutationnel et l'écart observé. Toutes ne se valent pas ; les mieux étayées n'ont pas de prétentions à l'universalité, les plus fragiles ou les plus débattues ne sont de toutes façons pas fondées à y prétendre, lorsqu'elles ne relèvent pas carrément d'un adaptationnisme naïf à la Pangloss.

William Clifford disait dans son *Éthique de la croyance* :

Il est mauvais dans tous les cas de croire sur des éléments de preuve insuffisants ; et où c'est présomption que de douter et chercher, là c'est pire présomption que celle de croire.<sup>34</sup>

Jusqu'à présent je n'ai pas trouvé d'éléments de preuve suffisants qui aient réussi à me convaincre que la sélection peut bien jouer le rôle qu'on veut lui faire jouer à l'égard de la composition en base ; sans doute alors n'est-il pas trop présomptueux de continuer à douter et chercher.

C'est bien pourquoi il me semble qu'il y a de la place pour la conversion génique biaisée, qui jusqu'alors n'a guère été étudié dans le contexte des Procaryotes, à quelques exceptions près. C'est l'objet de la section suivante de décrire ce que l'on sait de la conversion génique biaisée et des quelques indices que l'on a de son existence chez les bactéries.

33. Il y a de plus un autre problème expérimental qui n'est pas détaillé dans les articles cités, auquel j'ai été directement confronté, et qui a trait à la *randomisation* des manipulations. Je le détaille en annexe 9.4.

34. « It is wrong in all cases to believe on insufficient evidence ; and where it is presumption to doubt and to investigate, there it is worse than presumption to believe ». Cité par BOUVERESSE 2015.



## CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

---

Mais on ne gagne rien, dans une culture, à dénoncer sans cesse le *caractère hypothétique des hypothèses*, non plus qu'à se limiter à une affirmation inconditionnée sur la réalité des faits. Il faut que le problème de l'interprétation des faits soit toujours ouvert. Et tous les essais d'interprétation doivent laisser une trace dans la culture.

Gaston BACHELARD [1990]. *Le matérialisme rationnel*. Paris : Presses universitaires de France. 236 p., page 122.

La vie est un tourbillon continu, dont la direction, toute compliquée qu'elle est, demeure constante, ainsi que l'espèce des molécules qui y sont entraînées, mais non les molécules individuelles elles-mêmes, au contraire la matière actuelle du corps vivant n'y sera bientôt plus, et cependant elle est dépositaire de la force qui contraindra la matière future à marcher dans le même sens qu'elle. Ainsi la forme de ces corps leur est plus essentielle que leur matière, puisque celle-ci change sans cesse, tandis que l'autre se conserve, et que d'ailleurs ce sont les formes qui constituent les différences des espèces, et non les combinaisons de matières, qui sont presque les mêmes dans toutes.

Georges CANGUILHEM [1981]. *Idéologie et rationalité dans l'histoire des sciences de la vie : nouvelles études d'histoire et de philosophie des sciences*. Vrin. 146 p., p.187

La mutation est le matériau « ultime », au sens de MAYR [1961] sur lequel travaillent les différentes « forces » évolutives que sont la sélection naturelle et la dérive génétique [LYNCH et al. 2016; SUNG, ACKERMAN, MILLER et al. 2012]. Les travaux de ces cinquante dernières années ont contribué à montrer qu'il convient d'ajouter, à ces deux forces, celle que constituent les biais moléculaires [HURST 2019]. C'est à DOVER [1982] que l'on doit d'avoir popularisé ce concept de biais moléculaire (*molecular drive*, que l'on pourrait traduire par « élan moléculaire ») :

L'élan moléculaire n'est pas une alternative aux processus évolutifs de sélection naturelle et dérive génétique dans la fixation de variants mutationnels de gènes à copie unique (bien qu'ils ne soient en rien exclus de l'activité de

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

---

conversions dirigées et puissent être aussi soumis à l'élan); il constitue un *troisième mode d'évolution* qui, par nécessité, serait sujet à sélection naturelle et dérive génétique dans une variété intéressante de situations biologiques réelles.<sup>35</sup>

Parmi les biais moléculaires bien décrits chez les Eucaryotes, on compte la conversion génique biaisée (BGC). La *conversion génique* se produit notamment lors de la réparation des mésappariements dans les hétéroduplexes qui se forment au cours de la réparation des cassures double-brins par recombinaison homologue. La réparation de ces mésappariements donne lieu à la *conversion* de l'information génétique portée par l'un des brins par celle du brin matrice. Elle conduit à l'échelle d'un produit de méiose eucaryote, à une ségrégation non-mendélienne. À l'échelle de l'ensemble des produits de méiose, c'est-à-dire des gamètes d'un individu, cette conversion n'affecte pas la ségrégation si les deux haplotypes ont la même probabilité d'être les *donneurs* (l'haplotype matrice, qui convertit l'autre haplotype) de l'événement de conversion. Cependant, si l'un des haplotypes se trouve être plus fréquemment le donneur, la conversion génique de cette haplotype est dite *biaisée*.

On connaît à ce jour deux types de conversion génique biaisée :

**BGC DSB-dépendante (dBGC)** La conversion génique peut être biaisée si l'un des haplotypes est plus prompt à subir des cassures double-brins que l'autre. Il sera alors plus fréquemment réparé par l'haplotype homologue, donnant lieu à l'effacement progressif de l'haplotype « chaud » du point de vue des cassures au profit de l'haplotype « froid » [COOP et S. R. MYERS 2007]. Ce phénomène a des conséquences très intéressantes sur l'évolution des paysages de recombinaison notamment chez ces mammifères où la position des cassures double-brins est dirigée par la protéine PRDM9 [BAUDAT et al. 2010; Yann LESECQUE et al. 2014].

**BGC en faveur de GC (gBGC)** Il a été montré chez différents organismes, sur lesquels je reviens plus loin, que lorsqu'un mésappariement dans un hétéroduplexe oppose une base A ou T à une base G ou C, la réparation tend à ce que les bases G et C soient plus fréquemment les donneuses de l'événement de conversion. Ce biais de réparation en faveur des bases G et C peut avoir des conséquences fortes sur l'évolution de la composition en base, ce que je montrerai dans la partie 6.1.

Auparavant, je souhaiterais revenir un peu plus avant sur la conversion génique en en présentant le modèle classique chez les eucaryotes.

35. DOVER 1982, Je souligne.

## Conversion génique : définitions & mécanismes

WALSH [1983] définit la conversion génique comme le « transfert non-réciproque d'information génétique d'un variant à l'autre et résultant généralement en une déviation depuis la ségrégation Mendélienne normale. » En un sens plus général, on parle de conversion génique dès lors que l'information portée par une portion d'une molécule d'ADN est remplacée par celle d'une molécule homologue [James D. WATSON 2014, pp 368-374].

Une cassure double brin se produit sur l'un des haplotypes, ce qui entraîne la résection 3'-5' des extrémités de la cassure. Ces extrémités envahissent la double-hélice de la région homologue, et forment une double-jonction de Holliday. La resynthèse de l'ADN lésé sur la base du brin homologue conduit à la *conversion génique* du brin — par le brin — (en **a**).

La *résolution* des jonctions, ici en produit de non-crossover (NCO), conduit à des *hétéroduplexes*.

Ici un mésappariement dans cet hétéroduplexe entre une base G et une base A est représenté. Ce mésappariement est susceptible d'être réparé, ce qui là encore donne lieu à de la conversion génique (en **b**), soit d'un A par un G (à gauche), soit d'un G par un A (à droite). La conversion génique a ce locus est dite *biaisée vers GC* si le scénario de gauche se réalise plus fréquemment que celui de droite.

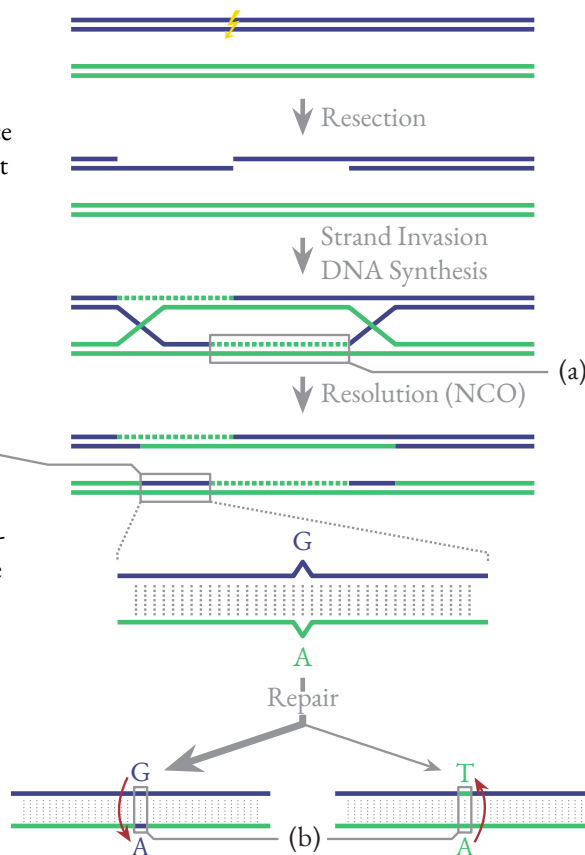


FIGURE 6.1 – Modèle classique de la conversion génique au cours de la réparation des DSB.

La conversion génique se produit lors de la correction des mésappariements dans les produits intermédiaires de la recombinaison appelés « hétéroduplexes » (figure 6.1). Elle a été décrite dans les 1960–1970 par HOLLIDAY [1964] puis raffinée par M. S. MESELSON et RADDING [1975], notamment chez les Ascomycètes du genre *Ascolobus* ou *Neurospora*, pour lesquels il est relativement aisé de discriminer les quatre produits de la méiose par dissection de tétrades. Un modèle de la manière dont la conversion génique peut se dérouler chez les eucaryotes, inspiré de

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

---

CHEN et al. [2007], est détaillé en figure 6.1. Brièvement, la conversion génique se produit au cours de la réparation des cassures double brins (*double-strand break*, désormais DSB) par recombinaison homologue. La détection d'une DSB conduit à la résection des brins bordants la lésion, générant des extrémités 3' simple-brin sortantes, qui « parcourent » le génome en quête d'une région homologue pour servir de matrice à la réparation. Lorsqu'elle est détectée, on dit de l'un des brins qu'il « envahit » la double hélice homologue, formant une structure en D caractéristique appelée *D-loop*. L'association de ces deux molécules d'ADN peut présenter des sites polymorphes, donc mésappariés, lorsque la région matrice correspond à un autre haplotype que celui de la région lésée : c'est un hétéroduplexe. La conversion génique a lieu lorsque ces sites mésappariés sont réparés par une machinerie moléculaire spécifique [CHEN et al. 2007]. Les sites convertis sont donc situés à proximité de la DSB [ORR-WEAVER et al. 1988].

Chez certains organismes, lorsque les mésappariements en question opposent une base A ou T à une base G ou C, la réparation tend à favoriser les bases G et C : on parle de conversion génique biaisée vers GC, ou gBGC. J'ai avancé plus haut que ce mécanisme peut avoir des conséquences fortes sur l'évolution de la composition en base, ce qui peut être prouvé par deux types d'éléments : 1) des éléments théoriques liés à la modélisation du destin évolutif d'un ou plusieurs locus soumis à de la conversion génique récurrente ; et 2) des preuves empiriques des conséquences de la gBGC sur l'évolution de la composition en base, notamment celle du génome humain.

### 6.1 Sélection, dérive & recombinaison

#### *Conséquences évolutives de la conversion biaisée : modèles théoriques*

Les travaux qui ont cherché à modéliser la trajectoire évolutive de locus soumis à de la conversion génique sont contemporains de ceux qui ont formalisé mathématiquement la dérive génétique. Ils ont montré qu'il s'agissait d'une force dont l'intensité n'est pas à négliger dans les modèles de génétique des populations [GUTZ et LESLIE 1976 ; B. C. LAMB et HELMI 1982]. NAGYLAKI [1983b] a ainsi modélisé l'évolution de locus multialléliques soumis à de la conversion génique, en grandes et petites populations.<sup>36</sup> WALSH [1983] s'est concentré sur le cas plus simple d'un locus biallélique dans une population finie. Je détaille ici son raisonnement (d'un abord plus facile que celui de NAGYLAKI), parce qu'il en découle une conclusion importante.

36. NAGYLAKI 1983a,b.



Si l'on considère un locus diallélique, aux allèles  $A$  et  $a$ , l'on peut définir la force du biais  $b$  par  $b = 2k - 1$ ,<sup>37</sup> où le ratio de ségrégation de  $A/a$  est  $k/1-k$ , à partir d'hétérozygotes  $Aa$ . Si une ségrégation normale  $1/1$  se produit ( $k = 1/2$ ),  $b = 0$ ; sinon  $b$  peut varier de  $1$  (seulement des gamètes  $A$  sont produits par les hétérozygotes) à  $-1$  (seulement des gamètes  $a$  sont produits). Soit  $\gamma$  la probabilité qu'un site soit impliqué dans un événement de conversion et  $\beta$  la probabilité qu'un site hétérozygote  $a : A$  impliqué dans un événement de conversion transmette l'allèle  $a$ . Alors la proportion  $(1 - \gamma)$  d'événements de ségrégations produisent autant de gamètes  $A$  et  $a$ , et  $\gamma\beta$  produisent seulement des gamètes  $A$  (impliquant que  $k = \frac{1-\gamma}{2} + \gamma\beta$ ) et  $b = \gamma(2\beta - 1)$ . Si la conversion génique est non-biaisée (c'est-à-dire qu'aucun allèle n'est favorisé par un événement de conversion inégal),  $\beta = 1/2$  et  $b = 0$ , indépendamment de la fréquence de conversion  $\gamma$ .

En l'absence de conversion génique biaisée, les fitness des différents génotypes  $AA : Aa : aa$  sont notées respectivement  $1 : 1 + h : 1 + s$ , où  $h$  est le coefficient de sélection associé à l'hétérozygotie et  $s$  celui associé à l'homozygote  $aa$ . Si le locus est soumis à de la conversion génique biaisée d'intensité  $b$ , la dynamique de fixation des allèles est identique à celle d'un locus soumis uniquement à la sélection avec des fitness respectives de  $1 + 2b : 1 + b + h : 1 + s$ . Il s'ensuit qu'un allèle sélectivement neutre ( $h = s = 0$ ) favorisé par la conversion biaisée équivaut à un allèle de fitness additive  $b$ . On ne peut qu'insister sur ce point : *un allèle favorisé par la conversion génique avec un coefficient  $b$  se comporte comme un allèle favorisé par la sélection naturelle avec un coefficient de sélection  $s = b$ , toutes choses égales par ailleurs.*

On peut alors appliquer à ce cas de figure la théorie standard pour la sélection additive [KIMURA 1957] : Ainsi, dans une population panmictique, un allèle défavorisé par la conversion ( $b < 0$ ) est perdu, et un allèle favorisé par elle ( $b > 0$ ) est fixé.

En postulant une taille efficace de population  $N_e$ , et un  $|b|$  faible, la probabilité  $P(U|p)$  qu'un allèle  $A$  soit fixé sachant sa fréquence initiale  $p$  est

$$P(U|p) \approx \frac{1 - e^{-4pN_e b}}{1 - e^{-4N_e b}} \quad (6.1)$$

En particulier, la probabilité  $P(U|\frac{1}{2N_a})$  est la probabilité de fixation de  $A$  sachant que  $A$  débute à l'état de simple copie dans une population de

37. WALSH utilise la notation de  $d$  pour quantifier la force du *drive*; par soucis de

cohérence avec la suite du manuscrit nous utiliserons plutôt la notation  $b$  pour *bias*.

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

---

$N_a$  diploïdes. Puisque  $|b| \ll 1$  et  $1/2N_a \ll 1$ , l'équation 6.1 donne :

$$P(U|\frac{1}{2N_a}) \approx \begin{cases} 1/2N_a & \text{si } 4N_e|b| \ll 1 \\ 2bN_e/N_a & \text{si } 4N_e b \gg 1 \\ (-2bN_e/N_a) e^{4N_e b} & \text{si } 4N_e b \ll -1 \end{cases} \quad (6.2)$$

Pour un allèle strictement neutre (en l'absence de tout biais de conversion ou de sélection),  $P(U|\frac{1}{2N_a}) = 1/(2N_a)$ , ce qui implique que la conversion est dépassée par la dérive quand  $4N_e|b| \ll 1$ . Par contre, quand  $4N_e b \gg 1$ , le biais de conversion surpasse la dérive génétique et peut constituer une force évolutive significative. En ce cas, de nouveaux mutants dans une situation de désavantage à l'égard de la conversion ( $b < 0$ ) sont très rarement fixés par rapport à un allèle strictement neutre, alors que les allèles avantageés par la conversion ( $b > 0$ ) ont une probabilité de fixation bien supérieure à celle d'allèles strictement neutres. On comprend ainsi que la probabilité de fixation d'un allèle à un locus soumis à de la conversion biaisée dépend essentiellement de ce facteur  $4N_e b$ , que l'on notera  $B$  dans la suite de ce manuscrit. Pour généraliser cette discussion aux espèces haploïdes — dont les Procaryotes —, je ferai intervenir la ploïdie  $\psi$ , qui vaut 1 pour les espèces haploïdes et 2 pour les espèces diploïdes, si bien que  $B$  peut s'écrire :

$$B = 2\psi N_e b \quad (6.3)$$

C'est vraiment de  $B$  que dépend *in fine* la probabilité de fixation d'un allèle à un locus soumis à de la conversion génique.

Par conséquent, le fait que la conversion génique puisse être biaisée à un locus donné conditionne fortement la probabilité de fixation de l'allèle favorisé par la conversion. Globalement, si la conversion tend à favoriser une catégorie comme les bases G et C, on s'attend à ce qu'ils soient préférentiellement transmis, donc à ce que les régions enclines à subir des conversions géniques comme les régions recombinantes s'enrichissent en bases G et C. C'est là le fondement théorique qui a permis d'avancer l'hypothèse gBGC pour expliquer la variabilité du taux de GC dans le génome humain.

La partie suivante est consacrée à détailler les étapes du raisonnement qui ont contribué à ce que la gBGC soit globalement acceptée comme l'hypothèse qui permet d'expliquer le mieux cette variabilité. Elle permettra en passant d'illustrer en quoi les biais moléculaires peuvent avoir des conséquences importantes sur l'évolution de la composition en base.

*Conséquences évolutives de la conversion biaisée vers GC :  
preuves empiriques – l'évolution des isochores*

Dans les années 1980, en étudiant la précipitation des fragments du génome humain dans le chlorure de césium, précipitation qui dépend de la composition en base, Giorgio BERNARDI, OLOFSSON et al. [1985] ont démontré que la composition en base du génome des vertébrés n'est pas homogène. Elle est structurée en large segments longs de plusieurs mégabases relativement homogènes en terme de taux de GC [EYRE-WALKER et HURST 2001]. Ces larges segments furent baptisés « isochores » [Giorgio BERNARDI, OLOFSSON et al. 1985].<sup>38</sup>

Giorgio BERNARDI et d'autres ont proposé que cette structuration résulte de pression de sélection ; lui propose par exemple que les isochores riches en GC aient évolué chez les vertébrés à sang chaud (oiseaux et mammifères) pour augmenter la stabilité de l'ADN [Giorgio BERNARDI 1993, 2000]. Cependant, la force évolutive à l'origine des isochores affecte aussi bien les régions fonctionnelles du génome que les régions non-fonctionnelles [ARNDT et al. 2005], une observation difficilement compatible avec l'hypothèse adaptationniste. Elle pourrait l'être si la sélection opérait sur la composition en base globale. Mais PIGANEAU et al. [2001] démontrent que les effets sélectifs additifs de sites en liaison ne surpassent les effets de la dérive qu'à condition que le nombre de sites sous sélection soit inférieur à la taille efficace de population — autrement dit la sélection ne peut opérer simultanément sur l'ensemble des sites d'un génome dès lors que ce génome est plus grand que la taille de population, ce qui est assurément le cas pour le génome humain (environ  $3,4 \cdot 10^9$  bp, pour une taille efficace de l'ordre de  $10^4$  selon les populations TENESA et al. [2007]).

Une autre hypothèse qui a été avancée pour expliquer la structure en isochores est que les isochores pourraient être soumis à des pressions mutationnelles différentes [FRANCINO et OCHMAN 1999]. Mais EYRE-WALKER [1999] montre en analysant le polymorphisme des gènes MHC chez les mammifères qu'il y a un excès de mutations vers AT par rapport au nombre de mutations vers GC : on ne peut donc pas expliquer les isochores par des différences de pressions mutationnelles, puisqu'elles sont uniformément biaisées vers AT. Il observe de plus un biais de fixation des mutations vers GC aux positions dites « silencieuses » (positions synonymes ou non-codantes), ce qui peut être expliqué par la sélection ou la conversion génique biaisée vers GC. C'est d'ailleurs EYRE-WALKER lui-même, dans un article précédent, qui avait proposé cette hypothèse.

38. DURET et GALTIER [2009] soulignent que la composition en base s'est avérée plus variable à une échelle plus fine

qu'attendue d'après ce terme d'isochores, sans pour autant vider de son sens le terme d'isochore.

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

---

EYRE-WALKER [1993] suggèrait en effet qu'il pourrait y avoir un lien entre la recombinaison et la composition en base dans le génome des mammifères, sur la base de corrélations indirectes entre le nombre de chiasme par chromosome (un phénomène associé aux crossing-over), donc indirectement du taux de recombinaison par chromosome, et leur taux de GC. Lui et HOLMQUIST [1992] formulent ainsi l'hypothèse gBGC sur les bases suivantes :

1. La composition en base et la recombinaison sont probablement liées dans les génomes des mammifères.
2. BROWN et JIRICNY [1987] rapportent que les mécanismes de correction des mésappariements n'opèrent pas avec la même efficacité selon les mésappariements, donc qu'ils sont susceptibles de générer des biais de conversion génique.
3. Les travaux théoriques déjà décrits de NAGYLAKI [1983a] entre autres montrent l'importance évolutive que peuvent revêtir les biais de conversion génique.

EYRE-WALKER expose l'hypothèse de la conversion génique biaisée vers GC en ces termes :

[Une autre] possibilité est que la variation des fréquences de recombinaison [...] *cause* une variation du contenu en GC. C'est une idée attrayante, puisqu'il existe au moins un mécanisme bien connu par le moyen duquel cela pourrait se produire : la conversion génique biaisée [Bernard C. LAMB 1986]. Au cours de la recombinaison, des mésappariements se forment dans l'ADN en hétéroduplexe, dont la réparation subséquente peut conduire à de la conversion génique biaisée. [...] Ainsi, si le biais incline vers G+C, les régions hautement recombinantes tendront à devenir riches en G+C, du fait que les mutations G·C sont stimulées (*driven*) à travers la population.<sup>39</sup>

FULLERTON et al. [2001] confirment ensuite la corrélation entre composition en base et taux de GC dans le génome humain, une fois celui-ci séquencé. En elle-même, cette corrélation ne prouve pas qu'il y ait bien une relation de causalité, c'est-à-dire que ce soit bien la conversion génique biaisée vers GC qui est à l'œuvre et pas une autre force. Par

39. EYRE-WALKER 1993. Je souligne. Je n'ai pas trouvé de traduction satisfaisante pour « *driven* » ; comme l'a montré PLUTYNSKI 2007, dès la formulation des premiers modèles de génétique des populations statistiques par Fischer, les généticiennes et généticiens ont eu

tendance à recourir à des analogies à la mécanique statistique, ou à des locutions issues de son champ lexical. Ici, le sens de *driven* qui conviendrait sans doute le mieux serait celui de « propulsé » ou « élané ».

exemple, CHARLESWORTH [1994] suggère que la relation entre taux de recombinaison et taux de GC pourrait s'expliquer par une meilleure efficacité de la sélection en faveur des bases G et C dans les régions recombinantes. On pourrait aussi imaginer que l'action mutagène de la recombinaison [Bjarni V. HALLDORSSON et al. 2019] elle-même conduise à un enrichissement en G et C des régions recombinantes, sans que cela soit dû à de la conversion biaisée.

Mais DURET et ARNDT [2008] montrent d'une part que chez l'Homme, il y a une corrélation entre le patron de *substitution* vers GC et le taux de recombinaison; d'autre part que la corrélation entre le patron de substitution et le taux de crossover est plus forte chez l'homme que la femme, ce qui peut difficilement être expliqué par la sélection : on voit mal quel avantage sélectif pourrait représenter un biais de substitution en faveur des bases G et C qui serait spécifiquement associé aux mâles. Enfin, GLÉMIN et al. [2015] démontrent que, dans le génome humain, la recombinaison augmente la probabilité de *fixation* des allèles G et C. C'est donc bien l'hypothèse gBGC qui semble la plus compatible avec les données empiriques pour expliquer la variabilité du taux de G et C dans le génome humain.

Enfin, si probantes qu'ils soient, ces indices ont définitivement été confirmés par la caractérisation directe de produits de recombinaison chez l'Homme par A. L. WILLIAMS et al. [2015] puis Bjarni V HALLDORSSON et al. [2016]. A. L. WILLIAMS et al. ont ainsi analysé le génotype de 98 produits de méiose, et identifié 103 régions correspondant à des produits de recombinaison dont la résolution est de type non-crossover (voir figure 6.1). Ils démontrent ainsi que  $(68 \pm 10)$  % des sites hétérozygotes AT : GC transmettent l'allèle GC. Cette valeur est étonnamment proche de celle de 67,6 % rapportée ensuite par Bjarni V HALLDORSSON et al., de sorte qu'il n'y plus vraiment de doutes sur l'existence de la gBGC chez l'Homme.

D'autres travaux ont montré que la gBGC ne se cantonnait ni à l'Homme, ni même aux mammifères. La gBGC a été directement quantifiée et caractérisée dans différents organismes eucaryotes, dont, à ma connaissance, les organismes listés dans la table 6.1.

De plus, des corrélations positives entre taux de GC et taux de recombinaison ont été décrites chez un grand nombre d'espèces eucaryotes, suggérant qu'il pourrait s'agir d'un phénomène très répandu chez les Eucaryotes [GALTIER, ROUX et al. 2018; PESSIA et al. 2012], même si l'on dispose de contre-exemples [H. LIU et al. 2017].

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

TABLE 6.1 – Organismes pour lesquels la gBGC a été directement mesurée.

Nom vernaculaire	Nom d'espèce	Référence
Homme	<i>Homo sapiens</i>	A. L. WILLIAMS et al. 2015 Bjarni V HALLDORSSON et al. 2016
Levure	<i>Saccharomyces cerevisiae</i>	Y. LESECQUE et al. 2013 MANCERA et al. 2008
Diamant mandarin ( <i>Zebra finch</i> )	<i>Taeniopygia guttata</i>	WEBER et al. 2014 BACKSTRÖM et al. 2010
Gobe-Mouche	<i>Ficedula albicollis</i>	SMEDS et al. 2016
Souris	<i>Mus musculus</i>	R. LI et al. 2019

### *La gBGC peut passer pour de (et s'opposer à) la sélection*

L'une des conséquences principales de la conversion génique biaisée que j'ai déjà soulignée plus haut est qu'à un locus biallélique donné, sujet à conversion génique inégale, l'avantage que possède l'un des deux allèles à l'égard de la conversion génique, se manifeste à l'échelle de la population de la même façon que se manifesterait un avantage sélectif. Dans le cadre précis de la gBGC, cela se traduit par un biais de fixation en faveur des bases G et C lorsqu'elles sont impliquées dans un mésappariement AT·GC.

Suivant NAGYLAKI [1983a], WALSH [1983], et DURET et GALTIER [2009], on peut modéliser la probabilité d'une fixation des allèles GC aux polymorphismes GC : AT par :

$$P_{AT \rightarrow GC} = \frac{1 - e^{-2b}}{1 - e^{-4N_e b}}; \quad P_{GC \rightarrow AT} = \frac{1 - e^{2b}}{1 - e^{4N_e b}}$$

où  $b$  est le coefficient de gBGC qui dépend

1. du taux de conversion  $c$ , c'est-à-dire la probabilité qu'un site donné soit impliqué dans un événement de conversion;
2. du biais de transmission  $k_{gc}$ , c'est-à-dire la probabilité que l'allèle GC soit transmis lorsqu'un site hétérozygote GC : AT est impliqué dans un événement de conversion.

Le taux de conversion  $c$  dépend lui-même du taux de recombinaison par paire de base (noté  $r$ ), et de la longueur des hétéroduplexes (notée  $L_h$ ). Du biais de transmission  $k_{gc}$  dépend la distortion des ratios de transmission des allèles GC notée  $b_0$ , avec  $b_0 = 2k_{gc} - 1$ .

De sorte que le coefficient de gBGC ( $b$ ) peut être décomposé en

$$\begin{aligned} b &= cb_0 \\ &= rL_h b_0 \end{aligned}$$



( $k_{gc}$  équivaut au  $k$  de WALSH, page 77.) Si la conversion est non-biaisée, alors  $k_{gc} = 1/2$  et  $b_0 = 0$ . Si la conversion est strictement biaisée vers GC, alors  $k_{gc} = 1$  et  $b_0 = 1$ .

Ce biais de fixation a deux conséquences liées : 1) il brouille les tests de sélection dans les génomes en faisant passer la gBGC pour de la sélection dirigée, [BERGLUND et al. 2009; RATNAKUMAR et al. 2010] et 2) il s'oppose à, ou renforce, l'action de la sélection, contribuant respectivement à ce que l'on sous-estime ou sur-estime son intensité [GALTIER, DURET et al. 2009; Katherine S. POLLARD et al. 2006a,b]. GALTIER et DURET [2007] en appellent donc à ce que la gBGC soit considérée comme une partie intégrante de la théorie neutre (ou quasi-neutre) de l'évolution moléculaire, qu'il convient de rejeter d'abord, pour ensuite pouvoir conclure à la sélection naturelle.

#### *Estimations de l'intensité de la gBGC*

Y. LESECQUE et al. [2013] ont estimé l'intensité du biais en faveur des bases G et C dans la réparation des mésappariements de produits de recombinaison obtenus par génotypage de tétrade chez la Levure. [MANCERA et al. 2008] Chez cet organisme,  $b_0$  a été estimé en moyenne dans les hétéroduplexes à 0,013, de sorte que  $k_{gc} = 0,507$ ; c'est-à-dire que lorsqu'un site hétérozygote AT : GC est impliqué dans un événement de conversion, l'allèle GC est le donneur dans 50,7 % des cas. À chaque méiose, 1 % du génome est affecté par la conversion. De plus, les levures se multiplient essentiellement par mitose, avec une méiose toutes les 1000 générations. Le taux de conversion effectif par base et par génération, noté  $\tau$  est donc de  $10^{-5} \text{ bp}^{-1} \text{ gen}^{-1}$ . Le coefficient de gBGC à l'échelle du génome  $b$  dépend de  $rL$  le taux de conversion par base, et de  $k_{gc}$  le biais de transmission des allèles G et C dans les sites hétérozygotes AT : GC soumis à de la conversion, et donc  $b = 10^{-5} \times b_0 = 1,3 \cdot 10^{-7}$ .

Ce biais peut sembler très faible, mais comme l'a montré l'équation 6.3, la probabilité de fixation des allèles GC à l'échelle populationnelle  $B$  dépend de  $4N_e b$ . Dès lors que  $B > 1$ , la conversion biaisée surpasse les effets de la dérive génétique, c'est-à-dire dès lors que  $N_e > \frac{1}{4} b^{-1}$ . Pour la levure, cela veut dire que  $B$  surpasse la dérive génétique dès lors que  $N_e$  est supérieure à  $1,9 \cdot 10^6$ , ce qui est compatible avec ce que l'on sait des tailles de populations chez la levure et pourrait expliquer les traces de gBGC détectées dans le genre *Saccharomyces* par HARRISON et CHARLESWORTH [2011].

En fait, dans les modèles où l'on connaît mieux le paysage de recombinaison, en particulier chez l'Homme, pareils indices masquent l'hétérogénéité de l'intensité de la gBGC le long du génome. Chez l'Humain, l'essentiel des DSB ont lieu à proximité de sites génomiques

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

---

spécifiques ciblés par la protéine à doigts de zinc PRDM9. [BAUDAT et al. 2010; S. MYERS et al. 2010] Ce type de déterminisme de la position des points chauds de recombinaison a pour conséquence d'enrichir fortement les régions porteuses de la séquence reconnue par PRDM9 en bases G et C. GLÉMIN et al. [2015] ont ainsi montré que l'intensité de la gBGC à l'échelle populationnelle peut, en moyenne, aller de  $B = 0,23$  en dehors des points chauds, de l'ordre de grandeur de la dérive génétique, à  $B = 2,96$  dans les points chauds (voir aussi l'équation 6.2 p. 78).

Pour conclure sur cette partie consacrée aux preuves empiriques dont on dispose à l'égard du rôle évolutif joué par les biais moléculaires, il apparaît que la conversion génique biaisée vers GC est un mécanisme probablement très répandu, et qui joue un rôle majeur sur la composition en base du génome des mammifères. Se pose alors la question des origines de ce biais : se peut-il qu'il ait lui-même été sélectionné, ou bien n'est-il que le signe de contraintes sélectives tierces ?

### *Quel rôle évolutif joue la gBGC ?*

Quelle peut-être l'explication de la gBGC ? Quel rôle évolutif joue-t-elle ? BIRDSELL [2002] émet l'hypothèse que la gBGC correspond à une « réponse » moléculaire au biais mutationnel qui incline vers les bases A et T :

Littéralement tous les organismes pour lesquels existe une corrélation entre recombinaison et contenu en GC semblent aussi posséder une machinerie de réparation biaisée vers GC. Je suggère que ce biais vers GC trans-règles dans les systèmes de réparation des mésappariements a évolué en réponse à un biais mutationnel vers AT résultant de propriétés fondamentales de l'ADN.

Les régions non-recombinantes du génome et les génomes non-recombinants seraient sujets à l'élan moléculaire causé par la BGC. Je suggère que les taux de GC bas, caractéristiques des génomes non-recombinants, pourraient résulter de trois processus : 1) un biais mutationnel vers AT prévalant, 2) la fixation aléatoire des types de mutations les plus communs par dérive génétique, et 3) l'absence de conversion génique biaisée vers GC qui, dans les organismes recombinants, permettrait d'inverser les formes les plus communes de mutations.<sup>40</sup>

40. BIRDSELL 2002.



Si la force mutationnelle contribue à décroître le taux de GC, il se pourrait qu'il soit avantageux dans ce contexte de « préférer » convertir un site AT·GC en GC plutôt que l'inverse.

Néanmoins, comme le signalent DURET et GALTIER [2009], la gBGC peut parfois contribuer à la fixation d'allèles délétères, et par conséquent diminuer la fitness. Qu'il faille expliquer le contrebalancement de ces coûts en fitness par des avantages évolutifs à plus long terme du type de ceux évoqués par BIRDELL n'est pour le moins pas du tout sûr. DURET et GALTIER émettent l'hypothèse que la gBGC constituerait un effet secondaire des pressions de sélection qui pèsent sur les mécanismes de réparation. Ceux-ci sont d'abord contraints par leur tâche première qu'est celle de réparer les lésions de l'ADN de cellules *somatiques*; or ce sont fréquemment des mutations liées à la déamination spontanée des cytosines, si bien que les mécanismes de réparation dans les cellules mitotiques sont extrêmement biaisés vers GC. [BROWN et JIRICNY 1987] « Une pression sélective qui restreindrait le taux de mutations somatiques devrait conduire à l'évolution d'une machinerie de réparation biaisée vers GC, ce qui aurait pour effet secondaire d'engendrer la gBGC lorsqu'elle opère sur les mésappariements dans les hétéroduplexes. » [DURET et GALTIER 2009]

Pour ma part, je trouve l'hypothèse de DURET et GALTIER plus économe de ses prétentions : d'une part parce qu'elle n'implique pas de recourir à une quelconque forme de pression de sélection du second ordre difficile à tester empiriquement; d'autre part parce qu'elle identifie clairement sur quoi pèsent les contraintes sélectives, et quelles en sont les conséquences.

La basilique Saint Marc à Venise présente la particularité architecturale de faire reposer ses dômes sur quatre arcs, dont l'intersection fait nécessairement apparaître des structures appelées pendentifs, qui ont été astucieusement exploités par les décorateurs. GOULD et R. C. LEWONTIN [1979], rappelant que les pendentifs sont moins produits d'une volonté des architectes que nécessaires conséquences d'une structure à quatre arcs, utilisent cette image pour prévenir des dérives des discours adaptationnistes. Si l'architecture génomique repose en partie sur les arcs de la machinerie de réparation, que la conversion génique soit biaisée vers GC n'en est peut-être après tout qu'un pendentif.

### *Résumés, limites et extensions*

Ces vingt dernières années ont contribué à installer la conversion génique biaisée vers GC dans le paysage des forces majeures qu'il convient de prendre en compte dans les modèles d'évolution des séquences chez les Eucaryotes. L'hypothèse gBGC a permis d'expliquer l'hétérogénéité

de la composition en base dans le génome des mammifères. Elle a été observée directement dans plusieurs modèles, et les corrélations positives entre composition en base et taux de recombinaison détectées chez un grand nombre d'espèce suggèrent qu'il s'agit d'un phénomène relativement répandu. Néanmoins, avant les années 2010, à de rares exceptions près, bien peu pensaient que la gBGC pouvait s'appliquer également aux procaryotes.

Dans la partie qui suit, je décris les éléments qui permettent de penser qu'il se pourrait que la gBGC soit aussi à l'œuvre chez les procaryotes.

### 6.2 Indices de la conversion génique biaisée chez les Procaryotes

Dans un premier temps, il pourrait sembler contre-intuitif que la gBGC puisse opérer chez les Procaryotes également. Puisqu'il s'agit d'un mécanisme intimement lié à la recombinaison méiotique, il n'était *a priori* pas envisageable de l'étendre aux Procaryotes, qui sont des organismes non-méiotiques. De plus, on pourrait de prime abord penser que la recombinaison est moins fréquente chez les Procaryotes que chez les Eucaryotes, notamment les organismes à reproduction sexuée, pour lesquels le déroulement complet de la méiose nécessite au moins un crossover par chromosome.<sup>41</sup>

Pour autant, les Procaryotes sont sujets à la recombinaison, au même titre que les Eucaryotes. Dans quelle mesure les taux de recombinaison chez les Eucaryotes et ceux chez les Procaryotes sont-ils commensurables ? Comme le précisent STUMPF et MCVEAN [2003], en génétique des populations les paramètres d'intérêt sont souvent proportionnels à  $2\psi N_e r$ , où  $\psi$  est le degré de ploïdie et  $N_e$  la taille efficace. Par exemple, chez l'Homme, l'estimation de  $\rho$ , le taux de recombinaison à l'échelle populationnelle tel que  $\rho = 4N_e r$ , est de l'ordre de  $4 \cdot 10^{-4} \text{ bp}^{-1} \text{ gen}^{-1}$ .

Peut-on avoir une estimation de ce paramètre  $\rho$ , le taux de recombinaison à l'échelle populationnelle, pour des Procaryotes ?

#### *Estimation du taux de recombinaison à l'échelle populationnelle chez les Procaryotes*

À ma connaissance, ce paramètre  $\rho = 2N_e r$  n'a pas été mesuré directement chez les Procaryotes, probablement parce qu'il y est difficile de quantifier directement le taux de recombinaison par génération, sans même parler de la taille efficace [Eduardo P. C. ROCHA 2018].

41. À ce propos, un interlocuteur m'a interpellé lors du congrès SMBE 2019 en

demandant : « But Bacteria do *not* recombine, right ? ». Wrong.

## Indices de la gBGC chez les Procaryotes

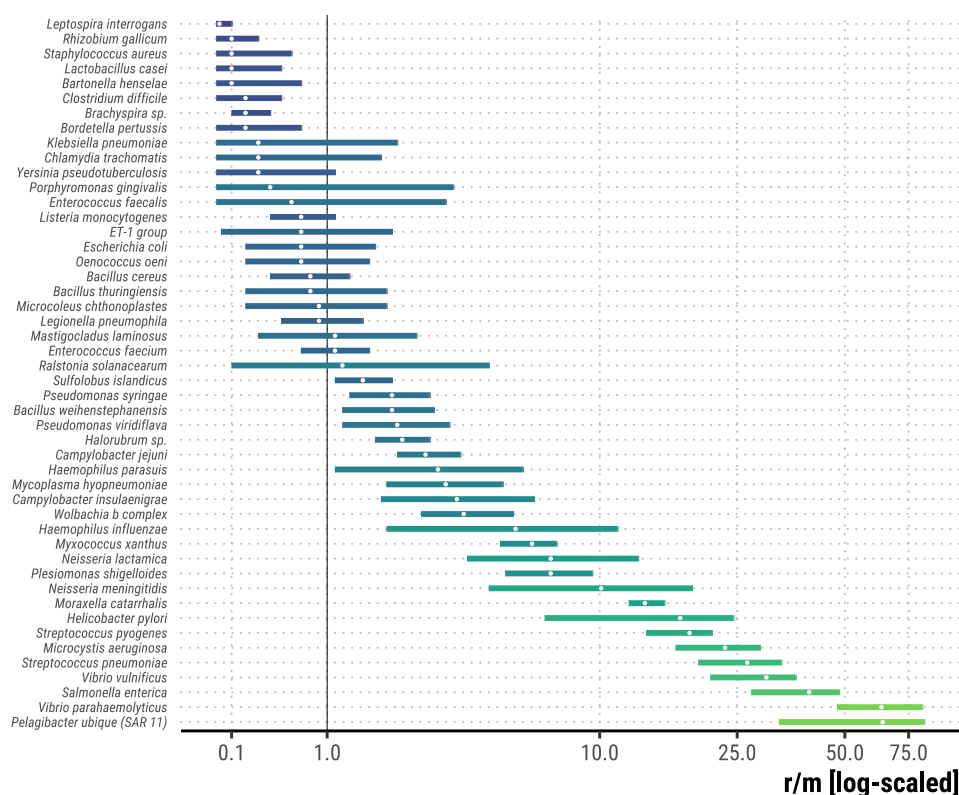


FIGURE 6.2 – Distribution du ratio  $r/\mu$  estimé pour 48 espèces de bactéries et d'archées

Concentrons-nous d'abord sur le paramètre  $r$ . L'inconvénient avec les espèces Procaryotes, c'est qu'elles ont tendance à échanger des allèles avec des partenaires distants, et ce sans qu'il y ait un seuil strict permettant de délimiter franchement une population qui satisferait les assomptions des modèles classiques de génétique des populations.<sup>42</sup> Ce qu'il est possible de faire, en revanche, c'est d'estimer les probabilités relatives qu'un changement de nucléotide soit le fait d'un événement de recombinaison  $r$  ou de mutation  $\mu$ , mesuré par leur ratio  $r/\mu$  [GUTTMAN et DYKHUIZEN 1994; MILKMAN et BRIDGES 1990; SHEPPARD et al. 2018]. Il peut être estimé de différentes manières; une méthode fréquemment utilisée, implémentée dans CLONALFRAME [DIDELOT et FALUSH 2007] puis CLONALFRAMEML, [DIDELOT et WILSON 2015] consiste à (essayer de) reconstruire l'ascendance clonale d'un ensemble de souches d'une part, et les événements de mutations et recombinaison qui se produisent le long de cette généalogie basée sur un modèle de coalescence [KINGMAN 1982]. VOS et DIDELOT [2009] ont ainsi estimé ce ratio pour 48 espèces de bactéries et d'archées (figure 6.2). On voit que le ratio  $r/\mu$

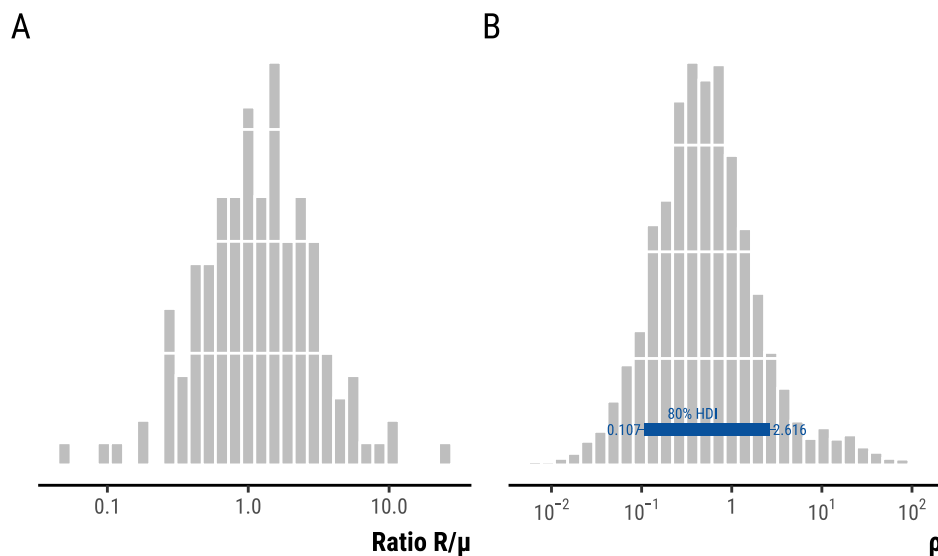
42. À ce sujet, voir la revue de Eduardo P. C. ROCHA [2018], qui mentionne notamment les travaux de MATIC et al. [1996], MAJEWSKI et COHAN

[1998], ou plus récemment ceux de OLIVEIRA et al. [2016]. Ce sont tous des travaux montrant que la divergence, au sens large, entre deux espèces réduit

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

varie fortement chez les Procaryotes : il va dans cette analyse de 0,02 à 63,60. BOBAY et OCHMAN [2018] ont conduit le même genre d'analyse avec un nombre d'espèces plus grand (153). La variabilité de  $r/\mu$  qu'ils observent est plus faible (de 0,244 à 21,032). (figure 6.3A, obtenue à partir des données que LM. BOBAY nous a fait parvenir.).

**FIGURE 6.3 – A.** Distribution du ratio  $r/\mu$  dans 153 espèces de Procaryotes. **B.** Distribution d'une estimation du taux de recombinaison populationnel  $\rho = 2N_e r$ . J'ai fait l'hypothèse que les valeurs de  $\mu$  sont distribuées comme celles des 153 espèces dont on connaît  $r/\mu$ , afin de générer une valeur aléatoire du paramètre  $r = r/\mu/\mu$ , que j'ai multiplié ensuite par l'estimation de  $2N_e$  de l'espèce correspondante, et ce  $10^4$  fois. On obtient ainsi une distribution de valeurs probables du paramètre  $\rho$  pour ces 153 espèces.



Sachant ce ratio  $r/\mu$ , il serait possible d'en déduire  $r$  sachant  $\mu$ . Mais le taux de mutation par base par génération n'a vraiment été déterminé que pour 22 espèces de Bactéries ; il va de  $1,07 \cdot 10^{-10}$  à  $9,78 \cdot 10^{-9} \text{ bp}^{-1} \text{ gen}^{-1}$ . Ne connaissant pas  $\mu$  pour les 153 Procaryotes du jeu de donnée de BOBAY et OCHMAN [2018], j'ai fait l'hypothèse que leurs taux de mutation étaient distribués comme ceux des 22 espèces de LONG et al. [2018]. En supposant que  $r$  et  $\mu$  sont indépendants, on peut estimer la distribution théorique de  $r$  sachant  $r/\mu$ .<sup>43</sup>

Mettons un instant de côté le paramètre  $r$ , pour considérer le paramètre  $N_e$ . Une façon de procéder pour estimer la taille efficace, employée par BOBAY et OCHMAN [2018] ou STUMPF et MCVÉAN [2003], consiste à utiliser l'estimateur de Wu et Watterson  $\theta$  [WATTERSON 1975], un estimateur du taux de mutation à l'échelle populationnelle dépendant du nombre de sites qui ségrègent dans la population. En effet,  $\theta = 2\psi N_e \mu$ ,

l'efficacité ou la probabilité d'un transfert, sans que cette probabilité soit jamais nulle pour autant.

43. Ça n'est vraisemblablement pas le cas que  $r$  et  $\mu$  sont indépendants. Sachant

notamment que le processus de recombinaison est lui-même mutagène,  $\mu$  est susceptible d'augmenter en raison de  $r$  [Bjarni V. HALLDORSSON et al. 2019].

de sorte que, connaissant  $\mu$  et  $\theta$ , on peut estimer  $N_e$ . Mais il est bien rare que l'on sache  $\mu$ , aussi est-il d'usage d'utiliser le ratio  $\frac{\pi_N}{\pi_S}$ , c'est-à-dire le ratio du nombre de substitutions non-synonymes et du nombre de substitutions synonymes qui ségrègent dans une population, comme un estimateur de  $N_e$ . NIELSEN et YANG [2003] montrent à ce titre que

$$\frac{\pi_N}{\pi_S} = \frac{2\psi N_e s}{1 - e^{-2\psi N_e s}} = \frac{S}{1 - e^{-S}},$$

avec  $S = 2\psi N_e s$  et  $s$  le coefficient de sélection. De sorte que, sachant  $\frac{\pi_N}{\pi_S}$  par l'analyse du polymorphisme et connaissant approximativement  $s$  par ailleurs, on peut en déduire une estimation de  $N_e$ . C'est précisément ce qu'ont utilisé BOBAY et OCHMAN [2018] pour estimer  $N_e$  chez 153 espèces de Procaryotes.

On dispose donc d'estimations de  $r$  et de  $N_e$ , ce qui nous permet de calculer une gamme de valeurs plausibles de  $\rho = 2N_e r$ , représenté dans la figure 6.3B. On voit que 80 % des valeurs simulées de  $\rho$  sont comprises entre 0,1 et 2,6. Ces estimations sont très approximatives. Quoi qu'il en soit, il apparaît tout à fait plausible qu'à l'échelle populationnelle, le taux de recombinaison des Procaryotes soit supérieur à celui estimé chez l'Homme, par exemple [STUMPF et MCVEAN 2003].

On ne peut donc pas vraiment dire que la recombinaison est moins fréquente chez les Procaryotes que chez les Eucaryotes, du moins pas au sens de « fréquent » pertinent pour la génétique des populations.

#### Relation entre $\tau/\mu$ et composition en base

De prime abord, on pourrait penser que l'« hypothèse gBGC » impliquerait une relation entre le taux de recombinaison et la composition en base du génome, pris globalement. Or les deux taux ne sont pas corrélés (figure 6.4). D'aucuns ont pris cette absence de corrélation pour un élément qui permettait de rejeter l'hypothèse gBGC [HILDEBRAND et al. 2010; WEISSMAN et al. 2019]. Il s'agit d'une erreur de raisonnement.

Pour le montrer, considérons deux cas simples. Mettons  $X$  une espèce à  $N_e = 10^8$  recombinaison à une fréquence de  $r = 10^{-10} \text{ bp}^{-1} \text{ gen}^{-1}$ , et un biais de transmission en faveur des allèles GC  $b = 0.5$  ( $k_{gc} = 0.75$ , 75 % des sites hétérozygotes AT : GC impliqués dans un événement de conversion sont convertis en GC.). À l'échelle populationnelle, le paramètre  $B$  pour l'espèce  $X$ , noté  $B_X$ , vaut  $B_X = 2N_e \tau b$ , où  $\tau$  est la fraction du génome impliquée dans un événement de conversion par génération. Cette fraction  $\tau$  dépend de la longueur des hétéroduplexes, fixée ici à  $1 \cdot 10^3 \text{ bp}$  et du taux de recombinaison par base  $r$ .

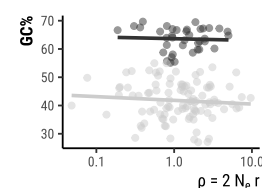
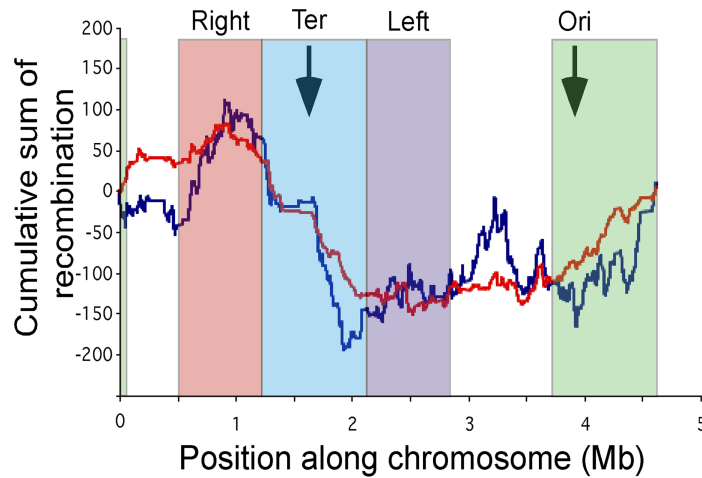


FIGURE 6.4 – Relation entre  $\rho = 2N_e r$  et taux de GC génomique. ( $n = 153$ , données extraites de BOBAY et OCHMAN 2018.)

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

FIGURE 6.5 – Somme cumulée du taux standardisé de conversion génique efficace (—), et du taux de GC (—), mesurés par fenêtre « glissante » de 3 kbp avec un pas de 500 bp le long du génome de l'espèce *E. coli*. Tirée de TOUCHON et al., sous license Creative Commons Attribution (CC BY)



Mettons maintenant Y une espèce identique à X à l'égard de  $N_e$ ,  $r$  et  $\tau$ , mais pour laquelle  $b$  vaut 0.005 ( $k_{gc} = 50,2\%$ ). On a ainsi :

$$B = 2N_e\tau b, \text{ où } \left. \begin{array}{l} B_X = 20b_x \quad \text{et } B_Y = 20b_y \\ 2N_e\tau = 2 \times 1 \cdot 10^8 \times (r \times 1 \cdot 10^3) = 20 \quad \left| \quad B_X = 10 \quad \text{et } B_Y = 0,1. \end{array} \right.$$

On voit donc que, à taux de recombinaison identiques, deux espèces pour lesquelles le paramètre  $b$  diffère ne présentent pas la même intensité de gBGC à l'échelle populationnelle. Dans le cas de l'espèce X,  $B_X > 1$  et donc la probabilité de fixation d'un allèle G et C surpasse les effets de la dérive (équation 6.3), ce qui n'est pas le cas pour l'espèce Y.

Ceci démontre que l'hypothèse gBGC ne prédit pas une corrélation positive entre le taux de recombinaison d'une espèce et sa composition en base. Autrement dit, il n'est pas possible d'exclure que la gBGC soit à l'œuvre seulement sur la base de corrélations « inter-génomiques ». En revanche, un attendu de cette hypothèse est qu'au sein d'une espèce, si la machinerie de réparation est biaisée vers GC, les régions les plus enclines à la recombinaison et à la conversion génique devraient être enrichies en bases G et C par rapport au reste du génome.

### Corrélations GC% et taux de Recombinaison

C'est précisément ce genre de relation entre taux de recombinaison et composition en base le long du génome qu'ont observés TOUCHON et al. [2009]. Les autrices et auteurs ont ainsi analysé les génomes complets de 20 souches de l'espèce *Escherichia coli*, et estimé le long du *core-genome* la probabilité qu'une base soit soumise à un événement de conversion génique (figure 6.5). Ces résultats montrent que la composition en base est corrélée au taux de conversion génique mesuré par fenêtre, au moins



à l'échelle des macro-domaines chromosomiques (représentés par les quatre blocs de couleur sur la figure).

TOUCHON et al. proposent deux explications possibles au fait que la région du terminus de réplication présente et un taux de GC inférieur et un taux de conversion génique inférieur :

1. ou bien c'est un indice d'une efficacité de la sélection réduite à ces locus, du fait d'une fréquence de recombinaison moindre ; ce qui se traduirait par plus grande incapacité de la recombinaison à éliminer les mutations délétères à cette région, sachant que la plupart des mutations sont plutôt GC→AT et que l'analyse du polymorphisme indique que la plupart des mutations non-synonymes sont délétères ;
2. ou bien c'est la conversion génique biaisée vers GC qui conduit à un biais de fixation des bases G et C dans les régions recombinantes, et donc à un taux de GC plus faible dans les régions qui le sont moins. <sup>44</sup>

En substance, c'est le même type d'analyse qu'ont conduit LASSALLE et al. [2015], en l'étendant à davantage de genres. Puisqu'il s'agit de l'article qui a probablement le plus contribué à ce que l'hypothèse gBGC soit étendue aux Procaryotes, j'en expose les conclusions importantes ici.

LASSALLE et al. se sont concentrés sur vingt genres bactériens et un genre archéen, pour lesquels ils disposait de plus de six génomes dans la base de donnée Hogenom. Pour chacune des familles de gènes — où deux gènes sont dits appartenants à la même « famille » sur la base de critères d'identité de séquence protéique — de ces génomes, il est possible de calculer d'une part le taux de GC, qu'il soit global ou aux différentes positions des codons, d'autre part un indice du degré de recombinaison qui affecte cette famille. (Il s'agit de l'indice  $\phi_w$  proposé par BRUEN [2005].) Pour s'affranchir de potentiels effets confondants liés à la paralogie, les auteurs se sont restreints aux gènes à copie unique dans le génome, de sorte que le polymorphisme d'un gène, au sein d'un genre, est bien le reflet d'événements de spéciation. <sup>45</sup>

Dans 11 des 14 genres pour lesquels il a été possible de détecter un signal de recombinaison moyen significatif (fourni par l'indicateur  $\phi_w$ ),

44. Il n'est pas inintéressant de souligner que l'on a dans cet article, sous une forme condensée, rigoureusement les mêmes débats et les mêmes raisonnements que ceux qui ont porté sur la structuration en GC des génomes mammifères ; on y retrouve par exemple quasiment telles

qu'elles les hypothèse de EYRE-WALKER [1993] ou de CHARLESWORTH [1994].

45. À la paralogie cachée près : par exemple, il se pourrait que deux gènes d'une même famille, à copie unique, dans l'espèce A et B soient le produit d'une duplication et de deux pertes.

## 6. CONVERSION GÉNIQUE BIAISÉE : QUELLE IMPORTANCE ?

---

les auteurs ont détecté un taux de GC dans les gènes recombinants supérieur à celui des gènes non-recombinants ; cette différence est d'autant plus forte lorsqu'on ne considère que la troisième position des codons (GC3).

Il se pourrait que la recombinaison favorise la fixation des codons optimaux en réduisant le déséquilibre de liaison entre sites sous sélections (un phénomène appelé « interférences de Hill-Robertson », FELSENSTEIN 1974; W. G. HILL et ROBERTSON 1966). Mais la sélection en faveur des codons optimaux ne peut pas expliquer la relation entre le taux de GC et le taux de recombinaison puisque d'une part les codons non-optimaux qui se terminent par GC sont plus fréquents dans les gènes recombinants, et d'autre part les codons optimaux qui se terminent par AT sont moins fréquents dans les gènes recombinants (à deux exceptions près). Ça n'est donc pas la sélection sur l'usage du code qui entraîne un taux de GC plus élevé dans les gènes recombinants.

De plus, le taux de GC des régions intergéniques proches des gènes détectés comme recombinants est supérieur en moyenne au taux de GC des régions intergéniques des gènes qui ne sont pas détectés comme recombinants. La force qui conduit à cette élévation du taux de GC dans les régions recombinantes agit donc uniformément sur les régions fonctionnelles (les gènes) et non-fonctionnelles du génome (les régions inter-géniques), ce qui correspond bien aux prédictions de l'hypothèse gBGC.

Enfin, si l'on classe les gènes en vingt groupes selon leur GC3 et qu'on calcule la proportion de gènes détectés comme recombinants dans cette classe, on obtient un indice  $P_{REC}$ , qui est proportionnel au taux de recombinaison moyen des gènes de cette classe.<sup>46</sup> LASSALLE et al. observent ainsi une corrélation positive significative entre le taux de GC en troisième position des codons et le taux de recombinaison moyen pour 11 des 14 espèces analysées. C'est là encore une observation qui correspond aux prédictions de la gBGC.

### Conclusions

En résumé, la conversion génique biaisée vers GC peut jouer un rôle significatif sur la composition en base, ce que cette partie s'est employée à prouver. En mobilisant des arguments théoriques, on comprend que la probabilité de fixation d'un allèle que favorise la conversion génique dépend de l'intensité du biais. De plus, pareil allèle se comporte comme

46. LASSALLE et al. montrent en effet que  $P_{REC}$  est proportionnel à la moyenne du paramètre  $\rho = 2N_e r$  pour chacun des gènes de la classe. Ce paramètre  $\rho$  est le taux de recombinaison à l'échelle

populationnelle, que l'on peut calculer avec *ClonalOrigin* [DIDELOT, LAWSON et al. 2010], mais beaucoup moins coûteux en temps de calcul.



un allèle que favoriserait la sélection avec une intensité identique à l'intensité du biais. En mobilisant maintenant des arguments empiriques, on a vu que la probabilité de fixation accrue des bases G et C dans le génome des mammifères s'explique le mieux par la conversion génique biaisée vers G et C. Ce mécanisme intervient au cours de la réparation des mésappariements dans les hétéroduplexes de recombinaison homologue, et augmente la probabilité que ce soit l'allèle GC qui soit transmis dans un hétérozygote AT : GC. Il a été expérimentalement démontré dans plusieurs modèles d'Eucaryotes.

L'une des propriétés intéressante de cette hypothèse est qu'elle a permis d'étoffer le débat entre sélectionnistes et neutralistes sur la question de la composition en base : ce que l'on aurait pu prendre à première vue comme un indice de sélection en faveur des bases G et C dans le génome humain s'est finalement avéré résulter d'un processus non-adaptatif, justifiant par conséquent les appels à une version étendue de la théorie neutre [GALTIER et DURET 2007].

Si initialement l'on pensait que la conversion génique biaisée vers GC était restreinte aux Eucaryotes, différents indices suggèrent qu'elle pourrait opérer également chez les Procaryotes. Le reste de ce manuscrit s'emploie à démontrer expérimentalement que c'est probablement le cas.



## **Deuxième partie**

# **Quantification de la Conversion Génique Biaisée chez les Bactéries**



## COMMENT MESURER LES FRÉQUENCES DE CONVERSION GÉNIQUE CHEZ DES BACTÉRIES ?

---

Comparant les données brutes de nos sens et ce concept extrêmement complexe et subtil que les mathématiciens appellent grandeur, nous sommes bien forcés de reconnaître une divergence ; ce cadre où nous voulons tout faire rentrer, c'est donc nous qui l'avons fait ; mais nous ne l'avons pas fait au hasard, nous l'avons fait pour ainsi dire sur mesure et c'est pour cela que nous pouvons y faire rentrer les faits sans dénaturer ce qu'ils ont d'essentiel.

Henri POINCARÉ [1902]. *La Science et l'hypothèse*.  
Bibliothèque de Philosophie Scientifique. Paris : Flammarion

La question à laquelle on cherche à répondre est la suivante : Existe-t-il un biais dans les fréquences de conversion de A ou T vers G ou C au cours de la recombinaison chez des Procaryotes ?

Si l'on veut y répondre expérimentalement, il faut (être en mesure de) mesurer lesdites fréquences, ce qui implique de caractériser des produits de recombinaison qui donnent lieu à de la conversion génique. Sachant par ailleurs la probable faible intensité du biais que l'on étudie, il nous faut pouvoir mesurer un nombre suffisant de produits de recombinaison pour pouvoir, le cas échéant, exclure l'absence de biais.

Je détaille dans ce chapitre la façon dont nous avons procédé pour générer ces produits de recombinaison, et expose également la méthode que j'ai mise au point pour étudier leur répartition sur le génome, et mesurer les fréquences de conversion géniques qui leur sont associées.

### 7.1 Description de la méthode d'obtention des produits de conversion

La méthode que nous avons utilisée pour obtenir le plus de produits de recombinaison possible est une méthode de transformation génomique, dérivée des travaux de CROUCHER, HARRIS et al. [2012] et MELL, Jae Yun LEE et al. [2014] et BUBENDORFER et al. [2016].

### *Transformation naturelle à l'échelle génomique*

Cette méthode repose sur la capacité que présentent certaines Bactéries et Archées à acquérir activement de l'ADN exogène par transformation naturelle. On dit de ces espèces qu'elles sont « naturellement compétentes » à la transformation. La transformation naturelle est, avec la conjugaison, la transduction et assimilés, l'un des moyens privilégiés de transfert de gène chez les Procaryotes [SOUCY et al. 2015]. (C'est d'ailleurs en ayant recours à la transformation naturelle chez *Streptococcus pneumoniae* que AVERY et al. [1944] ont démontré que l'ADN était probablement porteur d'information génétique.)

Il s'agit d'un mécanisme très répandu parmi les Bactéries et les Archées [CARVALHO et al. 2019 ; JOHNSTON, MARTIN, FICHANT et al. 2014]. Chez les Protéobactéries, elle repose sur la captation d'ADN exogène par des pili spécialisés dits « *pili* de compétence », dont la rétractation permet d'intégrer l'ADN exogène à travers des pores spécifiques [ELLISON et al. 2018]. Cette intégration s'accompagne d'une dégradation de l'un des deux brins, et d'une protection de l'ADN simple-brin (ssDNA) par des protéines de type DprA, qui protègent le ssDNA de la dégradation par des nucléases, et recrutent les protéines RecA, déjà décrites page 45, pour initier la recherche d'homologie. Lorsque des séquences homologues à l'ADN capté sont détectées, le ssDNA peut envahir le double-brin et être intégré au chromosome par recombinaison homologue [David DUBNAU et BLOKESCH 2019].

DE VRIES et WACKERNAGEL [2002] ont démontré chez *Acinetobacter baylyi* que de l'ADN hétérologue, pour lequel il n'existe pas de séquence homologue dans le génome, peut être intégré dans le génome s'il est flanqué de plus de 183 bp de séquences homologues. De l'ADN extrêmement fragmenté peut également être intégré par une autre voie que la voie classique dépendante de RecA à une fréquence très faible :  $7 \cdot 10^{-13}$  événements par cellule, d'après HÜLTER et WACKERNAGEL [2008].<sup>47</sup>

Les raisons évolutives pour lesquelles ces Bactéries et Archées maintiennent ces systèmes de transformation sont débattues depuis bien longtemps, sans que l'on ait, à ma connaissance, d'hypothèse tout à fait satisfaisante à ce jour. [BLOKESCH 2017 ; CARVALHO et al. 2019 ; CROUCHER, MOSTOWY et al. 2016 ; FALL et al. 2007 ; REDFIELD 1988, 1993].

De nombreux travaux ont documenté l'intégration de fragments d'ADN par la transformation naturelle, notamment dans les modèles classiques que sont *Streptococcus pneumoniae*, *Bacillus subtilis* pour les

47. OVERBALLE-PETERSEN et al. [2013] démontrent que cette méthode d'acquisition qui ne dépend pas de l'homologie peut permettre à

*Acinetobacter baylyi* d'acquérir de l'ADN aussi divergent et fragmenté que de l'ADN extrait d'un os de mammouth.

Firmicutes, et *Acinetobacter baylyi*, *Ralstonia solanacearum* ou encore *Vibrio cholerae* pour les Protéobactéries [COUPAT-GOUTALAND et al. 2011; DE VRIES, HERZFELD et al. 2004]. Généralement, ces études se concentraient sur des locus spécifiques du chromosome. Plus récemment, l'accessibilité accrue du séquençage a permis d'étudier la transformation à l'échelle génomique.

C'est notamment ce qu'ont utilisé MELL, SHUMILINA et al. [2011], les premiers à ma connaissance à avoir utilisé la méthode suivante : connaissant le génome de deux souches divergentes, dont l'une au moins est compétente à la transformation, on peut transformer la souche compétente par le génome extrait de la souche donneuse, séquencer le génome de souches transformantes et cartographier ainsi le paysage de conversion dans le génome de la souche receveuse. Dès lors qu'une région du génome de la souche transformée correspond à l'haplotype de la souche *donneuse*, on peut en déduire que la transformation a conduit à la conversion de cette région par l'haplotype de la souche donneuse.

J'appellerai par la suite ce type de protocole une expérience de transformation génomique (ou *genome-wide transformation*). Les études subséquentes de MELL, Jae Yun LEE et al. [2014], et plus encore celle de BUBENDORFER et al. [2016] ont démontré qu'il était possible de caractériser un grand nombre de produits de recombinaison par cette méthode. C'est donc vers elle que nous nous sommes tournés pour quantifier les fréquences de conversion chez *Acinetobacter baylyi* et *Bacillus subtilis*.

Je détaille dans la partie suivante la façon dont nous avons procédé pour générer ces produits de recombinaison chez *Acinetobacter baylyi*.

## 7.2 Transformation génomique chez *Acinetobacter baylyi*

*Acinetobacter baylyi* est un modèle de transformation naturelle bien décrit, qui a l'avantage d'acquérir de l'ADN exogène à une fréquence très élevée (de l'ordre de  $10^{-2}$  intégrations par cellule, d'après YOUNG et ORNSTON [2001]), et ce de façon « constitutive », c'est-à-dire quel que soit son état physiologique [ELLIOTT et NEIDLE 2011; PALMEN et HELLINGWERF 1997; PALMEN, VOSMAN, BUIJSMAN et al. 1993]. Le génome de la souche de référence ADP1 a déjà été séquencé [BARBE et al. 2004].

### *Séquençage du génome de la souche donneuse*

Néanmoins, pour pouvoir réaliser une expérience de transformation génomique, il faut être en possession d'une souche divergente de la

## 7. MESURE DES FRÉQUENCES DE CONVERSION

souche receveuse. En mars 2017, il n’existait dans les bases de données que quatre génomes de l’espèce *Acinetobacter baylyi*, dont celui de la souche de référence ADP1, dont aucun n’était suffisamment divergent de celle-ci pour pouvoir détecter la conversion avec suffisamment de précision.

La première étape a donc consisté à séquencer le génome de la souche 93A2, pour laquelle on disposait de la séquence nucléique du gène *mutS*, et dont le taux d’identité avec la séquence du gène *mutS* de la souche ADP1 est de 99,12%. Par ailleurs, la phylogénie de l’espèce *Acinetobacter baylyi* établie par VANECHOUTTE et al. [2006] semblait indiquer que cette souche était dans une gamme de divergence avec la souche de référence satisfaisante. Cette souche, isolée par CARR [2003] est la souche LMG 30072.

Cette souche a donc été purifiée par une culture de 12 h sur milieu LB+agar. Une colonie isolée a été cultivée 12 h en milieu LB liquide, dont 1 ml ont servi à l’extraction d’ADN sur membranes de silice NucleoSpin (Macherey Nagel). L’ADN extrait a été séquencé sur une machine Illumina HiSeq en lectures appariées de  $2 \times 150$  bp, avec une taille d’insert de 300 bp, par GATC Biotech. Les lectures de séquençage ont été utilisées pour l’assemblage du génome en utilisant les paramètres par défaut du programme Unicycler version 0.4.1 [WICK et al. 2017]. Pour améliorer l’assemblage, faisant l’hypothèse que la synténie entre la souche de référence et la souche ADP1 était conservée, les 55 contigs obtenus ont été ordonnés (*scaffolded*) par rapport au génome de la souche ADP1 en utilisant MeDuSa version 1.6 [BOSI et al. 2015]. J’ai ainsi obtenu un *scaffold* unique qui représente la totalité de la longueur du génome (table 7.1). Le génome assemblé et les lectures de séquençage ont été déposées à l’European Nucleotide Archive (ENA), sous le numéro d’accession PRJEB22271.

TABLE 7.1 – Description de l’assemblage du génome de *Acinetobacter baylyi* 93A2.

	Unicycler	MeDuSa
Contigs	55	1
Longueur	3 737 271	
N50	234 586	3 733 125
Segment le plus long	480 745	3 733 125

L’étape suivante a consisté à vérifier que le génome ainsi assemblé était suffisamment divergent de celui de la souche receveuse ADP1 pour que l’on soit capable de repérer les événements de conversion avec suffisamment de précision (figure 7.1). J’ai pour cela aligné les deux génomes



à l'aide de *parsnp* version 1.2, disponible dans la collection d'outils *harvest* [TREANGEN et al. 2014], inspecté visuellement l'alignement à l'aide de *gingr* et quantifié la divergence à l'aide du fichier VCF produit.

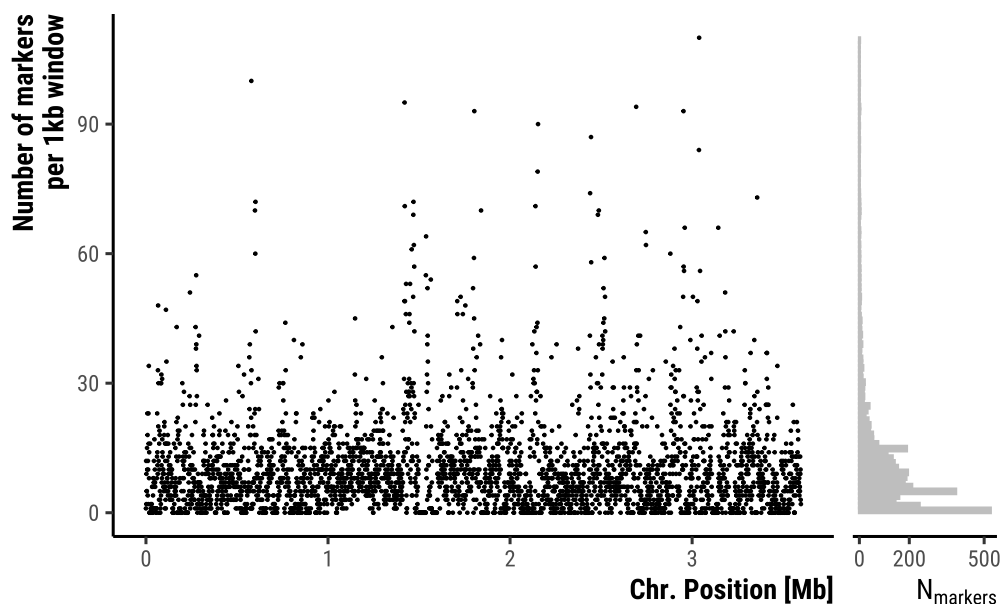


FIGURE 7.1 – Nombre de marqueurs polymorphes entre la souche *Acinetobacter baylyi* ADP1 et 93A2 par fenêtre de 1 kbp. Représentation le long du chromosome de la souche receveuse ADP1.

Avec un total de 42 408 sites polymorphes, ou marqueurs, entre ADP1 et 93A2, la distance médiane entre deux marqueurs est de 17 bp, la distance moyenne de 88 bp, pour une divergence de 1,18 %. L'identité nucléotidique moyenne (ANI) entre les deux souches calculée par le programme *fastANI* version 1.1 [JAIN et al. 2018] est de 98,67 %. Nous avons donc suffisamment de divergence pour pouvoir repérer précisément les événements de conversion.

#### Expérience de transformation génomique chez *Acinetobacter baylyi*

En s'appuyant sur les protocoles utilisés par BUBENDORFER et al. [2016], COWLEY et al. [2018], CROUCHER, HARRIS et al. [2012] et MELL, Jae Yun LEE et al. [2014], nous avons réalisé une expérience de transformation génomique de la souche *receveuse* *A. baylyi* ADP1 par le génome de la souche *donneuse* 93A2.

Le principe est le suivant : cinquante lignées isogéniques indépendantes de la souche receveuse ont été cultivées, dans un milieu LB riche et non-sélectif, en présence de l'ADN extrait de la souche donneuse, remises en culture dans du milieu frais, et ce pendant vingt cycles de transformation, de façon à maximiser le nombre de régions converties analysables. À l'issue de ces vingt cycles, pour chacune des lignées indépendantes, nous avons sélectionné une colonie ayant intégré au moins

## 7. MESURE DES FRÉQUENCES DE CONVERSION

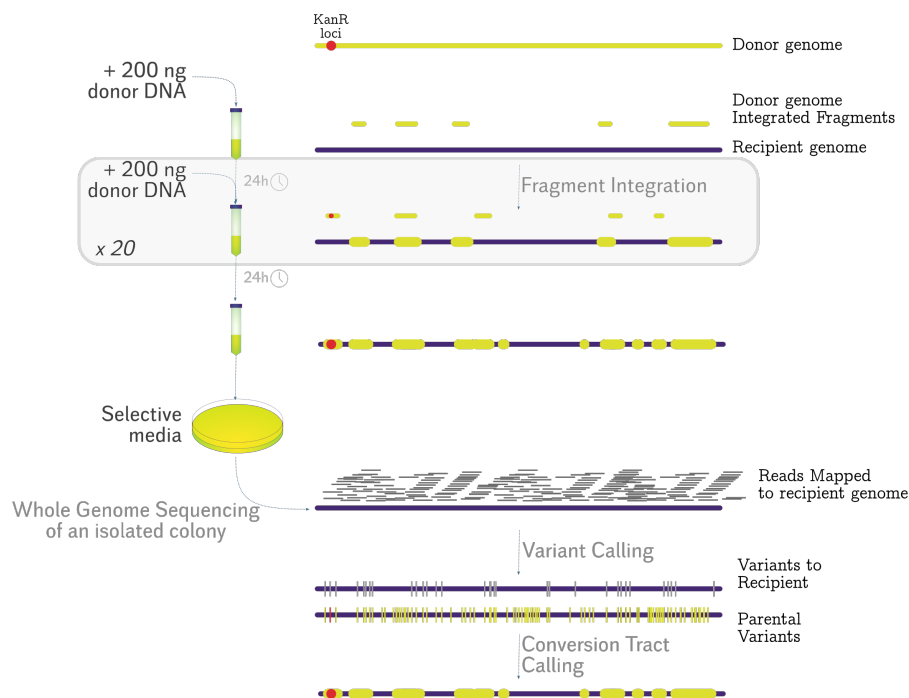


FIGURE 7.2 – Description de l'expérience de transformation génomique chez *Acinetobacter baylyi* pour une lignée individuelle. Donor DNA is repeatedly added to the recipient strain for twenty cycle, resulting in the accumulated integration of donor fragment in the recipient genome. La souche receveuse est exposée à l'ADN de la souche donneuse de façon répétée pendant vingt cycles, résultant en l'intégration cumulative de fragment d'ADN donneur dans le génome de la receveuse. Après vingt cycles, les souches ont été cultivées sur milieu sélectif pour sélectionner celles ayant intégré au moins le locus de résistance à la kanamycine. Une colonie isolée est choisie au hasard, purifiée, son ADN extrait et séquencé. Les lectures de séquençage ont ensuite été alignées (*mapped*) au génome de référence, de façon à déterminer les sites variants, et à vérifier qu'ils correspondent à l'haplotype du donneur. Un marqueur est dit *converti* si l'haplotype de la souche transformante correspond à celui du donneur.

un locus de la souche donneuse, que nous avons au préalable rendue résistante à la kanamycine, par culture sur milieu sélectif. Nous avons ensuite extrait leur génome et l'avons fait séquencer. Je décris ci-dessous le détail de cette expérience (figure 7.2).

La souche donneuse a d'abord été rendue résistante à la kanamycine par l'insertion chromosomique du gène *aphA3*, porté par le plasmide pAVA213-8 [PALMEN, VOSMAN, KOK et al. 1992], au locus compris entre 46,783 et 47,424 kbp, un locus annoté comme un « peptide signal putatif », classiquement utilisé pour la transformation chez *Acinetobacter baylyi* [DE VRIES, HERZFELD et al. 2004; PALMEN, VOSMAN, KOK et al. 1992]. Cette souche donneuse sera donc nommée souche *A. baylyi* 93A2 ACIAD0041::*aphA3*. Le génome d'une colonie isolée et purifiée a été extrait par la même méthode décrite plus haut.

Cinquante colonies isolées de la souche receveuse *Acinetobacter baylyi* ADP1 ont été cultivées 12 h dans 50 volumes séparés de 10 ml de milieu Luria Bertani (LB). Au matin, 2,5 ml de culture de nuit ont été mélangés à 22,5 ml de milieu LB frais; l'absorbance à 600 nm a été mesurée (figure 7.3); 1,1 ml ont été prélevés, puis exposés à 200 ng d'ADN génomique issue de la souche *A. baylyi* 93A2 ACIAD0041::*aphA3*; et incubés horizontalement pendant 4 h à 30 °C, sous agitation à 150 rpm. (La mesure d'absorbance permettait à la fois de vérifier qu'il n'y ait pas de contaminations des cultures et que le comportement de croissance soit bien stable d'un cycle de transformation à l'autre.) Après ces 4 h,

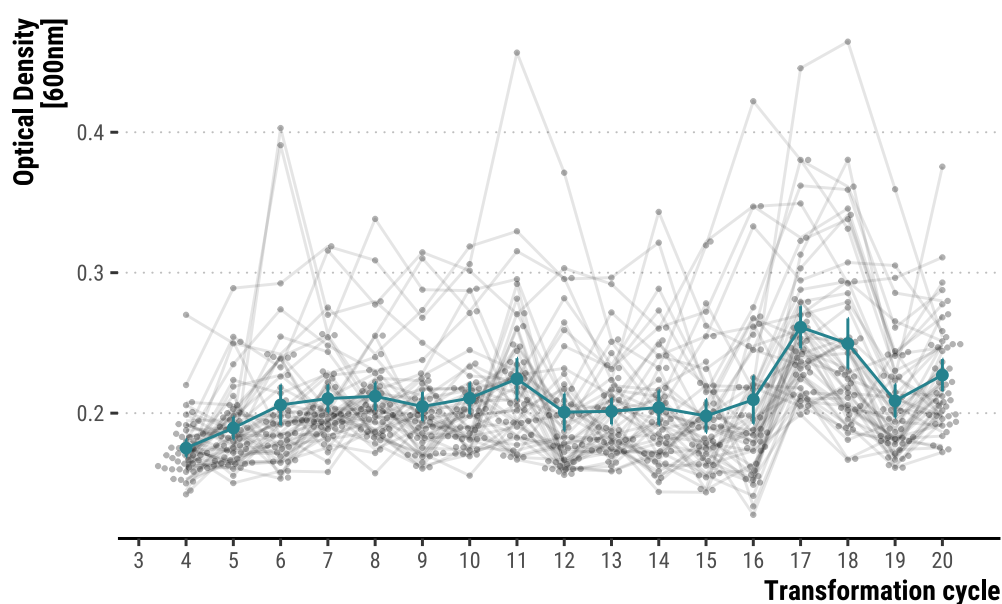


FIGURE 7.3 – Évolution de l'absorbance à 600 nm par cycle de transformation. Sous l'hypothèse qu'une contamination par un autre type de bactéries que le modèle *A. baylyi* augmenterait l'absorbance de la culture, nous avons vérifié que celle-ci ne déviait pas significativement de la valeur attendue de 0,2 pour chacune des lignées. En cas de doute, la pureté de la culture a été contrôlée par étalement sur milieu LB non-sélectif.

pour cinq lignées par cycle de transformation, 100 µl ont été dilués d'un facteur  $10^5$ ; deux dupliquats de 50 µl des dilutions  $10^{-5}$  et  $10^{-2}$  ont

## 7. MESURE DES FRÉQUENCES DE CONVERSION

été respectivement étalées sur milieu LB et milieu LB additionné de kanamycine à  $50 \mu\text{g ml}^{-1}$ , puis incubés 24 h à  $30^\circ\text{C}$ . L'efficacité de transformation a ensuite été calculée en rapportant le nombre de colonies résistantes à la kanamycine au nombre de colonies totales.

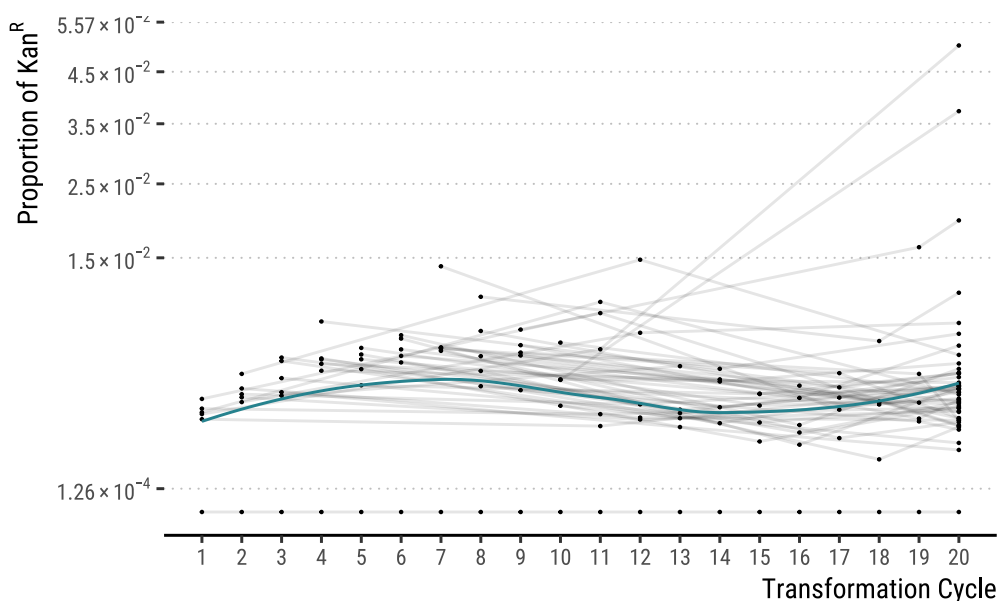
Pour l'ensemble des cinquante lignées, 1 ml du volume de culture ont été mélangés à 9 ml de milieu LB frais et cultivés 12 h à  $30^\circ\text{C}$  et 150 rpm. À l'issue de chaque cycle de transformation, 100  $\mu\text{l}$  des cinquante lignées ont été échantillonnés et dilués dans 100  $\mu\text{l}$  de glycérol à 70 % stérile, et conservés à  $-20^\circ\text{C}$ .

La lignée 50 n'a pas été exposée à l'ADN de la souche donneuse, et a systématiquement été contrôlée pour l'absence de colonies résistantes à la kanamycine, qui témoigneraient ou de résistances spontanément apparues, ou de contaminations entre lignées.

Après vingt cycles de transformation, pour chacune des cinquante lignées, une colonie résistante à la kanamycine a été échantillonnée au hasard, et purifiée par culture sur LB additionné de kanamycine. Une colonie isolée et purifiée a ensuite été cultivée 12 h en milieu liquide additionné de kanamycine à  $50 \mu\text{g m}^{-1}$ , à la suite de quoi 1 ml ont été prélevés pour extraire l'ADN de la souche transformée, pour les cinquante lignées. La même méthode déjà décrite d'extraction d'ADN a été utilisée.

Pour 41 lignées indépendantes, y compris la lignée contrôle 50, l'ADN a été séquencé sur la même plateforme décrite plus haut.

FIGURE 7.4 – Évolution du taux de colonies résistantes à la kanamycine par cycle de transformation.



J'ai représenté dans la figure 7.4 le taux de colonies résistantes à la

kanamycine pour les cinq lignées échantillonnées par cycle de transformation. Sous l'hypothèse que ce taux est le reflet de l'accumulation progressive de séquences issues du donneur, la proportion de colonies résistantes à la kanamycine devrait augmenter avec le nombre de cycle de transformation. Il s'avère que, si c'est bien ce que l'on observe pour les 8 premiers cycles de transformation ( $p = 1,7 \cdot 10^{-3}$ ,  $R^2 = 0.177$ ), la proportion de colonies résistantes à la kanamycine n'est *globalement* pas corrélée au nombre de cycle de transformation ( $p = 0,5$ ,  $R^2 \approx 0$ ).

Il n'est pas inintéressant de noter que la proportion de cellules transformées est globalement restée stable au cours des cycles de transformation. On aurait pu penser, de prime abord, qu'il serait corrélé. Une hypothèse que l'on pourrait proposer est que le coût en fitness représenté par l'insertion du locus de résistance contribue à ce que les cellules transformées à ce locus restent minoritaires (1 cellule transformée pour 1000 environ) et ce de façon relativement stable au cours des cycles.

Avant de décrire la façon dont ont été traitées les lectures de séquençage, je détaille ci-dessous l'adaptation du protocole que nous avons développé chez *Acinetobacter baylyi* au modèle *Bacillus subtilis*.

### 7.3 Transformation génomique chez *Bacillus subtilis*

*Bacillus subtilis* est un autre modèle de transformation naturelle particulièrement bien décrit [DUBNAU 1991 ; David DUBNAU et BLOKESCH 2019 ; JOHNSTON, MARTIN, FICHANT et al. 2014].<sup>48</sup> Pour ne pas restreindre notre analyse aux Protéobactéries, nous avons répété l'expérience de transformation génomique réalisée chez *Acinetobacter baylyi* avec la bactérie *Bacillus subtilis*, une Terrabactérie du phylum des Firmicutes.

#### *Souches receveuse et donneuse*

De façon à maximiser la fréquence des événements de transformation et à contrôler le plus directement possible l'état de compétence de la souche receveuse, nous avons utilisé la souche *Bacillus subtilis* 1A1276, qui dérive de la souche de référence 168. Elle a été rendue « super-compétente » à la transformation naturelle en rendant celle-ci inductible par culture en présence de mannitol [RAHMER et al. 2015]. (RAHMER et

48. Voir notamment à ce sujet la série d'articles classiques de David D. Dubnau : D. DUBNAU et CIRIGLIANO 1973 ;

D. DUBNAU et CIRIGLIANO 1972a,b,c ;  
D. DUBNAU et DAVIDOFF-ABELSON 1971

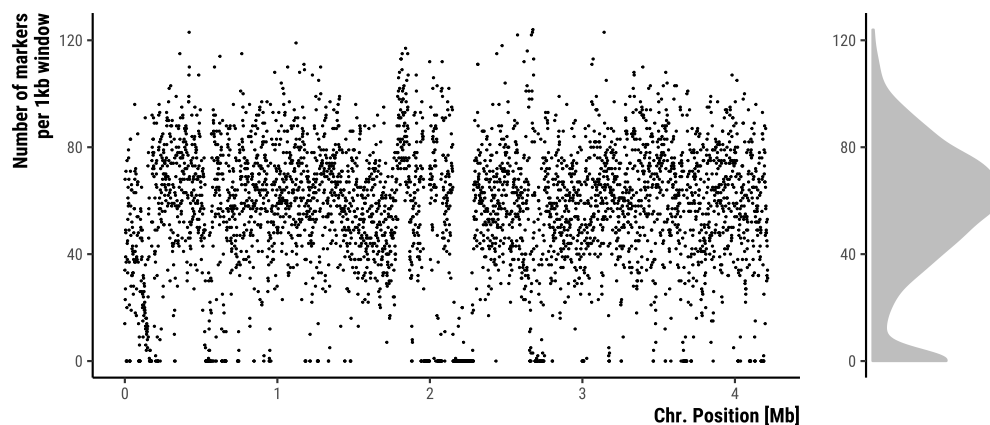
## 7. MESURE DES FRÉQUENCES DE CONVERSION

al. ont remplacé chacun des promoteurs des gènes clés de la transformation chez *Bacillus subtilis* par des promoteurs inductibles au mannitol.) Le génome de référence de la souche 168 dont elle dérive est disponible sur la base de donnée RefSeq, au numéro d'accèsion NC\_000964.3. Cette souche est disponible au *Bacillus* Genetics Stock Center [BGSC p. d.(a)].

La souche donneuse a été choisie sur la base : 1) de l'identité de séquence nucléotidique moyenne (*average nucleotide identity*, ANI) calculée avec PYANI version 0.2.7 [PRITCHARD et al. 2016]; 2) de sa disponibilité dans les collections. La souche retenue est la souche *Bacillus subtilis* 2A11T, qui porte le nom de *B. subtilis* sous-espèce *spizizenii* TU-B-10T; elle a été décrite par NAKAMURA et al. [1999] et séquencé par EARL et al. [2012]; elle est disponible au même BGSC [BGSC p. d.(b)]. Le génome de référence de cette souche est disponible sur la base de donnée RefSeq au numéro d'accèsion NC\_016047.1.

La figure 7.5 représente la quantité de marqueurs par fenêtre de 1 kbp. Il y a en tout 229 466 marqueurs entre la donneuse 2A11T et la receveuse dérivée de 168; la distance moyenne entre deux marqueurs est de 17 bp – la médiane de 8 bp —, ce qui représente une divergence de 5,32%. Là encore, la divergence était donc suffisante pour pouvoir cartographier précisément les événements de conversion.

FIGURE 7.5 – Nombre de marqueurs polymorphes entre la souche donneuse *Bacillus subtilis* 2A11T et receveuse dérivée de *B. subtilis* 168, par fenêtre de 1 kbp. Représentation le long du génome de la souche receveuse



La souche donneuse a été rendue résistante au chloramphénicol par l'insertion chromosomique, *via* la transformation naturelle, du gène *cat* porté par le plasmide pDG1662 (GUÉROUT-FLEURY et al. 1996, d'identifiant GenBank U46197), dans le gène *amyE*. Le génome de la souche donneuse a été extrait après isolement et purification par une méthode phénol/chloroforme suivie d'une précipitation dans l'éthanol.

*Expérience de transformation génomique  
chez Bacillus subtilis*

Une colonie isolée de la souche receveuse a été utilisée pour inoculer 100 ml de bouillon TS (TSB), incubés à 37 °C pendant 12 h, sous agitation à 150 rpm. Au matin 48 réplicats de 50 µl ont été mélangés à 9 ml de milieu TSB frais, auquel 1 ml de mannitol liquide à 5,5 % ont été ajoutés, puis incubés 1,5 h à 37 °C, sous agitation à 150 rpm. Les 48 réplicats ont ensuite été ajustés par dilution à une absorbance à 600 nm de 0,4 dans 1 ml de milieu TSB, et distribués dans les 48 puits d'une microplaque de 48 puits. Dans chacun des puits, excepté un puits servant de contrôle négatif pour la transformation, 0,5 µg d'ADN génomique de la souche donneuse ont été ajoutés. La plaque a ensuite été scellée hermétiquement avec un film adhésif stérile, et incubée pendant 1,5 h à 37 °C, sous agitation à 500 rpm (suivant le protocole de RAHMER et al. [2015]).

Après cette exposition à l'ADN de la souche donneuse, 500 µl de culture ont été mélangés à 9 ml de milieu TSB frais, et incubés une nuit à 37 °C à 45° d'inclinaison, sous agitation à 500 rpm. Avant le cycle de transformation suivant, 100 µl de culture de nuit ont été mélangés à 100 µl de glycérol à 70 % stérile, et conservés à -20 °C.

Après chaque cycle de transformation, immédiatement après l'exposition à l'ADN donneur, cinq lignées ont été étalées : 1) milieu TS-agar aux dilutions  $10^{-4}$  en trois réplicats pour dénombrer la population de souches receveuses totale, et 2) milieu TS-agar additionné de chloramphénicol à  $10 \mu\text{g ml}^{-1}$  (TSA+Cm<sub>10</sub>), aux dilutions  $10^{-1}$  en trois réplicats, pour dénombrer la fraction transformée de la culture. La lignée contrôle a été systématiquement contrôlée par étalement sur TSA et TSA+Cm<sub>10</sub> à chaque cycle de transformation pour vérifier qu'il n'y ait pas de contaminations croisées ou de mutants résistants spontanément au chloramphénicol.



## 7. MESURE DES FRÉQUENCES DE CONVERSION

FIGURE 7.6 – Évolution de la fréquence de colonies résistantes au chloramphénicol dans chaque lignée. Chaque panneau représente une lignée. (Les panneaux sont classés par ratio  $\delta_y/\delta_x$  croissants.)

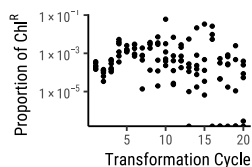
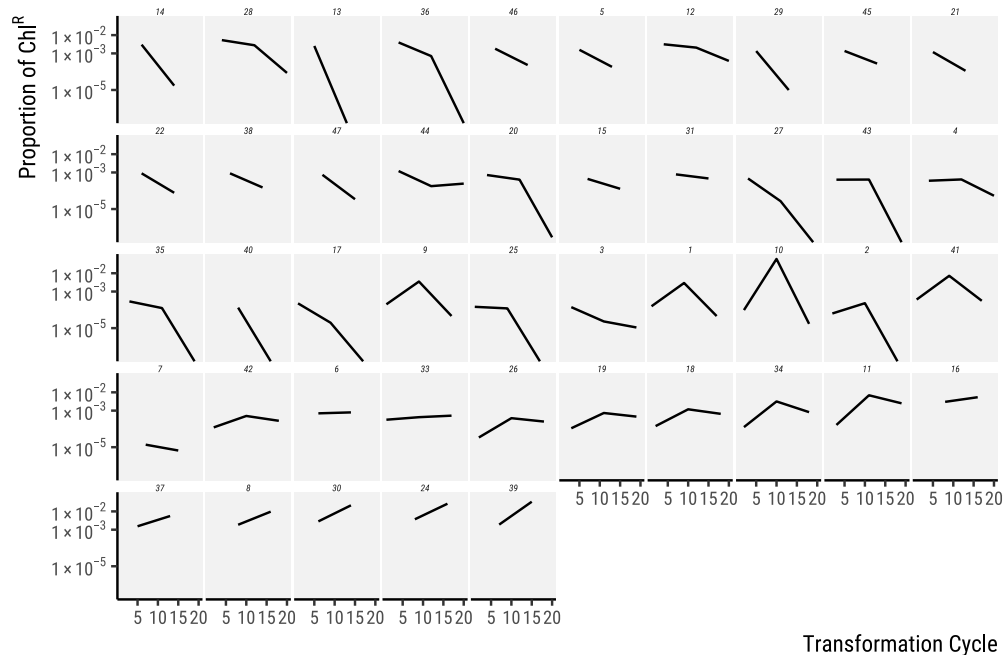


FIGURE 7.7 – Relation entre  $P_{RES}$  la fréquence de résistants au chloramphénicol et le nombre de cycle.

La proportion de colonies résistantes au chloramphénicol  $P_{RES}$  est décrite en figure 7.6. Dans 31 des 47 lignées transformées, cette fréquence est inférieure aux cycles tardifs par rapport aux cycles antérieurs ; c'est-à-dire qu'il y a moins de colonies résistantes au chloramphénicol dans ces lignées lorsqu'elles sont échantillonnées au cycle  $n$  qu'au cycle  $n - 10$ . Considérée globalement, il n'y a pas de corrélation entre  $P_{RES}$  et le nombre de cycle ( $p = 0,19$ ,  $R^2 = 7,4 \cdot 10^{-3}$ , figure 7.7). Initialement, nous avons prévu de séquencer les clones transformants après vingt cycles de transformation, comme nous l'avons fait pour *A. baylyi*. Cependant, le fait que la fréquence de transformation diminue au cours de l'expérience suggèrait que l'acquisition cumulative d'ADN donneur conduisait à la réduction de la compétence constitutive de la souche receveuse. Nous avons donc dans un premier temps séquencé le génome de clones transformants après 1, 5, 10, 15 et 20 cycles. Les résultats de ce séquençage ont montré qu'au-delà de cinq cycles de transformation, le nombre de segments du donneur intégrés était inférieur au nombre de segments intégrés après seulement cinq cycles. Par conséquent, il a été décidé d'extraire le génome des transformants après seulement cinq cycles de transformation. Ainsi, pour 45 des 48 lignées indépendantes (y compris la lignée contrôle), une colonie résistante au chloramphénicol (sauf pour la lignée contrôle) a été sélectionnée au hasard, isolée, purifiée et cultivée une nuit en milieu TSB additionné de chloramphénicol à  $10 \mu\text{g ml}^{-1}$ ,



duquel 1 ml a été prélevé pour réaliser une extraction d'ADN total et le séquençage tel que décrit plus haut.

Après ces expériences de transformation génomique, nous disposons des lectures de séquençage issues de 40 clones transformés indépendamment chez *Acinetobacter baylyi*, et 44 chez *Bacillus subtilis*, ainsi que d'une lignée contrôle non-transformée pour chacune des deux espèces. L'étape suivante pour quantifier les fréquences de conversion génique consiste à déterminer les sites qui ont été convertis pour chacun de ces transformants, et éventuellement à les regrouper par *régions converties*.

## 7.4 Mesure des fréquences de conversion

Je commencerai ici à décrire la méthode de détection des conversions pour le modèle *Acinetobacter baylyi*, qui a été appliquée ensuite aux autres modèles.

### Variant calling : *détection des sites variants*

Les lectures de séquençage (désormais *reads*) de chacun des 41 clones séquencés ont été préparées pour l'étape de *variant calling* telles que préconisé par les guides de bonne pratique de GATK [VAN DER AUWERA et al. 2013]. Ces *reads* ont été alignés, en utilisant `bwa mem` version 0.7.17 [H. LI et DURBIN 2009], au génome de référence de la souche receveuse ADP1, de même que les *reads* issus du séquençage de la souche donneuse. Les sites variants ont ensuite été déterminés par échantillon *via* GATK HaplotypeCaller version 4.0.0.0 [POPLIN et al. 2018]. Les variants ainsi déterminés ont été conjointement génotypés *via* l'outil GATK GenotypeGVCF, résultant en un fichier VCF version 4.2 [SAMTOOLS DEVELOPERS 2019], dont deux sites sont représentés dans la table 7.2. Ce fichier VCF est disponible en ligne [BARRETO 2019a]. Ce *pipeline* de *variant calling* que j'ai développé, avec des paramètres adaptés à l'analyse de résultats d'expérience de transformation génomique a été rassemblé dans l'outil `gatk-genotyper`, également disponible en ligne [BARRETO p. d.(a)].

### Conversion calling : *détection des sites convertis*

Ce fichier VCF a servi ensuite à détecter les marqueurs *convertis*. Pour déterminer ces sites convertis, nous avons développé un *package* écrit dans le langage R [R CORE TEAM 2018], et dépendant de l'environnement Bioconductor [HUBER et al. 2015], et notamment du *package* VariantAnnotation [OBENCHAIN et al. 2014]. Ce *package*, baptisé

## 7. MESURE DES FRÉQUENCES DE CONVERSION

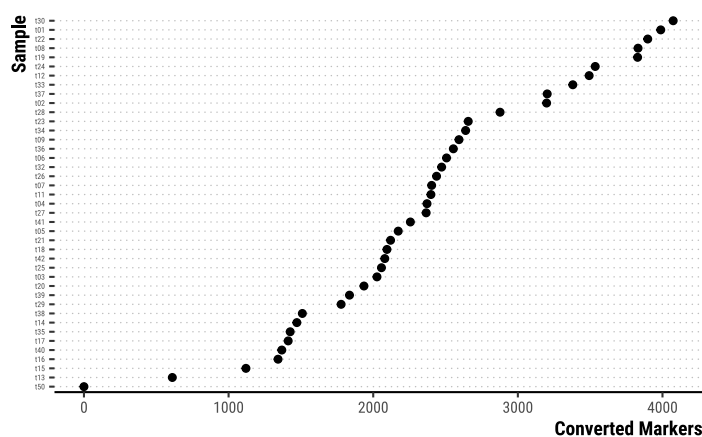
TABLE 7.2 – Extrait du fichier VCF contenant les sites variants. Il s’agit ici d’une version pivotée du fichier, qui contient en ligne les informations liées à un site donné, dont la position est donnée en ligne 2, et en colonne deux marqueurs. Le génotype du marqueur est représenté en rouge s’il correspond à l’haplotype donneur, en bleu sinon. Les marqueurs en position 590 et 1262 ont été convertis dans les souches t05 et t07.

chr. name	CR543861.1	CR543861.1
position	590	1262
reference allele	C	G
variant allele	T	A
site quality	71635.93	61539.93
INFO	AC=6;AF=0.071;...;SOR=0.321	AC=6;AF=0.071;...;SOR=0.415
FORMAT	GT:AD:DP:GQ:PL	GT:AD:DP:GQ:PL
donor	1/1:2,608:..99:25601,1812,0	1/1:0,554:..99:22917,1665,0
recipient	0/0:39,0:39:99:0,101,1542	0/0:39,0:39:99:0,101,1542
t01	0/0:36,0:36:99:0,99,1485	0/0:36,0:36:99:0,99,1485
t02	0/0:55,0:55:99:0,120,1800	0/0:55,0:55:99:0,120,1800
t05	1/1:1,536:..99:22186,1571,0	1/1:1,465:..99:18941,1352,0
t06	0/0:44,0:44:99:0,99,1485	0/0:44,0:44:99:0,99,1485
t07	1/1:1,578:..99:23958,1729,0	1/1:2,492:..99:19791,1426,0
...		

transformr est disponible en ligne [BARRETO p. d.(c)]. Toutes les étapes de détection des sites et des régions converties décrites ci-dessous ont été réalisées *via* les fonctions que j’ai implémentées dans ledit *package*.

Un marqueur est dit converti si : 1) il s’agit d’un site polymorphe entre les deux génomes de la souche donneuse et receveuse; 2) son haplotype correspond à l’haplotype du donneur; et 3) la qualité du génotypage exclue toute ambiguïté quant à l’haplotype de la souche transformante *et* de la donneuse. Nous avons ainsi obtenu un total de

FIGURE 7.8 –  
Nombre de sites convertis par lignée.



85 876 sites convertis, avec un nombre de sites convertis par échantillon compris entre 581 et 3661 (figure 7.8).

La figure 7.9 représente la distribution des marqueurs convertis de l’un des transformants (issue de la lignée 16), le long du génome de la souche receveuse. On voit que cette répartition n’est pas uniforme, et

## Mesure des fréquences de conversion

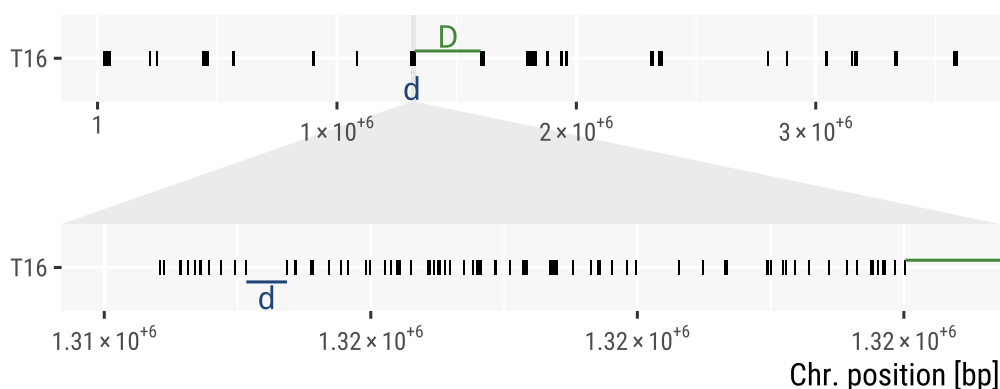


FIGURE 7.9 – Représentation des marqueurs convertis le long du génome de l'un des transformants (transformant 16). La partie basse de la figure est un grossissement de l'une des régions où la densité en marqueur convertis est élevée.

que les marqueurs convertis tendent à se regrouper en *clusters*, que l'on appellera désormais « régions converties », ou *conversion tracts*.

L'enjeu est de réussir à délimiter les bornes des régions converties, c'est-à-dire de regrouper en *régions converties* les marqueurs dont la conversion est vraisemblablement le fait d'un même événement de conversion. Deux marqueurs convertis proches l'un de l'autre peuvent l'être soit parce qu'ils ont été convertis conjointement, soit parce que deux événements successifs proches se sont produits. Par conséquent, en regroupant des marqueurs, deux types d'erreurs peuvent se produire : 1) deux marqueurs sont regroupés dans la même région converties alors qu'ils ont été convertis par deux événements indépendants ; 2) deux marqueurs sont séparés dans deux régions converties différentes alors qu'ils ont été convertis simultanément (par exemple à cause d'entrecouplements du produit de recombinaison par l'haplotype receveur).

On peut faire l'hypothèse que deux conversions indépendantes qui se produiraient simultanément sont davantage distantes les unes des autres que deux co-conversions. Sous cette hypothèse, la distribution de la distance entre marqueurs convertis *voisins* les plus proches est informative (figure 7.10).

Pour l'essentiel (96,9 %), ces distances sont celles qui séparent deux marqueurs convertis conjointement dans une même région (indiquée **d** dans la figure 7.9). Le reste de cette distribution traduit les distances entre marqueurs situés dans des régions probablement converties indépendamment (indiquée **D** dans la figure 7.9). Étant donnée la distribution de ces distances, nous avons placé à 1 kbp le seuil de distance au-delà duquel deux marqueurs convertis l'ont probablement été dans deux régions indépendantes (figure 7.10).

Cela m'a permis d'obtenir une liste de régions converties par transformant. Afin de tester que ce seuil de 1 kbp permette bien de distinguer différentes régions converties et minimise la probabilité que l'on a de se tromper lorsqu'on distingue deux régions indépendantes, j'ai représenté

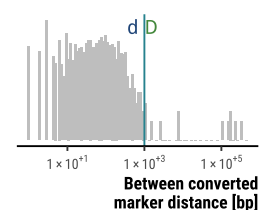
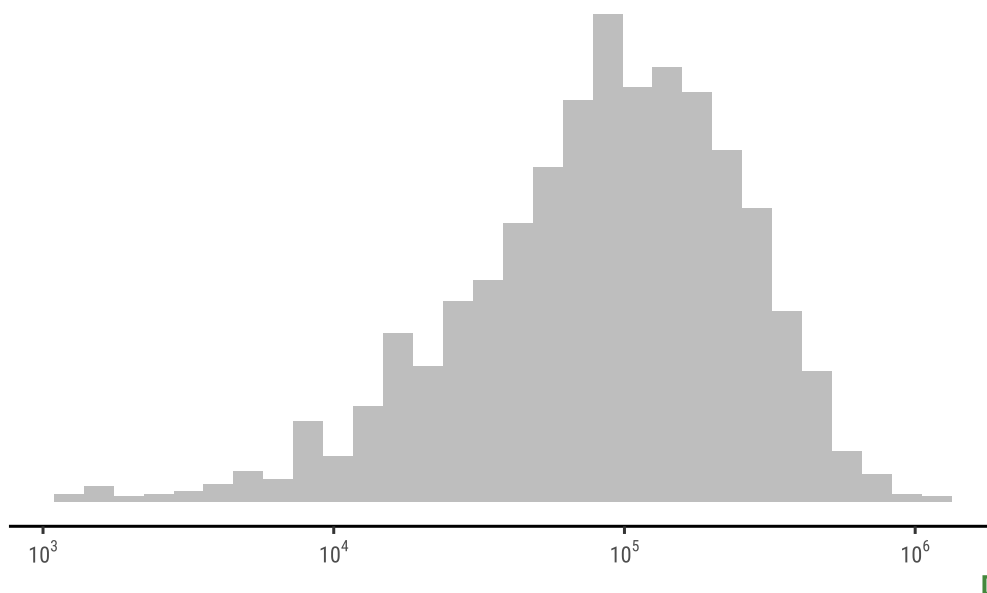


FIGURE 7.10 – Distribution de la distance entre marqueurs convertis.

la distribution des distances  $D$  qui séparent deux régions converties les plus proches au sein d'un transformant (figure 7.11). On voit que ce seuil

FIGURE 7.11 –  
Distribution de la  
distance entre ré-  
gions converties  
les plus proches.



de 1 kbp minimise bien cette probabilité : pour l'essentiel, les distances entre régions converties les plus proches sont comprises entre 10 kbp et 1 Mbp ; ceci indique bien que deux marqueurs convertis distants de moins de 1 kbp sont plus probablement le fait d'un même événement de conversion que d'événements indépendants successifs.

Je représente dans la partie suivante l'ensemble des régions converties obtenues chez *Acinetobacter baylyi*.

### *Obtention des régions converties chez Bacillus subtilis*

La même procédure a été appliquée au modèle *B. subtilis*. Les *reads* des transformants et de la lignée contrôle ont été alignés au génome de référence de la souche en utilisant `bwa mem` à nouveau. Comme l'on ne disposait pas des lectures de séquençage de la souche donneuse, et pour éliminer les régions où l'alignement des lectures de la souche donneuse ne permettrait pas de génotyper sans ambiguïté les marqueurs, j'ai simulé des *reads* de séquençage aux paramètres identiques à ceux des *reads* obtenus pour les transformants ( $2 \times 150$  bp avec une taille d'insert de 300 bp), à partir du génome de référence de la souche 2A11T, en utilisant `wgsim` [H. LI 2019]. Les sites variants ont ensuite été déterminés avec GATK HaplotypeCaller et génotypés conjointement, en utilisant GATK GenotypeGVCF. Le fichier VCF qui en résulte est disponible en ligne [BARRETO 2019b].

Chez *B. subtilis*, nous avons obtenu un total de 193 059 marqueurs convertis, avec un nombre de sites convertis par échantillon compris entre 240 et 9846 (figure 7.12).

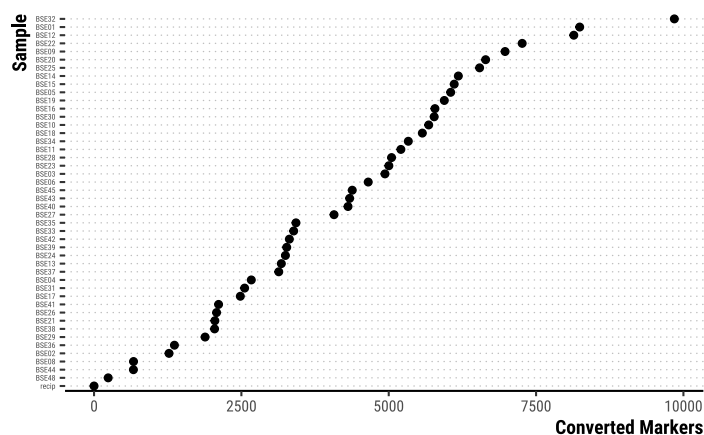


FIGURE 7.12 – Nombre de marqueurs convertis par lignée chez *B. subtilis*.

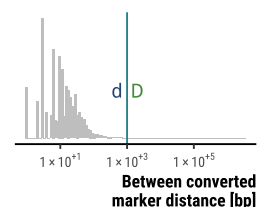


FIGURE 7.13 – Distribution de la distance entre marqueurs convertis chez *B. subtilis*.

De même que pour *A. baylyi*, la distribution de la distance entre marqueurs convertis a servi à regrouper les marqueurs ayant probablement été convertis par le même événement de conversion (figure 7.13).

Le même seuil de 1 kbp a été utilisé, de sorte que deux marqueurs convertis sont dits appartenir à deux régions converties différentes s'ils sont distants de plus de 1 kbp. Cela nous a permis d'obtenir une liste de régions converties dans cet organisme, que je détaille dans le chapitre suivant.

## 7.5 Ré-analyse des données publiques

À ce stade, on dispose donc des produits de conversion que nous avons générés expérimentalement chez *Acinetobacter baylyi* et *Bacillus subtilis*. De plus, des expériences similaires ont été réalisées par d'autres avec divers organismes, listés dans la table 7.3. J'ai ré-analysé les résultats

Organisme	Référence
<i>Helicobacter pylori</i>	BUBENDORFER et al. 2016
<i>Streptococcus pneumoniae</i>	COWLEY et al. 2018
	CROUCHER, HARRIS et al. 2012
<i>Haemophilus influenzae</i>	MELL, Jae Yun LEE et al. 2014
<i>Neisseria sp.</i>	ALFSNES et al. 2018
<i>Vibrio cholerae</i>	MATTHEY et al. 2019

TABLE 7.3 – Liste des organismes avec lesquels des expériences de transformation génomique ont été réalisées.

tats de ces expériences (à l'exception de *V. cholerae*). Je décris ici leurs

spécificités et la façon dont j'ai procédé pour obtenir une liste de régions converties pour chacun de ces organismes.

### *Transformation génomique chez Helicobacter pylori*

Les expériences de transformation ont été réalisées par BUBENDORFER et al. [2016]. En bref, les auteurs ont transformé la souche receveuse *H. pylori* 26695 par l'ADN génomique purifié issu de la souche J99-R3 porteuse du gène *cat* de résistance au chloramphénicol. Les deux génomes présentent une divergence  $\approx 4,5\%$ , avec en moyenne un polymorphisme toutes les 22 bp. BUBENDORFER et al. ont réalisé différents types de transformation. L'une a consisté à transformer la receveuse par l'ADN de la donneuse et à cribler les transformants par culture sur un milieu sélectif au chloramphénicol, puis à extraire et séquencer le génome de transformants sélectionnés au hasard. L'autre a consisté à exposer la souche receveuse à de l'ADN de la souche donneuse pendant 28 à 52 jours de culture en milieu non-sélectif, à sélectionner des colonies au hasard après ces cycles de transformation, puis à extraire et séquencer le génome de ces colonies potentiellement transformées.

Les données disponibles le sont sous forme de génomes assemblés, téléchargées depuis les projets ENA aux numéro d'accèsion PRJEB13893 et PRJEB13894. Le génome de la souche receveuse et de la souche donneuse a été téléchargé depuis respectivement GenBank et RefSeq aux numéros d'accèsion CP003904.1 et NC\_000921.1. J'ai amélioré la continuité de l'assemblage en utilisant MeDuSa [BOSI et al. 2015] tel que décrit plus haut. Les assemblages obtenus, ainsi que le génome de la souche donneuse, ont été alignés conjointement au génome de référence de la souche receveuse à l'aide de *par SNP*, et les sites variants ont été déterminés et exportés au format VCF en utilisant *harvest tools* [TREANGEN et al. 2014].

Ce fichier VCF a été ensuite utilisé pour déterminer les régions converties à l'aide du package *transformr*, de la même façon que décrit plus haut pour *Acinetobacter baylyi* et *Bacillus subtilis*. Toutes les étapes depuis le téléchargement des données brutes jusqu'à l'obtention du fichier VCF sont reproductibles à l'aide d'un *pipeline* dépendant de GNU Make ; ce *pipeline* est disponible en ligne [BARRETO p. d.(d)]. Le fichier VCF produit est également disponible en ligne [BARRETO 2019c].

Cette expérience représente 102 814 marqueurs convertis, allant de 7 à 5684 marqueurs par échantillon (figure 7.14)

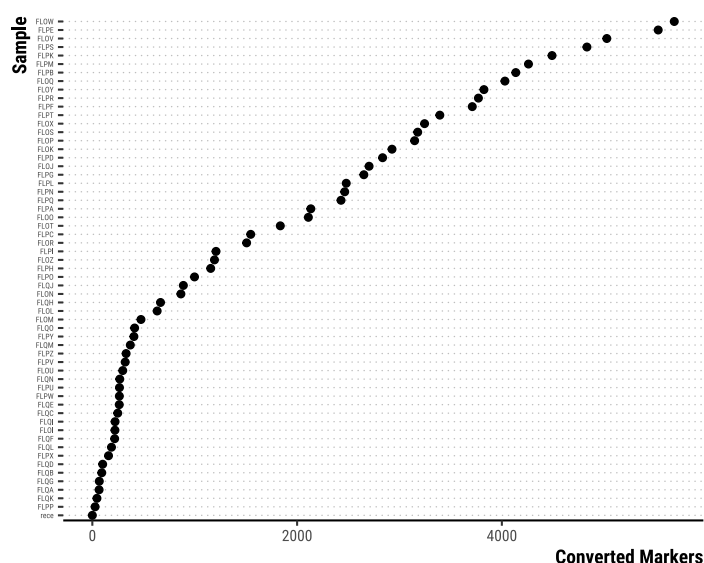


FIGURE 7.14 – Nombre de marqueurs convertis par transformants chez *Helicobacter pylori*.

La distribution de la distance entre marqueurs convertis les plus proches a permis de placer le seuil de distance entre deux régions converties à 200 bp, en accord avec le seuil utilisé par BUBENDORFER et al. [2016] (figure 7.15).

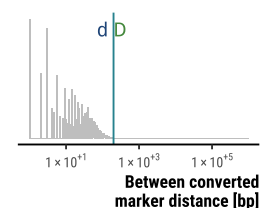


FIGURE 7.15 – Distribution de la distance entre marqueurs convertis chez *H. pylori*.

### Transformation génomique chez *Streptococcus pneumoniae*

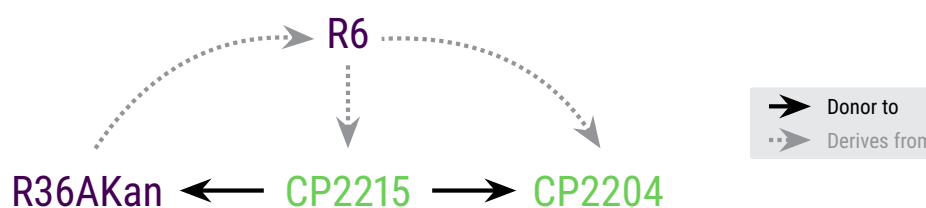
La transformation génomique chez cet organisme a été réalisée par COWLEY et al. [2018]. Les auteurs ont cherché à comprendre si la façon dont la transformation se déroulait avait une influence sur la longueur des régions intégrées, et ont donc réalisé des transformations d'une souche receveuse (la souche CP2204) soit par l'ADN génomique extrait d'une souche donneuse, soit par des contacts entre une souche compétente et une souche non-compétente au sein de biofilms, ou de conditions proches de celles d'un biofilm. L'expérience ne s'est pas déroulée de façon répétée, contrairement aux expériences précédemment décrites : les auteurs ont sélectionné après un cycle de transformation les colonies résistantes à la spectinomycine et à la novobiocine, résistances portées par le génome de la souche donneuse.

Les données produites sont disponibles sous formes de *reads* séquençage depuis le Sequence Read Archive BioProject PRJNA448170. Elles ont été téléchargées avec le SRA-toolkit du NCBI. Les auteurs ont utilisé deux souches receveuses différentes. L'une est la souche CP2204, dérivant de la souche de référence R6 ; l'autre est la souche R36AKan, dont dérive elle-même la souche R6. La souche donneuse est la souche CP2215, dérivant également de la souche R6 (figure 7.16).



## 7. MESURE DES FRÉQUENCES DE CONVERSION

FIGURE 7.16 – Schéma des relations entre souches dans l'expérience de COWLEY et al. 2018. Les noms de souches indiqués en vert sont des souches dont le génome a été reconstituée *in-silico* à partir du génome des souches indiquées en bleu.



La première étape a donc consisté à reconstituer le génome des souches receveuses et donneuses à partir du génome de référence de la souche R6 dont elles dérivent, et des *reads* de séquençage produits.

Le génome de référence de la souche R6, dont dérivent la souche receveuse CP2204 et la souche donneuse CP2215 a été téléchargé depuis l'ENA au numéro d'accèsion AE007317.

J'ai « corrigé » l'assemblage du génome de la souche de référence R6 par les *reads* de séquençage de la souche receveuse d'une part, de la souche donneuse d'autre part, en utilisant *Pilon* version 1.22 [WALKER et al. 2014], de façon à produire un assemblage du génome de la souche donneuse CP2215 et de la souche receveuse CP2204. Le génome de référence de la seconde souche receveuse R36AKan a été téléchargé depuis GenBank au numéro d'accèsion CP028436.1, puis corrigé de la même façon par les *reads* de séquençage de la souche receveuse utilisée dans cette étude.

Les *reads* issus du séquençage des souches transformées ont ensuite été alignés au génome de la souche receveuse correspondante, de façon à déterminer les sites variants, qui ont ensuite été génotypés de la même façon que décrit pour *Acinetobacter baylyi*. Toutes ces étapes ont été explicitées dans un *pipeline* reproductible dépendant de GNU Make, accessible en ligne [BARRETO p. d.(b)]; les fichiers VCF résultants sont également disponibles [BARRETO 2019d].

Cette expérience a généré 50 718 marqueurs convertis, avec entre 0 et 1026 marqueurs convertis par clone transformant (figure 7.17).

### *Transformation génomique chez Haemophilus influenzae*

L'expérience de transformation génomique a été réalisée par MELL, Jae Yun LEE et al. [2014]. Les auteurs ont séquencé 96 colonies de la souche receveuse RR722 transformées par une souche donneuse dérivée de la souche 86-028NP, et criblées par culture sur milieu sélectif pour la résistance à la novobiocine ou à l'acide nalidixique.



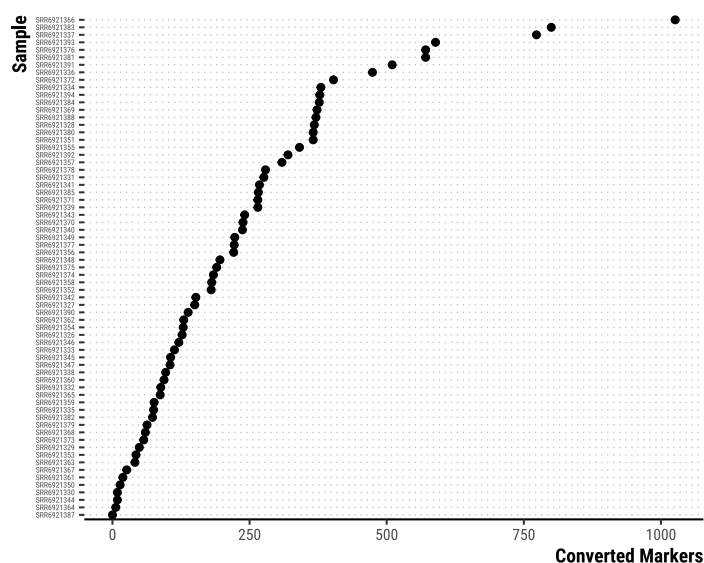


FIGURE 7.17 – Nombre de marqueurs convertis par clone transformants

Les données produites, sous forme de *reads* de séquençage, ont été téléchargées depuis le BioProject SRA SRR1165176. Le génome de référence de la souche receveuse a été téléchargé depuis RefSeq au numéro d'accèsion NC\_000907.

Les mêmes étapes pour déterminer les sites convertis que précédemment décrites ont ensuite été réalisées. Cette expérience a produit 14 239 marqueurs convertis, avec entre 3 et 625 marqueurs par clone transformant. De la même façon que décrit précédemment, j'ai pu déterminer les régions qui résultaient probablement des mêmes événements de conversion.

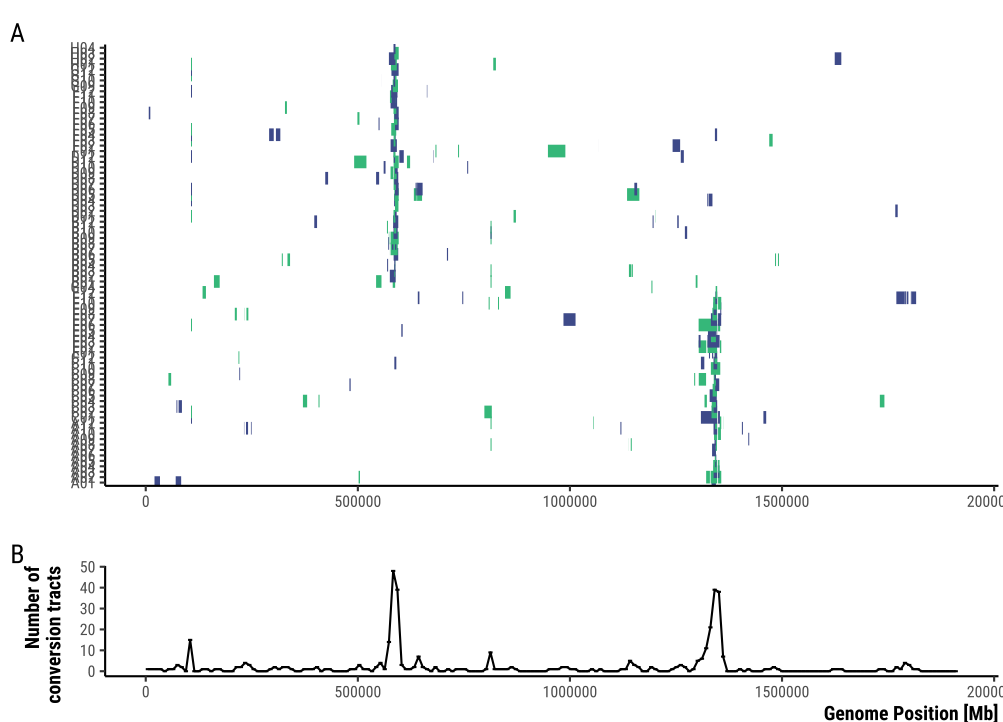
Néanmoins, nous avons fait le choix de ne pas inclure les données qui proviennent de ce modèle pour les raisons que je détaille ci-dessous.

La figure 7.18 décrit la répartition de ces régions converties le long du génome de la souche receveuse. On voit immédiatement que la majorité des régions converties se situe à proximité des locus qui confèrent la résistance à l'un des deux antibiotiques utilisés pour le criblage (54,6 % des régions converties, ce qui représente 58,1 % des marqueurs convertis). Or si l'on veut mesurer les fréquences de conversion, il ne faut pas qu'elles soient biaisées par leur liaison avec un marqueur dont l'intégration est forcée par les conditions expérimentales. Ce qui ne laisse que 104 régions converties qui ne soient pas associées à l'intégration du marqueur de résistance.

Il a donc été décidé de ne pas inclure les données qui proviennent de cet organisme dans les analyses ultérieures.

## 7. MESURE DES FRÉQUENCES DE CONVERSION

**FIGURE 7.18 –** Répartition des régions converties le long du génome de la souche receveuse chez *H. influenzae*. Les transformants ont été colorés alternativement en vert ou bleu pour mieux les distinguer les uns des autres. Les colonies sélectionnées pour leur résistance à la novobiocine sont représentés en haut de la figure, celles pour la résistance à l'acide nalidixique en bas.



### Transformation génomique chez *Neisseria meningitidis*

Enfin, le dernier organisme que j'aie analysé est *Neisseria meningitidis*, avec lequel des expériences de transformation ont été réalisées par ALF-SNES et al. [2018]. Les auteurs de cette étude ont transformé la souche receveuse MC58 par le génome d'une souche de *N. meningitidis* divergente FAM18, et par le génome d'une autre espèce du genre *Neisseria* : l'espèce *N. lactamica*, souche Z6793. Les deux génomes donneurs sont porteurs d'une mutation dans le gène *rpoB* qui introduit une résistance à la rifampycine dans les souches transformantes. De la même façon que MELL, Jae Yun LEE et al., les auteurs ont transformé la receveuse par l'ADN de la souche donneuse et criblé les transformants par culture sur milieu sélectif pour la résistance à la rifampycine après un cycle de transformation.

Le génome de référence de la souche receveuse MC58 et de la souche donneuse FAM18 sont disponibles sur GenBank aux numéro d'accèsion respectifs NC\_003112.2 et NC\_008767.1. Afin de reconstruire le génome de la souche receveuse utilisée pendant l'expérience, j'ai corrigé l'assemblage du génome de la souche de référence par les *reads* issus du séquençage de la souche receveuse par cette étude en utilisant Pilon comme décrit précédemment.

J'ai ensuite aligné les *reads* provenant de la souche FAM18 donneuse au génome de la souche receveuse ainsi obtenu, ainsi que tous les *reads*

provenant des colonies transformées par l'ADN génomique de la souche FAM18, de façon à déterminer les sites qui présentent l'haplotype de la souche donneuse. J'ai fait de même avec les *reads* provenant de l'espèce *Neisseria lactamica*, ainsi que de tous les transformants qui en découlent. Le génotypage a été réalisé avec *bcf tools* version 1.8 [H. LI, HANDSAKER et al. 2009].

Cette expérience a généré 4066 marqueurs convertis pour la transformation intra-spécifique, et 3669 pour la transformation inter-spécifique. Cependant, de la même façon que pour *Haemophilus influenzae*, l'essentiel des régions converties l'ont été à proximité du marqueur de résistance à la rifampicine (figure 7.19). Dans l'expérience de transfor-

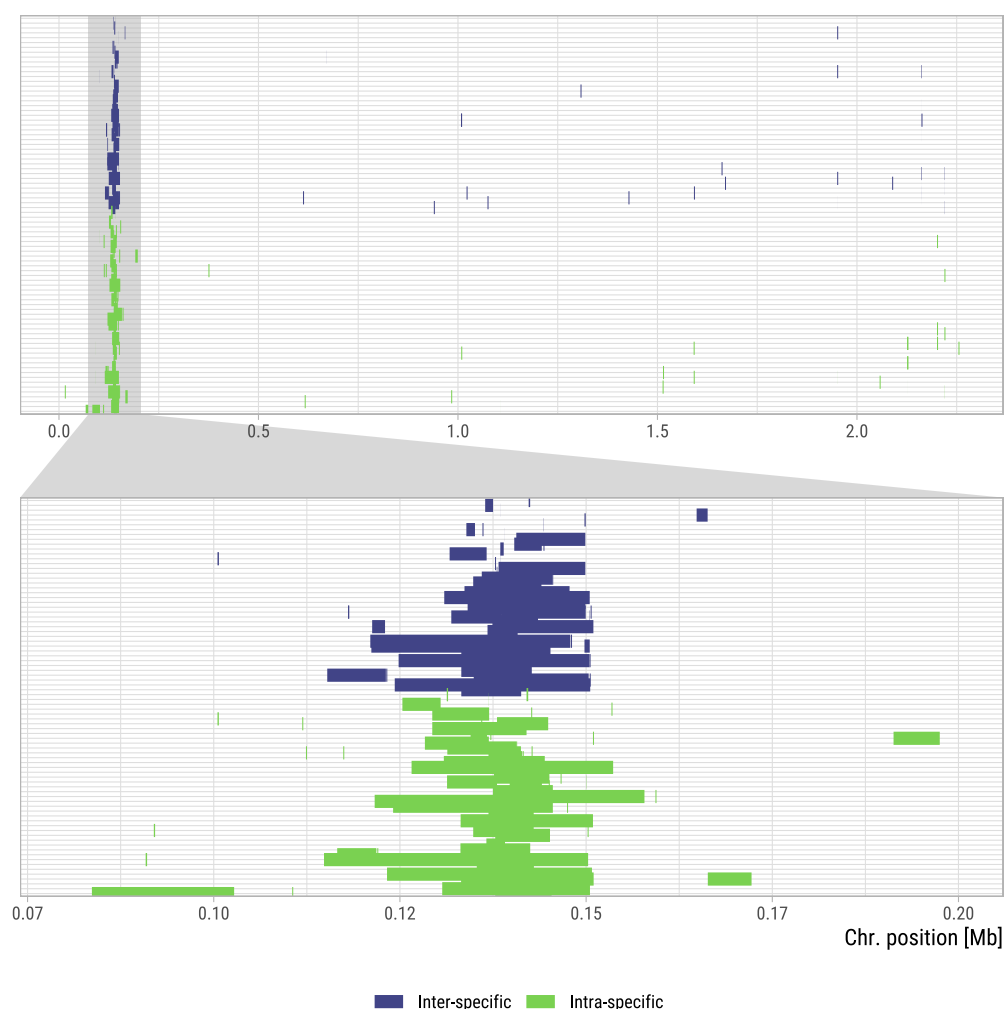


FIGURE 7.19 – Répartition des régions converties le long du génome de la souche receveuse chez *Neisseria*.

mation inter-spécifique, 98,1 % des marqueurs convertis se trouvent dans une région de 80 kbp centrée sur le marqueur de résistance à la rifampicine, et 96,7 % dans l'expérience intra-spécifique. Ceci ne laisse

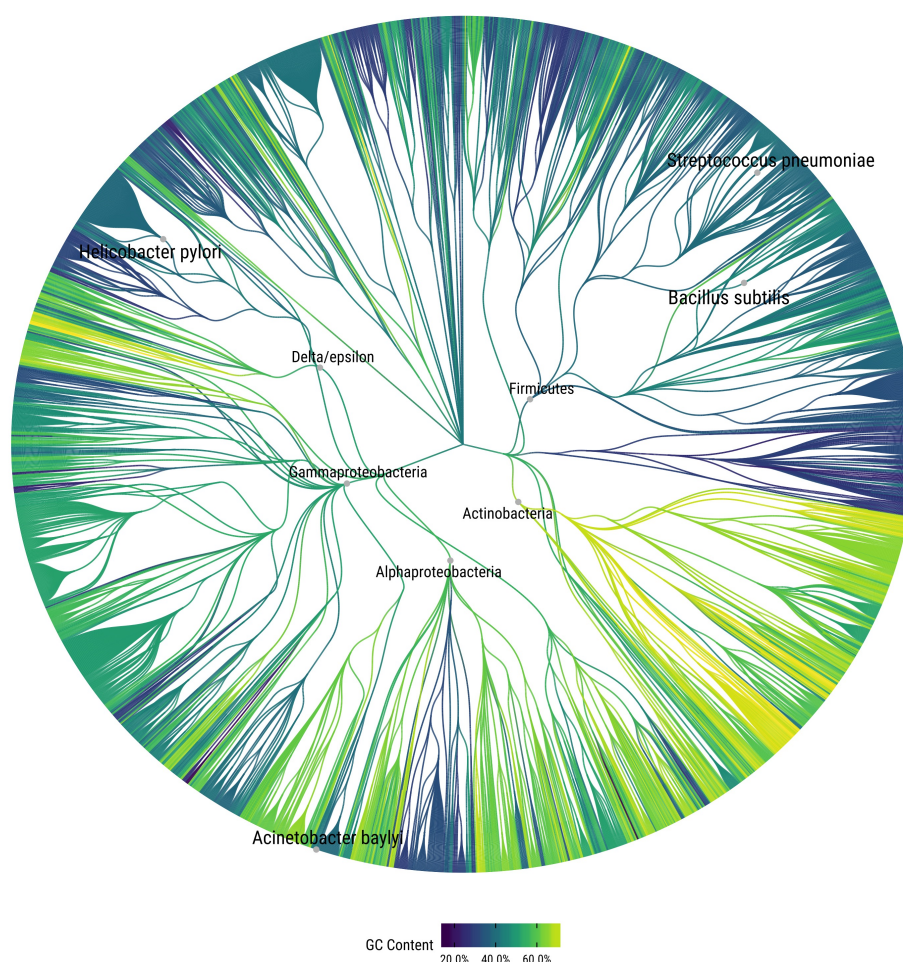
## 7. MESURE DES FRÉQUENCES DE CONVERSION

donc que 186 marqueurs dont la conversion soit indépendante de l'intégration du locus de résistance à la rifampicine. Nous n'avons donc pas inclus cet organisme dans nos analyses ultérieures.

### Résumé

Nous avons donc assemblé un jeu de données comprenant 2023 régions converties chez *Acinetobacter baylyi*, 789 chez *Bacillus subtilis*, 1084 chez *Streptococcus pneumoniae* et et 1759 chez *Helicobacter pylori*. Cela représente une diversité phylogénétique représentée dans la figure 7.20. Dans le chapitre suivant, je décris ces régions converties dans les dif-

**FIGURE 7.20 –** Distribution phylogénétique des espèces utilisées dans cette étude. Deux espèces sont des Protéobactéries : *A. baylyi* est une  $\gamma$ -Protéobactérie, *H. pylori* une  $\epsilon$ -Protéobactérie. Les deux autres espèces *S. pneumoniae* et *B. subtilis* font partie du groupe des Firmicutes, un sous-groupe des Terrabactéries (gram positives). La couleur des branches reflète le taux de GC génomique médian du taxon. La topologie correspond à celle de référence de la base de donnée NCBI Taxonomy [FEDERHEN 2012]. Les longueurs de branche n'ont pas de sens. Les génomes utilisés pour calculer le taux de GC sont référencés dans la base de donnée Hogenom version 7 (n = 1636).



férents modèles, en détaillant particulièrement l'espèce *Acinetobacter baylyi*, pour laquelle le jeu de donnée est le mieux adapté à la question que l'on pose : est-ce qu'il y a un biais dans les fréquences de conversion de AT vers GC et de GC vers AT?

## DÉMONSTRATION D'UN BIAIS DE CONVERSION GÉNIQUE VERS GC CHEZ DES BACTÉRIES

---

La nature confond les pyrrhoniens et la raison confond les dogmatiques. Que deviendrez-vous donc, ô homme qui cherchez quelle est votre véritable condition par votre raison naturelle? Vous ne pouvez fuir une de ces sectes ni subsister dans aucune.

Blaise PASCAL [1864]. *Oeuvres : pensées, lettres et opuscules divers*. Napoléon Chaix. Paris, page 143

J'en sais trop pour être Pyrrhonien, & j'en sais trop peu pour être Dogmatique.

Pierre BAYLE. Cité par POPKIN 2019

À ce stade, je dispose donc d'un ensemble de marqueurs convertis par transformation naturelle chez quatre espèces, regroupés en régions converties. Dans un premier temps, je montre comment se répartissent les régions converties sur le génome dans les deux espèces pour lesquelles nous avons généré nous-mêmes les données. Dans un second temps, je quantifie les fréquences de conversion chez ces deux espèces, ainsi que les deux espèces que nous avons retenues, pour lesquelles des expériences similaires ont été réalisées — *Streptococcus pneumoniae* et *Helicobacter pylori*.

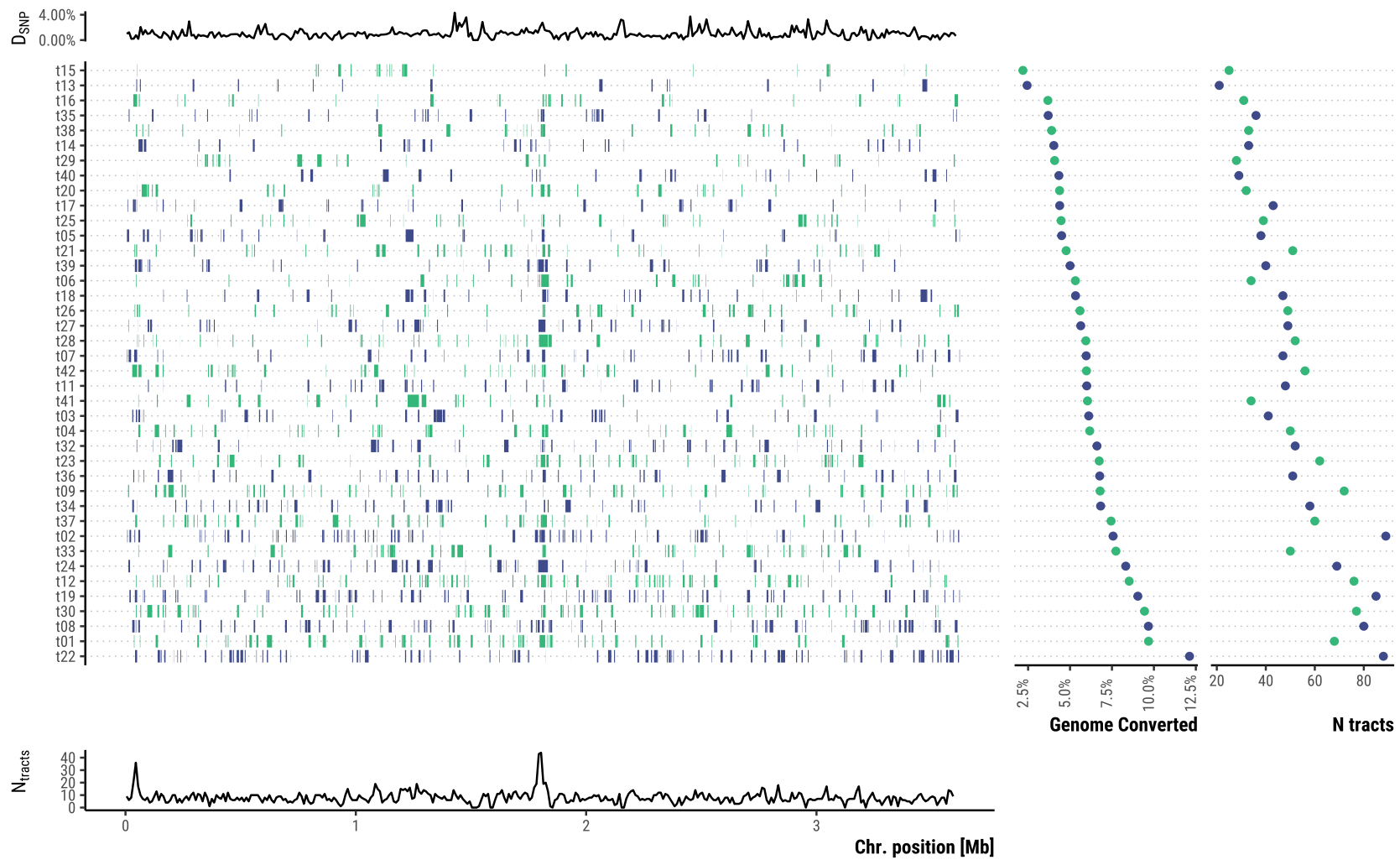


FIGURE 8.1 – Paysage d'intégration de segments donneurs dans le génome d'*Acinetobacter baylyi*. (Description page suivante.)

FIGURE 8.1 – **Haut.** Divergence mesurée en nombre de SNP par fenêtre de 10 kbp. **Milieu.** Distribution des régions converties le long du génome; chaque ligne représente un clone transformant séquencé, colorée en bleu ou vert pour contraster; taux de génome receveur converti; et nombre de régions converties par clone. **Bas.** Nombre de régions converties par fenêtre de 10 kbp.

### 8.1 Répartition des des régions converties sur le génome

#### *Acinetobacter baylyi*

Cette expérience de transformation génomique répétée a permis de générer 2023 produits de recombinaison (figure 8.1), avec entre 21 et 89 produits de recombinaison par clone (moyenne:  $50.6 \pm 18.2$ ). Comme observé chez *H. pylori* [BUBENDORFER et al. 2016], chaque transformant a importé au moins un segment de génome donneur, et a remplacé une fraction de génome receveur allant de 2,6 % à 15,7 % (moyenne:  $7.8 \pm 2.8$ ). Ces proportions de génome converti impliquent un taux de recombinaison par base par cycle de transformation de  $5,4 \cdot 10^{-5}$  à  $3,3 \cdot 10^{-4}$ .

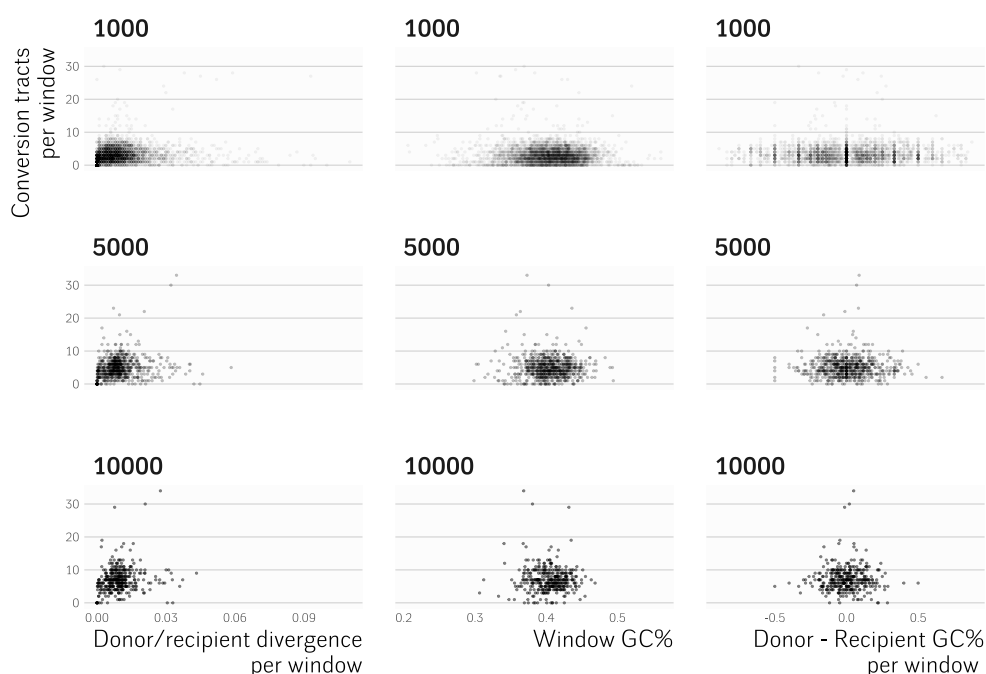
Comme attendu d'après les expériences de transformation génomique réalisées dans d'autres modèles, l'intégration de segment du donneur s'est produite sur toute la longueur du génome : chaque marqueur a été converti en moyenne chez  $3,6 \pm 2,6$  transformants différents. Deux régions à haute fréquence de transformation sont notables : la première est située entre 45,9 kbp et 47,0 kbp, et correspond à la région sélectionnée par la culture en présence de kanamycine (l'intégration du locus de résistance a été confirmée par détection des insertions en utilisant MindTheGap version 2.2.0 RIZK et al. 2014); la seconde région est située entre 1,79 Mbp et 1,80 Mbp, et est associée à l'excision d'une copie du transposon IS1236. Cette copie est absente des régions correspondantes du génome donneur. Dans 37 des 41 clones séquencés, le locus d'insertion de cette transposase a été éliminé par conversion *via* le segment donneur, dépourvu de l'IS à ce locus. Comme semblent l'indiquer d'autres travaux, cette fréquence d'excision élevée pourrait résulter d'effets sélectifs : les clones dépourvus d'IS1236 de la même souche ADP1 ont été démontrés comme ayant une fitness supérieure à la souche sauvage dans des milieux de culture courant [GENG et al. 2019; RENDA et al. 2015; SUÁREZ et al. 2017]. Une hypothèse alternative serait que les clones auxquels il manque cette copie de l'IS soient plus enclins à recombiner, potentiellement en vertu d'un effet putatif exercé par le gène interrompu par l'IS sur la compétence, et donc plus enclins



## 8. DÉMONSTRATION DE LA GBGC CHEZ DES BACTÉRIES

à acquérir la résistance à la kanamycine. Si c'était le cas, cette observation irait dans le sens d'une hypothèse proposée récemment selon laquelle la transformation naturelle permet de « guérir » le génome de ces éléments génétiques mobiles [CARVALHO et al. 2019; CROUCHER, MOSTOWY et al. 2016] : ces derniers tendent à perturber ou interrompre les gènes dont l'expression permettrait leur propre excision du génome par recombinaison homologue [DURIEUX et al. 2019].

**FIGURE 8.2** – (Absence d') Influence du contexte génomique local sur la probabilité de recombinaison. Pour chaque fenêtre (non-chevauchante) de 1, 5 ou 10 kbp du génome, le nombre de régions converties a été corrélié aux mesures de divergence entre le donneur et le receveur, de taux de GC et de différence de taux de GC entre donneur et receveur. Les relations sont représentées ici; aucune corrélation n'est significative.



En dehors de ces deux régions au sein desquelles une fréquence d'import élevée est observée, les régions converties se répartissent uniformément sur le génome. De plus, nous n'avons pas trouvé de corrélation significative entre d'une part le nombre de régions converties que l'on trouve dans un intervalle génomique donné, et la divergence, le taux de GC local ou la différence locale des taux de GC entre donneur et receveur d'autre part, quelle que soit l'échelle à laquelle ces paramètres sont mesurés (1, 5 ou 10 kbp, figure 8.2). Conjointement, ces observations suggèrent que la divergence n'est pas en elle-même une barrière à l'insertion chromosomique d'ADN homéologue d'origine étrangère, du moins pas à l'échelle de divergence que l'on considère.



## Répartition des régions converties sur le génome

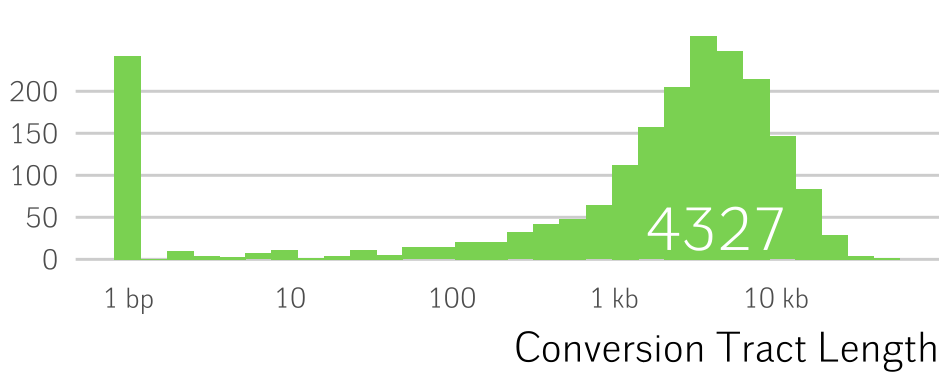


FIGURE 8.3 – Distribution de la longueur des régions converties. La longueur moyenne est de 4,327 kbp.

### Distribution de la longueur des régions converties

La longueur des régions converties est définie comme la plus grande distance entre deux marqueurs convertis d'une même région génomique, c'est-à-dire la distance entre les marqueurs convertis aux extrémités de la région. Cette longueur a une distribution nettement bimodale (figure 8.3). Cette bimodalité de la longueur des régions converties suggère que l'on peut distinguer deux classes de conversions (figure 8.4) :

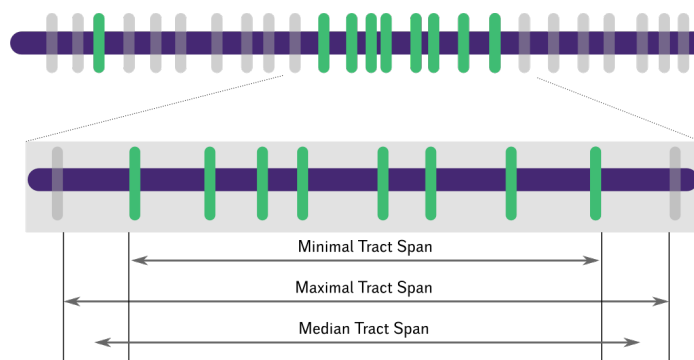


FIGURE 8.4 – Représentation schématique de deux produits de recombinaison

**les conversions isolées, SMC** les conversions d'un seul marqueur (*single marker conversions*), c'est-à-dire des conversions d'un seul marqueur générant des événements d'une longueur de 1 bp ;

**les conversions multiples, MMC** les régions converties comprenant des marqueurs multiples (*Multiple Marker Conversions*), éventuellement séparés par des marqueurs non-convertis, ou *restaurés*.

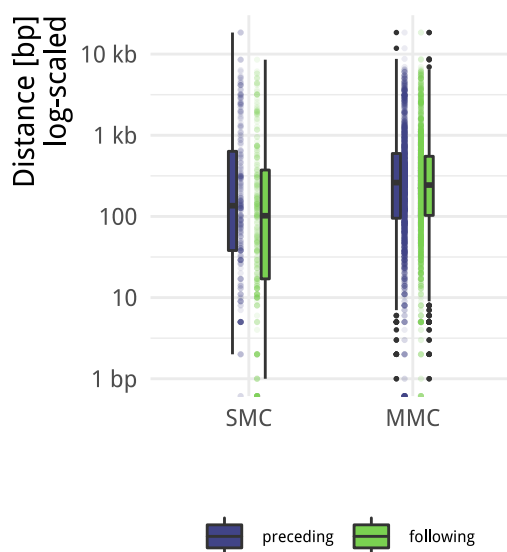
Nous avons dénombré 242 SMC, et 1781 événements de type MMC.

Une hypothèse qui pourrait expliquer que l'on observe ces SMC est qu'elles traduiraient une plus faible densité en marqueur dans les régions où elles se produisent : elles seraient moins le produit d'événements de conversion réels que de la plus faible résolution locale, et dériveraient donc d'une moindre capacité à déterminer les bornes de

## 8. DÉMONSTRATION DE LA GBGC CHEZ DES BACTÉRIES

ces événements. Une alternative serait que les mésappariements qui se produisent dans des régions de faible densité en marqueur seraient réparés différemment, générant des conversions isolées. Les deux hypothèses sont infirmées par le fait que les conversions d'un seul marqueur ne se produisent pas significativement plus loin d'un autre marqueur que les conversions impliquant plusieurs marqueurs (test t de Student,  $p$ .value = 0.15, voir figure 8.5).

**FIGURE 8.5** – Distance des régions converties aux marqueurs non-convertis les plus proches, pour les régions converties courtes (SMC) et longues (MMC). La comparaison de la moyenne des distances aux marqueurs non-convertis pour les régions de type SMC et de type MMC indique qu'elles ne sont pas significativement différentes ( $p = 0,15$ ).



La taille moyenne des conversions multiples, excluant les SMC, est de  $4,9 \text{ kbp} \pm 5,0 \text{ kbp}$ .

Cette distribution bimodale de la longueur des régions converties suggère que ces deux classes de conversions sont le produit de mécanismes de réparation différents : l'un tendrait à générer des produits de recombinaison courts, comme c'est le cas pour les mécanismes du *Short Patch Repair* (le BER notamment, décrit page 44 — KROKAN et BJØRÅS 2013); l'autre à générer des régions converties plus longues, comme c'est le cas pour les mécanismes du *mismatch repair* (MMR, décrit page 44 — SPIES et FISHEL 2015).

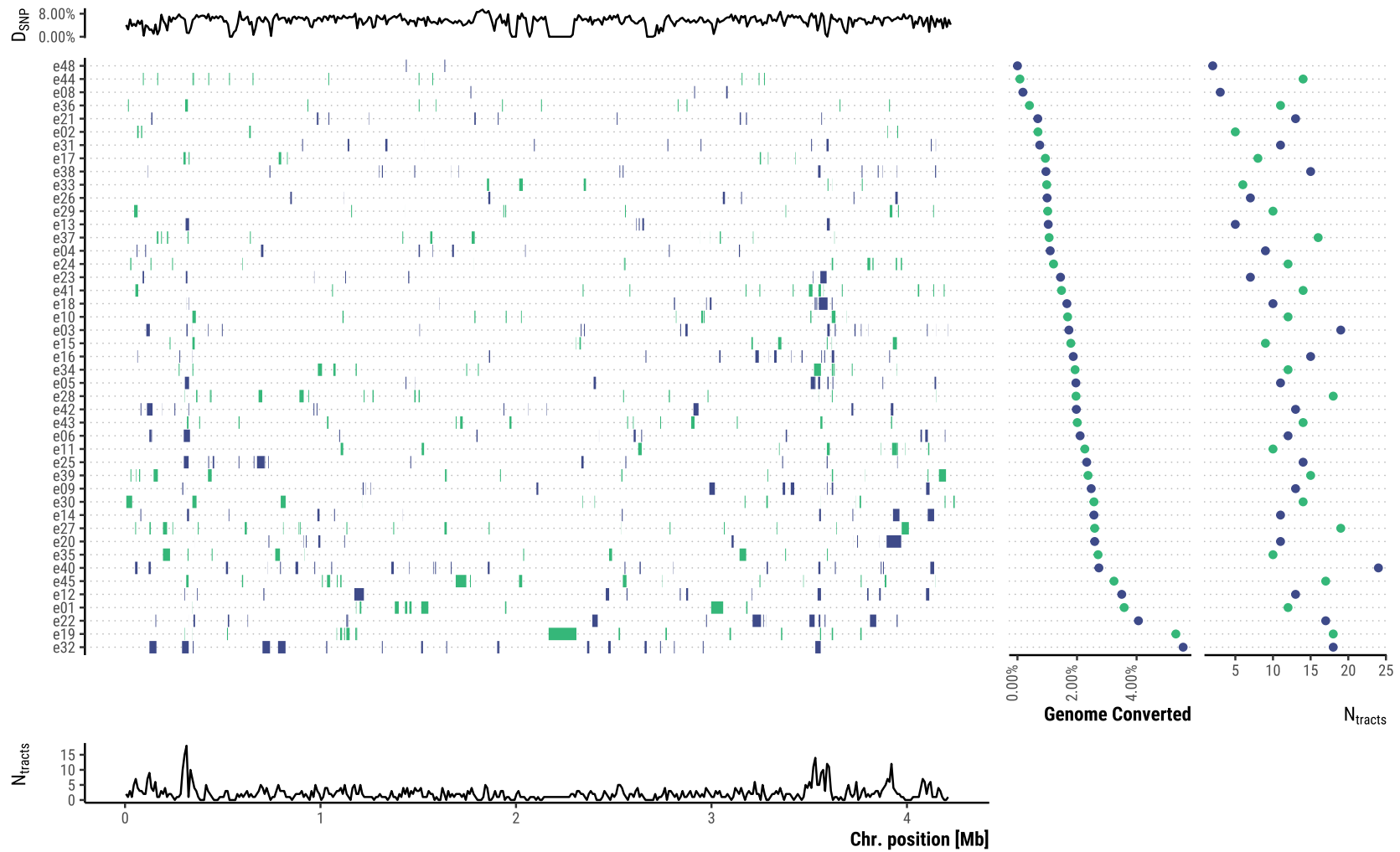


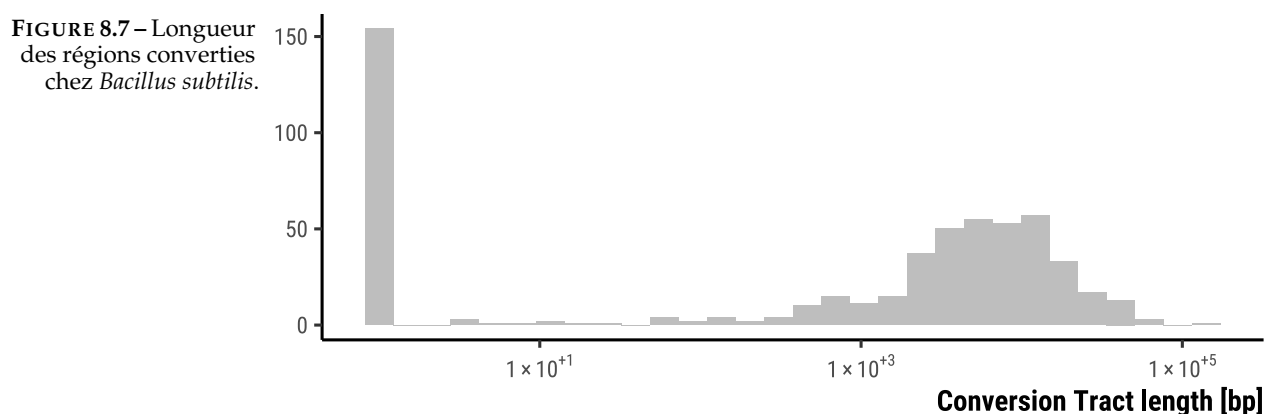
FIGURE 8.6 – Paysage d'intégration de segments donneurs dans le génome de *Bacillus subtilis*. (Description page suivante.)

## 8. DÉMONSTRATION DE LA GBGC CHEZ DES BACTÉRIES

FIGURE 8.6 – **Haut.** Divergence mesurée en nombre de SNP par fenêtre de 10 kbp. **Milieu.** Distribution des régions converties le long du génome; chaque ligne représente un clone transformant séquencé, colorée en bleu ou vert pour contraster; taux de génome receveur converti; et nombre de régions converties par clone. **Bas.** Nombre de régions converties par fenêtre de 10 kbp.

### Répartition et description des régions converties chez *Bacillus subtilis*

L'expérience de transformation génomique chez *B. subtilis* a permis de générer 549 régions converties (figure 8.6). Là encore, une distri-



bution bimodale de la longueur des régions converties a été observée (figure 8.7), avec 154 conversions isolées pour 395 conversions multiples. La longueur médiane des régions de type MMC est de 5,6 kbp, la longueur moyenne de 9,2 kbp.

## 8.2 *Acinetobacter baylyi* :

### Preuves de conversion génique biaisée vers GC dans les conversions isolées

Chez *A. baylyi*, les souches donneuses et receveuses partagent la même composition en base moyenne de 40,4 %. Parmi tous les marqueurs qui sont des polymorphismes GC/AT, le donneur est porteur de l'allèle GC dans 50,2 % des cas. Ainsi, en l'absence d'un biais de conversion quelconque, les proportions de conversions AT→GC sont censées correspondre à cette fréquence de 50,2 %.

En fait, la fréquence des conversions de AT→GC et GC→AT, prises globalement, c'est-à-dire définies comme le total des conversions de SNP (*single nucleotide polymorphisms*), ne s'écarte pas significativement

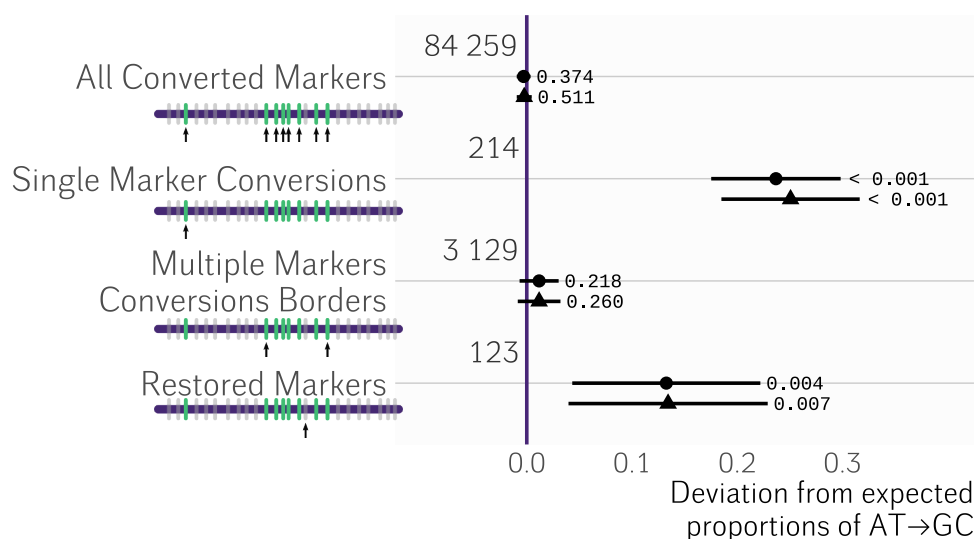


FIGURE 8.8 – Écart à la proportion attendue de conversions AT→GC. Chaque ligne représente la proportion de marqueurs de la catégorie indiquée sur l'axe des ordonnées qui sont de AT→GC — leur nombre est indiqué à gauche de la droite verticale —, normalisé par la proportion attendue, c'est-à-dire la proportion de marqueurs qui sont AT dans la souche receveuse et GC dans la donneuse. L'intervalle de confiance est donné par le test d'indépendance de  $\chi^2$ ; la *p-value* du test est indiquée à droite de l'intervalle de confiance. De haut en bas : l'ensemble des marqueurs convertis; les conversions d'un seul marqueur; les extrémités des régions de conversions multiples et les restaurations de marqueurs isolées. Les marqueurs restaurés isolés sont des marqueurs à l'intérieur d'une conversion plus longue et flanqués par deux marqueurs convertis. Ils pourraient être dûs à l'action du *Short Patch Repair* à l'intérieur de l'hétéroduplexe qui restaurerait l'haplotype receveur. Les triangles représentent les proportions observées quand les régions converties proches des deux régions à haute fréquence d'import sont exclues.

des proportions entre le donneur et le receveur (figure 8.8). Néanmoins, les fréquences de conversions de marqueurs isolés (SMC) montrent un excès significatif de conversions AT→GC (73,8 % des 214 SMC sont des conversions AT→GC, attendu 50,2 %;  $p$ -value du test d'indépendance de  $\chi^2 = 8,2 \cdot 10^{-12}$ ). Pour vérifier que cette observation ne soit pas due à un biais lié à des conversions à proximité des deux régions où la fréquence d'import est élevée, j'ai réalisé le même test en excluant les régions converties dans un périmètre de 100 kbp autour des deux régions en question. Ce test reste significatif quand les régions proches des deux régions où la fréquence d'import est plus élevée sont exclues (représentés par des triangles dans la figure 8.8). Chez la Levure, Y. LESECQUE et al. [2013] ont observé que le biais GC est spécifiquement associé aux extrémités des régions converties. Nous n'avons pas détecté un tel biais aux extrémités des régions de type MMC dans l'espèce *A. baylyi*.

Enfin, nous avons détecté des cas d'alternances d'haplotype qui génèrent des patrons de conversions mosaïques à l'intérieur des régions converties [KULICK et al. 2008], avec au moins une alternance entre l'haplotype du donneur et celui du receveur dans 13,5 % des régions converties. Ces marqueurs sont qualifiés de marqueurs *restaurés*, puisqu'ils ont vraisemblablement été inclus dans l'étendue de l'hétéroduplexe, et ont été restaurés vers l'haplotype du receveur. Parmi l'ensemble des marqueurs restaurés, il n'y a pas d'écart significatif aux proportions attendues de conversions AT→GC ( $p$ -value du test de  $\chi^2 \approx 0,55$ ). Cependant, les marqueurs restaurés de façon isolés, c'est-à-dire l'équivalent des marqueurs isolés dans le contexte des restaurations, sont plus fréquemment des restaurations AT→GC (où le donneur porte l'allèle AT et le receveur l'allèle GC) que des restaurations GC→AT. 63,4 % des 123 restaurations isolées sont des restaurations d'allèles GC quand le donneur porte l'allèle AT ( $p = 4,4 \cdot 10^{-3}$ ). Là encore, cela suggère qu'un mécanisme de réparation des mésappariements spécifique engendre ce patron de corrections isolées biaisées vers GC, ce qui correspond typiquement à la signature du *Short Patch Repair*.

### *Excès de transversions dans les conversions d'un seul marqueur*

Au sein des marqueurs convertis, davantage de conversions sont des transitions qu'attendu d'après les proportions de transitions et transversions entre le donneur et le receveur (figure 8.9). Cet excès est significatif, quoique très faible (24 % des conversions sont des transversions, contre 25 % des marqueurs). Cependant, les conversions isolées et les restaurations isolées sont toutes deux biaisées en faveur des transversions, suffisamment cette fois pour que cela soit biologiquement pertinent.

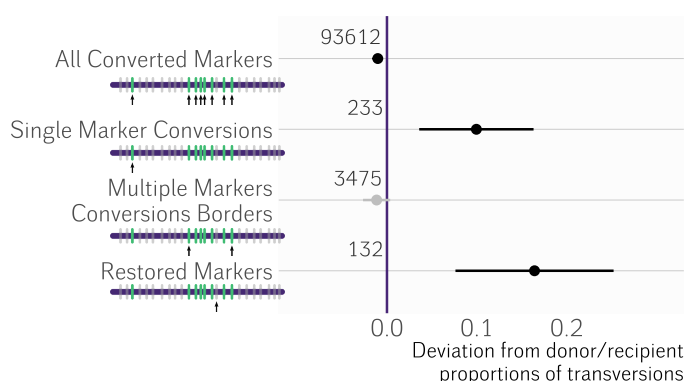


FIGURE 8.9 – Excès de transversions dans les conversions isolées.

(Respectivement 35 et 42 % des conversions isolées et des restaurations isolées sont des transversions, contre 25 % attendus.)

Ceci dit, les transitions et les transversions sont toutes deux biaisées vers GC dans le contexte des marqueurs isolés, bien que les transversions le soient davantage que les transitions (table 8.1). Cela suggère

			Écart aux proportions attendues.
<b>transitions</b>	GC→AT	AT→GC	
SMC	48	103	0.18 ± 0.08
Markers	12730	12882	
<b>transversions</b>	GC→TA	AT→CG	
SMC	8	55	0.38 ± 0.09
Markers	2485	2426	

TABLE 8.1 – Both transitions and transversions are GC-Biased.

que la propension d'une base mésappariée à ce que sa correction soit le fait d'une voie de réparation donnée dépend de la nature du mésappariement.

#### *Absence d'indices d'une asymétrie liée à l'orientation de la réplication dans le spectre de conversion*

De nombreux travaux ont montré que l'orientation des brins par rapport à la polarité de la réplication a une influence majeure sur la composition en base du brin et la façon dont elle varie le long du brin [DAUBIN et PERRIÈRE 2003; FOSTER, NICCUM et al. 2018; A. C FRANK et J. R LOBRY 1999; J. R. LOBRY 1996b; E. P. C. ROCHA 2004; SUEOKA 1995]. De plus, le devenir du brin intégré au cours de la transformation naturelle dépend probablement de la dynamique de réplication de la cellule compétente [JOHNSTON, MARTIN, FICHANT et al. 2014; JOHNSTON, MARTIN, POLARD et al. 2013]. Enfin, la correction des mésappariements liés à des *pré-mutations* qui surviennent au cours de la réplication est vraisemblablement biaisée à l'égard de l'asymétrie répliative,

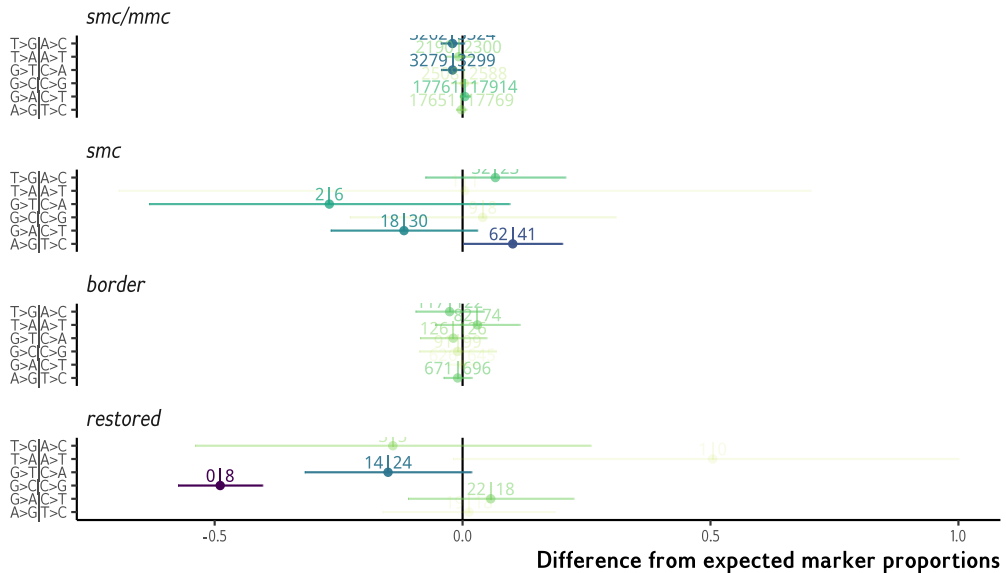
## 8. DÉMONSTRATION DE LA GBGC CHEZ DES BACTÉRIES

comme cela a été observé par des expériences d'accumulation de mutation [FOSTER, H. LEE et al. 2015 ; FOSTER, NICCUM et al. 2018]. À ce titre, il se pourrait que la correction des mésappariements liés à des hétéroduplexes de recombinaison homologue le soient aussi.

J'ai donc comparé la propension du brin sens (*leading*) à accumuler davantage de conversion d'un type donné par rapport à son complément. Ceci indiquerait un écart aux règles de parité en base (voir figure 8.10).

Si les mécanismes de la conversion sont biaisés par rapport au brin sur lequel ils opèrent, alors est attendu un excès d'un type de conversion — mettons A→G — sur un brin donné par rapport à la conversion complémentaire — T→C. J'ai donc séparé le génome en deux réplicores en fonction de leur asymétrie cumulative en GC (voir partie 4.2 page 53, A. C. FRANK et J. R. LOBRY 2000). Pour chacune des six catégories de conversions complémentaires (par exemple, A→G *versus* T→C, ou A→C *versus* T→G), la proportion de conversions a été comparée à celle des marqueurs entre donneur et receveur. Aucun écart aux proportions attendues n'est significatif, à l'exception du nombre de conversions qui sont des transversions G→C dans les conversions isolées; cet écart n'est pas significatif une fois appliquée une correction de Bonferonni pour les test multiples. Ces observations suggèrent que l'orientation de la

FIGURE 8.10 – Pas d'indice d'un biais lié à l'asymétrie réplivative



réplication n'a pas d'influence sur la manière dont sont corrigés les mésappariements liés aux hétéroduplexes de recombinaison homologue.



### Résumé

Ainsi, nous avons démontré dans cette partie que dans l'espèce *Acinetobacter baylyi*, la réparation des mésappariements qui se produisent dans les produits de recombinaison générant des conversions isolées sont biaisés en faveur des allèles G et C. Lorsqu'un site polymorphe AT : GC est impliqué dans un événement de conversion court, l'allèle GC est le donneur de l'événement de conversion dans 73,8% des cas (63,4% pour les restaurations).

Dans la partie qui suit, je mesure les fréquences de conversion pour les autres espèces : *Bacillus subtilis*, *Streptococcus pneumoniae* et *Helicobacter pylori*.

## 8.3 Preuves de gBGC dans deux autres espèces de Bactéries

Au delà de l'analyse des régions converties produites par la transformation génomique chez *A. baylyi*, nous avons répété le protocole de transformation avec *B. subtilis*. Nous avons aussi analysé les données produites par d'autres expériences de transformation génomique, et n'avons retenu que les modèles pour lesquelles il y avait suffisamment de marqueurs en dehors des régions intégrées à proximité du marqueur de sélection. Ceci nous a permis d'obtenir un jeu de données avec quatre espèces. La table 8.2 référence les aspects généraux de ces quatre espèces pour lesquelles j'ai pu mesurer les fréquences de conversion. Les

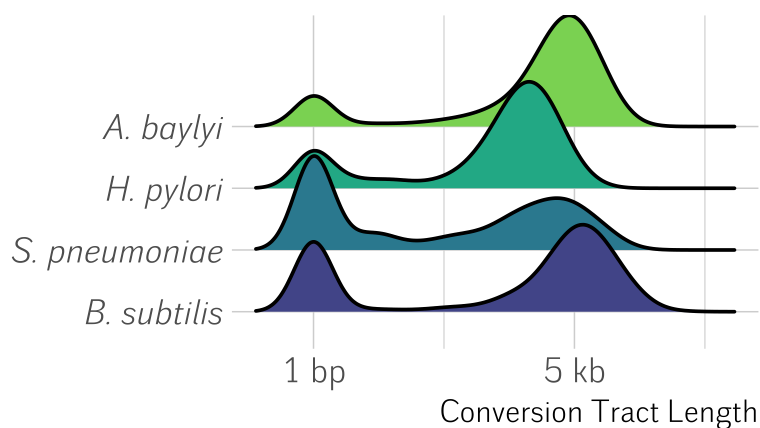
	GC%	Genome Size (mb)	Divergence (%)	MMC	SMC
<i>A. baylyi</i>	40,43	3,60	0,98	1781	242
<i>H. pylori</i>	38,87	1,67	4,22	1504	255
<i>S. pneumoniae</i>	39,72	2,04	0,86	687	136
<i>B. subtilis</i>	43,51	4,22	5,32	395	154

TABLE 8.2 – Description générale des espèces étudiées et des régions converties identifiées

quatre espèces sont des espèces à taux de GC relativement « bas », avec une taille de génome comprise entre 1,67 Mbp pour *H. pylori* à 4,22 Mbp pour *B. subtilis*. Au total, en incluant les régions converties chez *A. baylyi*, cela représente 5415 produits de recombinaison différents. La longueur médiane des régions converties, excluant les conversions isolées, est respectivement pour *A. baylyi*, *H. pylori*, *S. pneumoniae* et *B. subtilis* de 3,4, 0,9, 1,3 et 5,6 kbp (figure 8.11). (Mais voir COWLEY et al. [2018] pour plus de détails sur la longueur des produits de recombinaison chez *S. pneumoniae*.)

## 8. DÉMONSTRATION DE LA GBGC CHEZ DES BACTÉRIES

FIGURE 8.11 –  
Distribution de la  
longueur des régions  
converties pour quatre  
espèces de Bactéries.



Dans toutes les espèces étudiées, la distribution de la longueur des régions converties est bimodale, ce qui d'une part confirme qu'il y a bien deux catégories de régions converties, et d'autre part permet d'appliquer aux autres espèces les mêmes analyses que celles pratiquées dans l'espèce *A. baylyi*.

De façon similaire à l'espèce *A. baylyi*, aucun écart significatif aux proportions attendues de conversions AT→GC n'a été observé, quelle que soit l'espèce considérée, quand les conversions sont comptées conjointement (figure 8.12). Les conversions isolées ne sont biaisées en faveur des allèles G et C que dans l'espèce *A. baylyi* (figure 8.12).

Cependant, les *extrémités* des régions de conversions multiples sont significativement biaisées en faveur des allèles G et C dans les deux espèces *H. pylori* et *S. pneumoniae*. Ce patron correspond avec ce qui a déjà été observé chez la Levure [Y. LESECQUE et al. 2013].

Enfin, j'ai détecté des alternances d'haplotype dans les régions converties dans chaque espèce; ces régions converties sont dites *complexes* parce qu'elles comportent au moins deux alternances d'haplotype donneur/receveur. Les marqueurs restaurés vers l'haplotype receveur de manière isolée ne sont biaisés en faveur des allèles G et C que dans l'espèce *A. baylyi*.

L'espèce *B. subtilis* ne montre aucun signe de conversion génique biaisée vers GC, quel que soit le type de marqueur considéré.

Conjointement, ceci démontre qu'il existe une forme de gGBC dans trois des quatre espèces analysées. Dans la partie suivante, je quantifie

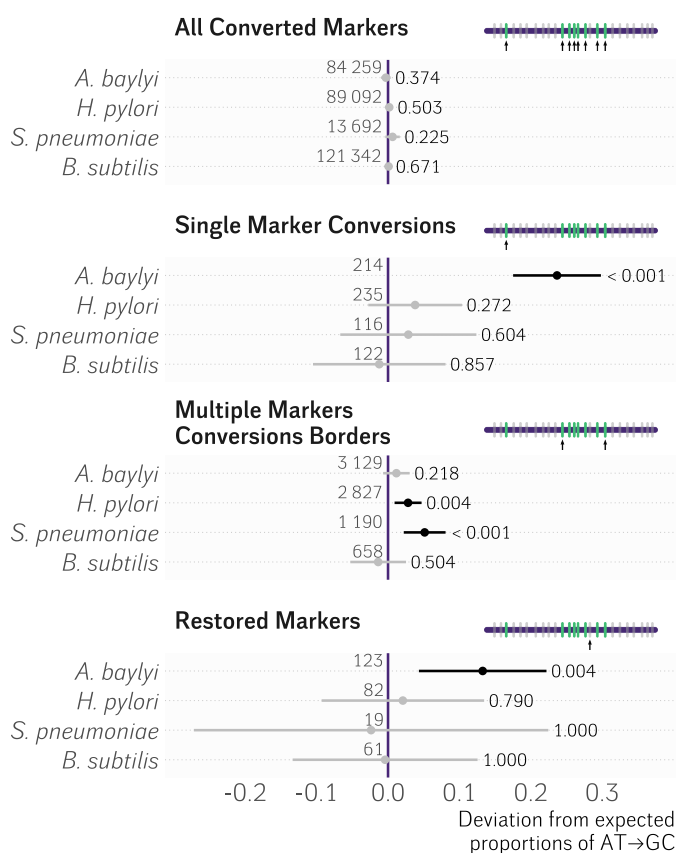


FIGURE 8.12 – Fréquence de conversions parmi quatre espèces de Bactéries. (*A. baylyi* a été incluse ici pour la comparaison.) Chaque point représente l'écart aux proportions attendues de conversions AT→GC, sachant les proportions de marqueurs qui portent l'allèle AT chez le donneur et l'allèle GC chez le receveur par rapport à l'inverse. Les points et l'intervalle de confiance associé sont colorés en noir quand le test d'indépendance est significatif (la valeur p est rapportée à droite de l'intervalle de confiance), en gris sinon. Le nombre de sites polymorphes pris en compte est indiqué à gauche de la ligne verticale d'équation  $y = 0$ .

l'influence du biais mesuré chez *A. baylyi* sur la probabilité de fixation des allèles G et C.

## 8.4 Estimation de l'intensité de la gBGC à l'échelle populationnelle

Intuitivement, il pourrait sembler qu'un mécanisme restreint à une certaine catégorie de polymorphisme n'ait qu'une influence mineure, sinon nulle, sur la composition en base globale. Cependant, comme je l'ai détaillé dans la partie 6.1 (page 83), ce qu'il convient de prendre en compte est le coefficient de gBGC à l'échelle populationnelle. En fait, comme je le montre dans cette partie, l'intensité du biais mesurée chez *A. baylyi* est étonnamment forte.

Pour rappel, l'effet populationnel moyen d'un biais de conversion B pour une espèce haploïde peut être résumé par

$$\begin{aligned}
 B &= 2N_e b \\
 &= 2N_e r L_h b_0
 \end{aligned}
 \tag{8.1}$$

où  $N_e$  est la taille efficace,  $r$  le taux de recombinaison par paire de base,  $L_h$  la longueur moyenne des hétéroduplexes, et  $b_0$  l'intensité du biais de réparation [DURET et GALTIER 2009; NAGYLAKI 1983a; WALSH 1983]. Pour le calculer, j'ai besoin d'une estimation du taux de recombinaison par paire de base, d'une estimation de la taille efficace et une mesure du biais de réparation. Ces trois paramètres peuvent être estimés pour l'espèce *A. baylyi*, aussi, dans un premier temps, je me restreindrai au calcul de  $B$  pour cette espèce.

Estimer  $N_e$  requerrait que nous disposions d'information sur le polymorphisme qui ségrège dans l'espèce *A. baylyi*, ce qui n'est pas possible pour le moment avec seulement trois génomes disponibles. Nous avons à la place utilisé l'estimation de la taille efficace réalisée par [BOBAY et OCHMAN 2018] pour l'espèce *Acinetobacter pittii*, ce qui donne  $N_e \approx 3,3 \cdot 10^8$ .<sup>49</sup> Le ratio  $r/\mu$ , où  $r$  est le taux de recombinaison par paire de base par génération, et  $\mu$  le taux de mutation par paire de base par génération, pour cette espèce a été estimé à  $r/\mu \approx 0.49$ . Ce ratio peut être utilisé pour estimer  $r$  si l'on connaît  $\mu$  par ailleurs. Une estimation de  $\mu$  a été mesurée par expérience d'accumulation de mutation pour *A. baylyi* réalisée par RENDA et al. [2015] et SUÁREZ et al. [2017]; de sorte que  $\mu \approx 7,9 \cdot 10^{-11} \text{ bp}^{-1} \text{ gen}^{-1}$ , et donc  $r \approx 3,9 \cdot 10^{-11} \text{ bp}^{-1} \text{ gen}^{-1}$ .

Venons-en maintenant à l'estimation de  $b_0$ . Nos travaux ont montré que la gBGC est spécifiquement associée aux conversions isolées (pour simplifier, je ne tiendrai pas compte de la restauration des marqueurs isolés). Ce biais de transmission en faveur des allèles GC, ou le biais de réparation associé aux événements de type SMC, noté  $b_S$ , correspond à la différence entre la probabilité pour un polymorphisme AT : GC d'être converti ou en AT (26 %) ou en GC (74 %). Ainsi,  $k_{GC} = 0.74$  et  $b_S = 2k_{GC} - 1 = 0,48$ .

Comme la gBGC résulte de la réparation biaisée des mésappariements associés aux hétéroduplexes de recombinaison, son intensité dépend de la probabilité pour une base d'être impliquée dans un hétéroduplexe. Cette probabilité, notée  $g$ , dépend donc de l'étendue moyenne d'un hétéroduplexe, notée  $L_h$ , et de  $r$  le taux de recombinaison par base par génération :  $g = r \cdot L_h$ . Puisque la gBGC chez *A. baylyi* est spécifiquement associée aux conversions isolées, le taux de conversion génique est donné par  $g_S = r L_h f$ , où  $f$  est la fraction des événements de recombinaison qui sont réparés par la voie du *Short Patch Repair*. Dans notre expérience, les conversions isolées représentent 242 des 2023 produits de recombinaison, et donc  $f = 0,12$ .

La taille de l'hétéroduplexe n'est pas connue *a priori*, mais DE VRIES et WACKERNAGEL [2002] ont montré que l'intégration de séquences

49. BOBAY et OCHMAN [2018] montrent à ce propos que des espèces « proches tendent à montrer des valeurs de  $N_e$  similaires ».

hétérologues dans le génome d'*A. baylyi* requiert deux séquences bordantes homologues d'environ 183 bp ; ce qui suggère que la longueur de l'hétéroduplexe est de l'ordre de 0,5 kbp. Si l'on postule une longueur moyenne d'hétéroduplexe de 0,5 kbp, la probabilité d'un événement de conversion génique isolée  $g_S$  est donc

$$\begin{aligned} g_S &= rL_h f \\ &= 3,9 \cdot 10^{-11} \times 0,5 \cdot 10^3 \times 0,12 \\ &= 2,4 \cdot 10^{-9} \end{aligned}$$

Ce terme  $g_S$  est la probabilité qu'une base soit soumise à un événement de conversion génique de type SMC ; il peut être substitué au terme  $r \cdot L_h$  dans l'équation 8.1.

Ces différentes estimations, si grossières soient-elles, de  $N_e$ ,  $g_S$  et  $b_S$  aboutissent à une intensité populationnelle de la gBGC de :

$$\begin{aligned} B &= 2 \cdot N_e \cdot g_S \cdot b_S \\ &= 2 \cdot 3,3 \cdot 10^8 \cdot 2,4 \cdot 10^{-9} \cdot 0,47 \\ &= 0,713 \end{aligned} \tag{8.2}$$

Or, comme je l'ai décrit dans l'équation 6.2 (page 78), quand  $B > 1$  (où  $B = 2\psi N_e b$ ), l'intensité de la conversion génique dépasse celle de la dérive génétique, et donc la probabilité de fixation d'un allèle GC soumis à de la conversion génique est supérieure à celle d'un allèle qui ne l'est pas.<sup>50</sup> La question qui se pose alors est de savoir si cette estimation du biais de fixation en faveur des bases G et C est suffisante pour expliquer l'écart entre la composition en base attendue à l'équilibre mutationnel et la composition en base observée ?

Pour rappel, le taux de GC à l'équilibre mutationnel est donné par

$$GC_n = \frac{1}{1 + m}$$

et le taux de GC à l'équilibre « substitutionnel », sachant un biais de fixation causé par la gBGC, noté B, par

$$GC_{eq} = \frac{1}{1 + m e^{-B}}$$

où  $m$  est le ratio du nombre de mutations GC→AT par rapport au nombre de mutations AT→GC, qui vaut 2,8 d'après les données de SUÁREZ et al. [2017]. Le taux de GC à l'équilibre mutationnel est donc

50. Il faut rappeler que je néglige ici le biais lié aux restaurations de marqueurs isolés.

Il est donc vraisemblable que B soit en fait supérieur à la valeur rapportée ici.

## 8. DÉMONSTRATION DE LA GBGC CHEZ DES BACTÉRIES

25,7 %, et le taux de GC à l'équilibre substitutionnel 42,1 %. Cependant, le taux de GC des positions des codons quatre fois dégénérées (GC4) est égal à 48 % dans l'espèce *Acinetobacter baylyi* ADP1. Ceci indique que l'effet populationnel de la force ou des forces, quelles qu'elles soient, qui influencent le taux de GC4, soit la sélection naturelle, soit la gBGC, soit les deux, doit être de l'ordre de 0,5 à 1 (figure 8.14). Ce qui est de l'ordre de grandeur d'une estimation de B à  $\approx 0,713$ .

Le fait que *Bacillus subtilis* ne montre pas de signal de gBGC pourrait s'expliquer assez bien par le fait qu'il s'agit de l'une de ces espèces pour lesquelles la composition en base correspond le mieux à celle attendue à l'équilibre mutationnel. Autrement dit, puisque le génome de *Bacillus subtilis* est quasiment à l'équilibre mutationnel, on ne devrait pas y trouver de biais de fixation en faveur des bases G et C, et par conséquent la force de la gBGC devrait y être relativement faible. Il faut bien admettre cependant qu'il est difficile de concevoir de quelle manière cette hypothèse pourrait bien être testée.

Ce travail démontre ainsi que, dans ce modèle, si l'on ne tient pas compte de l'incertitude associée à l'estimation des paramètres, la conversion génique biaisée vers GC suffirait à expliquer l'écart entre le taux de GC à l'équilibre mutationnel et le taux de GC aux positions les moins soumises à des contraintes sélectives.

### Estimation de l'incertitude autour de B

Évidemment les estimations des paramètres utilisées ici présentent plus ou moins d'incertitude. Dans ce qui suit, j'essaie de donner une « gamme » de B en faisant varier la longueur de l'hétéroduplexe, et les paramètres  $N_e$  et  $r$ .

L'incertitude dont il est question tient essentiellement à deux paramètres : la longueur de l'hétéroduplexe, qui n'est pas a priori pas connue chez *A. baylyi*, et la valeur de  $\rho$  l'intensité de la recombinaison à l'échelle populationnelle.

De la longueur de l'hétéroduplexe dépend la probabilité qu'une base soit soumise à un événement de conversion. L'intensité de la gBGC à l'échelle populationnelle augmente en fait linéairement avec cette longueur. Pour l'espèce *A. baylyi*, dès lors que la longueur de l'hétéroduplexe dépasse 690 bp, l'intensité de la gBGC à l'échelle populationnelle dépasse 1, lorsque  $N_e$ ,  $r$ ,  $f$  et  $b$  sont tenus constants (figure 8.13).

L'intensité de la recombinaison à l'échelle populationnelle  $\rho$  telle que  $\rho = 2N_e r$  n'est pas connue dans l'espèce *A. baylyi*, c'est pourquoi je l'ai approchée par celle d'une espèce du même genre. Or dans la partie 6.2 (page 86), j'ai tenté d'estimer la distribution de  $\rho = 2N_e r$  chez les Procaryotes. Pour rappel, en faisant l'hypothèse que le taux de

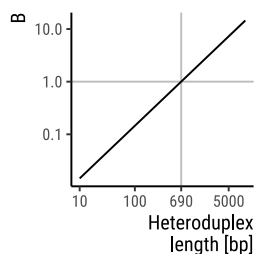


FIGURE 8.13 – Relation entre B et longueur de l'hétéroduplexe.

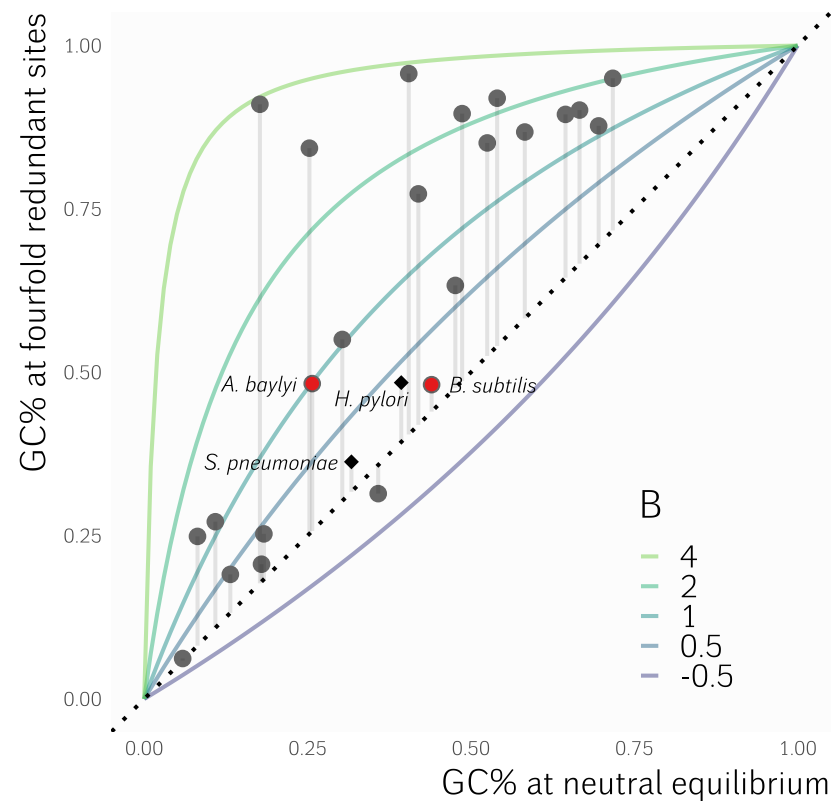


FIGURE 8.14 – Relation entre composition en base attendue à l'équilibre mutationnel et taux de GC observé aux positions quatre fois dégénérées. *A. baylyi* a été ajouté à cette figure en utilisant les données de SUÁREZ et al. [2017], et est souligné en rouge, tout comme *B. subtilis*. *S. pneumoniae* et *H. pylori*, pour lesquels je n'ai pas connaissance d'expériences d'accumulation de mutation permettant de connaître leur patron mutationnel, ont été ajoutés *via* une estimation du taux de GC à l'équilibre calculé par HILDEBRAND et al. [2010], représentés par des losanges noirs. J'ai représenté les isoclines décrivant le taux de GC attendu pour une valeur de B donnée, sachant m. Ceci indique que la valeur estimée de  $B \approx 0.726$  (pour  $L_h = 0,5$  kbp) ou  $B \approx 1.45$  (pour  $L_h = 1$  kbp) est une valeur plausible qui expliquerait l'écart entre GC4 observé et GC attendu à l'équilibre mutationnel.



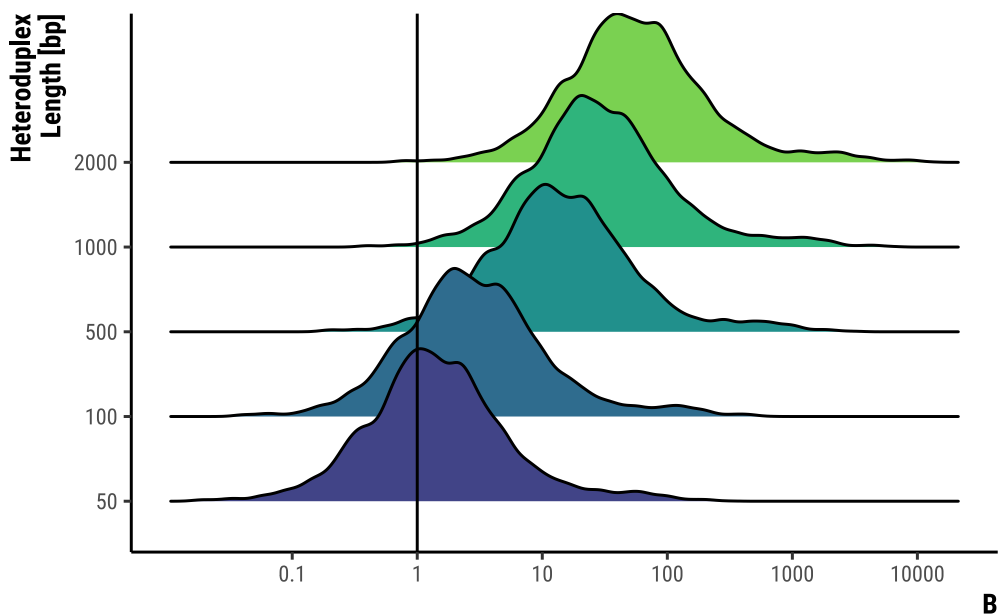
## 8. DÉMONSTRATION DE LA GBGC CHEZ DES BACTÉRIES

mutation  $\mu$  des Procaryotes est distribué comme le taux de mutation mesuré dans les 22 espèces de Bactéries rapportées par LONG et al. [2018], j'ai pu obtenir une distribution grossière de  $\rho$  sachant les valeurs de  $r/\mu$  et de  $N_e$  estimées par BOBAY et OCHMAN [2018] pour 153 espèces de Procaryotes. Faisons l'hypothèse, a priori fondée, que l'incertitude quant à la valeur de  $\rho$  de l'espèce *A. baylyi* est bien représentée par cette distribution de  $\rho$ ; autrement dit que la valeur de  $\rho$  de cette espèce est incluse dans cette distribution. Alors on peut réécrire l'équation 8.1 en

$$\begin{aligned} B &= (2N_e r) L_h b_0 \\ &= \rho \cdot L_h \cdot f \cdot b_S \\ &= \rho \cdot L_h \cdot 0.12 \cdot 0.47 \end{aligned}$$

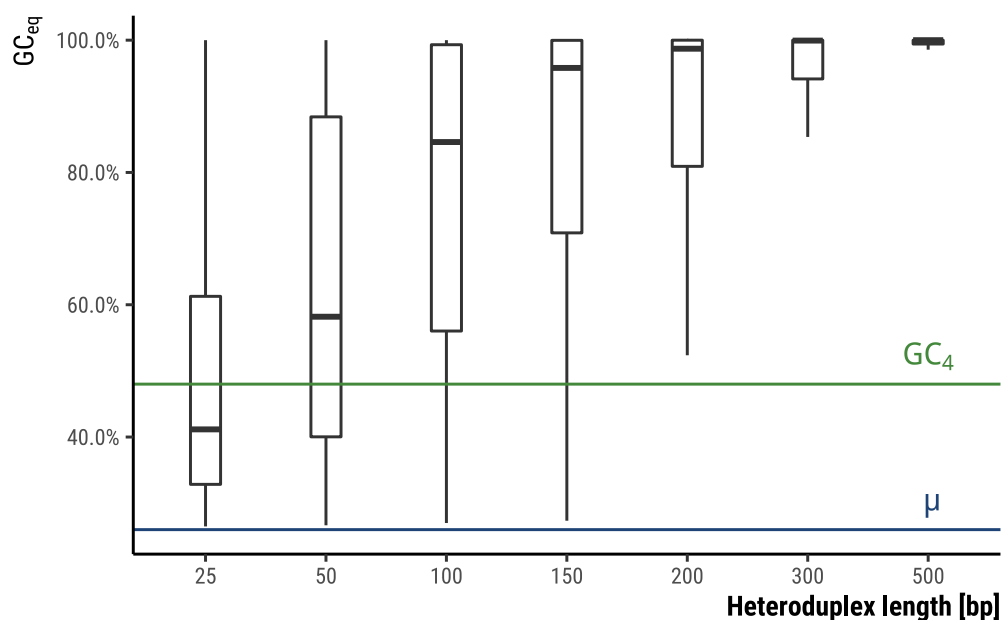
Sachant (grossièrement) la distribution de  $\rho$ , on peut donc obtenir une distribution de la valeur de  $B$  pour l'espèce *A. baylyi* en gardant constants  $f = 0.12$  et  $b_S = 0.47$  (figure 8.15).

FIGURE 8.15 – Distribution de la valeur de  $B$  dans l'espèce *Acinetobacter baylyi* sous l'hypothèse que l'incertitude quant à l'intensité de recombinaison populationnelle  $\rho$  de cette espèce est bien représentée par la distribution de la valeur de  $\rho$  chez 153 espèces de Bactéries. (Valeurs de  $\rho$  estimée à partir des données extraites de BOBAY et OCHMAN [2018] et LONG et al. [2018].)



On voit ainsi que, si l'incertitude quant à l'intensité de la recombinaison à l'échelle populationnelle, notée  $\rho$ , de l'espèce *Acinetobacter baylyi* est bien représentée par la distribution d'une estimation de  $\rho$  d'un ensemble de 153 Procaryotes, l'espace des paramètres dans lequel la valeur de  $B$  dépasse 1 est très grand. Autrement dit, l'intensité de la gBGC mesurée chez *A. baylyi* dans les conversions isolées est susceptible d'occasionner un biais de fixation des allèles GC très fort, étant données des valeurs plausibles des paramètres en jeu (figure 8.16).





**FIGURE 8.16** – Contenu en GC attendu à l'équilibre substitutionnel avec conversion génétique biaisée vers GC d'intensité B tel que  $GC_{eq} = \frac{1}{1 - m e^{-B}}$  en fonction de la longueur de l'hétéroduplexe. B a été estimé à partir de la distribution de  $\rho$  et des valeurs de  $b_0$  calculée chez *A. baylyi*. Les valeurs de  $GC_4$  et de taux de GC attendu à l'équilibre mutationnel sont représentées par les lignes horizontales.



## DISCUSSIONS

---

If I adopt the method of tenacity, and shut myself out from all influences, whatever I think necessary to doing this, is necessary according to that method. So with the method of authority : the state may try to put down heresy by means which, from a scientific point of view, seem very ill-calculated to accomplish its purposes ; but the only test *on that method* is what the state thinks ; so that it cannot pursue the method wrongly. [...] But with the scientific method the case is different. I may start with known and observed facts to proceed to the unknown ; and yet the rules which I follow in doing so may not be such as investigation would approve. The test of whether I am truly following the method is not an immediate appeal to my feelings and purposes, but on the contrary, itself involves the application of the method.

PEIRCE [1992b]

### 9.1 Est-ce que la sélection *cause* le biais en faveur de GC?

Dans trois des quatre espèces étudiées, nous avons observé un excès de conversion AT→GC par rapport à GC→AT dans les produits de conversion. Est-il possible que cet excès résulte de processus *sélectifs* qui se seraient déroulés au cours de l'expérience ? Il se peut en effet qu'il y ait de la compétition entre transformants dans les cultures en milieu liquide. Qu'une conversion donnée soit avantageuse conduirait à l'invasion des cultures par les clones chez lesquels elle se serait produite.

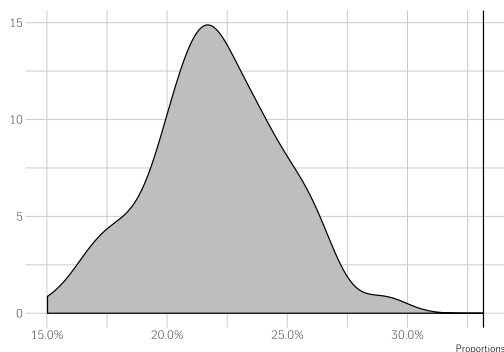
Néanmoins, j'estime cette hypothèse tout à fait improbable pour les raisons que j'expose maintenant.

Chez *A. baylyi*, où le signal de gBGC est le plus fort, les conversions isolées sont distribuées uniformément sur le génome de différentes lignées indépendantes. Or si la sélection promouvait la conversion d'un marqueur donné, l'on s'attendrait à ce que ce marqueur soit converti dans plusieurs lignées indépendantes. Seulement 17 marqueurs sont convertis dans deux clones indépendants. S'ils sont exclus du calcul de gBGC, il reste 208 marqueurs strictement uniques, dont 128 sont des conversions AT→GC (70,3 %,  $p = 9 \cdot 10^{-8}$ ). Ce ne sont donc pas des marqueurs fréquents parce que potentiellement avantageux qui causent le biais de transmission AT→GC.

## 9. DISCUSSIONS

Ceci dit, nous avons observé que les conversions isolées, lorsqu'elles se produisent dans des régions codantes, sont plus fréquemment non-synonymes que synonymes, et ce significativement plus que ce que l'on attendrait d'après la proportion de marqueurs non-synonymes ( $p < 10^{-3}$ , figure 9.1). Ce phénomène a également été observé dans les

**FIGURE 9.1 –**  
Distribution du nombre de conversions non-synonymes attendues en échantillonnant aléatoirement la quantité de conversion observée parmi l'ensemble des marqueurs. Le nombre observé est rapporté par une barre verticale.



expériences d'accumulation de mutation : les mutations non-synonymes sont plus fréquentes qu'attendu d'après un échantillonnage aléatoire de mutations le long du génome [FOSTER, H. LEE et al. 2015 ; H. LEE et al. 2012 ; SUNG, ACKERMAN, GOUT et al. 2015].

Cet excès de mutation non-synonyme indique que la diversité des conversions détectées au cours de l'expérience a pu être affectée par des effets sélectifs. Cependant, la proportion de conversions non-synonymes est à même parmi les SMC de type AT→GC et GC→AT (respectivement 31,8 % et 28,3 %).

Par ailleurs, l'excès de conversions de type SMC AT→GC est le même dans toutes les catégories de sites, qu'ils soient codants ou intergéniques, synonymes ou non (table 9.1). Ceci suggère que le processus qui favorise

**TABLE 9.1 –** Nombre de conversions de type SMC aux différentes position du génome

	AT→GC	GC→AT
codant	132 (73 %)	47
intergénique	26 (68 %)	12
non-synonymes	42 (76 %)	13
synonymes	90 (73 %)	33

les conversions AT→GC opère de la même façon quelle que soit la catégorie de site considérée. Il ne s'agit donc pas d'un processus sélectif spécifiquement lié au codage des protéines.

Est-il envisageable que cet excès de conversion AT→GC résulte d'une pression de sélection sur la composition en base globale du génome ? Si c'était le cas, il faudrait s'attendre à ce que la sélection favorise *tous* les

événements de conversion qui introduisent davantage de GC que de AT. Chez *A. baylyi*, en moyenne les conversions de type MMC introduisent 23,9 marqueurs AT et 23,7 marqueurs GC. La différence entre le nombre de marqueurs GC introduits et le nombre de marqueur AT introduits en moyenne est nulle. Cependant, l'écart type de cette différence est de 6,75, et 35,9 % des MMC introduisent plus d'un marqueur GC. Il existe donc

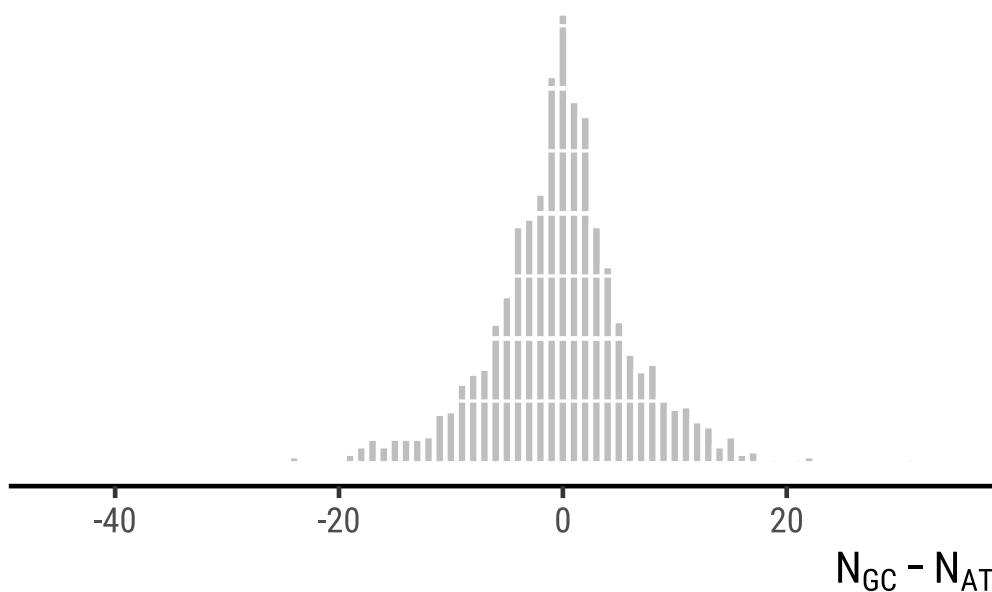


FIGURE 9.2 – Distribution de l'écart entre le nombre de marqueurs GC introduits et le nombre de marqueurs AT introduits par des événements de conversion de type MMC chez *A. baylyi*.

une variabilité du nombre de marqueurs GC introduits par les MMC sur laquelle la sélection est susceptible d'opérer. Or, si la sélection favorisait l'introduction d'allèles GC, cette pression de sélection devrait affecter les MMC plus fortement que les SMC qui n'apportent qu'un seul marqueur GC à la fois. Le fait que la gBGC soit observée uniquement dans les SMC chez *A. baylyi*, ou aux extrémités des régions converties chez *H. pylori* et *S. pneumoniae* est donc incompatible avec une hypothèse sélectionniste.

Il n'est donc pas exclu que quelques marqueurs isolément présentent un avantage sélectif quelconque, mais ça n'est vraisemblablement pas cela qui cause le biais en faveur des bases G et C.

## 9.2 Les mécanismes responsables de la gBGC

À ce jour, la gBGC a été observée par diverses études empiriques chez certains Eucaryotes [Bjarni V HALLDORSSON et al. 2016; Y. LESECQUE et al. 2013; R. LI et al. 2018; SMEDS et al. 2016; A. L. WILLIAMS et al. 2015], et désormais chez des Bactéries. Cependant, les mécanismes moléculaires qui le causent restent inconnus.

Nos travaux ont montré que la conversion génique est également biaisée vers GC dans trois espèces de Bactéries, ce qui constitue une opportunité unique d'investiguer plus avant sur cette question, du fait que les modèles bactériens sont généralement (beaucoup) plus simples à manipuler d'un point de vue génétique.

Quoiqu'indirectement, les résultats de nos travaux suggèrent qu'au moins deux classes de réparation biaisée sont à l'œuvre. On sait par ailleurs que les mécanismes de correction des mésappariements sont généralement séparés en mécanismes de type *Short Patch Repair* et en mécanismes de type *Long Patch Repair* [JOSEPH et al. 2006].

Le mécanisme prééminent de *Long Patch Repair* est le *mismatch repair*, qui fonctionne *via* l'action conjointe d'homologues de MutS, MutL et parfois MutH [SPIES et FISHEL 2015]. Le MMR intervient notamment, mais pas seulement, dans le contexte de la réparation des cassures double-brin (DSB) par recombinaison homologue, cassures que l'on sait être causées par l'invasion de l'ADN exogène [HÜLTER, SØRUM et al. 2017], ou par d'autres lésions spontanées. La réparation de ces DSB entraîne la formation d'hétéroduplexes, qui peuvent présenter des mésappariements. Ces mésappariements peuvent être corrigés par le MMR, qui entraîne la résection d'un long segment d'ADN suivie d'une re-synthèse utilisant le brin d'ADN intact pour matrice.

Un autre mode de correction des mésappariements est le *base excision repair* (BER), qui utilise des DNA-glycosylases relativement spécifiques de certaines classes de mésappariements [CAI et al. 2010; KISKER et al. 2013; KROKAN et BJØRÅS 2013]. Par exemple, la glycosylase MutY est une adénine DNA-glycosylase, qui excise la base A dans les mésappariements qui l'impliquent, laissant un site abasique, réparé ensuite par resynthèse via la DNA-polymérase I et ligature. Plusieurs éléments conduisent à penser que le BER intervient également dans la correction des mésappariements dans le contexte spécifique des hétéroduplexes de recombinaison [GU et al. 2002; KULICK et al. 2008; SCHANZ et al. 2009].

Les deux mécanismes génèrent des conversions géniques, en ce que l'information génétique portée par le brin matrice de la resynthèse est transmise au brin dégradé de façon unidirectionnelle. Il est possible que les deux classes de produits de conversion que nous avons observées soient dues à ces deux modes de réparation des mésappariements.

### *Hypothèses quant au mécanisme responsable de la gBGC dans les conversions isolées*

Chez *Acinetobacter baylyi*, le gBGC est observé parmi les produits de conversion courts, portant sur un seul marqueur isolé : les *single*

*marker conversions*. Ces conversions isolées sont plus probablement le fait de mécanismes de type BER que de type MMR, pour essentiellement plusieurs raisons que je détaille ici : la première est liée à la taille des produits de conversion, la deuxième au fait que l'inactivation du BER réduit le nombre de restaurations, la troisième au spectre de reconnaissance et réparation des pré-mutations que décrivent les expériences d'accumulation de mutation.

En quoi la taille des produits de conversion suggère-t-elle que c'est plus probablement le BER que le MMR qui est responsable des SMC ? Pour commencer, il est probable que ce genre de conversion isolées soit le fait d'un mécanisme de type *Short Patch Repair* : il n'est pas attendu du MMR qu'il génère des conversions isolées, étant donnée la taille des résections qu'il implique. Il se pourrait cependant que la résection consécutive au MMR soit plus courte que la distance moyenne entre deux marqueurs. Ceci est très peu probable. D'abord parce que la longueur moyenne de cette résection est de l'ordre de  $> 1$  kbp : chez la Levure, la résection dépendante de l'exonucléase ExoI s'étend sur 7,6 kbp dans les crossovers [YIN et PETES 2014]. Ensuite parce que la densité en marqueur chez *A. baylyi* est telle que 50 % des distances entre marqueurs sont inférieures à 17 bp. Enfin parce que nous avons montré que les conversions isolées ne se produisent pas préférentiellement dans les rares régions où la densité en marqueur est plus faible — ce qui aurait pu rendre la résection indétectable. Il est donc plus vraisemblable que ce soit le BER qui occasionne ces conversions isolées.

Un deuxième indice qui pointe vers le BER tient au fait que les restaurations isolées flanquées par deux marqueurs convertis montrent le même spectre de conversion biaisé vers GC et biaisé vers les transversions. Ce qui suggère que ces restaurations sont dues aux mêmes processus que les conversions isolées. Or, la sur-expression réalisée par KULICK et al. [2008] d'homologues de MutY chez *H. pylori* conduit à une augmentation du nombre d'alternance de matrice dans les régions converties, c'est-à-dire d'entrecouplements par des séquences du receveur (voir en particulier la figure 4 de KULICK et al. 2008 pour un modèle hypothétique explicite de l'entrecouplement dépendant de MutY). C'est-à-dire que la sur-expression de MutY dans cette espèce conduit à une augmentation du nombre de restaurations. Ainsi, MutY est un acteur clé du BER d'une part ; il contribue à augmenter le nombre de restauration dans l'espèce *H. pylori* d'autre part ; les restaurations dans l'espèce *A. baylyi* ont le même patron de restauration que les conversions isolées finalement.

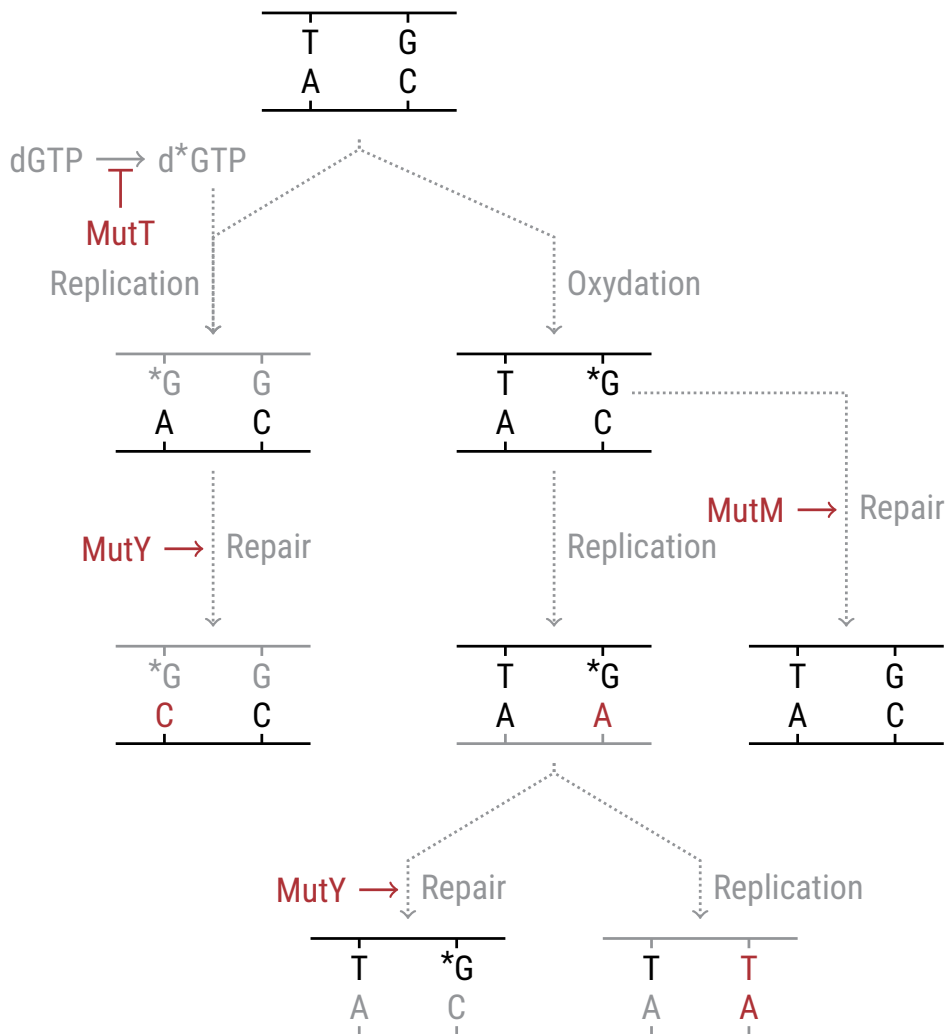
Un troisième indice qui suggère que le BER est responsable des SMC tient au fait que les expériences d'accumulation de mutation réalisées

## 9. DISCUSSIONS

chez différentes espèces ont montré que l'inactivation du MMR conduit à une augmentation drastique du nombre de *transitions* AT→GC [DETTMAN et al. 2016; FOSTER, NICCUM et al. 2018]. Autrement dit, le MMR lorsqu'il fonctionne tend à réduire le nombre de transitions AT→GC, d'une façon ou d'une autre. Or les conversions de type SMC chez *A. baylyi* sont à la fois biaisées en faveur des transversions, et biaisées en faveur des conversions AT→GC, ce qui suggère qu'il ne s'agit probablement pas d'un mécanisme de type MMR.

Un autre élément en ce sens est lié à ce que l'on sait du spectre de reconnaissance et de réparation des pré-mutations par le BER. En effet, la voie du BER traite en particulier des mésappariements liés à l'oxydation spontanée des guanines en 8-oxo-guanine [KROKAN et BJØRÅS 2013]. Elle peut se produire soit dans le pool de dNTP, soit dans les guanines intégrées dans la double-hélice (figure 9.3). L'oxy-

**FIGURE 9.3** – Modèle de réparation et mutagenèse dépendante de MutY. Au cours de la réplication, les 8-oxo-dGTP (d\*GTP) sont incorporés aussi bien en complément d'un A que d'un C. Si le site G·A est corrigé par MutY, la base A est excisée et remplacée par un C, ce qui donne lieu à une transversion A·T→C·G. L'oxydation peut aussi se produire directement dans la double-hélice. L'appariement \*G·C peut être corrigé par MutM, auquel cas il n'y a pas de mutation. Sinon la réplication peut incorporer un A en complément de la base \*G. Ce mésappariement \*G·A peut alors être réparé par MutY; s'il ne l'est pas la réplication subséquente peut conduire à une transversion G·C→T·A. (Adapté de KROKAN et BJØRÅS 2013.)





dation des guanines dans la double hélice conduit à une paire de base 8-oxo-guanine (\*G) et cytosine. La protéine MutM a pour rôle d'exciser la guanine oxydée, auquel cas elle est corrigée par une guanine, et la réparation est non-mutagène [MICHAELS et al. 1992]. Au cours de la réplication cependant, les polymérases intègrent, en complément d'une guanine oxydée, aussi bien une adénine qu'une cytosine. La fonction première de MutY dans le BER est de repérer et réparer ces mésappariements A·\*G en excisant l'adénine. Si cette réparation n'est pas effectuée, cela donne lieu à une transversion G·C→T·A à la prochaine réplication.

La fonction première de MutY étant d'exciser l'adénine impliquée dans un mésappariement A·\*G, elle est susceptible de conduire à une transversion lorsque la base insérée par erreur est la base guanine oxydée.<sup>51</sup>

Nous pensons que MutY est un bon candidat qui pourrait expliquer le patron de conversion observé dans les conversions de type SMC chez *A. baylyi*. En effet, des expériences d'accumulation de mutation, notamment dans l'espèce *E. coli*, ont montré que l'inactivation de gènes du BER conduit à une augmentation du nombre de *transversions* GC→AT [FOSTER, H. LEE et al. 2015]. Par exemple, l'inactivation de MutY conduit à une multiplication par 443 du nombre de transversions G·C →T·A; l'inactivation conjointe de MutY et MutM à une multiplication par 4223 de ce même nombre. Autrement dit, le BER dépendant de MutY ou MutM lorsqu'il fonctionne tend à réduire le nombre de transversions G·C →T·A. Ce qui indique que la fonction primaire de MutY et de MutM est de contrebalancer les effets de l'oxydation des guanines.

À ce titre, lorsqu'une base \*G est incorporée « par erreur » en complément d'une adénine, MutY excise la base adénine, ce qui donne lieu à une transversion A·T→C·G. Par conséquent, si MutY contribue à diminuer le nombre de transversions G·C→T·A, il le peut aussi à augmenter celui des transversions A·T→C·G.

De plus, si MutY a une affinité forte pour les mésappariements A·\*G, elle reconnaît aussi d'autres classes de mésappariement, et notamment les mésappariements A·G et A·C, dans lesquels c'est toujours la base adénine qui est excisée [PORELLO et al. 1998]. Or l'excision d'une adénine dans un mésappariement A·C conduit à une transition AT→GC. En ce sens, KIM et al. [2003] montrent d'une part que MutY

51. Les homologues de MutT ont d'ailleurs pour fonction de « nettoyer » le *pool* de dNTP des 8-oxo-dGTP, pour réduire la fréquence à laquelle elles sont insérées par erreur en complément d'une adénine. On voit donc que les effets des homologues de MutT et de MutY sont

antagonistes à l'égard du taux de transversion G·C vers T·A : MutY diminue le nombre de transversions G·C →T·A, et MutT celui des transversions T·A →G·C [COUCE et TENAILLON 2019].

opère également sur les mésappariements A·C, d'autre part que le taux de transitions A·T → G·C chez *E. coli* est 30 fois moindre lorsque MutY est inactivé.

Or, dans l'espèce *A. baylyi*, les conversions de type SMC et les restaurations conduisent préférentiellement à des transversions qu'à des transitions, ce qui correspond bien à l'affinité préférentielle de MutY pour les mésappariements A·G, dont la réparation conduit à des transversions. De plus, les transversions en question sont biaisées en faveur des transversions A·T → C·G, comme l'est le spectre de réparation de MutY. Enfin, parmi ces conversions, celles qui conduisent à des transitions sont également biaisées en faveur des transitions AT → GC, ce qui correspond tout autant au patron de réparation de MutY aux mésappariements A·C.

Par conséquent, chez *Acinetobacter baylyi*, il est vraisemblable que le patron de conversions biaisées dans les conversions de type SMC et dans les restaurations de marqueurs isolés résulte de l'intervention du BER dans le contexte spécifique de la correction des mésappariements qui surviennent dans les hétéroduplexes, et en particulier de l'activité de MutY [BARBE et al. 2004]. Le patron de reconnaissance et de réparation des mésappariements qui résultent de l'oxydation des guanines par cette protéine correspond bien au patron de *conversion* isolées et des restaurations dans l'espèce *A. baylyi*.

### *Hypothèse quant au mécanisme responsable du biais en faveur de GC dans les conversions aux extrémités des régions converties*

Venons-en maintenant au cas des extrémités des régions converties biaisées vers les bases G et C chez *S. pneumoniae* et *H. pylori*.

Une première hypothèse serait que l'initiation de la résection par le MMR est biaisée vers GC chez *S. pneumoniae* et *H. pylori*. Le système bien décrit de MMR chez *Escherichia coli* repose sur le fait que la protéine MutH excise le brin émergent non-méthylé au site spécifique GA<sup>me</sup>TC, et initie la résection de brin. Mais *S. pneumoniae*, comme la plupart des Bactéries à gram positif, possède un système d'initiation de la résection de brin indépendant de MutH [LENHART et al. 2016]. L'activité d'excision y est accomplie par l'enzyme MutL. Cependant, les mécanisme précis qui déclenchent l'excision sont encore débattus dans ces modèles. De plus, *H. pylori* est bien connu pour ne pas présenter le mécanisme canonique de MMR dépendant de MutLSH, puisqu'il lui manque à la fois l'homologue de MutL et l'homologue de MutH, encore qu'il se puisse que le paralogue MutS2 joue le rôle de MutH [DORER et al. 2011].

Autrement dit, dans l'espèce *E. coli* où l'on connaît le mieux l'initiation de la résection, elle est non-biaisée puisqu'elle est capable de discriminer le brin donneur du brin receveur. Mais pas grand chose n'est su de cette initiation chez *S. pneumoniae*; et chez *H. pylori*, on sait au moins que le modèle classique de type MutSLH ne s'applique pas.

Une hypothèse alternative, qui me semble plus crédible, est là encore liée à l'action du BER : Il se pourrait que les extrémités des régions de conversions multiples soient biaisées vers GC chez *H. pylori* et *S. pneumoniae* en raison d'une action du BER restreinte aux extrémités des régions converties.

1. Là encore, un mutant  $\Delta$ -MutY chez *S. pneumoniae* a montré un phénotype dit « hyper-recombinant » [SAMRAKANDI et PASTA 2000], et un mutant  $\Delta$ -MutY chez *H. pylori* présente des longueurs de régions converties plus longues que les souches sauvages [KULICK et al. 2008]. Ceci suggère que la correction par excision de base joue un rôle inhibiteur sur l'extension de la région convertie. Par exemple, la liaison de MutY à un mésappariement G·A pourrait encombrer la double hélice et empêcher la résection dépendante de UvrD de procéder plus avant; de sorte que la conversion à l'extrémité de la région convertie ne dépend finalement pas du MMR.
2. Différents travaux suggèrent qu'il y a une forme de « compétition » entre MutS et MutY/MutM pour la liaison aux mésappariements A·\*G. ZHAO et WINKLER [2000] ont montré que la sur-expression de MutS réduisait le nombre de transversions GC→TA dans un mutant  $\Delta$ -MutY chez *Escherichia coli*. KIM et al. [2003] ont rapporté que la sur-expression de MutS réduit le taux de transitions AT→GC lorsque MutY est inactivé, ce qui indique qu'une partie des mésappariements AT·CG est pris en charge par MutY.
3. D'autres travaux suggèrent également qu'il y a une forme de « coopération » entre le MMR et le *short patch repair* dans la correction des mésappariements des hétéroduplexes [POLOSINA et CUPPLES 2010]. En ce sens, BHAGWAT et LIEB [2002] démontrent que le MMR recrute probablement la protéine Vsr, impliquée dans la correction des mésappariements T·G résultants de la déamination spontanée des 5-méthylcytosine en thymine. Cette protéine corrige le mésappariement T·G en G·C. Or l'inactivation de MutS ou MutL réduit significativement le taux de réparation par Vsr, ce qui démontre que les deux voies agissent de concert pour la réparation de la déamination spontanée des cytosines. La protéine Vsr n'est pas impliquée dans le BER. Néanmoins, ceci démontre

*L'article de Polosina et Cupples s'est vu affublé d'un titre qui mériterait tout de même qu'on le mentionne, ne serait-ce que pour rendre justice à un très bel effort d'originalité : « Wot the 'L—Does MutL Do? »*

que le *mismatch repair* peut recruter d'autres systèmes de réparation, et n'est pas rigoureusement autonome dans la correction des mésappariements.

De plus, SCHANZ et al. [2009] ont également démontré dans des lignées cellulaires humaines que le MMR interagissait avec le BER dans un contexte spécifique. En effet, sous certaines conditions, l'activation de cytidine déaminase peut conduire à un accroissement local du nombre de mésappariements G·U (la déamination d'une cytosine non-méthylée conduit à une base uracile). Ces mésappariements G·U sont classiquement pris en charge par des uracile DNA-glycosylase (*ung* chez *E. coli*). Mais SCHANZ et al. [2009] démontrent que les « protéines du MMR activées par un mésappariement voisin interfèrent avec le traitement des uraciles [par le BER] pour générer de longs *gaps* simple-brins ». De plus, KULICK et al. [2008] indiquent que l'activité de MutY semble s'exprimer spécifiquement aux extrémités des régions converties chez *H. pylori*. On pourrait faire l'hypothèse que le même type de phénomène se produit dans le contexte des mésappariements liés aux hétéroduplexes.

En ce sens, il a été démontré d'abord chez l'Homme [GU et al. 2002; PARKER et al. 2001] puis chez *E. coli* [BAI et LU 2007] que MutS interagit étroitement avec MutY. Cette interaction est à la fois physique, comme le prouvent les expériences de co-immunoprécipitation réalisées, et fonctionnelle. En effet, la multiplication par 8 de la molarité de MutS multiplie par 8 l'activité de liaison de MutY aux mésappariements A·\*G. De plus, la fréquence de mutations AT→GC est deux à trois fois inférieure dans un double mutant *mutS*<sup>-</sup>-*mutY*<sup>-</sup> par rapport à un simple mutant *mutS*<sup>-</sup>.

- chez la Levure, Y. LESECQUE et al. [2013] ont observé également que la conversion génique biaisée vers GC est restreinte aux sites situés aux extrémités des produits de recombinaison de type crossover. Ils ont proposé l'hypothèse que ce soit l'introduction de *nick*, de cassures simple-brins qui soit responsable du biais observé. En effet, la reconnaissance d'un mésappariement par le MMR implique la recherche d'un *nick* à partir duquel opérer la résection de l'un des deux brins. Si le *nick* en question se produit plus fréquemment à proximité d'un marqueur GC que d'un marqueur AT, alors le marqueur GC se retrouve plus fréquemment à l'extrémité de la région convertie.

On pourrait proposer l'hypothèse alternative, en s'appuyant sur les points précédents, que c'est le BER (ou Vsr) qui introduit le *nick*, et qu'il tend à le faire préférentiellement en excisant les bases A

ou T, de sorte que le marqueur à l'extrémité de la région convertie se trouve être un G ou C.

Pour résumer, 1) MutY semble agir spécifiquement aux extrémités des régions converties chez *H. pylori*, 2) il existe peut-être une forme de compétition pour la liaison du MMR et du BER aux mésappariements reconnus par le BER mais surtout 3) MutS et MutY interagissent physiquement et l'affinité de MutY pour son substrat A.\*G dépend du MMR. Enfin, cette hypothèse pourrait expliquer une observation précédente chez la Levure d'un biais vers GC cantonné aux extrémités des régions converties.

Cette hypothèse butte néanmoins sur l'écueil que, si c'est le BER qui est responsable du biais aux extrémités des régions converties, alors on aurait pu s'attendre à ce que les marqueurs isolés qui résultent probablement tout autant de l'action du BER chez *S. pneumoniae* et *H. pylori* soient biaisés vers GC également. Or ça n'est pas le cas. Il se pourrait soit que la conversion des marqueurs isolés soit non-biaisée parce qu'ils découlent d'autres mécanismes que le BER, soit que les extrémités des régions converties soient biaisées vers GC en raison d'une activité du BER spécifique lorsqu'il est en interaction avec le MMR.

Tous ces indices semblent pointer vers la conclusion que le BER est sans doute un acteur moléculaire déterminant dans l'intensité du biais en faveur de G et C dans les corrections isolées, ou aux extrémités des régions converties. Une brève requête dans la base de donnée Swissprot des DNA-glycosylases bactériennes annotées indique qu'il n'y a ni de thymine ni de cytosine DNA glycosylase à proprement parler qui soit recensée chez les Bactéries (table 9.2). Ce qui semble suggérer que

Cible	Exemples	Nombre
Adénine	MutY	> 203
Uracile	UDG, MUG, UNG	> 370
Formamidopyrimidine	MutM	> 447
8-oxoguanine	?	6
Cytosine	?	0

TABLE 9.2 – Nombre de DNA-glycosylases bactériennes annotées dans Swissprot.

considéré globalement, le BER est un mode de réparation de l'ADN intrinsèquement biaisé vers GC.

(Il faut noter à ce propos l'observation intéressante de ASLAM et al. [2019], qui constatent que les Procaryotes strictement aérobies ont un taux de GC aux positions quatre fois dégénérées des codons significativement supérieur à celui des anaérobies, même en contrastant pour l'inertie phylogénétique. Les auteurs notent qu'il se pourrait que les

dommages oxydatifs les plus fréquents dans les organismes aérobies soient davantage liés à l'oxydation des guanines au stade dGTP, dont l'incorporation par erreur en complément d'une adénine conduit à des transversions A·T→C·G, qu'à l'oxydation des guanines dans la double hélice, en complément desquelles l'incorporation par erreur d'adénine conduit à des transversions G·C→T·A.

Une hypothèse alternative à cela serait que, si le dommage oxydatif le plus fréquent est bien lié à l'incorporation par erreur d'adénine en complément de guanine oxydée *dans* la double-hélice, la protéine MutY joue un rôle crucial pour contrebalancer ce genre de dommages. Un effet secondaire à cela serait que l'efficacité avec laquelle MutY répare les mésappariements A·G liés à l'oxydation des guanines conduirait à ce qu'elle le fasse également avec une grande efficacité dans le contexte des hétéroduplexes. Par conséquent, le coefficient  $b_0$  pourrait être supérieur dans les espèces les plus soumises aux dommages occasionnés par l'oxydation des guanines, ce qui pourrait expliquer pourquoi les anaérobies ont un taux de GC inférieur. Cette hypothèse pourrait être testée par exemple comparant l'intensité du biais de fixation en faveur des bases GC entre anaérobies et aérobies.)

### 9.3 Détecter la gBGC : une tâche délicate

Le fait que la gBGC, dans les modèles étudiés ici, est cantonnée à des types de conversions spécifiques rend sa détection particulièrement difficile à partir de données de polymorphisme.

En effet, la plupart des programmes employés pour détecter des régions recombinantes reposent sur l'augmentation locale du nombre d'homoplasies dans un alignement concaténé du *core-genome*, ou de la perturbation de la topologie de l'arbre que ces homoplasies génèrent [BOBAY et OCHMAN 2017; CROUCHER, PAGE et al. 2015; DIDELOT et WILSON 2015; MOSTOWY et al. 2017]. Cette augmentation est typiquement le signe d'un transfert de gène par recombinaison homologue entre deux souches de la même espèce.

Par exemple, le programme Gubbins [CROUCHER, PAGE et al. 2015] requiert qu'au moins trois substitutions corroborent un événement de recombinaison. BOBAY et OCHMAN [2017] spécifient dans la même veine qu'ils n'ont décidé de considérer que les événements impliquant le transfert de deux au moins sites polymorphes. Par conséquent, ces méthodes ne sont pas tout à fait optimales pour la détection d'événements de conversion très courts, comme le rapportent COWLEY et al. [2018], ce qui pourrait expliquer que certains auteurs n'aient pas détecté de gBGC



associé à la recombinaison dans l'analyse du polymorphisme [BOBAY et OCHMAN 2017; YAHARA et al. 2016].

BOBAY et OCHMAN [2017] par exemple ont avancé qu'il n'y a pas de preuves de gBGC chez les Bactéries, sur la base de ce que les bases détectées comme « recombinantes », c'est-à-dire correspondant à « des échanges de nucléotides introduits par la recombinaison », quoi que cela puisse être, « conduisent à l'enrichissement en A et T dans la plupart des Bactéries ». Il me semble qu'il y a à cela une explication très simple : c'est une erreur conceptuelle de discriminer entre des allèles qui résultent de *mutations*, et des allèles qui résultent d'*échanges de gènes*. Que peut bien être un allèle résultant d'un échange de gène ? L'échange de gène ne fait rien d'autre que littéralement *re-combiner*, c'est-à-dire créer de nouvelles combinaisons d'allèles qui découlent de mutations. Aussi est-il attendu d'emblée que le spectre de « changements de nucléotide » qui découlerait de pareils « échanges de gènes » soit identique à celui du spectre mutationnel.

Si bien que je trouve l'article de BOBAY et OCHMAN [2017] tout à fait peu probant à l'égard du rejet de l'hypothèse gBGC. D'abord parce que les événements de conversion courts sont indétectables avec leur méthode, ensuite parce que l'argument repose sur des bases logiques pour le moins floues. Enfin, quand bien même on admettrait le raisonnement suivi dans cet article, on s'apercevrait (en particulier dans les figures 2 et 4A de l'article) que les « changements de nucléotides » causés par l'échange de gène sont moins enrichis en GC→AT que ceux causés par des mutations ; or c'est précisément ce qui attendu si la gBGC augmentait la probabilité de fixation des allèles G et C dans les régions recombinantes.

## 9.4 La nature convergente de la gBGC

Quels que soient les mécanismes sous-jacents, il semble qu'il y ait une forme de convergence dans la réparation de l'ADN en faveur de l'introduction de bases G et C.

En ce sens, la gBGC a été détectée dans diverses espèces qui ont été étudiées expérimentalement (voir un contre-exemple dans H. LIU et al. 2017). Cependant, les mécanismes en jeu et leur intensité sont généralement différents d'une espèce à l'autre. En effet, la gBGC se manifeste par des patrons différents : chez la Souris et *Acinetobacter baylyi*, seules les conversions isolées sont affectées [R. LI et al. 2018] ; chez *Helicobacter pylori*, *Streptococcus pneumoniae* et la Levure [Y. LESECQUE et al. 2013], seules les extrémités des MMC le sont ; chez l'Homme, à la fois les SMC et les MMC sont affectées [Bjarni V HALLDORSSON et al. 2016].

## 9. DISCUSSIONS

La gBGC se manifeste également dans des contextes différents : chez la Levure, seuls les produits de recombinaison résolus en crossover sont affectés [Y. LESECQUE et al. 2013]; chez la Souris, ce sont seulement les produits de recombinaison résolus en non-crossover [R. LI et al. 2018]. Enfin, l'intensité à laquelle la gBGC se manifeste varie selon les espèces étudiées : l'intensité du biais de transmission  $b_0$  est faible chez la Levure (0,013), mais de l'ordre de 0,48 chez *Acinetobacter baylyi* (table 9.3).

TABLE 9.3 – Intensité du coefficient  $b_0$  selon les espèces.

Espèce	Intensité de $b_0$	Référence
Levure	0,013	Y. LESECQUE et al. 2013
Gobe-Mouche	0,18	SMEDS et al. 2016
Homme	0,3	Bjarni V HALLDORSSON et al. 2016
<i>Acinetobacter baylyi</i>	0,48	

Nos travaux ont montré que certaines Bactéries montrent également des preuves de biais en faveur des bases G et C au cours de la conversion génique, et que, en théorie, étant donnée une gamme de paramètres plausible, son intensité pourrait suffire à contrecarrer celle du biais en faveur de A et T dans le spectre mutationnel.

Comme c'est le cas chez les Eucaryotes, les mécanismes en jeu semblent diverger entre espèces. Ceci indique que les mécanismes de conversions évoluent, mais qu'ils le font d'une façon contrainte par leur capacité à contrer l'influence de la mutation à l'égard de la composition en base. Sans cela, la mutation, sans rien pour l'entraver, conduirait à une réduction drastique du contenu en GC.

En fait, ce scénario pourrait expliquer pourquoi les Bactéries endosymbiotiques tendent à avoir des génomes très enrichis en A et T [MCCUTCHEON et MORAN 2012; MORAN 1996], puisqu'ils ont généralement perdu la machinerie de réparation de l'ADN, et que leur taille efficace se situe dans la partie basse du spectre

L'on pourrait spéculer plus avant, et proposer que les espèces bactériennes avec les tailles efficaces  $N_e$  les plus grandes tendent à être riches en GC [BOBAY et OCHMAN 2018], à cause du terme  $2N_e$  dans l'effet à l'échelle populationnelle de la gBGC B; de sorte que même un biais de transmission  $b$  extrêmement faible (la probabilité qu'un polymorphisme AT : GC soit converti en GC) pourrait engendrer une intensité à l'échelle populationnelle de la gBGC très élevée.

Les contraintes sélectives qui pèsent sur les mécanismes de réparation constitueraient de la sorte un cas de sélection du second ordre [HERSHBERG 2015], au travers duquel la sélection opère moins sur des substitutions individuelles uniques, perdues dans un vaste génome, que sur les mécanismes qui globalement conduisent à un équilibre évolutivement stable du nombre de substitutions AT→GC et GC→AT. En



vérité, au vu de l'intensité de B que représenteraient chez certaines Bactéries les valeurs de b que j'ai mesurées chez *Acinetobacter baylyi* (figure 8.15), si sélection il doit y avoir, je me demande si elle ne pèserait pas davantage sur les mécanismes moléculaires qui génèrent la gBGC, de façon à contraindre B dans une gamme d'intensité soutenable, que sur la composition en base elle-même.

Une façon d'aborder la question consisterait à voir les entités génétiques responsables de la mutagenèse d'une part, celles responsables de la gBGC d'autre part, comme ayant des effets antagonistes sur le « phénotype génomique », dont une composante majeure est sa composition en base (figure 9.4). Ce phénotype est susceptible d'être soumis

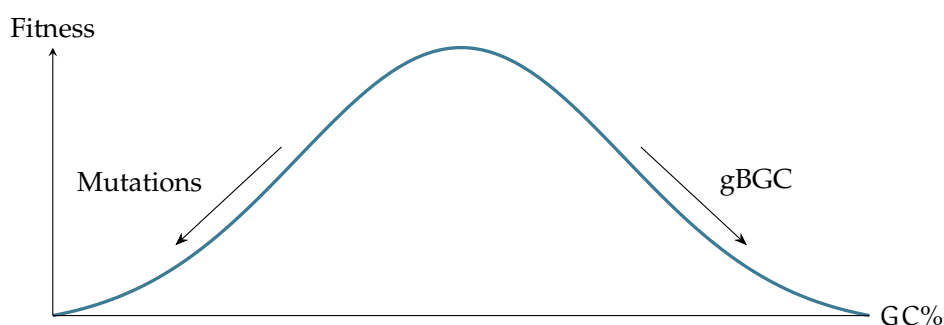


FIGURE 9.4 – Effets antagonistes de la gBGC et de la mutation sur le taux de GC d'un génome.

à l'action de la sélection naturelle. Si c'est le cas, il se pourrait que la sélection opère sur les mécanismes mutateurs et sur les mécanismes à l'origine de la gBGC de façon à minimiser leurs effets, qui tendent respectivement à *diminuer* et à *augmenter* le taux de GC.

Qu'il faille ou non invoquer la sélection naturelle dans ce genre de question est une aporie, par définition, difficile à résoudre, s'il nous faut éviter le piège des « contes *ad-hoc* », ce que Gould et Lewontin appellent les « "just-so"-stories », et le paradigme Panglossien de l'adaptationnisme [GOULD et R. C. LEWONTIN 1979]. Or étant données les conséquences délétères de l'oxydation des guanines sur le taux de mutation, la gBGC est peut-être un effet secondaire de pressions de sélection pesant sur l'efficacité des mécanismes de correction de l'oxydation des guanines, conduisant à ce que dans le contexte spécifique des hétéroduplexes, les adénines soient excisées par erreur, et que les bases G et C se trouvent être plus fréquemment donneuses qu'attendu. En fin de compte, il se pourrait bien que la gBGC ne soit qu'un pendentif des arches bâties sur la réparation de l'ADN, dans l'enchevêtrement alambiqué d'une cathédrale génomique.

Récemment, SCOTT et QUELLER [2019] ont donné aux conséquences des conflits évolutifs à long terme le nom de « course à l'armement Sysipheenne » (*Sysiphean arms race*). Or en moyenne, l'altitude du rocher

## 9. DISCUSSIONS

---

de Sysiphe correspond à la moitié de la hauteur de sa colline. Si la mutation est à la composition en base ce que la gravité est au rocher, la gBGC est peut-être à l'égard de celle-ci ce que Sysiphe est à celui-là. Plutôt qu'une course à l'armement Sysiphéenne, c'est d'un équilibre Sysiphéen qu'il faudrait parler dans ce cas.

D'aucuns pourraient voir d'un mauvais œil le genre de conclusions auxquelles conduit le raisonnement que l'on vient de tenir : certains traits des êtres vivants s'éloignent d'une hypothétique valeur idéale imperfectible. Mais, pour reprendre une phrase de Foucault déjà citée, « à la limite, la vie [...] c'est ce qui est capable d'erreur ». « Il faut imaginer Sysiphe heureux. »<sup>52</sup>

52. CAMUS 1985.

## PERSPECTIVES ET CONCLUSIONS GÉNÉRALES

---

Ce ne sont pas les faits mais une vision du monde en jeu, une divergence entre ceux qui, d'un côté, voient les processus dynamiques dans les populations comme essentiellement conservatifs, purificateurs et protégeant un *status quo* adapté et rationnel des forces non-adaptatives, corruptrices et irrationnelles de la mutation aléatoire, et ceux, de l'autre côté, pour qui la nature est un processus, et pour qui tout ordre existant ne peut qu'être instable à la longue, voyant comme Denis Diderot que « Tout change, tout passe, il n'y a que le tout qui reste. »

Richard C. LEWONTIN [1974]. *The Genetic Basis of Evolutionary Change*. Columbia University Press. 372 p., p. 157

Elles [les espèces] n'auroient dû leur première origine qu'à quelques productions fortuites, dans lesquelles les parties élémentaires n'auroient pas retenu l'ordre qu'elles tenoient dans les animaux peres & meres : chaque degré d'erreur auroit fait une nouvelle espece. [...] Des moyens différents des moyens ordinaires que la Nature emploie pour la production des animaux, loin d'être des objections contre ce système, lui sont indifférents, ou lui seroient plutôt favorables.

MAUPERTUIS, Oeuvres, 1756. Cité par Motoo KIMURA [déc. 1957]. « Some Problems of Stochastic Processes in Genetics ».

In : *Ann. Math. Statist.* 28.4, p. 882-901. DOI :

[10.1214/aoms/1177706791](https://doi.org/10.1214/aoms/1177706791).

Bien loin d'avancer que la conversion génique biaisée vers GC est l'unique force sous-tendant l'hétérogénéité de la composition en base du génome des Procaryotes, ce travail indique qu'elle pourrait servir de « force » explicative potentielle qu'il convient de prendre en compte, parmi d'autres forces [LASSALLE et al. 2015]. Je ne m'attends pas vraiment à ce que la gBGC contribue à une quelconque grandiose théorie unifiée de la composition en base, mais bien plutôt que des causes écologiques et génétiques, où « génétique » doit s'entendre en son sens historique et mécanistique, s'entremêlent de façons qui dépendent tellement du contexte qu'elles finissent par former de complexes entrelacs causaux (*a tangled bank*) qu'on ne peut vraiment démêler que par des travaux expérimentaux détaillés et orientés par la théorie.

Comme le disait un célèbre évolutionniste :

Notre ignorance en ce qui concerne les lois de la variation est bien profonde. Nous ne pouvons pas, une fois sur cent, prétendre indiquer les raisons d'une variation donnée. Cependant, toutes les fois que nous pouvons réunir les termes d'une comparaison, nous remarquons que les mêmes lois semblent avoir agi pour produire les petites différences qui existent entre les variétés d'une même espèce, et les grandes différences qui existent entre les espèces d'un même genre. Le changement des conditions ne produit généralement qu'une variabilité flottante, mais quelquefois aussi des effets directs et définis ; or, ces effets peuvent à la longue devenir très prononcés, bien que nous ne puissions rien affirmer, n'ayant pas de preuves suffisantes sur ce point.<sup>53</sup>

Une compréhension détaillée des mécanismes moléculaires qui sous-tendent la conversion génique biaisée vers GC nécessiterait une dissection détaillée des produits de recombinaison et des différents mécanismes en jeu. Cette caractérisation mériterait d'aller au delà de simple KO de gènes. Cette tâche sera d'autant plus à portée si elle est accomplie avec des modèles bactériens. En particulier, chez *A. baileyi* est disponible une collection de mutants KO pour chacun des gènes non-essentiels, qui pourrait servir de base pour une collection de mutants MMR/BER et de leurs différentes combinaisons [H. H. WANG et al. 2009].

Ce travail a montré expérimentalement que la conversion génique peut contribuer à expliquer la variabilité du taux de GC chez certaines Bactéries, ce qui incite maintenant à porter l'attention sur les Archées. Si la gBGC s'avérait se produire également chez les Archées, cela contribuerait à prouver qu'il s'agit d'un phénomène très répandu, un indice de plus qui inciterait à ce qu'on la considère correctement dans les modèles évolutifs [BORGES et al. 2019].

De plus, toutes les espèces utilisées ici ont la particularité de présenter un génome de composition en base moyenne autour de 40%. Il serait particulièrement intéressant de quantifier les fréquences de conversion dans des espèces à taux de GC plus élevé, et qui soient relativement faciles à manipuler au laboratoire, comme par exemple les Actinobactéries du genre *Streptomyces* [NAH et al. 2017]. Spécifiquement, les genres *Agrobacterium* et *Burkholderia* sont parmi ceux pour lesquels l'écart entre la composition attendue à l'équilibre mutationnel et le taux de GC observé aux positions quatre fois dégénérées est le plus grand. Si la gBGC permet de compenser la réduction en taux de GC occasionnée par la mutation, l'on devrait s'attendre à ce que son intensité soit la

53. DARWIN 1876, page 178.

---

plus forte dans ces espèces là. Ces deux genres comprennent aussi des espèces qui sont (relativement) faciles à manipuler du point de vue de leur biologie moléculaire. Cela comblerait les lacunes de la couverture phylogénétique que nous avons dans ce travail, et augmenterait d'autant la vraisemblance de l'hypothèse gBGC.

## Conclusions

L'une des grandes questions à laquelle s'affronte la Biologie depuis qu'elle a été unifiée par l'hypothèse de sélection naturelle et la théorie de l'évolution, consiste à essayer de comprendre ce qui, dans les organismes, contribue à leur *fitness*, et de quelle façon [GAYON 2019].

Pendant longtemps, on a pensé que la variabilité de la composition en base des génomes n'était le fait que de patrons de mutations différents entre espèces. Mais certains ont avancé que ce trait pouvait peut-être, après tout, être soumis en lui-même à des contraintes sélectives, c'est-à-dire qu'il contribue à la *fitness* de l'organisme. Selon les écoles de pensée, cette idée peut paraître tout à la fois parfaitement intuitive et complètement saugrenue. Il me semble que c'est justement là ce qui fait son intérêt ; car c'est probablement dans ce genre d'opposition des plus tranchées que le front de la science progresse [BACHELARD 1966].

Et précisément l'hypothèse de la conversion génique biaisée vers GC permet de nuancer cette opposition, et l'ouvrir vers une complexité plus subtile, et à mon sens plus intéressante que s'il s'agissait d'une vaine opposition entre hasard et nécessité. Dans la formulation de cette hypothèse, dans tous les travaux qui l'ont confortée, et plus forte raison encore au cours de mon travail de thèse, j'ai vu s'épanouir cet élan de la pensée scientifique que BACHELARD qualifie de philosophie du non :

Il nous faudra rappeler que la philosophie du non n'est pas psychologiquement un négativisme et qu'elle ne conduit pas, en face de la nature, à un nihilisme. Elle procède au contraire, en nous et hors de nous, d'une activité constructive. Elle prétend que l'esprit au travail est un facteur d'évolution. Bien penser le réel, c'est profiter de ses ambiguïtés pour modifier et alerter la pensée. Dialectiser la pensée, c'est augmenter la garantie de créer scientifiquement des phénomènes complets, de régénérer toutes les variables dégénérées ou étouffées que la science comme la pensée naïve, avait négligées dans sa première étude.<sup>54</sup>

Ce que BACHELARD appelle « dialectiser la pensée » ici, c'est cet effort du refus des dichotomies (hasard / nécessité, dérive / sélection)

*« Deux hommes, s'ils veulent s'entendre vraiment, ont dû d'abord se contredire. La vérité est fille de la discussion, non pas fille de la sympathie. » Bachelard 1966, page 134.*

54. BACHELARD 1966, page 17.

ou des vaines oppositions entre catégories, qui est au cœur de l'activité scientifique correctement menée. En quoi l'hypothèse gBGC est-elle venue « alerter la pensée » ? En rappelant que ce que l'on pourrait prendre à première vue comme le signe de l'action de la sélection naturelle pourrait bien, tout compte fait, n'être qu'un épiphénomène lié à la correction des mésappariements, un effet secondaire des mécanismes de réparation. Il ne s'agit pas d'une force purement stochastique, encore que ses effets ne se manifestent, comme ceux de la sélection naturelle, qu'à une échelle populationnelle. Il ne s'agit pas non plus d'une force « orientée » par ses effets sur la *fitness* des individus, en tout cas pas directement : à court terme évolutif, elle peut contribuer à la fixation d'allèles délétères.

Mon travail de thèse n'a finalement servi qu'à poser un deuxième jalon, après celui de LASSALLE et al. [2015], sur la longue route qui conduirait à prouver que la gBGC joue un rôle effectif sur la composition en base des Procaryotes. À ce stade, nous n'avons fait que montrer que la conversion génique est bien biaisée vers GC chez certaines Bactéries, et avec une intensité estimée suffisamment forte pour expliquer la différence entre la composition en base à l'équilibre mutationnel et la composition en base observée. Il resterait à prouver que c'est bien ce processus et pas un autre qui est en jeu. Cela nécessiterait que l'on ait une idée plus claire de la génétique des populations Procaryotes, un vaste pan de recherche qui reste mal formalisé, par rapport à celle des Eucaryotes.

Si cela s'avérait *in fine* n'être pas le cas, la démarche n'aura pas été inutile pour la simple raison que l'hypothèse de sélection naturelle n'en sortirait que renforcée : en fin de compte, je trouverais réellement fascinant que la sélection naturelle puisse opérer d'une façon ou d'une autre sur la composition en base, et ne demande qu'à voir émerger les travaux qui contribueraient (sérieusement) à exclure l'action de la gBGC. Si en revanche, comme j'ai pris le parti de le croire, le consensus s'établissait peu à peu autour de la conversion génique biaisée vers GC, ce serait là encore l'occasion pour l'esprit scientifique de revenir à lui-même mieux équipé, muni d'outils conceptuels plus à même de participer à cette « activité constructive » qui consiste à penser le réel biologique.

Darwin, dans *De la Variation*, parle de la sélection naturelle comme de la « force qui l'emporte » (*paramount power*) :

Throughout this chapter and elsewhere I have spoken of selection as the paramount power, yet its action absolutely depends on what we in our ignorance call spontaneous or accidental variability. Let an architect be compelled to build an edifice with uncut stones, fallen from a precipice. The

---

shape of each fragment may be called accidental; yet the shape of each has been determined by the force of gravity, the nature of the rock, and the slope of the precipice,—events and circumstances, all of which depend on natural laws; but there is no relation between these laws and the purpose for which each fragment is used by the builder. In the same manner the variations of each creature are determined by fixed and immutable laws; but these bear no relation to the living structure which is slowly built up through the power of selection, whether this be natural or artificial selection.

If our architect succeeded in rearing a noble edifice, using the rough wedge-shaped fragments for the arches, the longer stones for the lintels, and so forth, we should admire his skill even in a higher degree than if he had used stones shaped for the purpose. So it is with selection, whether applied by man or by nature; for though variability is indispensably necessary, yet, when we look at some highly complex and excellently adapted organism, variability sinks to a quite subordinate position in importance in comparison with selection, in the same manner as the shape of each fragment used by our supposed architect is unimportant in comparison with his skills.<sup>55</sup>

Au delà de cette belle métaphore de l'architecte,<sup>56</sup> la remarque de Darwin selon laquelle les variations entre individus sont déterminées par des lois fixes et immuables, mais qui n'ont aucune relation avec la structure vivante modelée peu à peu par la force de la sélection, est tout à fait pertinente ici. D'abord parce qu'il n'est plus du tout sûr que les lois en questions soient fixes et immuables — la gBGC par exemple semble avoir évolué de façon convergente chez différentes espèces. Ensuite parce que, concernant la composition en base, il se pourrait très bien que ce trait n'ait aucune relation avec la *fitness* de l'individu, et ne soit finalement qu'une autre pierre non taillée, sculptée par sa chute dans le précipice au gré des mutations et des conversions, avec laquelle la sélection (l'architecte) doit malgré tout composer.

Finalement, Darwin a probablement raison : je crois que j'admirerais d'autant plus la sélection naturelle d'être parvenue à composer cette infinie mélodie qu'est le vivant s'il s'avérait que les instruments dont elle disposait n'étaient *pas* fabriqués par elle et pour elle.

55. Charles DARWIN [1868]. *The Variation of Animals and Plants under Domestication*. 1st. T. 2. 2 t. London : John Murray, pages 248–249, cité par GAYON 2019, page 75

56. On peut raisonnablement se demander si Darwin n'a pas volontairement usé de

cette métaphore dans *De la Variation*, paru dans sa première édition 11 ans après de *l'Origine des Espèces*, pour parer aux assauts des théologiens en brouillant les pistes.



Il me faut maintenant conclure sur mon expérience de thèse, d'un point de vue plus général. À certains, j'ai déjà dit n'avoir vraiment eu que trois questions en tête au cours de ce travail :

1. Quels sont les processus qui gouvernent l'évolution de la composition en base des Procaryotes ?
2. Quelles sont les conséquences de la recombinaison sur l'évolution des génomes ?
3. Quelles sont les raisons pour lesquelles je trouve les deux questions précédentes intéressantes ?

Dans ce manuscrit, j'ai tenté d'apporter, sinon une réponse, du moins un éclairage à la première question. En partie pour répondre à la troisième question, j'ai entrepris d'en fouiller sommairement l'histoire, pour mieux voir sortir du terreau pré-scientifique le germe théorique autour duquel s'est articulé l'opposition entre neutralistes et adaptationnistes. Les *progrès* technologiques — et il faut bien parler de progrès ici, puisqu'il n'y probablement pas beaucoup d'autres domaines où la notion de progrès est plus prégnante qu'en science — les progrès technologiques donc ont permis d'explorer toujours plus avant la diversité du vivant. De sorte qu'il rentre sans cesse plus de lumière par les fenêtres de cette obscure demeure qu'est le vivant, à l'obscurité de laquelle Chargaff disait s'être progressivement déshabitué. Plus la diversité du vivant se fait jour, plus clairement il apparaît que celle de la composition en base entre génomes des organismes est d'abord le fait de patrons de mutation qui varient d'un organisme à l'autre. Subsistent néanmoins des zones d'ombre qu'il convient d'expliquer. En effet, chez les Procaryotes la plupart des génomes ne sont pas à l'équilibre du point de vue de leur composition en base. Si certains ont proposé d'explorer l'hypothèse sélectionniste, nous avons pris le parti d'une hypothèse neutraliste et avons longuement argumenté en sa faveur. Il faut bien l'admettre, je l'ai fait probablement sans faire justice à l'ensemble des hypothèses sélectionnistes ; probablement aussi sans avoir fait preuve de l'objectivité la plus complète concernant lesdites hypothèses. Mais après tout en science comme ailleurs, il est vain de vouloir sans cesse retenir son jugement. « A vrai dire, disait encore Bachelard, en n'importe quel temps, on a des difficultés à s'installer dans une position purement rationaliste. <sup>57</sup> »

L'hypothèse neutraliste aux côtés de laquelle nous nous sommes rangés défend que la composition en base des génomes n'est pas directement liée à la *fitness* des organismes, et dérive au gré des patrons

57. BACHELARD 1972.



---

de mutations, auxquels il faut rajouter au moins un processus non-adaptatif : la conversion génique biaisée vers GC. Nous avons contribué à montrer que ce processus est probablement à l'œuvre chez certains Procaryotes, qu'il est susceptible d'y jouer un rôle important, et avons tenté d'en disséquer, autant que possible, les mécanismes moléculaires. Pour cela il a fallu en passer par des approches diverses, allant de la biologie moléculaire à la microbiologie expérimentale, de la bioinformatique à la génétique des populations, ou de la génétique moléculaire à la génomique comparative. Probablement que d'autres l'auraient fait plus vite, mieux, différemment, de façon plus « productive » ; dans mon cas c'est en résistant aux injonctions à la publication à tout crin, ou en refusant plus ou moins sciemment de diffuser des travaux qui me semblaient incomplets que j'ai pu aboutir à une contribution que j'estime mûre désormais.

(Le plus pernicieux dans le système scientifique actuel, c'est qu'il n'a pas besoin de prendre la forme d'une direction de thèse dictatoriale pour peser fortement sur le comportement de publication des doctorantes et des doctorants ; dans mon cas précis je suis à peu près certain d'avoir eu la chance de faire l'expérience d'une direction de thèse irréprochable à tout point de vue, et à ce point de vue en particulier. J'ignore si les trois personnes qui m'ont encadré pendant cette thèse pensent qu'ils s'agissait d'une erreur de ma part, et que j'aurais dû « publier plus » ; mais c'est tout le point, je l'ignore. Si elles le pensaient, je n'en ai rien su, et elles m'ont laissé l'opportunité d'errer. De ce que j'en ai vu par ailleurs, cette opportunité est de plus en plus précieuse.)

Qu'en retenir de cette contribution ?

La première leçon qu'il me semble falloir en retirer, pour sortir du cadre strictement scientifique, est qu'une bonne partie des états de choses dans le monde ne sont pas telles pour les *raisons* que l'on croit. C'est le travail de la pensée scientifique de mettre au jour ces raisons cachées, ces « variables dégénérées » oubliées en première étude.

La seconde leçon, pour moi la plus importante, pourrait se résumer par : « le hasard est réel ». Le hasard agit sur le vivant et sur le monde, et n'est pas seulement le nom que l'on donne à « l'ignorance où nous sommes des véritables causes » [LAPLACE 1825].

J'ai toujours trouvé tout à fait singulier que, lorsqu'on demande à un évolutionniste, Spencérien ou Darwinien, ou quelle que soit l'école à laquelle il appartient, quelles sont les entités (*agencies*) qui ont conduit à l'évolution, il fasse mention de divers faits déterminés et de lois, mais pas une fois ne mentionne la *Chance*.<sup>58</sup>

58. Design and Chance, PEIRCE 1992a, page 219.

Cela peut paraître bien banal à n'importe quelle personne de science digne de ce nom désormais ; mais il faut se rappeler qu'il y a là entre elle et le reste de la population (occidentale du moins) un point de départ important.

Pour répondre à la question de PEIRCE [1992a], le grand philosophe américain de la fin du XVIII<sup>e</sup> déjà posée au début de ce manuscrit, « à quoi sert de tenter de former une conception de l'univers sinon de rendre les choses intelligibles ? » Personnellement, je trouve cet univers nettement moins inintelligible après ces trois ans passés à tenter de comprendre que la contingence pure *explique* déjà beaucoup.

À la deuxième question posée plus haut — quelles sont les conséquences de la recombinaison sur l'évolution des génomes ? —, j'ai aussi en partie contribué à répondre dans ce travail de thèse, en montrant que la recombinaison était susceptible d'influencer la composition en base des Procaryotes. Ça n'est toutefois pas là que s'arrête cette vaste question, ni pour la science, ni pour moi, puisqu'il est prévu que j'y consacre les trois prochaines années de ma vie scientifique.

À la troisième question — quelles sont les raisons pour lesquelles je trouve les deux questions précédentes intéressantes —, je crois qu'il faudrait pour y répondre se lancer dans une vaste entreprise de socio-analyse qui n'aurait d'intérêt pour pas beaucoup d'autres que le principal concerné. J'ai pour autant mis très longtemps à me convaincre qu'il y avait un sens à ce que je faisais ; et tout aussi longtemps à comprendre que rares étaient ceux qui ne posaient pas les mêmes questions. J'ai lu les sceptiques, les relativistes, les rationalistes, les pragmatistes et autres empiristes, les philosophes, les sociologues, les logiciens et les historiens, pour tenter de comprendre si tout cela n'était pas complètement vain.

Si la science n'était qu'un agréable passe-temps, un jeu pour les oisifs, un ornement de luxe, une fantaisie d'amateur, la *moins vaine des vanités* en un mot, il y aurait des jours où le savant devrait dire avec le poète :

« Honte à qui peut chanter, pendant que Rome brûle. »

Mais si la science est une chose sérieuse, si les destinées de l'humanité et la perfection de l'individu y sont attachées, si elle est une religion, elle a, comme les choses religieuses, une valeur de tous les jours et de tous les instants. Ne donner à l'étude et à la culture intellectuelle que les moments de calme et de loisir, c'est faire injure à l'esprit humain, c'est supposer qu'il y a quelque chose de plus important que la recherche de la vérité. Or s'il en était ainsi, si la science ne constituait

---

qu'un intérêt de second ordre, l'homme qui a voué sa vie au parfait, qui veut pouvoir dire à ses derniers instants : J'ai accompli ma fin, devrait-il y consacrer une heure, quand il saurait que des devoirs plus élevés le réclament ?<sup>59</sup>

Rome brûle à grands feux, et pourtant je suis maintenant aussi fermement convaincu qu'on puisse l'être qu'il n'y a pas beaucoup d'espoir à chercher ailleurs que dans la diffusion la plus large possible des valeurs scientifiques. Si la science n'est pas si innocente de l'état actuel du monde qu'elle voudrait bien l'admettre, il reste qu'il *faut* distinguer les valeurs scientifiques de ce qui a été fait sous leur couvert. Pour citer Bertrand Russell :

Il y a une joie pure dans la connaissance intrépide de notre véritable place dans le monde, et un drame plus profond que n'importe quel problème posé par ceux qui se cachent derrière les murs clos d'un mythe. Il existe des « mers périlleuses » dans l'univers de la pensée, où ne peuvent s'embarquer que ceux qui sont prêts à affronter leur propre impuissance physique. Et, ce qui compte plus que tout, on se libère ainsi de la tyrannie de la Peur qui souille la lumière du jour et rend les hommes vils et cruels. Aucun homme n'est libéré de la peur s'il n'ose concevoir sa place dans le monde telle qu'elle est en réalité, aucun homme ne peut accomplir les grandes choses dont il est capable sans reconnaître sa petitesse.<sup>60</sup>

J'ai eu la chance de pouvoir m'embarquer sur ces mers périlleuses, et doit bien admettre que jamais je n'avais fait d'expérience plus libératrice. Si je ne suis pas tout à fait certain de concevoir ma place dans le monde telle qu'elle est en réalité, j'y travaille. À cela je n'ai pas trouvé de meilleures méthodes que celles de la science.

59. RENAN 1890, page 32.

60. RUSSELL 2011.

Je concluerai avec ces aphorismes de Wittgenstein :

6.37 — Rien ne contraint quelque chose à arriver du fait qu'autre chose soit arrivé. Il n'est de nécessité que *logique*.

6.371 — Toute la vision moderne du monde repose sur l'illusion que les prétendues lois de la nature sont des explications des phénomènes de la nature.

6.372 — Aussi se tiennent-ils devant les lois de la nature comme devant quelque chose d'intouchable, comme les Anciens devant Dieu et le Destin. Et les uns et les autres ont en effet raison et tort. Cependant les Anciens ont assurément une idée plus claire en ce qu'ils reconnaissent une limitation, tandis que dans le système nouveau il doit sembler que *tout* est expliqué.

6.373 — Le monde est indépendant de ma volonté.<sup>61</sup>

61. Ludwig WITTGENSTEIN [1972]. *Tractatus Logico-Philosophicus*. Avec la coll. de Bertrand RUSSELL. Trad. par

Gilles G. GRANGER. 2001<sup>re</sup> éd. Tel 311. Gallimard. 128 p.

It is the same! — For, be it joy or sorrow,  
The path of its departure still is free :  
Man's yesterday may ne'er be like his morrow ;  
Nought may endure but mutability !

Percy B. Shelley, *Mutability*, 1816



# Appendices





## SÉLECTION SUR LA COMPOSITION EN BASE ET TAUX DE CROISSANCE

---

Il est bien remarquable de trouver, à la base d'un des phénomènes d'adaptation moléculaire les plus exquisement précis qu'on connaisse, une source au hasard.

Jacques MONOD [23 oct. 2014]. *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*. Points. 256 p.

Dans cette annexe, je décris une expérience que nous avons réalisée pour tenter de déterminer s'il existe, dans le milieu de culture LB que nous avons utilisé pour nos expériences de transformation chez *A. baylyi*, une forme de sélection en faveur des allèles G et C.

Pour cela nous avons construit par transformation naturelle des souches d'*A. baylyi* qui possède des versions d'un gène plus ou moins enrichies en GC, à la façon de RAGHAVAN et al. [2012].

Pour les construire, nous avons cloné dans le plasmide pGEM-T la séquence du génome d'*A. baylyi* au même locus décrit page 103 (région annotée « ancre » dans la figure 1). Nous avons ensuite accolé à cette séquence le gène de résistance à la kanamycine aphA-3, ainsi que deux versions de la séquence en amont contenant 23 marqueurs séparés par environ 30 bp. La première version de cette séquence, la version S,

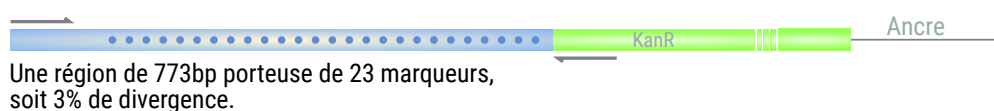


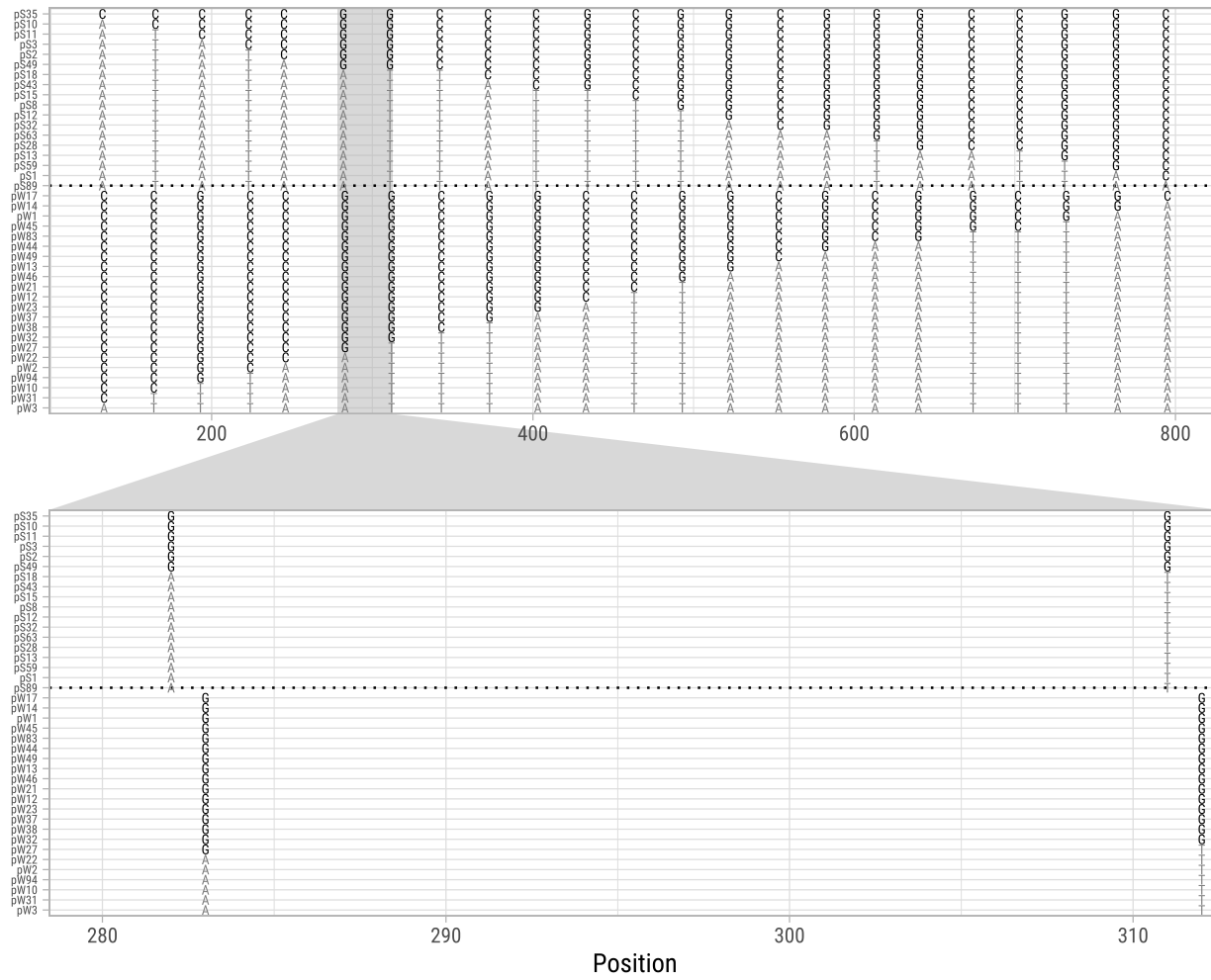
FIGURE 1 –  
Construction utilisée pour le clonage

introduit 23 marqueurs C ou G à des positions qui sont A ou T dans le génome de la souche ADP1. La seconde version de cette séquence, la version W, introduit 23 marqueurs A ou T à des positions qui sont 1) C ou G dans le génome de la souche ADP1, et 2) immédiatement voisines des marqueurs de la version S.

Nous avons ensuite transformé la souche ADP1 en l'exposant séparément aux deux versions du plasmide donneur W et S. Après quoi nous avons sélectionné les transformants par culture en présence de Kanamycine. Nous avons ensuite amplifié par PCR la séquence du gène dans lequel nous avons introduit des marqueurs pour 96 colonies isolées, que nous avons séquencé par la technique de Sanger.

Cela nous a permis d'obtenir 91 clones transformés par le plasmide S, ayant intégré de 0 à 23 marqueurs C ou G; et 92 clones transformés par le plasmide W, ayant intégré de 0 à 23 marqueurs A ou T. Parmi ceux-là,

A



B

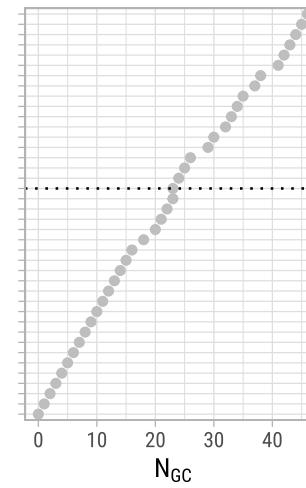


FIGURE 2 – Génotype des souches dont la croissance a été étudiée. (Description page suivante.)

FIGURE 2 – **A.** Alignement de la séquence des clones transformés à ce locus, classés de haut en bas par nombre de marqueurs GC intégrés décroissants. La partie basse de la figure illustre que les marqueurs AT intégrés le sont à une base de décalage des positions où les marqueurs GC sont intégrés. La ligne verticale sépare en haut les clones ayant intégré un nombre croissant de marqueurs GC, en bas ceux ayant intégré un nombre croissant de marqueurs AT. **B.** Nombre de marqueurs GC.

nous en avons sélectionné 23 transformants en fonction du nombre de marqueurs qu'ils avaient intégrés pour les deux versions du plasmide (figure 2).

Aux deux extrêmes, il y a le clone pW3 pour lequel les 23 marqueurs C ou G ont été convertis en A ou T, et le clone pS35, pour lequel les 23 marqueurs A ou T ont été convertis en C ou G. De sorte que pW3 ne possède aucun marqueur GC, et que pS35 en possède 46.

Chacun de ces clones a ensuite été cultivé pendant une nuit en milieu LB liquide, puis ajusté à une absorbance à 600 nm de 0,1 et relancé dans 10 ml de milieu LB neuf. Toutes les heures pendant 6 h, l'absorbance à 600 nm de la culture a été mesurée (figure 3). Les mesures ont été effectuées manuellement, en suivant l'ordre croissant du nombre de marqueurs GC. On peut estimer à > 10 min l'intervalle entre la dernière et première mesure d'absorbance.

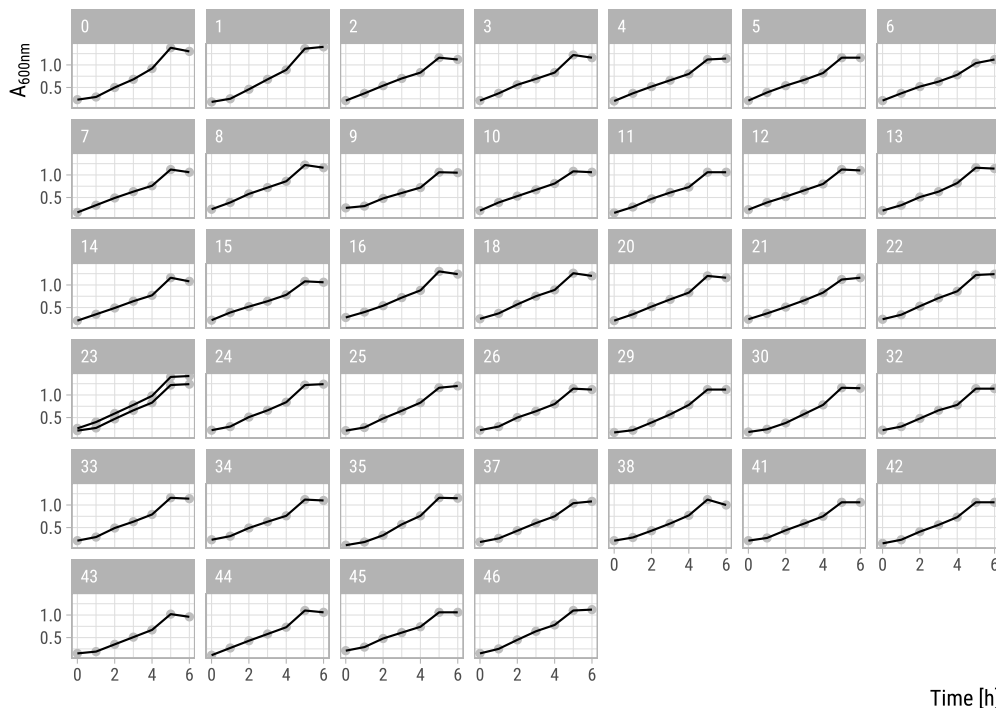


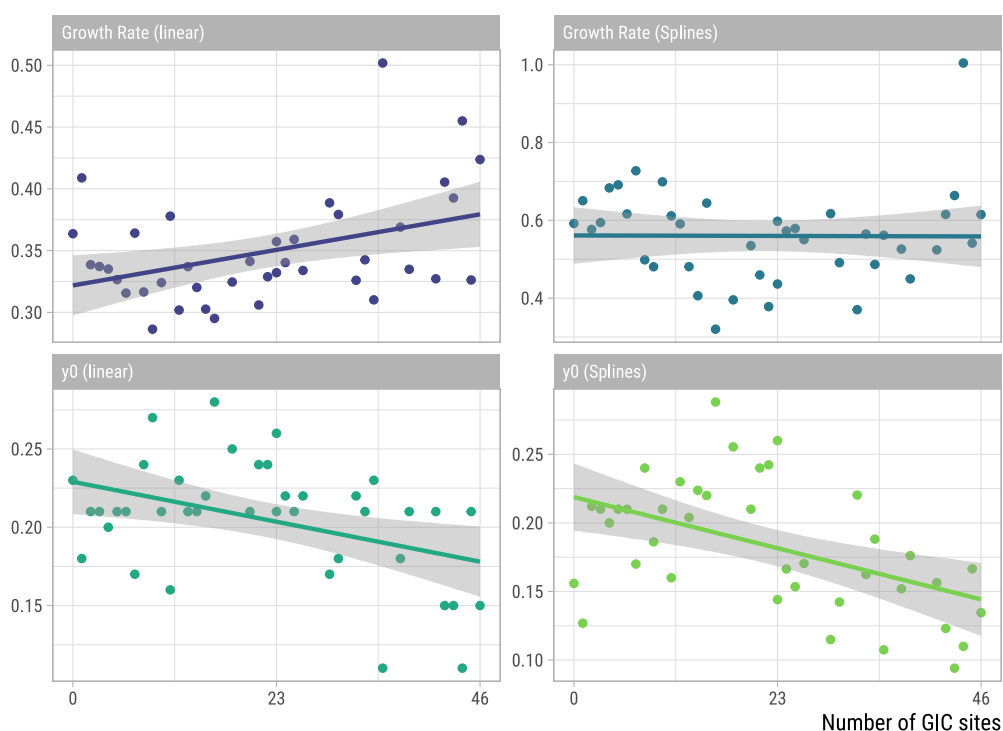
FIGURE 3 – Augmentation de l'absorbance en fonction du temps et du nombre de marqueurs GC. Chaque panneau correspond à un clone ayant intégré N marqueurs GC, N étant indiqué en haut du panneau.

J'ai ensuite mesuré le taux de croissance pour chacun de ces clones

en utilisant deux modèles, un modèle de croissance linéaire dans l'intervalle temporel où la croissance est maximale, et un modèle non-paramétrique [HALL et al. 2014] implémenté dans le package R PETZOLDT [2018]. Le taux de croissance d'une culture estimé par un modèle linéaire peut être résumé par les coefficients de régression de la droite au cours de la phase de croissance exponentielle, c'est-à-dire  $\mu_{\max}$  et  $y_0$ .

La figure 4 décrit la relation entre d'une part ces deux paramètres  $\mu_{\max}$  et  $y_0$ , et d'autre part le nombre de sites C et G intégrés par les transformants. On voit qu'un modèle linéaire indique une corrélation positive

FIGURE 4 – Relation entre les coefficients de croissance  $\mu_{\max}$  et  $y_0$  et la quantité de marqueurs GC convertis.



significative entre le taux de croissance et le nombre de marqueurs GC. En moyenne, l'acquisition d'un marqueur GC en plus augmente le taux de croissance de  $1,3 \cdot 10^{-3}$  ( $p = 1,1 \cdot 10^{-2}$ ). Ce qui pourrait suggérer qu'il y a une forme de sélection en faveur de l'intégration de davantage de marqueurs GC.

Ce serait cependant une affirmation mal soutenue statistiquement, pour les raisons suivantes :

1. D'une part parce que le modèle linéaire en question n'explique que 13,7 % de la variance totale ( $R^2 = 0,137$ ).
2. Ensuite le taux de croissance estimé avec un autre modèle de type non-paramétrique n'est *pas* significativement corrélé avec le nombre de marqueurs GC ( $p = 0,2$ ,  $R^2 = 4,5 \cdot 10^{-2}$ .)

Pour vérifier que cette relation ne soit pas un artéfact lié à l'expérimentation, nous avons répété l'expérience en mesurant la croissance en deux microplaques 100 puits, en utilisant cinq réplicats par clone transformant et en mesurant l'absorbance à 600 nm via un appareil à lecture d'absorbance automatisé de type BioScreen. Nous avons prêté un soin particulier à manipuler les différents clones de façon « aléatoire », par exemple en répartissant aléatoirement les différents clones sur les deux plaques (voir la répartition du nombre de sites C et G intégrés sur les deux plaques en figure 5).

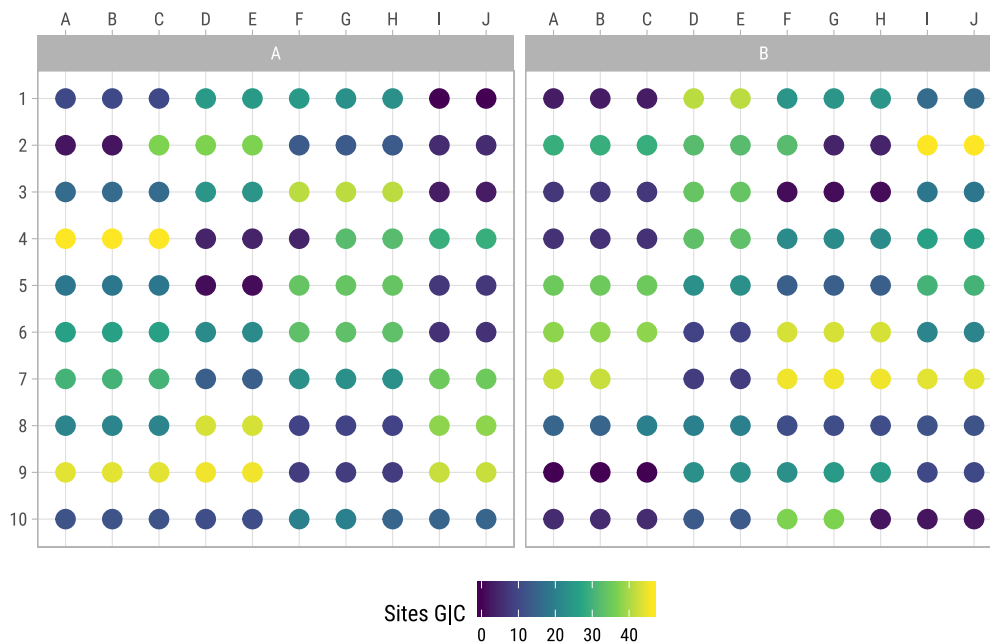


FIGURE 5 – Répartition des clones sur les deux plaques de culture en fonction du nombre de sites C et G intégrés. Cette répartition a été pensée comme un compromis entre la minimisation des effets artéfactuels et celle du taux d'erreur lié à la faisabilité de l'expérience.

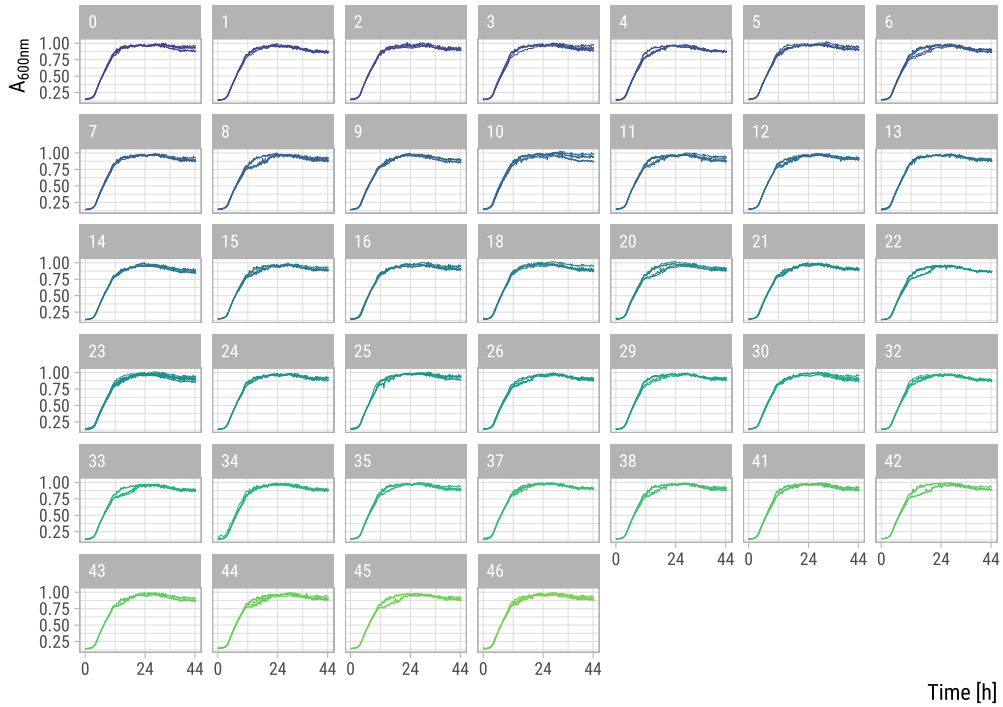
Les deux plaques de cultures ont été incubées pendant 44 h à 30 °C, et l'absorbance à 600 nm mesurée dans chaque puits toutes les 20 min. Nous avons obtenu les courbes de croissance représentée en figure 6.

J'ai ensuite calculé le taux de croissance maximal par réplicat et par transformant, et l'ai représenté en fonction du nombre de marqueurs C et G (figure 7).

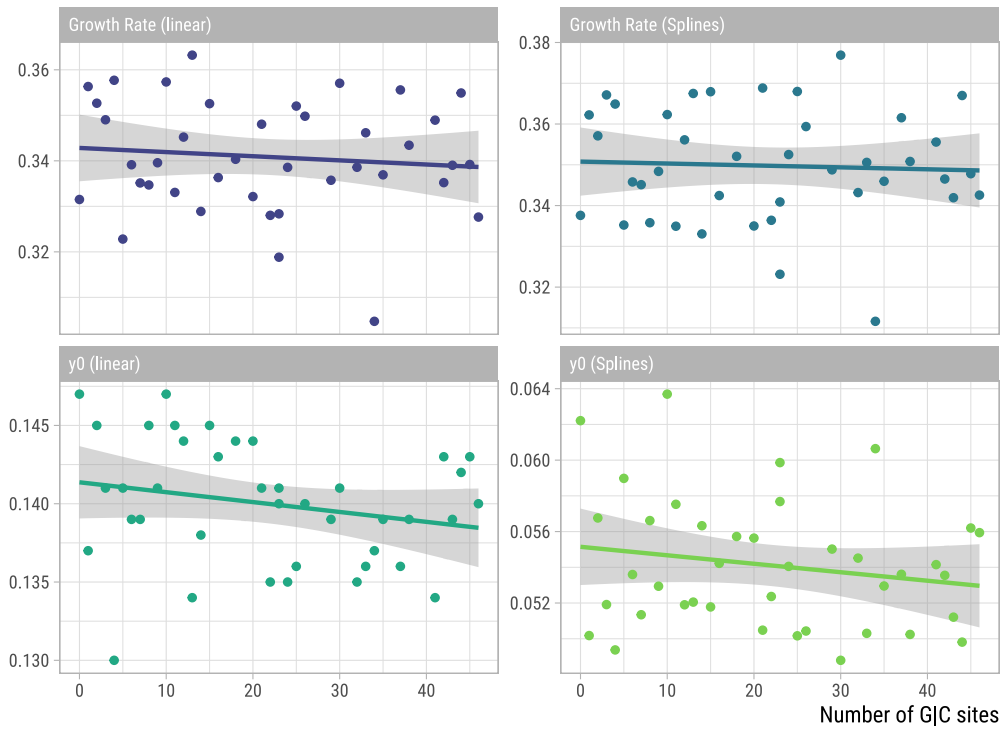
Il n'y a pas de corrélation significative entre le taux de croissance et le nombre de marqueurs C et G, quel que soit le modèle de croissance utilisé pour calculer le taux de croissance ( $p = 0,52$  et  $R^2 = 0.01$  pour un modèle linéaire,  $p = 0.77$  et  $R^2 = 0.52$  pour un modèle de croissance non-paramétrique).

TAUX DE CROISSANCE ET TAUX DE GC%

**FIGURE 6 –**  
Augmentation de l'absorbance à 600 nm en fonction du temps de culture et du nombre de marqueurs C et G. (Les conventions sont les mêmes que pour la figure 3.)



**FIGURE 7 –** Relation entre le taux de croissance et le nombre de marqueurs C et G.



---

## Discussion & conclusion

Cette brève expérience avait pour but de vérifier si les conclusions de RAGHAVAN et al. [2012] et KELKAR et al. [2015] concernant la relation entre taux de croissance et taux de GC d'un gène s'appliquaient à notre modèle *Acinetobacter baylyi*. Pour rappel, ces deux travaux mentionnent une relation linéaire positive significative entre le taux de croissance et la composition en base d'un gène codant la GFP, intégré sur un plasmide multi-copies. L'hypothèse proposée par eux était que la sélection naturelle opérant à l'échelle de la composition en base globale, sur le « phénotype génomique », pouvait contribuer à accroître le taux de GC d'un génome parce que les individus qui possèdent une version plus riche en GC d'un même gène croissent plus rapidement.

Nous nous sommes trouvés en mesure de répliquer une expérience du même type, en utilisant cette fois les versions de composition en base croissante d'un gène chromosomique. Il faut noter toutefois que l'on ne sait rien par ailleurs du niveau d'expression de ce locus. Il est accolé à un marqueur de résistance à la kanamycine, qui dans le contexte de cette expérience précise n'a pas été contraint de s'exprimer, puisque les cultures se déroulaient en milieu non-sélectif.

Dans un premier temps, nous avons bien observé une relation entre le taux de croissance et la composition en base. En effectuant les mesures d'absorbance manuellement, une culture après l'autre, dans un ordre pratique qui a conduit à ce que les mesures des cultures de clones enrichis en G et C soient effectuées les dernières, il est compréhensible que ces dernières cultures aient eu un temps de développement plus long à chaque mesure (environ 10 min pour 40 mesures d'absorbance).

Nous avons répété l'expérience en la « randomisant », de façon à pallier au biais qui vient d'être exposé. Cette fois aucune corrélation entre composition en base et taux de croissance n'a été observée.

Il n'est pas improbable qu'un biais du même genre se soit appliqué aux cultures réalisées par RAGHAVAN et al. [2012]. Il n'est pas fait mention d'une « randomisation » quelconque dans l'article. Or il n'est pas tout à fait évident, à première vue, qu'il faille s'imposer de remplir une microplaque de 100 puits dans le désordre.

Au vu des arguments déjà exposés plus haut et des résultats de cette expérience, il me semble que, quelle que soit l'origine de cet écart en taux de croissance entre les souches qui possèdent une version enrichie en GC de la GFP et les souches qui en possèdent une version enrichie en AT, elle est plutôt à chercher du côté des artéfacts expérimentaux que du côté de la sélection naturelle.





## BIBLIOGRAPHIE

---

- ABIR-AM, Pnina (1980). « From Biochemistry To Molecular Biology : DNA And The Acculturated Journey Of The Critic Of Science Erwin Chargaff ». In : *Hist. Philos. Life Sci.* 2.1, p. 3-60. JSTOR : [23328242](#).
- (1<sup>er</sup> mar. 1985). « Themes, Genres and Orders of Legitimation in the Consolidation of New Scientific Disciplines : Deconstructing the Historiography of Molecular Biology ». In : *Hist Sci* 23.1, p. 73-117. DOI : [10.1177/007327538502300103](#).
- ACOSTA, Sharlene et al. (9 jan. 2015). « DNA Repair Is Associated with Information Content in Bacteria, Archaea, and DNA Viruses ». In : *J Hered* 106.5, p. 644-659. DOI : [10.1093/jhered/esv055](#). pmid : [26320243](#).
- AGASHE, Deepa et Nachiket SHANKAR (nov. 2014). « The Evolution of Bacterial DNA Base Composition : Evolution of Bacterial GC-Content ». In : *J. Exp. Zool. B Mol. Dev. Evol.* 322.7, p. 517-528. DOI : [10.1002/jev.b.22565](#).
- ALFSNES, Kristian et al. (2018). « A Genomic View of Experimental Intraspecies and Interspecies Transformation of a Rifampicin-Resistance Allele into *Neisseria Meningitidis* ». In : *Microb. Genomics*. DOI : [10.1099/mgen.0.000222](#).
- ARNDT, Peter F., Terence HWA et Dmitri A. PETROV (2005). « Substantial Regional Variation in Substitution Rates in the Human Genome : Importance of GC Content, Gene Density, and Telomere-Specific Effects ». In : *J. Mol. Evol.* 60.6, p. 748-763.
- ASLAM, Sidra et al. (28 jan. 2019). « Aerobic Prokaryotes Do Not Have Higher GC Contents than Anaerobic Prokaryotes, but Obligate Aerobic Prokaryotes Have ». In : *BMC Evolutionary Biology* 19.1, p. 35. DOI : [10.1186/s12862-019-1365-8](#).
- AVERY, Oswald T., Colin M. MACLEOD et Maclyn MCCARTY (1<sup>er</sup> fév. 1944). « Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type III ». In : *J. Exp. Med.* 79.2, p. 137-158. DOI : [10.1084/jem.79.2.137](#). pmid : [19871359](#).
- BACHELARD, Gaston (1966). *La Philosophie Du Non*. Presses universitaires de France Paris.
- (1972). *L'Engagement Rationaliste*. Avec la coll. de Georges CANGUILHEM. 1<sup>re</sup> éd. Bibliothèque de Philosophie Contemporaine. Paris : Presses Universitaires de France.
- (1990). *Le matérialisme rationnel*. Paris : Presses universitaires de France. 236 p.
- BACKSTRÖM, Niclas et al. (4 jan. 2010). « The Recombination Landscape of the Zebra Finch *Taeniopygia Guttata* Genome ». In : *Genome Res.* 20.4, p. 485-495. DOI : [10.1101/gr.101410.109](#). pmid : [20357052](#).
- BAI, Haibo et A.-Lien LU (1<sup>er</sup> fév. 2007). « Physical and Functional Interactions between *Escherichia Coli* MutY Glycosylase and Mismatch Repair Protein MutS ». In : *J. Bacteriol.* 189.3, p. 902-910. DOI : [10.1128/JB.01513-06](#). pmid : [17114250](#).
- BALBÁS, Paulina et Francisco BOLÍVAR (2004). « Back to Basics : pBR322 and Protein Expression Systems in *E. Coli* ». In : *Recombinant Gene Expression : Reviews and Protocols*. Sous la dir. de Paulina BALBÁS et Argelia LORENCE. Methods in Molecular Biology. Totowa, NJ : Humana Press, p. 77-90. DOI : [10.1385/1-59259-774-2:077](#).
- BARBE, Valérie et al. (1<sup>er</sup> oct. 2004). « Unique Features Revealed by the Genome Sequence of *Acinetobacter* Sp. ADP1, a Versatile and Naturally Transformation Competent Bacterium ». In : *Nucleic Acids Res* 32.19, p. 5766-5779. DOI : [10.1093/nar/gkh910](#).
- BARRETO, Samuel (24 sept. 2019a). *Acinetobacter VCF File*. DOI : [10.6084/m9.figshare.9896063.v1](#).

## BIBLIOGRAPHIE

---

- BARRETO, Samuel (24 sept. 2019b). *Bacillus Subtilis Genome-Wide Transformation*. DOI : [10.6084/m9.figshare.9896009.v2](https://doi.org/10.6084/m9.figshare.9896009.v2).
- (24 sept. 2019c). *Helicobacter Pylori VCF*. DOI : [10.6084/m9.figshare.5419501.v1](https://doi.org/10.6084/m9.figshare.5419501.v1).
- (24 sept. 2019d). *S.Pneumoniae Recombination Products*. DOI : [10.6084/m9.figshare.6893132.v1](https://doi.org/10.6084/m9.figshare.6893132.v1).
- (p. d.[a]). *Bacterial-Gbgc / Gatk-Genotyper*.
- (p. d.[b]). *Bacterial-Gbgc / Strepto-Cowley*.
- (p. d.[c]). *Bacterial-Gbgc / Transformr*.
- (p. d.[d]). *Files · Dataset · Bacterial-Gbgc / Helicobacter-Pylori-Tracts*.
- BATEMAN, A. J. (1<sup>er</sup> jan. 1959). « The Viability of Near-Normal Irradiated Chromosomes ». In : *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* 1.2, p. 170-180. DOI : [10.1080/09553005914550241](https://doi.org/10.1080/09553005914550241).
- BATUT, Bérénice et al. (déc. 2014). « Reductive Genome Evolution at Both Ends of the Bacterial Population Size Spectrum ». In : *Nat Rev Micro* 12.12, p. 841-850. DOI : [10.1038/nrmicro3331](https://doi.org/10.1038/nrmicro3331).
- BAUDAT, F. et al. (12 fév. 2010). « PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice ». In : *Science* 327.5967, p. 836-840. DOI : [10.1126/science.1183439](https://doi.org/10.1126/science.1183439). pmid : [20044539](https://pubmed.ncbi.nlm.nih.gov/20044539/).
- BELOZERSKY, A. N. et A. S. SPIRIN (juil. 1958). « A Correlation between the Compositions of Deoxyribonucleic and Ribonucleic Acids ». In : *Nature* 182.4628, p. 111-112. DOI : [10.1038/182111a0](https://doi.org/10.1038/182111a0).
- BERGLUND, Jonas, Katherine S POLLARD et Matthew T WEBSTER (27 jan. 2009). « Hotspots of Biased Nucleotide Substitutions in Human Genes ». In : *PLoS Biol* 7.1, e1000026. DOI : [10.1371/journal.pbio.1000026](https://doi.org/10.1371/journal.pbio.1000026).
- BERNARDI, Giorgio (1<sup>er</sup> jan. 1993). « The Vertebrate Genome : Isochores and Evolution. » In : *Mol Biol Evol* 10.1, p. 186-204. DOI : [10.1093/oxfordjournals.molbev.a039994](https://doi.org/10.1093/oxfordjournals.molbev.a039994).
- (4 jan. 2000). « Isochores and the Evolutionary Genomics of Vertebrates ». In : *Gene* 241.1, p. 3-17. DOI : [10.1016/S0378-1119\(99\)00485-0](https://doi.org/10.1016/S0378-1119(99)00485-0).
- BERNARDI, Giorgio et Giacomo BERNARDI (1<sup>er</sup> déc. 1986). « Compositional Constraints and Genome Evolution ». In : *J Mol Evol* 24.1, p. 1-11. DOI : [10.1007/BF02099946](https://doi.org/10.1007/BF02099946).
- BERNARDI, Giorgio, B. OLOFSSON et al. (24 mai 1985). « The Mosaic Genome of Warm-Blooded Vertebrates ». In : *Science* 228.4702, p. 953-958. DOI : [10.1126/science.4001930](https://doi.org/10.1126/science.4001930). pmid : [4001930](https://pubmed.ncbi.nlm.nih.gov/4001930/).
- BGSC (p. d.[a]). *Bacillus Genetics Stock Center : Details for 1A1276*.
- (p. d.[b]). *Bacillus Genetics Stock Center : Details for 2A11T*.
- BHAGWAT, Ashok S., Weilong HAO et al. (23 fév. 2016). « Strand-Biased Cytosine Deamination at the Replication Fork Causes Cytosine to Thymine Mutations in Escherichia Coli ». In : *PNAS* 113.8, p. 2176-2181. DOI : [10.1073/pnas.1522325113](https://doi.org/10.1073/pnas.1522325113). pmid : [26839411](https://pubmed.ncbi.nlm.nih.gov/26839411/).
- BHAGWAT, Ashok S. et Margaret LIEB (2002). « Cooperation and Competition in Mismatch Repair : Very Short-Patch Repair and Methyl-Directed Mismatch Repair in Escherichia Coli ». In : *Mol. Microbiol.* 44.6, p. 1421-1428. DOI : [10.1046/j.1365-2958.2002.02989.x](https://doi.org/10.1046/j.1365-2958.2002.02989.x).
- BIRDSELL, John A. (7 jan. 2002). « Integrating Genomics, Bioinformatics, and Classical Genetics to Study the Effects of Recombination on Genome Evolution ». In : *Mol Biol Evol* 19.7, p. 1181-1197. pmid : [12082137](https://pubmed.ncbi.nlm.nih.gov/12082137/).
- BLOKESCH, Melanie (1<sup>er</sup> août 2017). « In and out—Contribution of Natural Transformation to the Shuffling of Large Genomic Regions ». In : *Current Opinion in Microbiology*. Mobile Genetic Elements and HGT in Prokaryotes \* Microbiota 38, p. 22-29. DOI : [10.1016/j.mib.2017.04.001](https://doi.org/10.1016/j.mib.2017.04.001).

- BLOUNT, Zachary D., Richard E. LENSKI et Jonathan B. LOSOS (9 nov. 2018). « Contingency and Determinism in Evolution : Replaying Life's Tape ». In : *Science* 362.6415, eaam5979. DOI : [10.1126/science.aam5979](https://doi.org/10.1126/science.aam5979). pmid : 30409860.
- BOBAY, Louis-Marie et Howard OCHMAN (1<sup>er</sup> oct. 2017). « Impact of Recombination on the Base Composition of Bacteria and Archaea ». In : *Mol Biol Evol* 34.10, p. 2627-2636. DOI : [10.1093/molbev/msx189](https://doi.org/10.1093/molbev/msx189).
- (12 oct. 2018). « Factors Driving Effective Population Size and Pan-Genome Evolution in Bacteria ». In : *BMC Evolutionary Biology* 18.1, p. 153. DOI : [10.1186/s12862-018-1272-4](https://doi.org/10.1186/s12862-018-1272-4).
- BOHLIN, Jon et al. (2010). « Analysis of Intra-Genomic GC Content Homogeneity within Prokaryotes ». In : *BMC Genomics* 11, p. 464. DOI : [10.1186/1471-2164-11-464](https://doi.org/10.1186/1471-2164-11-464).
- BONHAM, Kevin S., Benjamin E. WOLFE et Rachel J. DUTTON (23 juin 2017). « Extensive Horizontal Gene Transfer in Cheese-Associated Bacteria ». In : *eLife Sciences* 6, e22144. DOI : [10.7554/eLife.22144](https://doi.org/10.7554/eLife.22144).
- BORGES, Rui, Gergely SZÖLLÖSI et Carolin KOSIOL (30 mai 2019). « Quantifying GC-Biased Gene Conversion in Great Ape Genomes Using Polymorphism-Aware Models ». In : *Genetics*, genetics.302074.2019. DOI : [10.1534/genetics.119.302074](https://doi.org/10.1534/genetics.119.302074). pmid : 31147380.
- BOSI, Emanuele et al. (1<sup>er</sup> août 2015). « MeDuSa : A Multi-Draft Based Scaffold ». In : *Bioinformatics* 31.15, p. 2443-2451. DOI : [10.1093/bioinformatics/btv171](https://doi.org/10.1093/bioinformatics/btv171).
- BOUVERESSE, Jacques (4 nov. 2015). « L'éthique de La Croyance et La Question Du « Poids de l'autorité » ». In : *L'éthique de La Croyance et La Question Du « Poids de l'autorité »*. Philosophie de La Connaissance. Paris : Collège de France.
- BROWN, Thomas C. et Josef JIRICNY (11 sept. 1987). « A Specific Mismatch Repair Event Protects Mammalian Cells from Loss of 5-Methylcytosine ». In : *Cell* 50.6, p. 945-950. DOI : [10.1016/0092-8674\(87\)90521-6](https://doi.org/10.1016/0092-8674(87)90521-6). pmid : 3040266.
- BRUEN, T. C. (15 déc. 2005). « A Simple and Robust Statistical Test for Detecting the Presence of Recombination ». In : *Genetics* 172.4, p. 2665-2681. DOI : [10.1534/genetics.105.048975](https://doi.org/10.1534/genetics.105.048975).
- BUBENDORFER, Sebastian et al. (22 juin 2016). « Genome-Wide Analysis of Chromosomal Import Patterns after Natural Transformation of *Helicobacter Pylori* ». In : *Nat Commun* 7, p. 11995. DOI : [10.1038/ncomms11995](https://doi.org/10.1038/ncomms11995).
- BULMER, M. (1<sup>er</sup> nov. 1991). « The Selection-Mutation-Drift Theory of Synonymous Codon Usage. » In : *Genetics* 129.3, p. 897-907. pmid : 1752426.
- CAI, Yuqin et al. (23 août 2010). « Base Sequence Context Effects on Nucleotide Excision Repair ». In : *J Nucleic Acids* 2010. DOI : [10.4061/2010/174252](https://doi.org/10.4061/2010/174252). pmid : 20871811.
- CAMUS, Albert (21 fév. 1985). *Le mythe de Sisyphe*. Paris : Gallimard. 169 p.
- CANGUILHEM, Georges (1981). *Idéologie et rationalité dans l'histoire des sciences de la vie : nouvelles études d'histoire et de philosophie des sciences*. Vrin. 146 p.
- CARR, E. L. (1<sup>er</sup> juil. 2003). « Seven Novel Species of *Acinetobacter* Isolated from Activated Sludge ». In : *Int. J. Syst. Evol. Microbiol.* 53.4, p. 953-963. DOI : [10.1099/ijss.0.02486-0](https://doi.org/10.1099/ijss.0.02486-0).
- CARVALHO, Gabriel et al. (24 fév. 2019). « Bacterial Transformation Buffers Environmental Fluctuations through the Reversible Integration of Mobile Genetic Elements ». In : *bioRxiv*, p. 557462. DOI : [10.1101/557462](https://doi.org/10.1101/557462).
- CHARGAFF, E., R. LIPSHITZ et al. (sept. 1951). « The Composition of the Deoxyribonucleic Acid of Salmon Sperm ». In : *J. Biol. Chem.* 192.1, p. 223-230. pmid : 14917668.

## BIBLIOGRAPHIE

---

- CHARGAFF, E. et E. VISCHER (jan. 1949). « The Composition of the Desoxyribose Nucleic Acids of Thymus and Spleen ». In : *J. Biol. Chem.* 177.1, p. 405-416. pmid : [18107444](#).
- CHARGAFF, Erwin (1<sup>er</sup> juin 1950). « Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation ». In : *Experientia* 6.6, p. 201-209. DOI : [10.1007/BF02173653](#).
- (juil. 1951). « Some Recent Studies on the Composition and Structure of Nucleic Acids ». In : *J. Cell. Comp. Physiol.* 38.S1, p. 41-59. DOI : [10.1002/jcp.1030380406](#).
- (1963). *Essays on Nucleic Acids*. Amsterdam, New York, Elsevier Pub. Co. 236 p.
- (1976). « The Path to the Double Helix by Robert Olby (Review) ». In : *Perspect. Biol. Med.* 19.2, p. 289-290. DOI : [10.1353/pbm.1976.0030](#).
- (1978). *Heraclitean Fire : Sketches from a Life Before Nature*. Rockefeller Univ. Press. 270 p.
- CHARLESWORTH, Brian (1<sup>er</sup> fév. 1994). « Genetic Recombination : Patterns in the Genome ». In : *Current Biology* 4.2, p. 182-184. DOI : [10.1016/S0960-9822\(94\)00045-X](#).
- CHEN, Jian-Min et al. (oct. 2007). « Gene Conversion : Mechanisms, Evolution and Human Disease ». In : *Nat Rev Genet* 8.10, p. 762-775. DOI : [10.1038/nrg2193](#).
- COOP, Graham et Simon R MYERS (9 mar. 2007). « Live Hot, Die Young : Transmission Distortion in Recombination Hotspots ». In : *PLoS Genet* 3.3, e35. DOI : [10.1371/journal.pgen.0030035](#).
- COUCE, Alejandro, Larissa Viraphong CAUDWELL et al. (24 oct. 2017). « Mutator Genomes Decay, despite Sustained Fitness Gains, in a Long-Term Experiment with Bacteria ». In : *PNAS* 114.43, E9026-E9035. DOI : [10.1073/pnas.1705887114](#). pmid : [29073099](#).
- COUCE, Alejandro et Olivier TENAILLON (15 juil. 2019). « Mutation Bias and GC Content Shape Antimutator Invasions ». In : *Nat. Commun.* 10.1, p. 3114. DOI : [10.1038/s41467-019-11217-6](#).
- COUPAT-GOUTALAND, Bénédicte et al. (avr. 2011). « *Ralstonia Solanacearum* Virulence Increased Following Large Interstrain Gene Transfers by Natural Transformation ». In : *Mol. Plant Microbe Interact.* 24.4, p. 497-505. DOI : [10.1094/MPMI-09-10-0197](#). pmid : [21190441](#).
- COWLEY, Lauren A. et al. (13 juin 2018). « Evolution via Recombination : Cell-to-Cell Contact Facilitates Larger Recombination Events in *Streptococcus Pneumoniae* ». In : *PLOS Genetics* 14.6, e1007410. DOI : [10.1371/journal.pgen.1007410](#).
- CROUCHER, Nicholas J., Simon R. HARRIS et al. (14 juin 2012). « A High-Resolution View of Genome-Wide Pneumococcal Transformation ». In : *PLOS Pathogens* 8.6, e1002745. DOI : [10.1371/journal.ppat.1002745](#).
- CROUCHER, Nicholas J., Rafal MOSTOWY et al. (2 mar. 2016). « Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict ». In : *PLOS Biol* 14.3, e1002394. DOI : [10.1371/journal.pbio.1002394](#).
- CROUCHER, Nicholas J., Andrew J. PAGE et al. (18 fév. 2015). « Rapid Phylogenetic Analysis of Large Samples of Recombinant Bacterial Whole Genome Sequences Using Gubbins ». In : *Nucl. Acids Res.* 43.3, e15-e15. DOI : [10.1093/nar/gku1196](#). pmid : [25414349](#).
- DAHM, Ralf (1<sup>er</sup> jan. 2008). « Discovering DNA : Friedrich Miescher and the Early Years of Nucleic Acid Research ». In : *Hum Genet* 122.6, p. 565-581. DOI : [10.1007/s00439-007-0433-0](#).
- DARWIN, Charles (1868). *The Variation of Animals and Plants under Domestication*. 1st. T. 2. 2 t. London : John Murray.

- (1876). *L'origine des espèces au moyen de la sélection naturelle, ou la lutte pour l'existence dans la nature*. Trad. par Edmond BARBIER. 6e. Paris : C. Reinward et cie.
- DAUBIN, Vincent et Guy PERRIÈRE (4 jan. 2003). « G+C3 Structuring Along the Genome : A Common Feature in Prokaryotes ». In : *Mol Biol Evol* 20.4, p. 471-483. DOI : [10.1093/molbev/msg022](https://doi.org/10.1093/molbev/msg022). pmid : 12654929.
- DE LEY, Jozef (1<sup>er</sup> déc. 1964). « Effect of Mutation on DNA-Composition of Some Bacteria ». In : *Antonie van Leeuwenhoek* 30.1, p. 281-288. DOI : [10.1007/BF02046734](https://doi.org/10.1007/BF02046734).
- (1<sup>er</sup> sept. 1968a). « DNA Base Composition and Hybridization in the Taxonomy of Phytopathogenic Bacteria ». In : *Annu. Rev. Phytopathol.* 6.1, p. 63-90. DOI : [10.1146/annurev.py.06.090168.000431](https://doi.org/10.1146/annurev.py.06.090168.000431).
- (1968b). « Molecular Biology and Bacterial Phylogeny ». In : *Evolutionary Biology*. Springer, Boston, MA, p. 103-156. DOI : [10.1007/978-1-4684-8094-8\\_3](https://doi.org/10.1007/978-1-4684-8094-8_3).
- DE LEY, Jozef et J. Van MUYLEM (1<sup>er</sup> déc. 1963). « Some Applications of Deoxyribonucleic Acid Base Composition in Bacterial Taxonomy ». In : *Antonie van Leeuwenhoek* 29.1, p. 344-358. DOI : [10.1007/BF02046087](https://doi.org/10.1007/BF02046087).
- DE VRIES, Johann, Thilo HERZFELD et Wilfried WACKERNAGEL (1<sup>er</sup> juil. 2004). « Transfer of Plastid DNA from Tobacco to the Soil Bacterium *Acinetobacter* Sp. by Natural Transformation ». In : *Mol. Microbiol.* 53.1, p. 323-334. DOI : [10.1111/j.1365-2958.2004.04132.x](https://doi.org/10.1111/j.1365-2958.2004.04132.x).
- DE VRIES, Johann et Wilfried WACKERNAGEL (2002). « Integration of Foreign DNA during Natural Transformation of *Acinetobacter* Sp. by Homology-Facilitated Illegitimate Recombination ». In : *Proc. Natl. Acad. Sci.* 99.4, p. 2094-2099.
- DETMAN, Jeremy R., Jacqueline L. SZTEPANACZ et Rees KASSEN (2016). « The Properties of Spontaneous Mutations in the Opportunistic Pathogen *Pseudomonas Aeruginosa* ». In : *BMC Genomics* 17, p. 27. DOI : [10.1186/s12864-015-2244-3](https://doi.org/10.1186/s12864-015-2244-3).
- DIDELOT, Xavier et Daniel FALUSH (1<sup>er</sup> mar. 2007). « Inference of Bacterial Microevolution Using Multilocus Sequence Data ». In : *Genetics* 175.3, p. 1251-1266. DOI : [10.1534/genetics.106.063305](https://doi.org/10.1534/genetics.106.063305). pmid : 17151252.
- DIDELOT, Xavier, Daniel LAWSON et al. (12 jan. 2010). « Inference of Homologous Recombination in Bacteria Using Whole-Genome Sequences ». In : *Genetics* 186.4, p. 1435-1449. DOI : [10.1534/genetics.110.120121](https://doi.org/10.1534/genetics.110.120121). pmid : 20923983.
- DIDELOT, Xavier et Daniel J. WILSON (12 fév. 2015). « ClonalFrameML : Efficient Inference of Recombination in Whole Bacterial Genomes ». In : *PLOS Computational Biology* 11.2, e1004041. DOI : [10.1371/journal.pcbi.1004041](https://doi.org/10.1371/journal.pcbi.1004041).
- DILLON, Marcus M., Way SUNG, Michael LYNCH et al. (1<sup>er</sup> juil. 2015). « The Rate and Molecular Spectrum of Spontaneous Mutations in the GC-Rich Multichromosome Genome of *Burkholderia Cenocepacia* ». In : *Genetics* 200.3, p. 935-946. DOI : [10.1534/genetics.115.176834](https://doi.org/10.1534/genetics.115.176834). pmid : 25971664.
- DILLON, Marcus M., Way SUNG, Robert SEBRA et al. (15 oct. 2016). « Genome-Wide Biases in the Rate and Molecular Spectrum of Spontaneous Mutations in *Vibrio Cholerae* and *Vibrio Fischeri* ». In : *Mol Biol Evol*, msw224. DOI : [10.1093/molbev/msw224](https://doi.org/10.1093/molbev/msw224). pmid : 27744412.
- DOBZHANSKY, Theodosius (1941). *Genetics And The Origin Of Species*. 2nd.
- DOBZHANSKY, Theodosius, Max K. HECHT et William C. STEERE (1968). *Evolutionary Biology : Volume 2*. Springer US.
- DORER, Marion S., Tate H. SESSLER et Nina R. SALAMA (2011). « Recombination and DNA Repair in *Helicobacter Pylori* ». In : *Annu. Rev. Microbiol.* 65.1, p. 329-348. DOI : [10.1146/annurev-micro-090110-102931](https://doi.org/10.1146/annurev-micro-090110-102931). pmid : 21682641.
- DOVER, Gabriel (sept. 1982). « Molecular Drive : A Cohesive Mode of Species Evolution ». In : *Nature* 299.5879, p. 111. DOI : [10.1038/299111a0](https://doi.org/10.1038/299111a0).



## BIBLIOGRAPHIE

---

- DRAKE, John W. (1<sup>er</sup> déc. 1991). « Ernst Freese (1925–1990) ». In : *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 251.2, p. 165-169. DOI : [10.1016/0027-5107\(91\)90071-U](https://doi.org/10.1016/0027-5107(91)90071-U).
- (1<sup>er</sup> avr. 2012). « Contrasting Mutation Rates from Specific-Locus and Long-Term Mutation-Accumulation Procedures ». In : *G3 Genes Genomes Genet.* 2.4, p. 483-485. DOI : [10.1534/g3.111.001842](https://doi.org/10.1534/g3.111.001842). pmid : [22540039](https://pubmed.ncbi.nlm.nih.gov/22540039/).
- DUBNAU, D (sept. 1991). « Genetic Competence in Bacillus Subtilis. » In : *Microbiol Rev* 55.3, p. 395-424. pmid : [1943994](https://pubmed.ncbi.nlm.nih.gov/1943994/).
- DUBNAU, D. et C. CIRIGILIANO (1973). « Fate of Transforming DNA Following Uptake by Competent Bacillus Subtilis. VI. Non-Covalent Association of Donor and Recipient DNA ». In : *Mol. Gen. Genet. MGG* 120.2, p. 101-106.
- DUBNAU, D. et C. CIRIGILIANO (1972a). « Fate of Transforming Deoxyribonucleic Acid after Uptake by Competent Bacillus Subtilis : Size and Distribution of the Integrated Donor Segments ». In : *J. Bacteriol.* 111.2, p. 488-494.
- (28 fév. 1972b). « Fate of Transforming DNA Following Uptake by Competent Bacillus Subtilis. III. Formation and Properties of Products Isolated from Transformed Cells Which Are Derived Entirely from Donor DNA ». In : *J. Mol. Biol.* 64.1, p. 9-29. pmid : [4622632](https://pubmed.ncbi.nlm.nih.gov/4622632/).
- (28 fév. 1972c). « Fate of Transforming DNA Following Uptake by Competent Bacillus Subtilis. IV. The Endwise Attachment and Uptake of Transforming DNA ». In : *J. Mol. Biol.* 64.1, p. 31-46. pmid : [4622631](https://pubmed.ncbi.nlm.nih.gov/4622631/).
- DUBNAU, D. et R. DAVIDOFF-ABELSON (14 mar. 1971). « Fate of Transforming DNA Following Uptake by Competent Bacillus Subtilis. I. Formation and Properties of the Donor-Recipient Complex ». In : *J. Mol. Biol.* 56.2, p. 209-221. pmid : [4994568](https://pubmed.ncbi.nlm.nih.gov/4994568/).
- DUBNAU, David et Melanie BLOKESCH (2019). « Mechanisms of DNA Uptake by Naturally Competent Bacteria ». In : *Annu. Rev. Genet.* 53.1, null. DOI : [10.1146/annurev-genet-112618-043641](https://doi.org/10.1146/annurev-genet-112618-043641). pmid : [31433955](https://pubmed.ncbi.nlm.nih.gov/31433955/).
- DUHEM, Pierre (1913-1959). *Le système du monde : histoire des doctrines cosmologiques de Platon à Copernic*. T. 1.
- DURET, Laurent et Peter F. ARNDT (9 mai 2008). « The Impact of Recombination on Nucleotide Substitutions in the Human Genome ». In : *PLoS Genet* 4.5, e1000071. DOI : [10.1371/journal.pgen.1000071](https://doi.org/10.1371/journal.pgen.1000071).
- DURET, Laurent et Nicolas GALTIER (sept. 2009). « Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes ». In : *Annu. Rev. Genomics Hum. Genet.* 10.1, p. 285-311. DOI : [10.1146/annurev-genom-082908-150001](https://doi.org/10.1146/annurev-genom-082908-150001).
- DURIEUX, Isabelle et al. (10 sept. 2019). « Diverse Conjugative Elements Silence Natural Transformation in Legionella Species ». In : *PNAS* 116.37, p. 18613-18618. DOI : [10.1073/pnas.1909374116](https://doi.org/10.1073/pnas.1909374116). pmid : [31455740](https://pubmed.ncbi.nlm.nih.gov/31455740/).
- EARL, A. M. et al. (1<sup>er</sup> mai 2012). « Whole-Genome Sequences of Bacillus Subtilis and Close Relatives ». In : *J. Bacteriol.* 194.9, p. 2378-2379. DOI : [10.1128/JB.05675-11](https://doi.org/10.1128/JB.05675-11).
- ELEZ, Marina, Miroslav RADMAN et Ivan MATIC (2007). « The Frequency and Structure of Recombinant Products Is Determined by the Cellular Level of MutL ». In : *Proc. Natl. Acad. Sci.* 104.21, p. 8935-8940.
- ELLIOTT, Kathryn T. et Ellen L. NEIDLE (déc. 2011). « Acinetobacter Baylyi ADP1 : Transforming the Choice of Model Organism ». In : *IUBMB Life* 63.12, p. 1075-1080. DOI : [10.1002/iub.530](https://doi.org/10.1002/iub.530).
- ELLISON, Courtney K. et al. (juil. 2018). « Retraction of DNA-Bound Type IV Competence Pili Initiates DNA Uptake during Natural Transformation in Vibrio Cholerae ». In : *Nat Microbiol* 3.7, p. 773-780. DOI : [10.1038/s41564-018-0174-y](https://doi.org/10.1038/s41564-018-0174-y).

- EYRE-WALKER, Adam (22 juin 1993). « Recombination and Mammalian Genome Evolution ». In : *Proc. R. Soc. Lond. B Biol. Sci.* 252.1335, p. 237-243. DOI : [10.1098/rspb.1993.0071](https://doi.org/10.1098/rspb.1993.0071). pmid : 8394585.
- (6 jan. 1999). « Evidence of Selection on Silent Site Base Composition in Mammals : Potential Implications for the Evolution of Isochores and Junk DNA ». In : *Genetics* 152.2, p. 675-683. pmid : 10353909.
- EYRE-WALKER, Adam et Laurence D. HURST (juil. 2001). « The Evolution of Isochores ». In : *Nat Rev Genet* 2.7, p. 549-555. DOI : [10.1038/35080577](https://doi.org/10.1038/35080577).
- FALL, Saliou et al. (24 oct. 2007). « Horizontal Gene Transfer Regulation in Bacteria as a “Spandrel” of DNA Repair Mechanisms ». In : *PLoS ONE* 2.10, e1055. DOI : [10.1371/journal.pone.0001055](https://doi.org/10.1371/journal.pone.0001055).
- FEDERHEN, Scott (1<sup>er</sup> jan. 2012). « The NCBI Taxonomy Database ». In : *Nucleic Acids Res* 40.D1, p. D136-D143. DOI : [10.1093/nar/gkr1178](https://doi.org/10.1093/nar/gkr1178).
- FELSENSTEIN, Joseph (1974). « The Evolutionary Advantage of Recombination ». In : *Genetics* 78.2, p. 737-756.
- FILIPSKI, J. (1990). « Evolution of DNA Sequence Contributions of Mutational Bias and Selection to the Origin of Chromosomal Compartments ». In : *Advances in Mutagenesis Research*. Sous la dir. de Günter OBE. Advances in Mutagenesis Research. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 1-54. DOI : [10.1007/978-3-642-75599-6\\_1](https://doi.org/10.1007/978-3-642-75599-6_1).
- FOERSTNER, Konrad U. et al. (déc. 2005). « Environments Shape the Nucleotide Composition of Genomes ». In : *EMBO Rep.* 6.12, p. 1208-1213. DOI : [10.1038/sj.embor.7400538](https://doi.org/10.1038/sj.embor.7400538). pmid : 16200051.
- FOSTER, Patricia L., Heewook LEE et al. (11 mar. 2015). « Determinants of Spontaneous Mutation in the Bacterium *Escherichia Coli* as Revealed by Whole-Genome Sequencing ». In : *PNAS* 112.44, E5990-E5999. DOI : [10.1073/pnas.1512136112](https://doi.org/10.1073/pnas.1512136112). pmid : 26460006.
- FOSTER, Patricia L., Brittany A. NICCUM et al. (1<sup>er</sup> août 2018). « Determinants of Base-Pair Substitution Patterns Revealed by Whole-Genome Sequencing of DNA Mismatch Repair Defective *Escherichia Coli* ». In : *Genetics* 209.4, p. 1029-1042. DOI : [10.1534/genetics.118.301237](https://doi.org/10.1534/genetics.118.301237). pmid : 29907647.
- FOWLER, Robert G. et al. (3 fév. 2003). « Interactions among the *Escherichia Coli* mutT, mutM, and mutY Damage Prevention Pathways ». In : *DNA Repair* 2.2, p. 159-173. DOI : [10.1016/S1568-7864\(02\)00193-3](https://doi.org/10.1016/S1568-7864(02)00193-3).
- FRANCINO, M. Pilar, Lin CHAO et al. (5 avr. 1996). « Asymmetries Generated by Transcription-Coupled Repair in Enterobacterial Genes ». In : *Science* 272.5258, p. 107-109. DOI : [10.1126/science.272.5258.107](https://doi.org/10.1126/science.272.5258.107). pmid : 8600517.
- FRANCINO, M. Pilar et Howard OCHMAN (1<sup>er</sup> juin 1997). « Strand Asymmetries in DNA Evolution ». In : *Trends in Genetics* 13.6, p. 240-245. DOI : [10.1016/S0168-9525\(97\)01118-9](https://doi.org/10.1016/S0168-9525(97)01118-9).
- (juil. 1999). « Isochores Result from Mutation Not Selection ». In : *Nature* 400.6739, p. 30. DOI : [10.1038/21804](https://doi.org/10.1038/21804).
- FRANK, A. C. et J. R. LOBRY (1<sup>er</sup> juin 2000). « Oriloc : Prediction of Replication Boundaries in Unannotated Bacterial Chromosomes ». In : *Bioinformatics* 16.6, p. 560-561. DOI : [10.1093/bioinformatics/16.6.560](https://doi.org/10.1093/bioinformatics/16.6.560).
- (30 sept. 1999). « Asymmetric Substitution Patterns : A Review of Possible Underlying Mutational or Selective Mechanisms ». In : *Gene* 238.1, p. 65-77. DOI : [10.1016/S0378-1119\(99\)00297-8](https://doi.org/10.1016/S0378-1119(99)00297-8).
- FREEMAN, James M. et al. (20 mar. 1998). « Patterns of Genome Organization in Bacteria ». In : *Science* 279.5358, p. 1827-1827. DOI : [10.1126/science.279.5358.1827a](https://doi.org/10.1126/science.279.5358.1827a).

## BIBLIOGRAPHIE

---

- FREESE, Ernst (1<sup>er</sup> juil. 1962). « On the Evolution of the Base Composition of DNA ». In : *Journal of Theoretical Biology* 3.1, p. 82-101. DOI : [10.1016/S0022-5193\(62\)80005-8](https://doi.org/10.1016/S0022-5193(62)80005-8).
- FULLERTON, Stephanie M., Antonio BERNARDO CARVALHO et Andrew G. CLARK (1<sup>er</sup> juin 2001). « Local Rates of Recombination Are Positively Correlated with GC Content in the Human Genome ». In : *Mol Biol Evol* 18.6, p. 1139-1142. DOI : [10.1093/oxfordjournals.molbev.a003886](https://doi.org/10.1093/oxfordjournals.molbev.a003886).
- GALTIER, Nicolas et Laurent DURET (juin 2007). « Adaptation or Biased Gene Conversion? Extending the Null Hypothesis of Molecular Evolution ». In : *Trends Genet.* 23.6, p. 273-277. DOI : [10.1016/j.tig.2007.03.011](https://doi.org/10.1016/j.tig.2007.03.011).
- GALTIER, Nicolas, Laurent DURET et al. (jan. 2009). « GC-Biased Gene Conversion Promotes the Fixation of Deleterious Amino Acid Changes in Primates ». In : *Trends in Genetics* 25.1, p. 1-5. DOI : [10.1016/j.tig.2008.10.011](https://doi.org/10.1016/j.tig.2008.10.011).
- GALTIER, Nicolas et J. R. LOBRY (1997). « Relationships between Genomic G+ C Content, RNA Secondary Structures, and Optimal Growth Temperature in Prokaryotes ». In : *J. Mol. Evol.* 44.6, p. 632-636.
- GALTIER, Nicolas, Camille ROUX et al. (1<sup>er</sup> mai 2018). « Codon Usage Bias in Animals : Disentangling the Effects of Natural Selection, Effective Population Size, and GC-Biased Gene Conversion ». In : *Mol Biol Evol* 35.5, p. 1092-1103. DOI : [10.1093/molbev/msy015](https://doi.org/10.1093/molbev/msy015).
- GARCIA-GONZALEZ, Aurian, Ruben RIVERA-RIVERA et Steven E. MASSEY (2012). « The Presence of the DNA Repair Genes mutM, mutY, mutL, and mutS Is Related to Proteome Size in Bacterial Genomes ». In : *Front. Genet.* 3. DOI : [10.3389/fgene.2012.00003](https://doi.org/10.3389/fgene.2012.00003).
- GAYON, Jean (7 août 2018). *La Connaissance de la Vie Aujourd'hui*. Avec la coll. de Victor PETIT. Hermes Science Publishing Ltd.
- (juin 2019). *Darwin et l'après-Darwin, Une Histoire de l'hypothèse de Sélection Naturelle*. 2e édition revue et corrigée. Éditions Matériologiques. 550 p.
- GENG, Peng et al. (31 jan. 2019). « Synthetic Genome Defenses against Selfish DNA Elements Stabilize Engineered Bacteria against Evolutionary Failure ». In : *ACS Synth. Biol.* DOI : [10.1021/acssynbio.8b00426](https://doi.org/10.1021/acssynbio.8b00426).
- GLÉMIN, Sylvain et al. (8 jan. 2015). « Quantification of GC-Biased Gene Conversion in the Human Genome ». In : *Genome Res.* 25.8, p. 1215-1228. DOI : [10.1101/gr.185488.114](https://doi.org/10.1101/gr.185488.114). pmid : 25995268.
- GOULD, S. J. et R. C. LEWONTIN (21 sept. 1979). « The Spandrels of San Marco and the Panglossian Paradigm : A Critique of the Adaptationist Programme ». In : *PNAS* 205.1161, p. 581-598. DOI : [10.1098/rspb.1979.0086](https://doi.org/10.1098/rspb.1979.0086). pmid : 42062.
- GROOTHUIZEN, Flora S. et al. (2015). « MutS/MutL Crystal Structure Reveals That the MutS Sliding Clamp Loads MutL onto DNA ». In : *Elife* 4, e06744.
- GU, Yesong et al. (29 mar. 2002). « Human MutY Homolog, a DNA Glycosylase Involved in Base Excision Repair, Physically and Functionally Interacts with Mismatch Repair Proteins Human MutS Homolog 2/Human MutS Homolog 6 ». In : *J. Biol. Chem.* 277.13, p. 11135-11142. DOI : [10.1074/jbc.M108618200](https://doi.org/10.1074/jbc.M108618200). pmid : 11801590.
- GUÉROUT-FLEURY, Anne-Marie, Niels FRANDBSEN et Patrick STRAGIER (1<sup>er</sup> jan. 1996). « Plasmids for Ectopic Integration in Bacillus Subtilis ». In : *Gene* 180.1, p. 57-61. DOI : [10.1016/S0378-1119\(96\)00404-0](https://doi.org/10.1016/S0378-1119(96)00404-0).
- GUTTMAN, D. S. et D. E. DYKHUIZEN (25 nov. 1994). « Clonal Divergence in Escherichia Coli as a Result of Recombination, Not Mutation ». In : *Science* 266.5189, p. 1380-1383. DOI : [10.1126/science.7973728](https://doi.org/10.1126/science.7973728). pmid : 7973728.



- GUTZ, Herbert et John F. LESLIE (1<sup>er</sup> août 1976). « Gene Conversion : A Hitherto Overlooked Parameter in Population Genetics ». In : *Genetics* 83.4, p. 861-866. pmid : [971809](#).
- HAECKEL, Ernst (1866). *Generelle Morphologie Der Organismen*. T. 2. Verlag von Georg Reimer.
- (1877). *Histoire de la création des êtres organisés d'après les lois naturelles*. Trad. par Charles LETOURNEAU.
- HALL, Barry G. et al. (jan. 2014). « Growth Rates Made Easy ». In : *Mol. Biol. Evol.* 31.1, p. 232-238. DOI : [10.1093/molbev/mst187](#). pmid : [24170494](#).
- HALLDORSSON, Bjarni V. et al. (25 jan. 2019). « Characterizing Mutagenic Effects of Recombination through a Sequence-Level Genetic Map ». In : *Science* 363.6425, eaau1043. DOI : [10.1126/science.aau1043](#).
- HALLDORSSON, Bjarni V et al. (19 sept. 2016). « The Rate of Meiotic Gene Conversion Varies by Sex and Age ». In : *Nat. Genet.* 48.11, p. 1377-1384. DOI : [10.1038/ng.3669](#).
- HARRISON, Richard J. et Brian CHARLESWORTH (1<sup>er</sup> jan. 2011). « Biased Gene Conversion Affects Patterns of Codon Usage and Amino Acid Usage in the Saccharomyces Sensu Stricto Group of Yeasts ». In : *Mol Biol Evol* 28.1, p. 117-129. DOI : [10.1093/molbev/msq191](#).
- HELLWEGER, Ferdi L., Yongjie HUANG et Haiwei LUO (12 jan. 2018). « Carbon Limitation Drives GC Content Evolution of a Marine Bacterium in an Individual-Based Genome-Scale Model ». In : *ISME J.* P. 1. DOI : [10.1038/s41396-017-0023-7](#).
- HERSHBERG, Ruth (9 jan. 2015). « Mutation—The Engine of Evolution : Studying Mutation and Its Role in the Evolution of Bacteria ». In : *Cold Spring Harb Perspect Biol* 7.9, a018077. DOI : [10.1101/cshperspect.a018077](#). pmid : [26330518](#).
- HERSHBERG, Ruth et Dmitri A. PETROV (2010). « Evidence That Mutation Is Universally Biased towards AT in Bacteria ». In : *PLoS Genet.*
- HERSHEY, A. D. et Martha CHASE (20 sept. 1952). « Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage ». In : *J. Gen. Physiol.* 36.1, p. 39-56. DOI : [10.1085/jgp.36.1.39](#). pmid : [12981234](#).
- HILDEBRAND, Falk, Axel MEYER et Adam EYRE-WALKER (2010). « Evidence of Selection upon Genomic GC-Content in Bacteria ». In : *PLoS Genet* 6.9, e1001107.
- HILL, L. R. (1966). « An Index to Deoxyribonucleic Acid Base Compositions of Bacterial Species ». In : *Microbiology* 44.3, p. 419-437. DOI : [10.1099/00221287-44-3-419](#).
- HILL, W. G. et Alan ROBERTSON (déc. 1966). « The Effect of Linkage on Limits to Artificial Selection ». In : *Genet. Res.* 8.3, p. 269-294. DOI : [10.1017/S0016672300010156](#).
- HOLLIDAY, Robin (juil. 1964). « A Mechanism for Gene Conversion in Fungi ». In : *Genet. Res.* 5.2, p. 282-304. DOI : [10.1017/S001667230001233](#).
- HOLMQUIST, G. P. (juil. 1992). « Chromosome Bands, Their Chromatin Flavors, and Their Functional Features. » In : *Am J Hum Genet* 51.1, p. 17-37. pmid : [1609794](#).
- HUBER, Wolfgang et al. (fév. 2015). « Orchestrating High-Throughput Genomic Analysis with Bioconductor ». In : *Nat Meth* 12.2, p. 115-121. DOI : [10.1038/nmeth.3252](#).
- HÜLTER, Nils, Vidar SØRUM et al. (15 fév. 2017). « Costs and Benefits of Natural Transformation in *Acinetobacter Baylyi* ». In : *BMC Microbiology* 17.1, p. 34. DOI : [10.1186/s12866-017-0953-2](#).
- HÜLTER, Nils et Wilfried WACKERNAGEL (mar. 2008). « Double Illegitimate Recombination Events Integrate DNA Segments through Two Different Mechanisms during Natural Transformation of *Acinetobacter Baylyi* ». In : *Mol. Microbiol.* 67.5, p. 984-995. DOI : [10.1111/j.1365-2958.2007.06096.x](#).
- HURST, Laurence D. (juil. 2019). « A Century of Bias in Genetics and Evolution ». In : *Heredity* 123.1, p. 33. DOI : [10.1038/s41437-019-0194-2](#).

## BIBLIOGRAPHIE

---

- HURST, Laurence D. et Alexa R. MERCHANT (7 mar. 2001). « High Guanine–Cytosine Content Is Not an Adaptation to High Temperature : A Comparative Analysis amongst Prokaryotes ». In : *Proc. R. Soc. Lond. B Biol. Sci.* 268.1466, p. 493-497. DOI : [10.1098/rspb.2000.1397](https://doi.org/10.1098/rspb.2000.1397). pmid : 11296861.
- JACOBS, Angelika L. et Primo SCHÄR (1<sup>er</sup> fév. 2012). « DNA Glycosylases : In DNA Repair and Beyond ». In : *Chromosoma* 121.1, p. 1-20. DOI : [10.1007/s00412-011-0347-4](https://doi.org/10.1007/s00412-011-0347-4).
- JAIN, Chirag et al. (30 nov. 2018). « High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries ». In : *Nat. Commun.* 9.1, p. 5114. DOI : [10.1038/s41467-018-07641-9](https://doi.org/10.1038/s41467-018-07641-9).
- JOHNSTON, Calum, Bernard MARTIN, Gwennaele FICHANT et al. (mar. 2014). « Bacterial Transformation : Distribution, Shared Mechanisms and Divergent Control ». In : *Nat Rev Micro* 12.3, p. 181-196. DOI : [10.1038/nrmicro3199](https://doi.org/10.1038/nrmicro3199).
- JOHNSTON, Calum, Bernard MARTIN, Patrice POLARD et al. (oct. 2013). « Postreplication Targeting of Transformants by Bacterial Immune Systems? » In : *Trends Microbiol.* 21.10, p. 516-521. DOI : [10.1016/j.tim.2013.08.002](https://doi.org/10.1016/j.tim.2013.08.002).
- JONES, Mary Ellen (sept. 1953). « Albrecht Kossel, A Biographical Sketch ». In : *Yale J Biol Med* 26.1, p. 80-97. pmid : 13103145.
- JOSEPH, Nimesh, Viswanadham DUPPATLA et Desirazu N. RAO (2006). « Prokaryotic DNA Mismatch Repair ». In : *Prog. Nucleic Acid Res. Mol. Biol.* 81, p. 1-49.
- KANVAH, Sriram et Gary B. SCHUSTER (1<sup>er</sup> sept. 2005). « The Sacrificial Role of Easily Oxidizable Sites in the Protection of DNA from Damage ». In : *Nucleic Acids Res* 33.16, p. 5133-5138. DOI : [10.1093/nar/gki801](https://doi.org/10.1093/nar/gki801).
- KARKAS, J. D., R. RUDNER et E. CHARGAFF (1<sup>er</sup> juil. 1968). « Separation of B. Subtilis DNA into Complementary Strands. II. Template Functions and Composition as Determined by Transcription with RNA Polymerase ». In : *PNAS* 60.3, p. 915-920. DOI : [10.1073/pnas.60.3.915](https://doi.org/10.1073/pnas.60.3.915). pmid : 4970113.
- KELKAR, Yogeshwar D., Daniel S. PHILLIPS et Howard OCHMAN (1<sup>er</sup> juin 2015). « Effects of Genic Base Composition on Growth Rate in G+C-Rich Genomes ». In : *G3 Genes Genomes Genet.* 5.6, p. 1247-1252. DOI : [10.1534/g3.115.016824](https://doi.org/10.1534/g3.115.016824). pmid : 25897009.
- KENNAWAY, Ernest (31 déc. 1952). « Some Recollections of Albrecht Kossel : Professor of Physiology in Heidelberg, 1901–1924 ». In : *Ann. Sci.* 8.4, p. 393-397. DOI : [10.1080/00033795200200292](https://doi.org/10.1080/00033795200200292).
- KERSTERS, Karel (1998). « In Memoriam Jozef De Ley 1924-1997 ». In : *Int. J. Syst. Evol. Microbiol.* 48.3, p. ix-ix. DOI : [10.1099/00207713-48-3-ix](https://doi.org/10.1099/00207713-48-3-ix).
- KIM, Mandy, Tiffany HUANG et Jeffrey H. MILLER (1<sup>er</sup> août 2003). « Competition between MutY and Mismatch Repair at A · C Mispairs In Vivo ». In : *J. Bacteriol.* 185.15, p. 4626-4629. DOI : [10.1128/JB.185.15.4626-4629.2003](https://doi.org/10.1128/JB.185.15.4626-4629.2003). pmid : 12867475.
- KIMURA, Motoo (déc. 1957). « Some Problems of Stochastic Processes in Genetics ». In : *Ann. Math. Statist.* 28.4, p. 882-901. DOI : [10.1214/aoms/1177706791](https://doi.org/10.1214/aoms/1177706791).
- (fév. 1961). « Natural Selection as the Process of Accumulating Genetic Information in Adaptive Evolution ». In : *Genet. Res.* 2.1, p. 127-140. DOI : [10.1017/S0016672300000616](https://doi.org/10.1017/S0016672300000616).
- (fév. 1968). « Evolutionary Rate at the Molecular Level ». In : *Nature* 217.5129, p. 624. DOI : [10.1038/217624a0](https://doi.org/10.1038/217624a0).
- KINGMAN, J. F. C. (1<sup>er</sup> sept. 1982). « The Coalescent ». In : *Stochastic Processes and their Applications* 13.3, p. 235-248. DOI : [10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4).
- KISKER, Caroline, Jochen KUPER et Bennett Van HOUTEN (3 jan. 2013). « Prokaryotic Nucleotide Excision Repair ». In : *Cold Spring Harb Perspect Biol* 5.3, a012591. DOI : [10.1101/cshperspect.a012591](https://doi.org/10.1101/cshperspect.a012591). pmid : 23457260.

- KOWALCZYKOWSKI, Stephen C. (11 jan. 2015). « An Overview of the Molecular Mechanisms of Recombinational DNA Repair ». In : *Cold Spring Harb Perspect Biol* 7.11, a016410. DOI : [10.1101/cshperspect.a016410](https://doi.org/10.1101/cshperspect.a016410). pmid : 26525148.
- KREITMAN, Martin (août 1983). « Nucleotide Polymorphism at the Alcohol Dehydrogenase Locus of *Drosophila Melanogaster* ». In : *Nature* 304.5925, p. 412. DOI : [10.1038/304412a0](https://doi.org/10.1038/304412a0).
- KROKAN, Hans E. et Magnar BJØRÅS (4 jan. 2013). « Base Excision Repair ». In : *Cold Spring Harb Perspect Biol* 5.4, a012583. DOI : [10.1101/cshperspect.a012583](https://doi.org/10.1101/cshperspect.a012583). pmid : 23545420.
- KUDLA, Grzegorz et al. (10 avr. 2009). « Coding-Sequence Determinants of Gene Expression in *Escherichia Coli* ». In : *Science* 324.5924, p. 255-258. DOI : [10.1126/science.1170160](https://doi.org/10.1126/science.1170160). pmid : 19359587.
- KULICK, Stefan et al. (24 nov. 2008). « Mosaic DNA Imports with Interspersions of Recipient Sequence after Natural Transformation of *Helicobacter Pylori* ». In : *PLOS ONE* 3.11, e3797. DOI : [10.1371/journal.pone.0003797](https://doi.org/10.1371/journal.pone.0003797).
- KUNKEL, Thomas A. et Dorothy A. ERIE (1<sup>er</sup> juin 2005). « DNA Mismatch Repair ». In : *Annu. Rev. Biochem.* 74.1, p. 681-710. DOI : [10.1146/annurev.biochem.74.082803.133243](https://doi.org/10.1146/annurev.biochem.74.082803.133243).
- LAMB, B. C. et S. HELMI (avr. 1982). « The Extent to Which Gene Conversion Can Change Allele Frequencies in Populations ». In : *Genet. Res.* 39.2, p. 199-217. DOI : [10.1017/S0016672300020887](https://doi.org/10.1017/S0016672300020887).
- LAMB, Bernard C. (1<sup>er</sup> oct. 1986). « Gene Conversion Disparity : Factors Influencing Its Direction and Extent, with Tests of Assumptions and Predictions in Its Evolutionary Effects ». In : *Genetics* 114.2, p. 611-632. pmid : 3770472.
- LAMERS, Meindert H. et al. (oct. 2000). « The Crystal Structure of DNA Mismatch Repair Protein MutS Binding to a G-T Mismatch ». In : *Nature* 407.6805, p. 711-717. DOI : [10.1038/35037523](https://doi.org/10.1038/35037523).
- LAPLACE, Pierre-Simon de (1825). *Essai philosophique sur les probabilités*. 5e.
- LASSALLE, Florent et al. (6 fév. 2015). « GC-Content Evolution in Bacterial Genomes : The Biased Gene Conversion Hypothesis Expands ». In : *PLOS Genet.* 11.2. Sous la dir. de Dmitri A. PETROV, e1004941. DOI : [10.1371/journal.pgen.1004941](https://doi.org/10.1371/journal.pgen.1004941).
- LEE, Heewook et al. (10 sept. 2012). « Rate and Molecular Spectrum of Spontaneous Mutations in the Bacterium *Escherichia Coli* as Determined by Whole-Genome Sequencing ». In : *PNAS* 109.41, E2774-E2783. DOI : [10.1073/pnas.1210309109](https://doi.org/10.1073/pnas.1210309109). pmid : 22991466.
- LEE, Ja Yil et al. (28 août 2015). « Base Triplet Stepping by the Rad51/RecA Family of Recombinases ». In : *Science* 349.6251, p. 977-981. DOI : [10.1126/science.aab2666](https://doi.org/10.1126/science.aab2666). pmid : 26315438.
- LEE, Ki Yong, R. WAHL et E. BARBU (1956). « Contenu En Bases Puriques et Pyrimidiques Des Acides Désoxyribonucléiques Des Bactéries ». In : *Annales de l'institut Pasteur*. T. 91. Masson Editeur 120 Blvd Saint-Germain, 75280 Paris 06, France, p. 212-224.
- LENHART, Justin S. et al. (1<sup>er</sup> jan. 2016). « Mismatch Repair in Gram-Positive Bacteria ». In : *Research in Microbiology* 167.1, p. 4-12. DOI : [10.1016/j.resmic.2015.08.006](https://doi.org/10.1016/j.resmic.2015.08.006).
- LESECQUE, Y., D. MOUCHIROUD et L. DURET (1<sup>er</sup> juin 2013). « GC-Biased Gene Conversion in Yeast Is Specifically Associated with Crossovers : Molecular Mechanisms and Evolutionary Significance ». In : *Mol. Biol. Evol.* 30.6, p. 1409-1419. DOI : [10.1093/molbev/mst056](https://doi.org/10.1093/molbev/mst056).
- LESECQUE, Yann et al. (13 nov. 2014). « The Red Queen Model of Recombination Hotspots Evolution in the Light of Archaic and Modern Human Genomes ». In : *PLoS Genet* 10.11, e1004790. DOI : [10.1371/journal.pgen.1004790](https://doi.org/10.1371/journal.pgen.1004790).

## BIBLIOGRAPHIE

---

- LEVENE, P. A. et W. A. JACOBS (1909). « Über Die Hefe-Nucleinsäure ». In : *Berichte Dtsch. Chem. Ges.* 42.2, p. 2474-2478.
- LEWONTIN, Richard C. (1974). *The Genetic Basis of Evolutionary Change*. Columbia University Press. 372 p.
- LI, Heng (15 oct. 2019). *Wgsim*.
- LI, Heng et Richard DURBIN (15 juil. 2009). « Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform ». In : *Bioinformatics* 25.14, p. 1754-1760. DOI : [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324).
- LI, Heng, Bob HANDSAKER et al. (15 août 2009). « The Sequence Alignment/Map Format and SAMtools ». In : *Bioinformatics* 25.16, p. 2078-2079. DOI : [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352).
- LI, Ran et al. (27 sept. 2018). « A High-Resolution Map of Non-Crossover Events in Mice Reveals Impacts of Genetic Diversity on Meiotic Recombination ». In : *bioRxiv*, p. 428987. DOI : [10.1101/428987](https://doi.org/10.1101/428987).
- (29 août 2019). « A High-Resolution Map of Non-Crossover Events Reveals Impacts of Genetic Diversity on Mammalian Meiotic Recombination ». In : *Nat Commun* 10.1, p. 1-15. DOI : [10.1038/s41467-019-11675-y](https://doi.org/10.1038/s41467-019-11675-y).
- LIND, Peter A. et Dan I. ANDERSSON (18 nov. 2008). « Whole-Genome Mutational Biases in Bacteria ». In : *PNAS* 105.46, p. 17878-17883. DOI : [10.1073/pnas.0804445105](https://doi.org/10.1073/pnas.0804445105). pmid : 19001264.
- LIU, Haoxuan et al. (20 nov. 2017). « Tetrad Analysis in Plants and Fungi Finds Large Differences in Gene Conversion Rates but No GC Bias ». In : *Nat. Ecol. Evol.* DOI : [10.1038/s41559-017-0372-7](https://doi.org/10.1038/s41559-017-0372-7).
- LIU, Jiaquan et al. (16 nov. 2016). « Cascading MutS and MutL Sliding Clamps Control DNA Diffusion to Activate Mismatch Repair ». In : *Nature*. DOI : [10.1038/nature20562](https://doi.org/10.1038/nature20562).
- LOBRY, J. R. (1<sup>er</sup> mar. 1995). « Properties of a General Model of DNA Evolution under No-Strand-Bias Conditions ». In : *J Mol Evol* 40.3, p. 326-330. DOI : [10.1007/BF00163237](https://doi.org/10.1007/BF00163237).
- (1<sup>er</sup> jan. 1996a). « A Simple Vectorial Representation of DNA Sequences for the Detection of Replication Origins in Bacteria ». In : *Biochimie* 78.5, p. 323-326. DOI : [10.1016/0300-9084\(96\)84764-X](https://doi.org/10.1016/0300-9084(96)84764-X).
- (1<sup>er</sup> mai 1996b). « Asymmetric Substitution Patterns in the Two DNA Strands of Bacteria. » In : *Mol Biol Evol* 13.5, p. 660-665. DOI : [10.1093/oxfordjournals.molbev.a025626](https://doi.org/10.1093/oxfordjournals.molbev.a025626).
- LONG, Hongan et al. (1<sup>er</sup> jan. 2018). « Evolutionary Determinants of Genome-Wide Nucleotide Composition ». In : *Nat. Ecol. Evol.* P. 1. DOI : [10.1038/s41559-017-0425-y](https://doi.org/10.1038/s41559-017-0425-y).
- LOZUPONE, Catherine A. et al. (10 jan. 2013). « Meta-Analyses of Studies of the Human Microbiota ». In : *Genome Res.* 23.10, p. 1704-1714. DOI : [10.1101/gr.151803.112](https://doi.org/10.1101/gr.151803.112). pmid : 23861384.
- LYNCH, Michael et al. (2016). « Genetic Drift, Selection and the Evolution of the Mutation Rate ». In : *Nat. Rev. Genet.* 17.11, p. 704-714.
- MAHARJAN, Ram P. et Thomas FERENCI (8 juin 2017). « A Shifting Mutational Landscape in 6 Nutritional States : Stress-Induced Mutagenesis as a Series of Distinct Stress Input–Mutation Output Relationships ». In : *PLOS Biology* 15.6, e2001477. DOI : [10.1371/journal.pbio.2001477](https://doi.org/10.1371/journal.pbio.2001477).
- MAJEWSKI, Jacek et Frederick M. COHAN (1<sup>er</sup> jan. 1998). « The Effect of Mismatch Repair and Heteroduplex Formation on Sexual Isolation in *Bacillus* ». In : *Genetics* 148.1, p. 13-18. pmid : 9475717.

- MANCERA, Eugenio et al. (24 juil. 2008). « High-Resolution Mapping of Meiotic Crossovers and Non-Crossovers in Yeast ». In : *Nature* 454.7203, p. 479-485. DOI : [10.1038/nature07135](https://doi.org/10.1038/nature07135).
- MARMUR, J. et P. DOTY (1<sup>er</sup> juil. 1962). « Determination of the Base Composition of Deoxyribonucleic Acid from Its Thermal Denaturation Temperature ». In : *Journal of Molecular Biology* 5.1, p. 109-118. DOI : [10.1016/S0022-2836\(62\)80066-7](https://doi.org/10.1016/S0022-2836(62)80066-7).
- MATIC, Ivan, François TADDEI et Miroslav RADMAN (1<sup>er</sup> fév. 1996). « Genetic Barriers among Bacteria ». In : *Trends in Microbiology* 4.2, p. 69-73. DOI : [10.1016/0966-842X\(96\)81514-9](https://doi.org/10.1016/0966-842X(96)81514-9).
- MATTHEY, Noémie et al. (3 sept. 2019). « Neighbor Predation Linked to Natural Competence Fosters the Transfer of Large Genomic Regions in *Vibrio Cholerae* ». In : *eLife* 8. Sous la dir. de Târn MIGNOT et Detlef WEIGEL, e48212. DOI : [10.7554/eLife.48212](https://doi.org/10.7554/eLife.48212).
- MAYR, Ernst (1961). « Cause and Effect in Biology ». In : *Science* 134.3489, p. 1501-1506.
- MCCUTCHEON, John P. et Nancy A. MORAN (jan. 2012). « Extreme Genome Reduction in Symbiotic Bacteria ». In : *Nat Rev Micro* 10.1, p. 13-26. DOI : [10.1038/nrmicro2670](https://doi.org/10.1038/nrmicro2670).
- MELL, Joshua Chang, Jae Yun LEE et al. (1<sup>er</sup> avr. 2014). « Extensive Cotransformation of Natural Variation into Chromosomes of Naturally Competent *Haemophilus Influenzae* ». In : *G3 Genes Genomes Genet.* 4.4, p. 717-731. DOI : [10.1534/g3.113.009597](https://doi.org/10.1534/g3.113.009597). pmid : 24569039.
- MELL, Joshua Chang, Svetlana SHUMILINA et al. (28 août 2011). « Transformation of Natural Genetic Variation into *Haemophilus Influenzae* Genomes ». In : *PLOS Pathogens* 7.7, e1002151. DOI : [10.1371/journal.ppat.1002151](https://doi.org/10.1371/journal.ppat.1002151).
- MESELSON, M. S. et C. M. RADDING (1<sup>er</sup> jan. 1975). « A General Model for Genetic Recombination ». In : *PNAS* 72.1, p. 358-361. DOI : [10.1073/pnas.72.1.358](https://doi.org/10.1073/pnas.72.1.358). pmid : 1054510.
- MICHAELS, M. L. et al. (1<sup>er</sup> août 1992). « Evidence That MutY and MutM Combine to Prevent Mutations by an Oxidatively Damaged Form of Guanine in DNA ». In : *PNAS* 89.15, p. 7022-7025. DOI : [10.1073/pnas.89.15.7022](https://doi.org/10.1073/pnas.89.15.7022). pmid : 1495996.
- MIESCHER, Friedrich (1871). *Ueber Die Chemische Zusammensetzung Der Eiterzellen*.
- MILKMAN, R. et M. M. BRIDGES (1<sup>er</sup> nov. 1990). « Molecular Evolution of the *Escherichia Coli* Chromosome. III. Clonal Frames. » In : *Genetics* 126.3, p. 505-517. pmid : 1979037.
- MONOD, Jacques (23 oct. 2014). *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*. Points. 256 p.
- MORAN, Nancy A. (1996). « Accelerated Evolution and Muller's Ratchet in Endosymbiotic Bacteria ». In : *Proc. Natl. Acad. Sci.* 93.7, p. 2873-2878.
- MORGAN, Thomas Hunt (4 juin 1934). « The Relation of Genetics to Physiology and Medicine ». Nobel Lecture (Stockholm).
- MOSTOWY, Rafal et al. (11 fév. 2017). « Efficient Inference of Recent and Ancestral Recombination within Bacterial Populations ». In : *Mol Biol Evol.* DOI : [10.1093/molbev/msx066](https://doi.org/10.1093/molbev/msx066).
- MUKAI, Terumi (1964). « The Genetic Structure of Natural Populations of *Drosophila Melanogaster*. I. Spontaneous Mutation Rate of Polygenes Controlling Viability ». In : *Genetics* 50.1, p. 1.
- MULLINS, Elwood A. et al. (1<sup>er</sup> sept. 2019). « Emerging Roles of DNA Glycosylases and the Base Excision Repair Pathway ». In : *Trends in Biochemical Sciences* 44.9, p. 765-781. DOI : [10.1016/j.tibs.2019.04.006](https://doi.org/10.1016/j.tibs.2019.04.006).



## BIBLIOGRAPHIE

---

- MUSTO, Héctor et al. (2004). « Correlations between Genomic GC Levels and Optimal Growth Temperatures in Prokaryotes ». In : *FEBS Lett.* 573.1-3, p. 73-77. DOI : [10.1016/j.febslet.2004.07.056](https://doi.org/10.1016/j.febslet.2004.07.056).
- MUTO, Akira et Syozo OSAWA (1987). « The Guanine and Cytosine Content of Genomic DNA and Bacterial Evolution ». In : *Proc. Natl. Acad. Sci.* 84.1, p. 166-169.
- MYERS, Simon et al. (12 fév. 2010). « Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination ». In : *Science* 327.5967, p. 876-879. DOI : [10.1126/science.1182363](https://doi.org/10.1126/science.1182363). pmid : 20044541.
- NAGYLAKI, Thomas (1983a). « Evolution of a Finite Population under Gene Conversion ». In : *PNAS* 80.20, p. 6278-6281.
- (1<sup>er</sup> oct. 1983b). « Evolution of a Large Population under Gene Conversion ». In : *PNAS* 80.19, p. 5941-5945. DOI : [10.1073/pnas.80.19.5941](https://doi.org/10.1073/pnas.80.19.5941). pmid : 6577463.
- NAH, Hee-Ju et al. (2017). « Cloning and Heterologous Expression of a Large-Sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species ». In : *Front. Microbiol.* 8. DOI : [10.3389/fmicb.2017.00394](https://doi.org/10.3389/fmicb.2017.00394).
- NAKAMURA, L. K., Michael S. ROBERTS et Frederick M. COHAN (1999). « Note : Relationship of Bacillus Subtilis Clades Associated with Strains 168 and W23 : A Proposal for Bacillus Subtilis Subsp. Subtilis Subsp. Nov. and Bacillus Subtilis Subsp. Spizizenii Subsp. Nov. ». In : *Int. J. Syst. Evol. Microbiol.* 49.3, p. 1211-1215. DOI : [10.1099/00207713-49-3-1211](https://doi.org/10.1099/00207713-49-3-1211).
- NAYA, Hugo et al. (1<sup>er</sup> sept. 2002). « Aerobiosis Increases the Genomic Guanine Plus Cytosine Content (GC%) in Prokaryotes ». In : *J. Mol. Evol.* 55.3, p. 260-264. DOI : [10.1007/s00239-002-2323-3](https://doi.org/10.1007/s00239-002-2323-3).
- NIELSEN, Rasmus et Ziheng YANG (1<sup>er</sup> août 2003). « Estimating the Distribution of Selection Coefficients from Phylogenetic Data with Applications to Mitochondrial and Viral DNA ». In : *Mol Biol Evol* 20.8, p. 1231-1239. DOI : [10.1093/molbev/msg147](https://doi.org/10.1093/molbev/msg147).
- OBENCHAIN, Valerie et al. (15 juil. 2014). « VariantAnnotation : A Bioconductor Package for Exploration and Annotation of Genetic Variants ». In : *Bioinformatics* 30.14, p. 2076-2078. DOI : [10.1093/bioinformatics/btu168](https://doi.org/10.1093/bioinformatics/btu168). pmid : 24681907.
- OLBY, Robert (oct. 1969). « Cell Chemistry in Miescher's Day ». In : *Med. Hist.* 13.4, p. 377-382. DOI : [10.1017/S0025727300014800](https://doi.org/10.1017/S0025727300014800).
- (13 juin 2012). *The Path to the Double Helix : The Discovery of DNA*. reprint edition. New York : Dover Publications. 560 p.
- OLIVEIRA, Pedro H., Marie TOUCHON et Eduardo P. C. ROCHA (17 mai 2016). « Regulation of Genetic Flux between Bacteria by Restriction-Modification Systems ». In : *PNAS* 113.20, p. 5658-5663. DOI : [10.1073/pnas.1603257113](https://doi.org/10.1073/pnas.1603257113). pmid : 27140615.
- ORR-WEAVER, T. L., A. NICOLAS et J. W. SZOSTAK (1<sup>er</sup> déc. 1988). « Gene Conversion Adjacent to Regions of Double-Strand Break Repair. ». In : *Mol. Cell. Biol.* 8.12, p. 5292-5298. DOI : [10.1128/MCB.8.12.5292](https://doi.org/10.1128/MCB.8.12.5292). pmid : 3072478.
- OVERBALLE-PETERSEN, Søren et al. (12 mar. 2013). « Bacterial Natural Transformation by Highly Fragmented and Damaged DNA ». In : *PNAS* 110.49, p. 19860-19865. DOI : [10.1073/pnas.1315278110](https://doi.org/10.1073/pnas.1315278110). pmid : 24248361.
- PALMEN, Ronald et Klaas J HELLINGWERF (11 juin 1997). « Uptake and Processing of DNA by Acinetobacter Calcoaceticus – a Review1 ». In : *Gene* 192.1, p. 179-190. DOI : [10.1016/S0378-1119\(97\)00042-5](https://doi.org/10.1016/S0378-1119(97)00042-5).
- PALMEN, Ronald, Ben VOSMAN, Pieter BUIJSMAN et al. (1993). « Physiological Characterization of Natural Transformation in Acinetobacter Calcoaceticus ». In : *Microbiology* 139.2, p. 295-305.
- PALMEN, Ronald, Ben VOSMAN, Ruben KOK et al. (1992). « Characterization of Transformation-Deficient Mutants of Acinetobacter Calcoaceticus ». In : *Mol. Microbiol.* 6.13, p. 1747-1754. DOI : [10.1111/j.1365-2958.1992.tb01347.x](https://doi.org/10.1111/j.1365-2958.1992.tb01347.x).

- PARKER, Antony et al. (23 fév. 2001). « Human Homolog of the MutY Repair Protein (hMYH) Physically Interacts with Proteins Involved in Long Patch DNA Base Excision Repair ». In : *J. Biol. Chem.* 276.8, p. 5547-5555. DOI : [10.1074/jbc.M008463200](https://doi.org/10.1074/jbc.M008463200). pmid : [11092888](https://pubmed.ncbi.nlm.nih.gov/11092888/).
- PASCAL, Blaise (1864). *Oeuvres : pensées, lettres et opuscules divers*. Napoléon Chaix. Paris.
- PASTEUR, Louis (1922). *Oeuvres de Pasteur réunies par Pasteur Vallery-Radot*. Sous la dir. de Pasteur VALLERY-RADOT. Paris, Masson et cie, DOI : [10.5962/bhl.title.1989](https://doi.org/10.5962/bhl.title.1989).
- PEIRCE, Charles Sanders (1992a). *The Essential Peirce : Selected Philosophical Writings*. T. 2. 2 t. Indiana University Press. 626 p.
- (1992b). « The Fixation of Belief ». In : *The Essential Peirce : Selected Philosophical Writings*. T. 1. 2 t. Indiana University Press.
- PENEL, Simon et al. (16 juin 2009). « Databases of Homologous Gene Families for Comparative Genomics ». In : *BMC Bioinformatics* 10.6, S3. DOI : [10.1186/1471-2105-10-S6-S3](https://doi.org/10.1186/1471-2105-10-S6-S3).
- PESSIA, E. et al. (16 juil. 2012). « Evidence for Widespread GC-Biased Gene Conversion in Eukaryotes ». In : *Gen Biol Evol* 4.7, p. 675-682. DOI : [10.1093/gbe/evs052](https://doi.org/10.1093/gbe/evs052).
- PETZOLDT, Thomas (2018). *Growthrates : Estimate Growth Rates from Experimental Data*. Version 0.7.2. R package version 0.7.2.
- PIGANEAU, Gwenaél et al. (oct. 2001). « Multiplicative versus Additive Selection in Relation to Genome Evolution : A Simulation Study ». In : *Genet. Res.* 78.2, p. 171-175. DOI : [10.1017/S0016672301005249](https://doi.org/10.1017/S0016672301005249).
- PLUTYNSKI, Anya (1<sup>er</sup> juin 2007). « Drift : A Historical and Conceptual Overview ». In : *Biol Theory* 2.2, p. 156-167. DOI : [10.1162/biot.2007.2.2.156](https://doi.org/10.1162/biot.2007.2.2.156).
- POINCARÉ, Henri (1902). *La Science et l'hypothèse*. Bibliothèque de Philosophie Scientifique. Paris : Flammarion.
- POLLARD, Katherine S. et al. (14 sept. 2006a). « An RNA Gene Expressed during Cortical Development Evolved Rapidly in Humans ». In : *Nature* 443.7108, p. 167-172. DOI : [10.1038/nature05113](https://doi.org/10.1038/nature05113).
- POLLARD, Katherine S et al. (13 oct. 2006b). « Forces Shaping the Fastest Evolving Regions in the Human Genome ». In : *PLoS Genet* 2.10, e168. DOI : [10.1371/journal.pgen.0020168](https://doi.org/10.1371/journal.pgen.0020168).
- POLOSINA, Yaroslava Y. et Claire G. CUPPLES (1<sup>er</sup> déc. 2010). « Wot the 'L—Does MutL Do? » In : *Mutation Research/Reviews in Mutation Research*. This Issue Includes the Special Issue Section : Omics : Tools for Understanding and Studying Systems Toxicology 705.3, p. 228-238. DOI : [10.1016/j.mrrev.2010.07.002](https://doi.org/10.1016/j.mrrev.2010.07.002).
- POPA, Ovidiu et al. (4 jan. 2011). « Directed Networks Reveal Genomic Barriers and DNA Repair Bypasses to Lateral Gene Transfer among Prokaryotes ». In : *Genome Res.* 21.4, p. 599-609. DOI : [10.1101/gr.115592.110](https://doi.org/10.1101/gr.115592.110). pmid : [21270172](https://pubmed.ncbi.nlm.nih.gov/21270172/).
- POPKIN, Richard (4 oct. 2019). *Histoire du scepticisme*. Trad. par Benoit GAULTIER. Banc d'essais. Marseille : Agone. 912 p.
- POPLIN, Ryan et al. (24 juil. 2018). « Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples ». In : *bioRxiv*, p. 201178. DOI : [10.1101/201178](https://doi.org/10.1101/201178).
- PORELLO, Silvia L., Aquiles E. LEYES et Sheila S. DAVID (1<sup>er</sup> oct. 1998). « Single-Turnover and Pre-Steady-State Kinetics of the Reaction of the Adenine Glycosylase MutY with Mismatch-Containing DNA Substrates ». In : *Biochemistry* 37.42, p. 14756-14764. DOI : [10.1021/bi981594+](https://doi.org/10.1021/bi981594+).
- PRITCHARD, Leighton et al. (2016). « Genomics and Taxonomy in Diagnostics for Food Security : Soft-Rotting Enterobacterial Plant Pathogens ». In : *Anal. Methods* 8.1, p. 12-24. DOI : [10.1039/C5AY02550H](https://doi.org/10.1039/C5AY02550H).

## BIBLIOGRAPHIE

---

- R CORE TEAM (2018). *R : A Language and Environment for Statistical Computing*. Vienna, Austria.
- RADMAN, Miroslav et R. WAGNER (1986). « Mismatch Repair in Escherichia Coli ». In : *Annu. Rev. Genet.* 20.1, p. 523-538. DOI : [10.1146/annurev.ge.20.120186.002515](https://doi.org/10.1146/annurev.ge.20.120186.002515). pmid : [2949693](https://pubmed.ncbi.nlm.nih.gov/2949693/).
- RAGHAVAN, Rahul, Yogeshwar D. KELKAR et Howard OCHMAN (9 avr. 2012). « A Selective Force Favoring Increased G+C Content in Bacterial Genes ». In : *PNAS* 109.36, p. 14504-14507. DOI : [10.1073/pnas.1205683109](https://doi.org/10.1073/pnas.1205683109). pmid : [22908296](https://pubmed.ncbi.nlm.nih.gov/22908296/).
- RAGUNATHAN, Kaushik, Cheng LIU et Taekjip HA (13 déc. 2012). « RecA Filament Sliding on DNA Facilitates Homology Search ». In : *eLife* 1, e00067. DOI : [10.7554/eLife.00067](https://doi.org/10.7554/eLife.00067). pmid : [23240082](https://pubmed.ncbi.nlm.nih.gov/23240082/).
- RAHMER, Regine, Kambiz MORABBI HERAVI et Josef ALTENBUCHNER (2015). « Construction of a Super-Competent Bacillus Subtilis 168 Using the PmtIA-comKS Inducible Cassette ». In : *Front. Microbiol.* 6. DOI : [10.3389/fmicb.2015.01431](https://doi.org/10.3389/fmicb.2015.01431).
- RATNAKUMAR, A. et al. (27 août 2010). « Detecting Positive Selection within Genomes : The Problem of Biased Gene Conversion ». In : *Philos. Trans. R. Soc. B Biol. Sci.* 365.1552, p. 2571-2580. DOI : [10.1098/rstb.2010.0007](https://doi.org/10.1098/rstb.2010.0007).
- RAVENHALL, Matt et al. (28 mai 2015). « Inferring Horizontal Gene Transfer ». In : *PLOS Computational Biology* 11.5, e1004095. DOI : [10.1371/journal.pcbi.1004095](https://doi.org/10.1371/journal.pcbi.1004095).
- REDFIELD, R. J. (1<sup>er</sup> mai 1988). « Evolution of Bacterial Transformation : Is Sex with Dead Cells Ever Better than No Sex at All? ». In : *Genetics* 119.1, p. 213-221. pmid : [3396864](https://pubmed.ncbi.nlm.nih.gov/3396864/).
- (1<sup>er</sup> avr. 1993). « Evolution of Natural Transformation : Testing the DNA Repair Hypothesis in Bacillus Subtilis and Haemophilus Influenzae. ». In : *Genetics* 133.4, p. 755-761. pmid : [8462839](https://pubmed.ncbi.nlm.nih.gov/8462839/).
- REICHENBERGER, Erin R. et al. (5 jan. 2015). « Prokaryotic Nucleotide Composition Is Shaped by Both Phylogeny and the Environment ». In : *Genome Biol Evol* 7.5, p. 1380-1389. DOI : [10.1093/gbe/evv063](https://doi.org/10.1093/gbe/evv063). pmid : [25861819](https://pubmed.ncbi.nlm.nih.gov/25861819/).
- RENAN, Ernest (1890). *L'avenir de la science : pensées de 1848*. Paris : Calmann-Lévy. 582 p.
- RENDA, Brian A. et al. (3 jan. 2015). « Genome Instability Mediates the Loss of Key Traits by Acinetobacter Baylyi ADP1 during Laboratory Evolution ». In : *J. Bacteriol.* 197.5, p. 872-881. DOI : [10.1128/JB.02263-14](https://doi.org/10.1128/JB.02263-14). pmid : [25512307](https://pubmed.ncbi.nlm.nih.gov/25512307/).
- RIZK, Guillaume et al. (15 déc. 2014). « MindTheGap : Integrated Detection and Assembly of Short and Long Insertions ». In : *Bioinformatics* 30.24, p. 3451-3457. DOI : [10.1093/bioinformatics/btu545](https://doi.org/10.1093/bioinformatics/btu545).
- ROBERT, Lydia et al. (16 mar. 2018). « Mutation Dynamics and Fitness Effects Followed in Single Cells ». In : *Science* 359.6381, p. 1283-1286. DOI : [10.1126/science.aan0797](https://doi.org/10.1126/science.aan0797).
- ROCHA, E. P. C. (1<sup>er</sup> juin 2004). « The Replication-Related Organization of Bacterial Genomes ». In : *Microbiology* 150.6, p. 1609-1627. DOI : [10.1099/mic.0.26974-0](https://doi.org/10.1099/mic.0.26974-0).
- ROCHA, Eduardo P. C. (1<sup>er</sup> juin 2018). « Neutral Theory, Microbial Practice : Challenges in Bacterial Population Genetics ». In : *Mol Biol Evol* 35.6, p. 1338-1347. DOI : [10.1093/molbev/msy078](https://doi.org/10.1093/molbev/msy078).
- ROCHA, Eduardo P. C. et Antoine DANCHIN (1<sup>er</sup> juin 2002). « Base Composition Bias Might Result from Competition for Metabolic Resources ». In : *Trends in Genetics* 18.6, p. 291-294. DOI : [10.1016/S0168-9525\(02\)02690-2](https://doi.org/10.1016/S0168-9525(02)02690-2).
- ROCHA, Eduardo P. C. et Edward J. FEIL (9 sept. 2010). « Mutational Patterns Cannot Explain Genome Composition : Are There Any Neutral Sites in the Genomes of Bacteria? ». In : *PLoS Genet* 6.9, e1001104. DOI : [10.1371/journal.pgen.1001104](https://doi.org/10.1371/journal.pgen.1001104).



- ROCHA, Eduardo P. C., Marie TOUCHON et Edward J. FEIL (12 jan. 2006). « Similar Compositional Biases Are Caused by Very Different Mutational Effects ». In : *Genome Res.* 16.12, p. 1537-1547. DOI : [10.1101/gr.5525106](https://doi.org/10.1101/gr.5525106). pmid : 17068325.
- ROMERO, Héctor et al. (1<sup>er</sup> août 2009). « Oxygen and Guanine–Cytosine Profiles in Marine Environments ». In : *J Mol Evol* 69.2, p. 203-206. DOI : [10.1007/s00239-009-9230-9](https://doi.org/10.1007/s00239-009-9230-9).
- ROSA, Hartmut (29 août 2013). *Accélération : Une critique sociale du temps*. La Découverte. 486 p.
- RUDNER, R., J. D. KARKAS et E. CHARGAFF (1<sup>er</sup> juin 1968a). « Separation of B. Subtilis DNA into Complementary Strands. I. Biological Properties ». In : *PNAS* 60.2, p. 630-635. DOI : [10.1073/pnas.60.2.630](https://doi.org/10.1073/pnas.60.2.630). pmid : 4973487.
- (1<sup>er</sup> juil. 1968b). « Separation of B. Subtilis DNA into Complementary Strands. III. Direct Analysis ». In : *PNAS* 60.3, p. 921-922. DOI : [10.1073/pnas.60.3.921](https://doi.org/10.1073/pnas.60.3.921). pmid : 4970114.
- RUSSELL, Bertrand (18 jan. 2011). *Essais sceptiques*. Réd. par Jean-Claude ZYLBERSTEIN. Trad. par André BERNARD. Le goût des idées. Paris : Les Belles Lettres. 264 p.
- SAMRAKANDI, Moulay Mustapha et Franck PASTA (15 juin 2000). « Hyperrecombination in Streptococcus Pneumoniae Depends on an Atypical mutY Homologue ». In : *J. Bacteriol.* 182.12, p. 3353-3360. DOI : [10.1128/JB.182.12.3353-3360.2000](https://doi.org/10.1128/JB.182.12.3353-3360.2000). pmid : 10852864.
- SAMTOOLS DEVELOPPERS (2019). *The Variant Call Format Specification (v4.2)*.
- SAPP, Jan (3 mar. 2005). *Microbial Phylogeny and Evolution : Concepts and Controversies*. Oxford University Press, USA. 349 p.
- SCHANZ, Silvia et al. (7 avr. 2009). « Interference of Mismatch and Base Excision Repair during the Processing of Adjacent U/G Mispairs May Play a Key Role in Somatic Hypermutation ». In : *PNAS* 106.14, p. 5593-5598. DOI : [10.1073/pnas.0901726106](https://doi.org/10.1073/pnas.0901726106). pmid : 19307563.
- SCHILDKRAUT, Carl L., Julius MARMUR et Paul DOTY (1<sup>er</sup> juin 1962). « Determination of the Base Composition of Deoxyribonucleic Acid from Its Buoyant Density in CsCl ». In : *Journal of Molecular Biology* 4.6, p. 430-443. DOI : [10.1016/S0022-2836\(62\)80100-4](https://doi.org/10.1016/S0022-2836(62)80100-4).
- SCHROEDER, Jeremy W. et al. (7 mar. 2016). « The Effect of Local Sequence Context on Mutational Bias of Genes Encoded on the Leading and Lagging Strands ». In : *Current Biology* 26.5, p. 692-697. DOI : [10.1016/j.cub.2016.01.016](https://doi.org/10.1016/j.cub.2016.01.016). pmid : 26923786.
- SCOTT, Trey J. et David C. QUELLER (4 sept. 2019). « Long-term Evolutionary Conflict, Sisyphian Arms Races, and Power in Fisher's Geometric Model ». In : *Ecol. Evol.* DOI : [10.1002/ece3.5625](https://doi.org/10.1002/ece3.5625).
- SEBALD, M., F. GASSER et H. WERNER (1965). « Teneur GC% et Classification. Application Au Groupe Des Bifidobacteries et à Quelques Genres Voisins ». In : *Annales de l'Institut Pasteur*. T. 109. 120 Blvd Saint-Germain, 75280 Paris 06, France : Masson Editeur, p. 251-&.
- SHEPPARD, Samuel K., David S. GUTTMAN et J. Ross FITZGERALD (4 juil. 2018). « Population Genomics of Bacterial Host Adaptation ». In : *Nat. Rev. Genet.* P. 1. DOI : [10.1038/s41576-018-0032-z](https://doi.org/10.1038/s41576-018-0032-z).
- SMEDS, Linnéa et al. (24 mai 2016). « High-Resolution Mapping of Crossover and Non-Crossover Recombination Events by Whole-Genome Re-Sequencing of an Avian Pedigree ». In : *PLOS Genetics* 12.5, e1006044. DOI : [10.1371/journal.pgen.1006044](https://doi.org/10.1371/journal.pgen.1006044).

## BIBLIOGRAPHIE

---

- SOUICY, Shannon M., Jinling HUANG et Johann Peter GOGARTEN (août 2015). « Horizontal Gene Transfer : Building the Web of Life ». In : *Nat. Rev. Genet.* 16.8, p. 472-482. DOI : [10.1038/nrg3962](https://doi.org/10.1038/nrg3962).
- SPIES, Maria et Richard FISHEL (3 jan. 2015). « Mismatch Repair during Homologous and Homeologous Recombination ». In : *Cold Spring Harb Perspect Biol* 7.3, a022657. DOI : [10.1101/cshperspect.a022657](https://doi.org/10.1101/cshperspect.a022657). pmid : 25731766.
- STUMPF, Michael P. H. et Gilean A. T. MCVEAN (déc. 2003). « Estimating Recombination Rates from Population-Genetic Data ». In : *Nat Rev Genet* 4.12, p. 959-968. DOI : [10.1038/nrg1227](https://doi.org/10.1038/nrg1227).
- SUÁREZ, Gabriel A. et al. (9 jan. 2017). « Reduced Mutation Rate and Increased Transformability of Transposon-Free *Acinetobacter Baylyi* ADP1-ISx ». In : *Appl. Environ. Microbiol.* 83.17, e01025-17. DOI : [10.1128/AEM.01025-17](https://doi.org/10.1128/AEM.01025-17). pmid : 28667117.
- SUEOKA, Noboru (1<sup>er</sup> fév. 1961). « Variation and Heterogeneity of Base Composition of Deoxyribonucleic Acids : A Compilation of Old and New Data ». In : *Journal of Molecular Biology* 3.1, 31-IN15. DOI : [10.1016/S0022-2836\(61\)80005-3](https://doi.org/10.1016/S0022-2836(61)80005-3).
- (15 avr. 1962). « On the Genetic Basis of Variation and Heterogeneity of DNA Base Composition ». In : *Proc. Natl. Acad. Sci. U.S.A.* 48, p. 582-592. pmid : 13918161.
- (avr. 1988). « Directional Mutation Pressure and Neutral Molecular Evolution ». In : *PNAS* 85.8, p. 2653-2657. pmid : 3357886.
- (1<sup>er</sup> août 1993). « Directional Mutation Pressure, Mutator Mutations, and Dynamics of Molecular Evolution ». In : *J Mol Evol* 37.2, p. 137. DOI : [10.1007/BF02407349](https://doi.org/10.1007/BF02407349).
- (1<sup>er</sup> mar. 1995). « Intrastrand Parity Rules of DNA Base Composition and Usage Biases of Synonymous Codons ». In : *J Mol Evol* 40.3, p. 318-325. DOI : [10.1007/BF00163236](https://doi.org/10.1007/BF00163236).
- SUNG, Way, Matthew S. ACKERMAN, Jean-François GOUT et al. (1<sup>er</sup> juil. 2015). « Asymmetric Context-Dependent Mutation Patterns Revealed through Mutation-Accumulation Experiments ». In : *Mol Biol Evol* 32.7, p. 1672-1683. DOI : [10.1093/molbev/msv055](https://doi.org/10.1093/molbev/msv055).
- SUNG, Way, Matthew S. ACKERMAN, Samuel F. MILLER et al. (6 nov. 2012). « Drift-Barrier Hypothesis and Mutation-Rate Evolution ». In : *Proc. Natl. Acad. Sci.* 109.45, p. 18488-18492. DOI : [10.1073/pnas.1216223109](https://doi.org/10.1073/pnas.1216223109).
- SYMINGTON, Lorraine S. (8 jan. 2014). « End Resection at Double-Strand Breaks : Mechanism and Regulation ». In : *Cold Spring Harb Perspect Biol* 6.8, a016436. DOI : [10.1101/cshperspect.a016436](https://doi.org/10.1101/cshperspect.a016436). pmid : 25085909.
- TENESA, Albert et al. (4 jan. 2007). « Recent Human Effective Population Size Estimated from Linkage Disequilibrium ». In : *Genome Res.* 17.4, p. 520-526. DOI : [10.1101/gr.6023607](https://doi.org/10.1101/gr.6023607). pmid : 17351134.
- THAM, Khek-Chian, Nicolaas HERMANS et al. (8 août 2013). « Mismatch Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA Structures ». In : *Molecular Cell* 51.3, p. 326-337. DOI : [10.1016/j.molcel.2013.07.008](https://doi.org/10.1016/j.molcel.2013.07.008).
- THAM, Khek-Chian, Roland KANAAR et Joyce H. G. LEBBINK (1<sup>er</sup> fév. 2016). « Mismatch Repair and Homeologous Recombination ». In : *DNA Repair. DNA Mismatch Repair* 38, p. 75-83. DOI : [10.1016/j.dnarep.2015.11.010](https://doi.org/10.1016/j.dnarep.2015.11.010).
- TOUCHON, Marie et al. (23 jan. 2009). « Organised Genome Dynamics in the *Escherichia Coli* Species Results in Highly Diverse Adaptive Paths ». In : *PLOS Genet* 5.1, e1000344. DOI : [10.1371/journal.pgen.1000344](https://doi.org/10.1371/journal.pgen.1000344).
- TREANGEN, Todd J. et al. (2014). « The Harvest Suite for Rapid Core-Genome Alignment and Visualization of Thousands of Intraspecific Microbial Genomes ». In : *Genome Biology* 15, p. 524. DOI : [10.1186/s13059-014-0524-x](https://doi.org/10.1186/s13059-014-0524-x).
- VAN DER AUWERA, Geraldine A. et al. (2013). « From FastQ Data to High Confidence Variant Calls : The Genome Analysis Toolkit Best Practices Pipeline ». In : *Curr*

- Protoc Bioinformatics* 43, p. 11.10.1-33. DOI : [10.1002/0471250953.bi1110s43](https://doi.org/10.1002/0471250953.bi1110s43). pmid : 25431634.
- VANEECHOUTTE, M. et al. (1<sup>er</sup> jan. 2006). « Naturally Transformable *Acinetobacter* Sp. Strain ADP1 Belongs to the Newly Described Species *Acinetobacter* Baylyi ». In : *Appl. Environ. Microbiol.* 72.1, p. 932-936. DOI : [10.1128/AEM.72.1.932-936.2006](https://doi.org/10.1128/AEM.72.1.932-936.2006).
- VIEIRA-SILVA, Sara et Eduardo P. C. ROCHA (1<sup>er</sup> sept. 2008). « An Assessment of the Impacts of Molecular Oxygen on the Evolution of Proteomes ». In : *Mol Biol Evol* 25.9, p. 1931-1942. DOI : [10.1093/molbev/msn142](https://doi.org/10.1093/molbev/msn142).
- VISCHER, Ernst, Stephen ZAMENHOF et Erwin CHARGAFF (1<sup>er</sup> jan. 1949). « Microbial Nucleic Acids : The Desoxypentose Nucleic Acids of Avian Tubercle Bacilli and Yeast ». In : *J. Biol. Chem.* 177.1, p. 429-438. pmid : 18107446.
- VOS, Michiel et Xavier DIDELOT (2 oct. 2009). « A Comparison of Homologous Recombination Rates in Bacteria and Archaea ». In : *ISME J* 3.2, p. 199-208. DOI : [10.1038/ismej.2008.93](https://doi.org/10.1038/ismej.2008.93).
- WAGNER, R. et M. MESELSON (nov. 1976). « Repair Tracts in Mismatched DNA Heteroduplexes ». In : *Proc. Natl. Acad. Sci. U.S.A.* 73.11, p. 4135-4139. pmid : 1069303.
- WALKER, Bruce J. et al. (19 nov. 2014). « Pilon : An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement ». In : *PLOS ONE* 9.11, e112963. DOI : [10.1371/journal.pone.0112963](https://doi.org/10.1371/journal.pone.0112963).
- WALSH, James Bruce (1<sup>er</sup> oct. 1983). « Role of Biased Gene Conversion in One-Locus Neutral Theory and Genome Evolution ». In : *Genetics* 105.2, p. 461-468. pmid : 17246166.
- WANG, Harris H. et al. (13 août 2009). « Programming Cells by Multiplex Genome Engineering and Accelerated Evolution ». In : *Nature* 460.7257, p. 894-898. DOI : [10.1038/nature08187](https://doi.org/10.1038/nature08187).
- WANG, Huai-Chun, Edward SUSKO et Andrew J. ROGER (14 avr. 2006). « On the Correlation between Genomic G+C Content and Optimal Growth Temperature in Prokaryotes : Data Quality and Confounding Factors ». In : *Biochemical and Biophysical Research Communications* 342.3, p. 681-684. DOI : [10.1016/j.bbrc.2006.02.037](https://doi.org/10.1016/j.bbrc.2006.02.037).
- WATSON, J. D. et F. H. C. CRICK (avr. 1953). « Molecular Structure of Nucleic Acids : A Structure for Deoxyribose Nucleic Acid ». In : *Nature* 171.4356, p. 737-738. DOI : [10.1038/171737a0](https://doi.org/10.1038/171737a0).
- WATSON, James D., éd. (2014). *Molecular Biology of the Gene*. Seventh edition. Boston : Pearson. 872 p.
- WATTERSON, G. A. (1<sup>er</sup> avr. 1975). « On the Number of Segregating Sites in Genetical Models without Recombination ». In : *Theoretical Population Biology* 7.2, p. 256-276. DOI : [10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9).
- WEBER, Claudia C. et al. (11 déc. 2014). « Evidence for GC-Biased Gene Conversion as a Driver of between-Lineage Differences in Avian Base Composition ». In : *Genome Biology* 15.12, p. 549. DOI : [10.1186/s13059-014-0549-1](https://doi.org/10.1186/s13059-014-0549-1).
- WEISSMAN, Jake L., William F. FAGAN et Philip L. F. JOHNSON (8 fév. 2019). « Potential Link between Selection for High GC Content and Repair of Double Strand Breaks in Prokaryotic Genomes ». In : *bioRxiv*, p. 544924. DOI : [10.1101/544924](https://doi.org/10.1101/544924).
- WESOŁOWSKA-ANDERSEN, Agata et al. (5 juin 2014). « Choice of Bacterial DNA Extraction Method from Fecal Material Influences Community Structure as Evaluated by Metagenomic Analysis ». In : *Microbiome* 2.1, p. 19. DOI : [10.1186/2049-2618-2-19](https://doi.org/10.1186/2049-2618-2-19).
- WICK, Ryan R. et al. (8 juin 2017). « Unicycler : Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads ». In : *PLOS Computational Biology* 13.6, e1005595. DOI : [10.1371/journal.pcbi.1005595](https://doi.org/10.1371/journal.pcbi.1005595).

## BIBLIOGRAPHIE

---

- WIELGOSS, Sébastien et al. (1<sup>er</sup> août 2011). « Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With *Escherichia Coli* ». In : *G3 Genes Genomes Genet.* 1.3, p. 183-186. DOI : [10.1534/g3.111.000406](https://doi.org/10.1534/g3.111.000406). pmid : [22207905](https://pubmed.ncbi.nlm.nih.gov/22207905/).
- WILLIAMS, Amy L. et al. (25 mar. 2015). « Non-Crossover Gene Conversions Show Strong GC Bias and Unexpected Clustering in Humans ». In : *eLife* 4, e04637. DOI : [10.7554/eLife.04637](https://doi.org/10.7554/eLife.04637). pmid : [25806687](https://pubmed.ncbi.nlm.nih.gov/25806687/).
- WILLIAMS, George C. (30 oct. 2018). *Adaptation and Natural Selection*. Avec la coll. de Richard DAWKINS. Princeton University Press. 335 p.
- WITTGENSTEIN, Ludwig (1972). *Tractatus Logico-Philosophicus*. Avec la coll. de Bertrand RUSSELL. Trad. par Gilles G. GRANGER. 2001<sup>re</sup> éd. Tel 311. Gallimard. 128 p.
- WOESE, Carl R. (6 jan. 2004). « A New Biology for a New Century ». In : *Microbiol. Mol. Biol. Rev.* 68.2, p. 173-186. DOI : [10.1128/MMBR.68.2.173-186.2004](https://doi.org/10.1128/MMBR.68.2.173-186.2004). pmid : [15187180](https://pubmed.ncbi.nlm.nih.gov/15187180/).
- WU, Chung-I. et Nobuyo MAEDA (mai 1987). « Inequality in Mutation Rates of the Two Strands of DNA ». In : *Nature* 327.6118, p. 169. DOI : [10.1038/327169a0](https://doi.org/10.1038/327169a0).
- WU, Hao et al. (2012). « On the Molecular Mechanism of GC Content Variation among Eubacterial Genomes ». In : *Biol Direct* 7.2, p. 6150-7.
- XIA, Xuhua et al. (1<sup>er</sup> août 2002). « Genomic Changes in Nucleotide and Dinucleotide Frequencies in *Pasteurella Multocida* Cultured Under High Temperature ». In : *Genetics* 161.4, p. 1385-1394. pmid : [12196387](https://pubmed.ncbi.nlm.nih.gov/12196387/).
- YAHARA, Koji et al. (2 jan. 2016). « The Landscape of Realized Homologous Recombination in Pathogenic Bacteria ». In : *Mol Biol Evol* 33.2, p. 456-471. DOI : [10.1093/molbev/msv237](https://doi.org/10.1093/molbev/msv237). pmid : [26516092](https://pubmed.ncbi.nlm.nih.gov/26516092/).
- YIN, Yi et Thomas D. PETES (1<sup>er</sup> août 2014). « The Role of Exo1p Exonuclease in DNA End Resection to Generate Gene Conversion Tracts in *Saccharomyces Cerevisiae* ». In : *Genetics* 197.4, p. 1097-1109. DOI : [10.1534/genetics.114.164517](https://doi.org/10.1534/genetics.114.164517). pmid : [24835424](https://pubmed.ncbi.nlm.nih.gov/24835424/).
- YOUNG, David M. et L. Nicholas ORNSTON (12 jan. 2001). « Functions of the Mismatch Repair Gene *mutS* from *Acinetobacter Sp.* Strain ADP1 ». In : *J. Bacteriol.* 183.23, p. 6822-6831. DOI : [10.1128/JB.183.23.6822-6831.2001](https://doi.org/10.1128/JB.183.23.6822-6831.2001). pmid : [11698371](https://pubmed.ncbi.nlm.nih.gov/11698371/).
- ZAMENHOF, Stephen, George BRAWERMAN et Erwin CHARGAFF (1<sup>er</sup> jan. 1952). « On the Desoxyribose Nucleic Acids from Several Microorganisms ». In : *Biochimica et Biophysica Acta* 9, p. 402-405. DOI : [10.1016/0006-3002\(52\)90184-4](https://doi.org/10.1016/0006-3002(52)90184-4).
- ZHAO, Jingyong et Malcolm E. WINKLER (9 jan. 2000). « Reduction of GC →TA Transversion Mutation by Overexpression of MutS in *Escherichia Coli* K-12 ». In : *J. Bacteriol.* 182.17, p. 5025-5028. DOI : [10.1128/JB.182.17.5025-5028.2000](https://doi.org/10.1128/JB.182.17.5025-5028.2000). pmid : [10940054](https://pubmed.ncbi.nlm.nih.gov/10940054/).

# TABLE DES MATIÈRES

---

<b>I Introduction</b>	<b>1</b>
<b>1 Prologue</b>	<b>3</b>
<b>2 Aspects historiques du GC%</b>	<b>9</b>
2.1 Découvertes des acides nucléiques . . . . .	11
Haeckel, Hoppe–Seyler et la chimie physiologique · 12, Friedrich Miescher, la découverte de la <i>nucléine</i> et des acides nucléiques · 13, Albrecht Kossel, caractérisation des cinq bases nucléotidiques · 15, Phoebus Levene, l’hypothèse tétranucléotidique · 16, Erwin Chargaff, vers la quantification du GC% · 18	
2.2 Expliquer la variabilité du GC%? . . . . .	20
GC% et taxonomie moléculaire · 20, Théorie neutraliste de la composition en base · 22, L’hypothèse de la balance mutationnelle · 23, Séquençage et composition en base · 25	
<b>3 Distribution des taux de GC entre génomes</b>	<b>27</b>
3.1 Diversité inter-génomique . . . . .	27
Distribution phylogénétique du GC% · 28	
3.2 Diversité intra-génomique . . . . .	29
<b>4 Processus mutationnels</b>	<b>35</b>
4.1 Balance mutationnelle . . . . .	36
Analyse du polymorphisme · 36, Expériences d’accumulation de mutations · 38, Biais de fixation en faveur de GC · 40	
4.2 Déterminants de la balance mutationnelle . . . . .	42
Correction des mésappariements : mécanismes · 43, Mutagénèse et mécanismes mutateurs · 46, Approches pour l’étude du spectre mutationnel · 47, Spectre mutationnel de mutants MMR · 48, Spectre mutationnel de mutants BER · 50, Biais mutationnels vers GC · 51, Environnement, stress et mutation · 51, Asymétrie entre brins liée à la réplication · 53	
4.3 Distribution phylogénétique des gènes impliqués . . . . .	58
Conclusion · 60	
<b>5 Sélection sur la composition en base (?)</b>	<b>63</b>
5.1 Corrélations GC% et traits environnementaux . . . . .	63
Température optimale de croissance · 63, Environnements différents et taux de GC différents? · 65, Réquisits en oxygène · 67, Ressources métaboliques et fitness · 68, Résumé · 69	
5.2 Indices expérimentaux de sélection sur le GC% . . . . .	70
Taux de croissance, GFP et GC% · 70, Et les autres modèles? · 70	
<b>6 Conversion Génique Biaisée : quelle importance?</b>	<b>73</b>
Conversion génique : définitions & mécanismes · 75	
6.1 Sélection, dérive & recombinaison . . . . .	76

## TABLE DES MATIÈRES

---

	Modèles théoriques de l'importance évolutive des biais de conversion génique · 76, Preuves empiriques des conséquences évolutives de la conversion génique biaisée vers GC : évolution de la structure en isochore chez les mammifères · 79, La gBGC peut passer pour de (et s'opposer à) la sélection · 82, Estimations de l'intensité de la gBGC · 83, Explications proposées pour l'origine évolutive de la gBGC · 84, Résumés, limites et extensions · 85	
6.2	Indices de la gBGC chez les Procaryotes . . . . .	86
	Estimation indirectes du taux de recombinaison au niveau de la population dans diverses espèces de Procaryotes · 86, L'absence de relation entre le ratio $\tau/\mu$ de l'espèce et la composition en base du génome n'est pas un indicateur de l'absence de gBGC · 89, Corrélations GC% et taux de Recombinaison · 90, Conclusions · 92	
<b>II</b>	<b>Quantification de la gBGC chez les Bactéries</b>	<b>95</b>
<b>7</b>	<b>Mesure des fréquences de conversion</b>	<b>97</b>
7.1	Comment faire? . . . . .	97
	Transformation naturelle à l'échelle génomique · 98	
7.2	Transformation génomique d' <i>Acinetobacter baylyi</i> . . . . .	99
	Séquençage du génome de la souche donneuse · 99, Expérience de transformation génomique chez <i>Acinetobacter baylyi</i> · 101	
7.3	Transformation génomique de <i>Bacillus subtilis</i> . . . . .	105
	Souches receveuse et donneuse · 105, Expérience de transformation génomique chez <i>Bacillus subtilis</i> · 107	
7.4	Mesure des fréquences de conversion . . . . .	109
	<i>Variant calling</i> : détection des sites variants · 109, <i>Conversion calling</i> : détection des sites convertis · 109, Obtention des régions converties chez <i>Bacillus subtilis</i> · 112	
7.5	Ré-analyse des données publiques . . . . .	113
	Transformation génomique chez <i>Helicobacter pylori</i> · 114, Transformation génomique chez <i>Streptococcus pneumoniae</i> · 115, Transformation génomique chez <i>Haemophilus influenzae</i> · 116, Transformation génomique chez <i>Neisseria meningitidis</i> · 118, Résumé · 120	
<b>8</b>	<b>Démonstration de la gBGC chez des Bactéries</b>	<b>121</b>
8.1	Répartition des régions converties sur le génome . . . . .	123
	<i>Acinetobacter baylyi</i> · 123, Distribution de la longueur des régions converties · 125, Répartition et description des régions converties chez <i>Bacillus subtilis</i> · 128	
8.2	<i>A. baylyi</i> : Preuves de gBGC dans les conversions isolées .	128
	Excès de transversions dans les conversions isolées · 130, Absence d'asymétrie liée à la réplication dans les conversions · 131, Résumé · 133	
8.3	Preuves de gBGC dans deux autres espèces de Bactéries .	133
8.4	Intensité populationnelle de la gBGC . . . . .	135
	Estimation de l'incertitude autour de B · 138	
<b>9</b>	<b>Discussions</b>	<b>143</b>
9.1	Est-ce que la sélection <i>cause</i> le biais en faveur de GC? . .	143
9.2	Les mécanismes responsables de la gBGC . . . . .	145

Proposition d'une hypothèse pour le spectre des conversions isolées biaisé en faveur des bases G et C · 146, Proposition d'une hypothèse pour le spectre des extrémités des régions converties biaisées en faveur des bases G et C · 150	
9.3 Détecter la gBGC : une tâche délicate . . . . .	154
9.4 La nature convergente de la gBGC . . . . .	155
<b>Perspectives et conclusions générales</b>	<b>159</b>
<b>Appendices</b>	<b>I</b>
<b>Taux de croissance et taux de GC%</b>	<b>III</b>
<b>Bibliographie</b>	<b>XI</b>
<b>Table des matières</b>	<b>XXXI</b>