



Année : 2015-2016

Étude expérimentale de la  
conversion génique en faveur de GC  
au cours de la recombinaison homologue  
chez *Acinetobacter baylyi*

Nom, prénom :

BARRETO Samuel

Laboratoire d'accueil :

UMR5557, Laboratoire d'Écologie Microbienne

UMR5558, Laboratoire de Biométrie et Biologie Évolutive

Nom du Maître de stage :

BERTOLLA Franck

DURET Laurent

Composition du jury :

CONDEMINÉ Guy

LAURENT Frédéric

LEMAIRE Marc

LOUKIADIS Estelle

MOENNE-LOCCOZ Yvan

OTTMANN Michèle

POTIER Patrick

RICHAUME Agnès



Université Claude Bernard



Lyon 1





## Remerciements

En premier lieu, je tiens à remercier messieurs Yvan Moëgne-Loccoz et Manolo Gouy, responsables respectifs de mes deux terrains de stage, qui rendent nos stages possibles à grands coups de planning surchargés et de réunions sans fin.

Je remercie également l'ensemble de l'équipe pédagogique du Master Écosciences et Microbiologie, qui a réussi à faire de ce master un lieu où il fait bon apprendre, malgré les coupes budgétaires et contraintes administratives en tout genre.

Viennent ensuite celles et ceux qui rendent notre vie à la paillasse beaucoup plus simple : Mme Céline Lavire, qui donne beaucoup de son temps pour animer cette équipe éparpillée aux quatre coins de la Doua ; M. Xavier Nesme ; Audrey Dubost, qui prend soin de notre santé au Mendel ; Corinne Sannaire, qui prend soin de la santé de nos paillasses et nous évite bien des heures à remplir des boîtes de cônes ; David Chapuliot, qui prend soin de laisser ses consommables bien en vue et nous évite ainsi d'avoir à utiliser les nôtres ; Raphaël Masse, qui prend soin de laisser nos consommables pas très très bien en vue et s'évite ainsi d'avoir à utiliser les siens ; Aurélie, qui prend soin de notre sens moral et nous incite à passer très vite devant les rayons boucherie ; Thibault, qui ne prend pas du tout soin de notre sens olfactif ; Rosa, qui n'accorde pas plus d'attention à nos oreilles que Thibault à nos narines ; Quentin, qui prend soin de nous, petits étudiants ; Yoann, qui prend soin d'eux, vieux thésards ; Jordan Senior, qui accorde un soin tout particulier à être le moins drôle possible ; tous les thésards qu'on a pu côtoyer, et enfin les M2, qui ont rendu ce stage très agréable à force de croissants, miels et thés en tout genre.

Viennent après ceux qui rendent ma vie à moi beaucoup plus simple : ma Lu, qui me supporte depuis bientôt trois ans et *a fortiori* depuis janvier ; ma grand-mère, à qui je dois beaucoup ; ma mère, qui me supporte depuis vingt trois ans et sans qui je ne serais pas grand-chose ; mon père, que j'espère rendre fier de là haut, tu m'as appris à "ne pas me satisfaire de ne pas comprendre le monde", pour citer Dawkins.

Viennent enfin ceux qui ont rendu ma vie beaucoup plus dure : Vincent, que je remercierai dans un style qu'il comprendra certainement et qui, alors que tout le prédisposait à être impliqué dans ce projet à 100%, n'a montré qu'un dédain souverain et un manque d'investissement à la limite de la bienséance, manifeste flagrant du peu d'intérêt qu'il porte à la Science ; Franck,

pour ses conseils et ses coups de main à la paillasse, ses relectures avisées, ses critiques détaillées, perfectionnistes mais toujours constructives, et surtout sa persévérance à mes côtés lorsque tout le poussait à rester chez lui pour se soigner ; et enfin Laurent, qui devra attendre encore un peu.



# Sommaire

<b>Introduction</b>	<b>1</b>
<b>1 Matériels &amp; Méthodes</b>	<b>3</b>
1.1 Conception des séquences de synthèse . . . . .	3
1.2 Constructions des plasmides . . . . .	3
1.3 Transformations d' <i>Acinetobacter baylyi</i> . . . . .	4
1.4 Alignements . . . . .	5
<b>2 Résultats</b>	<b>6</b>
2.1 Validation de la démarche expérimentale . . . . .	6
2.2 Validation de la qualité des séquences . . . . .	6
2.3 Description des régions recombinantes . . . . .	7
2.4 Comparaison de la longueur des régions converties . . . . .	8
2.5 Restaurations de l'haplotype sauvage . . . . .	8
2.6 Estimation des fréquences de conversion en faveur de GC . . . . .	9
<b>3 Discussions</b>	<b>10</b>
3.1 Distribution des points de recombinaison . . . . .	10
3.2 Des preuves de l'existence d'une conversion génique biaisée vers GC chez <i>A. baylyi</i> . . . . .	11
3.3 Conséquences évolutives de la conversion biaisée vers GC . . . . .	13
<b>Conclusion</b>	<b>15</b>
<b>1 Annexes</b>	<b>ii</b>
1.2 Amorces utilisées . . . . .	ii
1.3 Conditions de PCR . . . . .	ii
1.4 Carte des constructions donneuses . . . . .	ii
1.5 Confirmations des restaurations des haplotypes sauvages . . . . .	iii
1.6 Traces de conversions . . . . .	iii



# Abbréviations

**BER** Base Excision Repair

**gBGC** GC-biased gene conversion : conversion génique biaisée vers GC

**GC%** Contenu en GC

**MMR** Mismatch repair : machinerie de réparation des mésappariements

**VSP** Very Short Patch repair : réparations courtes portées

## Table des figures

1	Mécanismes moléculaires de conversion génique . . . . .	1
2	Constructions moléculaires . . . . .	2
3	Protocoles de transformation . . . . .	3
4	Analyses des zones de recombinaison . . . . .	4
5	Marqueur montrant des traces de contaminations . . . . .	5
6	Des contaminations dues au hasard ? . . . . .	5
7	Zones de recombinaison détaillée . . . . .	6
8	Distribution de la position du dernier marqueur . . . . .	7
9	Taux de GC et Taux de Recombinaison . . . . .	9
10	Liste des amorces utilisées . . . . .	ii
11	Confirmation des restaurations . . . . .	iii
12	Régions converties . . . . .	iv

## Liste des tableaux

1	Fréquences de transformation . . . . .	5
2	Dénombrement des cas de restauration . . . . .	7
3	Dénombrement des cas de conversion . . . . .	8





# Introduction

Chez les procaryotes, le taux de guanine et cytosine (GC%) varie de 16,5 % à 75 % dans les génomes séquencés à ce jour. Est-ce que ces différences de composition en base des génomes résultent de pressions de sélection ou est-ce qu'elles résultent de processus non-adaptatif ? Autrement dit, est-ce le taux de GC joue un rôle dans le bon fonctionnement de l'organisme, ou est-ce qu'il est déterminé uniquement par des processus neutres, résultant de variations des patrons de mutations ou de réparation de l'ADN ? L'un des mécanismes essentiels de réparation de l'ADN est la recombinaison homologe, impliquée dans la réparation des cassures doubles brins. Au cours de la réparation de l'ADN par recombinaison homologe, deux molécules différentes d'ADN monocaténares sont appariées au sein d'un hétéroduplex. Ces deux brins appariés peuvent présenter des mésappariements. Leur correction entraîne de la *conversion génique*<sup>1</sup>, c'est-à-dire une transmission à sens unique de l'information portée par un brin vers l'autre. Chez les mammifères et probablement chez un grand nombre d'eucaryotes, la correction des mésappariements est biaisée vers l'introduction des bases G et C<sup>2</sup>. Le biais de conversion génique vers GC (gBGC<sup>a</sup>) tend à augmenter localement le taux de GC des régions fortement recombinantes<sup>3,4,5</sup>. C'est un processus similaire à la sélection puisqu'il augmente la probabilité de fixation des allèles G et C au fil des générations. Il est cependant non-adaptatif car il agit indépendamment de l'éventuel effet des allèles sur la fitness de l'individu.

En 2010, deux études simultanées<sup>6,7</sup> ont démontré chez les procaryotes que les bases G et C ont une probabilité de fixation plus élevée, probablement sous l'effet d'un processus à l'action semblable à celle de la sélection naturelle. Hildebrand *et al.* ont avancé que le GC% est en soi un trait soumis à la sélection, et ont rejeté l'hypothèse d'un biais de conversion génique biaisé vers GC chez les procaryotes. Cependant, les analyses récentes de Lassalle *et al.*<sup>8</sup> et de Yahara *et al.*<sup>9</sup> ont montré que les zones recombinantes des génomes procaryotes ont un taux de GC plus élevé que les régions non-recombinantes. Elles ont également montré que la fixation préférentielle des allèles GC va à l'encontre de la fixation des allèles optimaux des codons. Ces observations correspondent précisément aux prédictions du modèle gBGC. Ces résultats sont importants car ils suggèrent que le gBGC pourrait être un processus universel affectant tous les organismes sexués ou parasexués. Cependant, ces observations sont basées sur des corrélations et ne permettent pas de démontrer formellement que le gBGC est à l'œuvre chez les procaryotes.

a. gBGC pour GC-Biased Gene Conversion

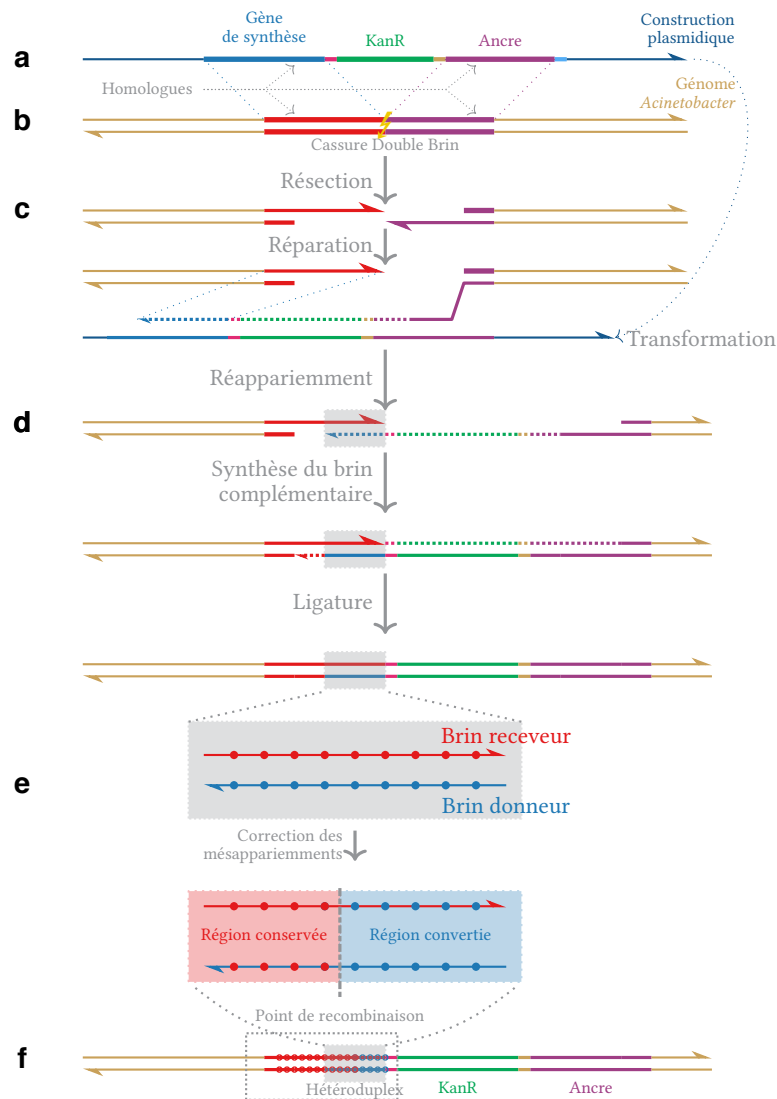


Figure 1 : **Mécanismes de conversion génique au cours de la transformation naturelle par les constructions plasmidiques**

Les constructions plasmidiques que nous avons conçues (a) comportent un gène de synthèse introduisant des mésappariements avec la séquence sauvage receveuse, une cassette de résistance à la kanamycine (KanR) et une zone très recombino-gène (Ancre) qui permet d'augmenter la fréquence de transformation.

Lors d'une cassure double brin (b), l'extrémité de la cassure subit une résection, qui est réparée sur la base d'une matrice d'ADN homologue (c). La recombinaison homologue permet de réparer ces cassures en utilisant l'ADN exogène comme matrice. Le réappariement (d) avec le brin parental conduit à la formation d'un hétéroduplex (zone grisée) entre le donneur (en bleu) et le receveur (en rouge). Les mésappariements au sein de cet hétéroduplex sont réparés (e), du donneur vers le receveur dans la région convertie, du receveur vers le donneur dans la région conservée.

Nous avons séquencé la région encadrée (f) chez un grand nombre de recombinants.

L'objectif de ce travail était de mettre en place un protocole expérimental permettant de quantifier les fréquences de conversion vers GC chez les bactéries. Nous avons choisi comme modèle la bactérie *Acinetobacter baylyi* ADP1<sup>a</sup> pour deux raisons. 1) Le genre *Acinetobacter* montre une corrélation entre taux de GC et taux de recombinaison<sup>8</sup> qui suggère que ce genre pourrait être soumis au gBGC. 2) Cette bactérie est naturellement transformable : elle développe un stade de compétence pendant la phase exponentielle de croissance au cours duquel elle intègre de l'ADN exogène monocaténaire par un mécanisme contrôlé génétiquement<sup>11</sup>. Cet ADN peut être intégré dans le génome par recombinaison homologue. Nous avons utilisé ces propriétés pour forcer la recombinaison homologue à un locus neutre, en introduisant artificiellement des mésappariements entre la séquence donneuse — nos constructions — et la séquence receveuse — le génome (voir figure 1). La réparation de ces mésappariements donne lieu à la conversion génique d'un brin par l'autre. On distingue donc la région convertie, dont le génotype correspond à celui de l'haplotype donneur, de la région conservée, dont le génotype correspond à celui de l'haplotype receveur. Elles sont séparées par le point de recombinaison (voir figure 1). Les mésappariements présents sur la séquence donneuse permettent 1) de localiser la position du point de recombinaison et 2) de déterminer les fréquences de conversion en faveur de GC chez *A. baylyi*.

Nous avons validé le protocole expérimental et obtenu 384 transformants, dont l'analyse a permis de décrire les régions de conversion et de quantifier les fréquences de conversion en faveur de GC. Nous avons trouvé un excès significatif de transmission des allèles G et C par rapport aux allèles A et T.

a. *A. baylyi* était appelée *A. calcoaceticus* avant 1995, mais la taxonomie du genre a fait l'objet d'une révision<sup>10</sup>

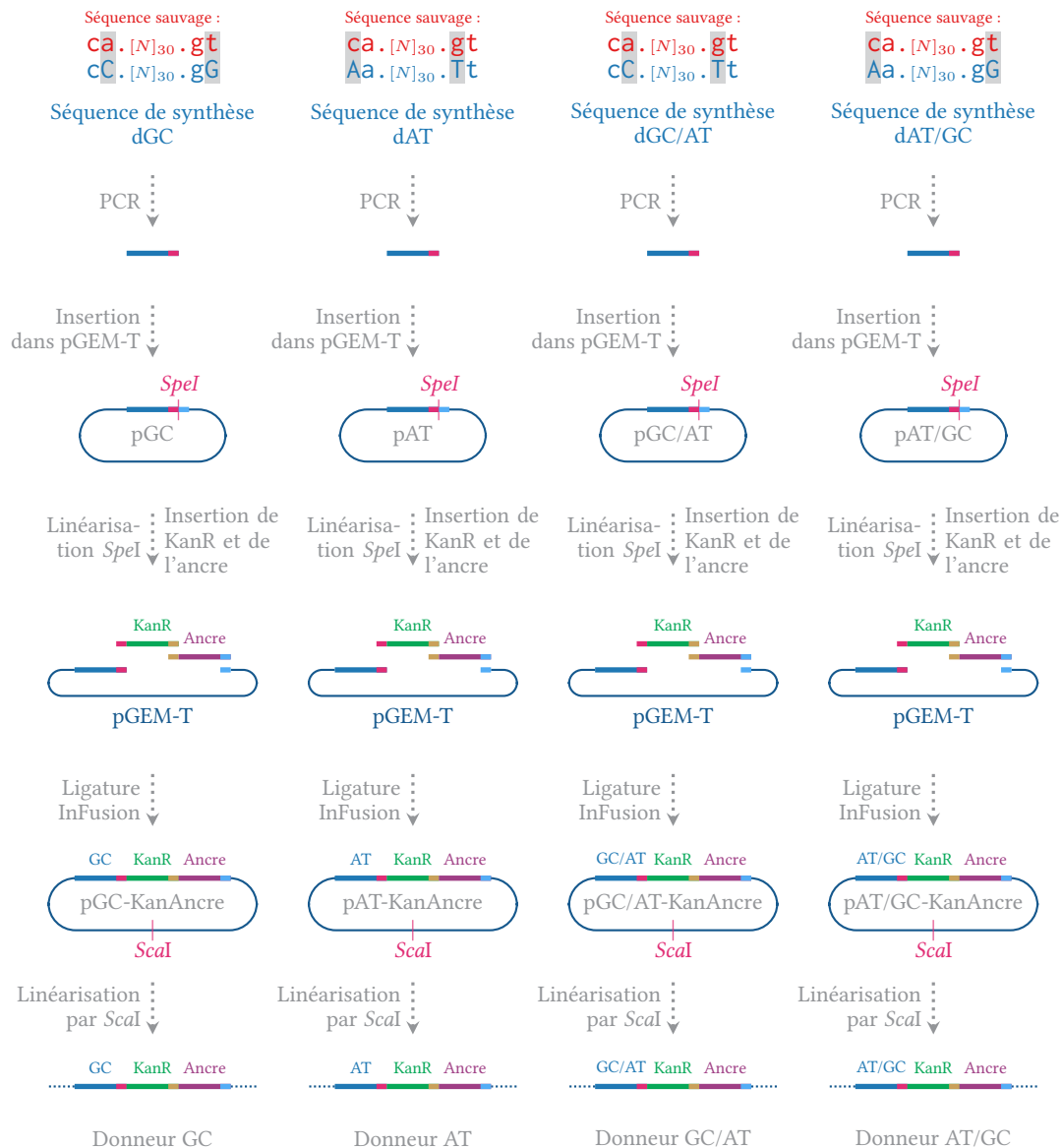


Figure 2 : **Constructions moléculaires et préparation de l'ADN recombinant**

Les gènes de synthèse sont amplifiés par PCR spécifique. Les fragments sont ensuite ligaturés dans le plasmide pGEM-T. Les plasmides obtenus sont linéarisés par *SpeI*. La cassette de résistance à la kanamycine et l'ancre sont insérés au site de restriction. Ces fragments sont préalablement amplifiés par PCR en utilisant des amorces porteuses de régions 3' flottantes, complémentaires des extrémités des fragments voisins. La ligation entre les trois fragments obtenus est réalisée à l'aide du kit InFusion. Les constructions plasmidiques obtenues sont amplifiées par culture, extraites et linéarisées par digestion enzymatique *ScaI*.

# 1 Matériels & Méthodes

## 1.1 Conception des séquences de synthèse

Pour étudier la correction des mésappariements au cours de la recombinaison homologue chez *A. baylyi*, nous avons conçu des séquences contenant des sites variant par rapport au génome receveur. Nous avons choisi un locus neutre du génome connu au laboratoire pour permettre de fortes efficacités de transformation. Il code une 3-oxoacyl-ACP reductase, impliquée dans la synthèse des acides gras<sup>12</sup> et est localisé à la position 47 184 du génome de référence. Ce locus mesure 800 paires de bases dans lesquelles nous avons substitué 23 positions réparties toutes les 30 paires de bases. La densité en position que nous avons choisie correspond à un compromis entre l'efficacité de la recombinaison et la précision de la détection des événements de conversion ; une densité plus forte diminue les fréquences de transformation, une densité plus faible réduit le nombre d'évènements détectables.

Nous avons conçu 4 gènes de synthèse, nommés respectivement dGC<sup>a</sup>, dAT, dGC/AT et dAT/GC (voir figure 2). Le gène dAT introduit uniquement des bases A et T ; le gène dGC introduit uniquement des bases G et C. Les sites variants dans le gène dAT sont adjacents aux sites variants dans le gène GC. Pour s'affranchir d'un éventuel effet dû à l'écart de GC% entre les séquences de synthèse et la séquence sauvage, nous avons également conçu deux séquences qui introduisent en alternance un mésappariement GC | AT et AT | GC. Le gène de synthèse dGC/AT introduit d'abord un mésappariement GC | AT, puis un mésappariement AT | GC, le gène de synthèse dAT/GC suit l'ordre inverse. (voir figure 2). Les quatre séquences ont été synthétisées par ThermoFischer (Waltham, États-Unis).

## 1.2 Constructions des plasmides

De façon à pouvoir sélectionner les clones recombinants au locus d'intérêt, nous avons construit les plasmides représentés dans la figure 2. Le gène synthétique est associé avec une cassette de résistance à la kanamycine, ainsi qu'une région "ancrage". Cette région recombinogène est complètement identiques à la séquence en 3' du locus d'intérêt et permet d'augmenter les fréquences de transformation<sup>13,14</sup>.

a. dGC pour donneur GC.

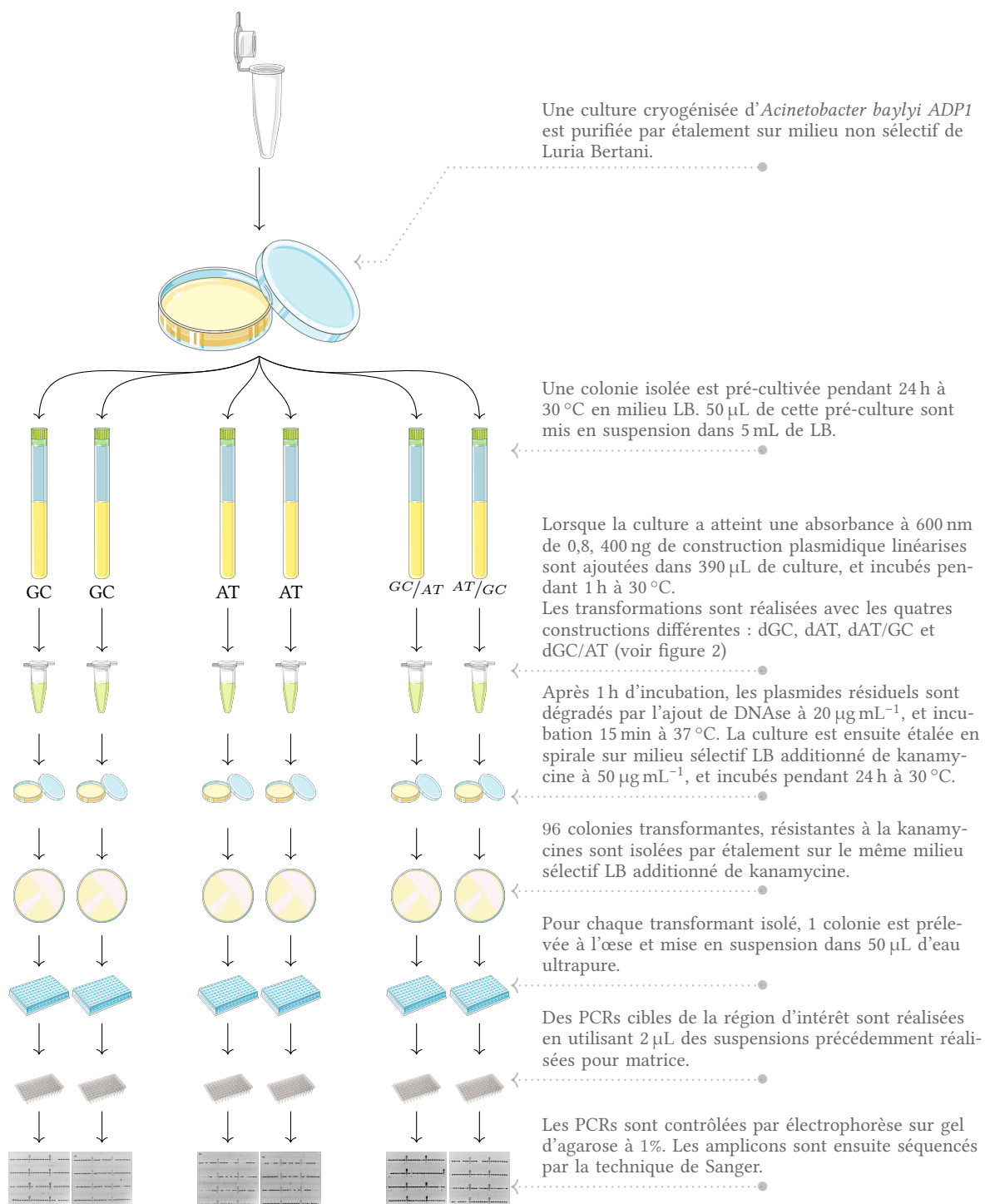


Figure 3 : **Protocole de transformation et d'obtention des amplicons des zones de recombinaison.**

Les gènes synthétiques ont d'abord été amplifiés par PCR <sup>a</sup> avec les amorces 1392 et 1393. Les plasmides obtenus ont été insérés dans la souche *E.coli* OneShot<sup>®</sup> TOP-10 (Invitrogen, Carlsbad, États-Unis), suivant le protocole du fabricant. Le sens d'insertion du gène synthétique dans le plasmide a été vérifié par PCR avec les amorces M13R et 1392, et l'absence de mutation a été vérifiée par séquençage (GATC Biotech, Constance, Allemagne). Les plasmides des clones validés par séquençage ont été extraits par le kit Nucleospin-Plasmid (Macherey-Nagel, Düren, Allemagne), et linéarisés par l'enzyme *SpeI* (ThermoFischer).

La cassette de résistance à la kanamycine *aphA3* et l'ancre ont été respectivement amplifiés par PCR avec le couple d'amorce 1408 / 1409 et 1410 / 1411. Les deux amplicons obtenus ont été ligaturés dans le plasmide pGEM-T porteurs des gènes de synthèse. La ligation a été réalisée simultanément par le kit InFusion (Takara Clontech, Saint Germain en Layes, France). Le produit de ligation obtenu a été inséré dans la souche optimisée pour la chimio-compétence *E.coli* Stellar (Takara Clontech). Les transformants ont été sélectionnés sur milieu LB solide additionné d'ampicilline à 75  $\mu\text{g mL}^{-1}$  de kanamycine à 50  $\mu\text{g mL}^{-1}$ . Les transformants ont été confirmés par PCR spécifique de l'insert avec les amorces 1393 et 1411.

### 1.3 Transformations d'*Acinetobacter baylyi*

Un microgramme de plasmide a été extrait et linéarisé par l'enzyme *ScaI* (ThermoFisher). L'enzyme a été ensuite inactivée par incubation 10 min à 70 °C. 390  $\mu\text{L}$  d'une culture pure d'*Acinetobacter baylyi* ADP1, avec une absorbance de 0,8 à 600 nm ont été incubés pendant 1 h à 28 °C en présence de 200 ng de plasmide linéarisé. Les suspensions ont été ensuite incubées 15 min à 37 °C en présence de DNase à 20  $\mu\text{mol L}^{-1}$  pour éliminer les plasmides résiduels. Les cellules recombinantes ont été sélectionnées par étalement en spirales (InterScience, St Nom la Bretèche, France) sur milieu LB solide additionné de kanamycine à 50  $\mu\text{g mL}^{-1}$  et incubées 24 h à 30 °C. Les dénombrements ont été effectués via un compteur automatique Scan<sup>®</sup>1200 (InterScience). Les fréquences de transformations ont été calculées en divisant le nombre de cellules transformantes, résistantes à la kanamycine par le nombre de cellules totales. 96 clones recombinants ont été ensuite isolés et incubés 24 h à 30 °C sur milieu LB solide additionné de kanamycine à 50  $\mu\text{mol mL}^{-1}$ . Une colonie isolée par clone a été mise en suspension dans 50  $\mu\text{L}$  d' $\text{H}_2\text{O}$  ultra-pure. 2  $\mu\text{L}$  de ces suspensions ont servi de matrice pour amplifier par PCR les régions recombinantes avec la Taq polymérase haute fidélité Phusion <sup>b</sup> (ThermoFischer,

a. Voir l'annexe 1.2 pour le détail des amorces et l'annexe 1.3 pour les conditions de PCR

b. Voir conditions de PCR dans l'annexe 1.3



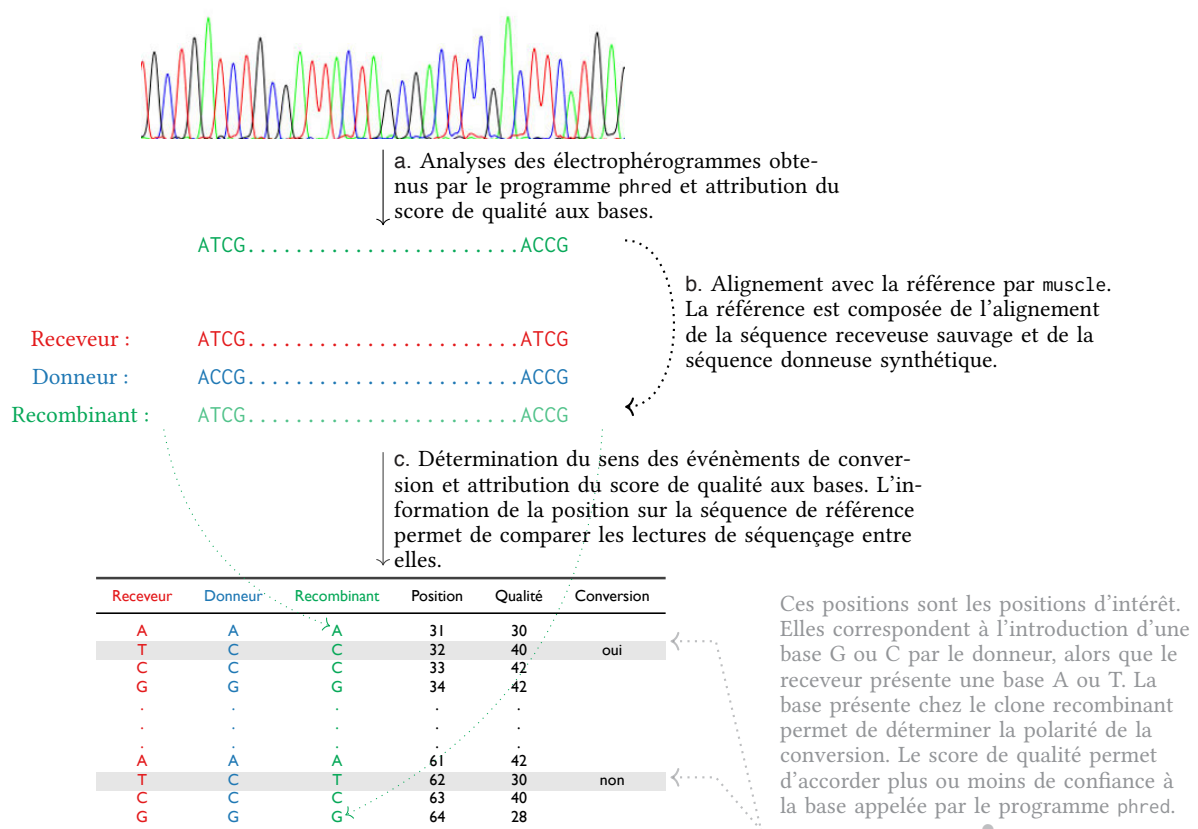


Figure 4 : **Exemple d'analyse de la zone de recombinaison pour un clone transformant**

Les électrophérogrammes de séquençage obtenus ont été analysés en utilisant un programme permettant d'attribuer à chaque position un score de qualité (a). Les séquences obtenues ont été alignées à la référence (b), ce qui a permis d'inférer pour chaque site polymorphe le sens de la conversion du recombinant (c).

Waltham, États-Unis). Les amplicons ont été vérifiés par migration sur gel d'agarose à 1% et séquencés par la technique de Sanger<sup>15</sup> (GATC Biotech).

## 1.4 Alignements

Les spectrogrammes de séquençage reçus au format propriétaire abi (Applied Biosystem, Foster City, États-Unis) ont été analysés par le programme phred<sup>16</sup> et convertis en format universel FASTA (voir figure 4). Les séquences obtenues ont été alignées aux références par muscle v3.8.31<sup>17</sup>. Les références en question correspondent à la séquence sauvage et la séquence du gène de synthèse, respectivement *receveur* et *donneur* de l'évènement de recombinaison. Un programme Python<sup>18</sup> a été développé pour analyser les alignements obtenus. Il détermine les positions des SNPs d'intérêt dans l'alignement de référence et infère le génotype du clone séquencé. Les alignements par paire en colonne obtenus ont été analysés par R 3.2.3<sup>19</sup>. Les programmes développés sont accessibles à l'adresse [https://github.com/sam217pa/gbc-seq\\_mars](https://github.com/sam217pa/gbc-seq_mars). Les données formatées et les fonctions d'analyse ont été assemblées dans le package R *gcbiasr* disponible à l'adresse <https://github.com/sam217pa/gbc-gcbiasr>.

**Table 1 : Fréquences de transformation chez *A. baylyi***

Construction donneuse	Fréquences de transformation
GC	$6,53 \times 10^{-5}$
AT	$2,42 \times 10^{-5}$
AT/GC	$4,97 \times 10^{-5}$
GC/AT	$1,74 \times 10^{-4}$

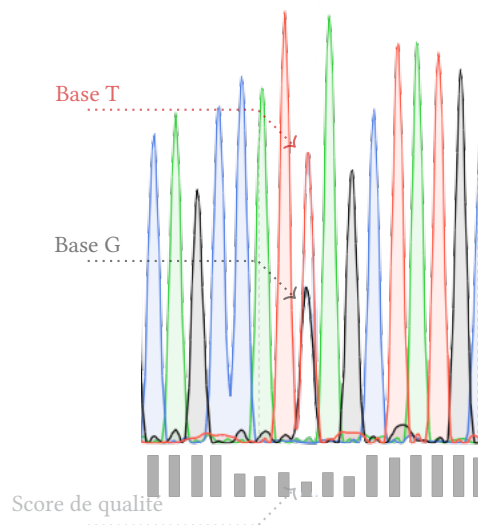


Figure 5 : **Exemple de marqueur montrant des traces de contaminations**

Cet électrophérogramme montre les bases autour du marqueur à la position 200. Dans une région de qualité moyenne élevée (bases en 5' et en 3'), le marqueur présente une trace de contamination par une autre base. La base déterminée est la base T mais une base G est présente dans la population d'amplicon séquencée.

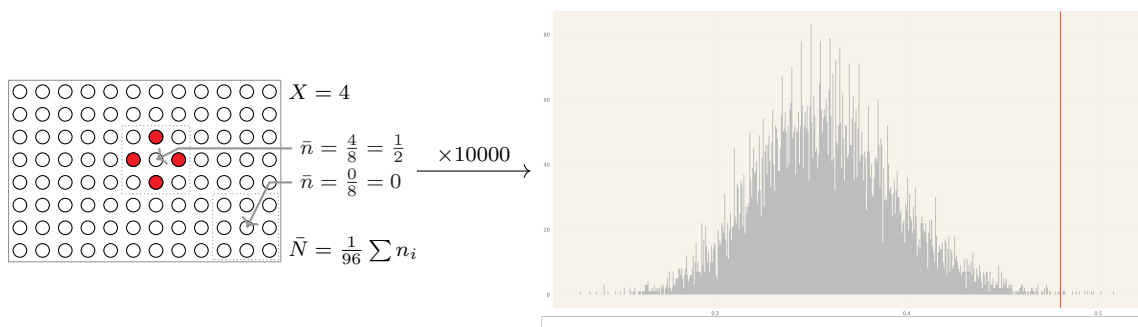


Figure 6 : **Des contaminations dues au hasard ?**

Par plaque de 96 puits, nous avons déterminé  $X$  le nombre de puits dont la séquence montre des traces de pics secondaires (voir figure 5) et mesuré  $\bar{n}$  la moyenne du nombre de puits voisins contaminés.  $\bar{N}$  est la moyenne des 96  $\bar{n}$  obtenus par plaque. Nous avons simulé  $1 \times 10^4$  plaques avec  $X$  puits contaminés répartis aléatoirement, mesuré  $\bar{N}$  et comparé la valeur expérimentale de  $\bar{N}$  (trait vertical rouge) avec la distribution des  $1 \times 10^4$   $\bar{N}$  (en gris). Seules 78/10000 plaques simulées montrent un  $\bar{N}$  supérieur à la valeur expérimentale : la répartition des séquences contaminées dans les plaques ne peut pas être attribuée au hasard.

## 2 Résultats

### 2.1 Validation de la démarche expérimentale

Nous avons transformé une suspension d'*Acinetobacter baylyi* par des constructions dont l'intégration dans le génome par recombinaison homologue entraîne la réparation des mésappariements.

Les fréquences de transformations obtenues, représentées dans le tableau 1 sont de l'ordre de  $1 \times 10^{-5}$  et ont permis d'obtenir un grand nombre de recombinants, avec une moyenne de  $9,2 \times 10^2$  transformants par boîte étalée, soit près de 3000 clones par transformation. Ce résultat indique que le protocole expérimental est suffisamment au point pour permettre d'obtenir le nombre de recombinants souhaité. Pour chaque gène de synthèse donneur, 96 clones ont été séquencés, soit un total de 384 recombinants.

### 2.2 Validation de la qualité des séquences

Deux séquences trop courtes pour être alignées ont été exclues de l'analyse. Neuf autres séquences présentaient une qualité médiane significativement inférieure à celle des autres séquences et ont également été exclues. Au total, 8561 marqueurs correspondant à des positions variantes entre le donneur et le receveur ont été analysées. Globalement, la qualité des 373 séquences suit un profil attendu : la qualité médiane est de 51 tout au long de la séquence, à l'exclusion des bases situées aux extrémités 5' et 3'.

Cependant, 9,8% des marqueurs sont d'une qualité de séquençage inférieure à la qualité de la région environnante. L'examen des électrophérogrammes a révélé la présence de pics secondaires (voir figure 5). Des pics secondaires apparaissent quand la population d'amplicon séquencée n'est pas homogène au site considéré. Or les deux pics présents à un marqueur donné correspondent toujours soit à la base sauvage, soit à la base synthétique introduite. Cette donnée peut être interprétée de deux façons.

1) Cette hétérogénéité pourrait résulter d'un signal biologique authentique : si les mésappariements de l'hétéroduplex ne sont pas réparés, la division de la cellule mère conduirait à une colonie hétérogène au site considéré et donc à la ségrégation simultanée des deux allèles.

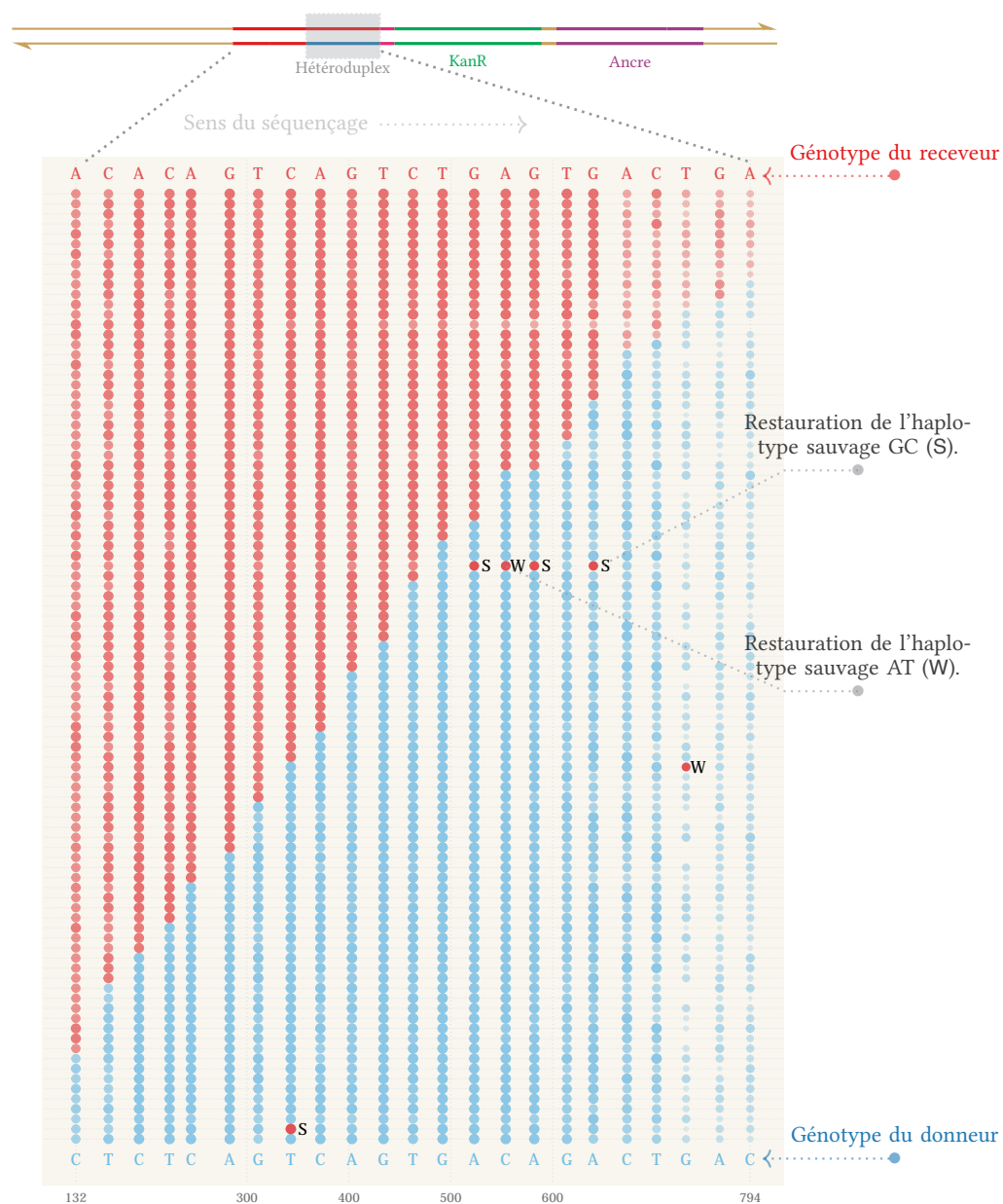


Figure 7 : **Zones de recombinaison entre un locus génomique d'*Acinetobacter baylyi* et un gène synthétique donneur d'allèles GC et AT.**

Chaque ligne horizontale représente une séquence. Les points représentent les positions des marqueurs sur les séquences. L'intensité et la taille des points représentent le score de qualité de séquençage du site. Les points sont bleus lorsque le site est dans la région convertie : ils correspondent à l'haplotype du donneur. Ils sont rouges lorsque le site correspond à l'haplotype du receveur. La première alternance rouge / bleu en partant de l'extrémité 5' marque la transition de l'haplotype converti à l'haplotype receveur : le point de recombinaison est localisé entre ces deux marqueurs. Certaines séquences sont des cas de conversion complexes, avec une ou plusieurs alternances donneur / receveur (rouge / bleu). Ce sont des cas de conversion accompagnés de restaurations de l'haplotype du receveur. Les séquences ont été triées par longueur de région convertie.

Pour tester cette hypothèse, nous avons séquencé à nouveau 31 isolats issus d'un clone séquencé en premier lieu et qui montrait des pics secondaires. Aucun des sous-clones séquencé ne montrent l'allèle correspondant au pic secondaire (voir figure 11). Ceci indique que le clone initial n'était pas hétérogène.

2) Une hypothèse alternative est que cette hétérogénéité pourrait être due à de la contamination entre échantillons. Nous avons donc analysé la répartition dans les plaques des puits dont la séquence montre des traces de contamination. Cette analyse montre que les séquences présentant des pics secondaires sont plus souvent voisines avec une autre séquence affectée que si elles étaient réparties aléatoirement (voir figure 6). Ce résultat indique que ces pics secondaires correspondent très probablement à de la contamination. Pour éviter les artéfacts liés à ces contaminants, nous avons décidé de n'étudier que les marqueurs dont la qualité est supérieure à 40 : cela correspond à la probabilité  $1 \times 10^{-4}$  d'étudier une base déterminée à tort, et discrimine clairement les bases litigieuses des autres. Nous avons donc obtenu un ensemble de 6818 marqueurs de conversion, soit 79,6 % de l'ensemble des marqueurs obtenus.

## 2.3 Description des régions recombinantes

La figure 7 représente le détail des zones de recombinaison obtenues avec une construction donneuse alternant GC et AT. Les zones de recombinaisons des clones transformés par les donneurs GC, AT et AT/GC sont détaillées en annexe 1.6. En moyenne, à notre locus cible,  $393 \pm 228$  nucléotides ont été transférés et intégrés dans le génome. La position du dernier marqueur converti indique le point de recombinaison (voir figure 1 et 7) qui se situe entre les 30 bases le séparant du premier marqueur conservé. Nous avons donc estimé la position du point de recombinaison par celle du dernier marqueur converti, dont la distribution est représentée dans la figure 8. Les points de recombinaison se répartissent de façon assez uniforme tout le long de la séquence. La majorité des recombinants correspondent à des cas de conversion simple, avec une seule alternance entre l'haplotype du donneur et celui du receveur. Cependant, 12 recombinants montrent des cas de conversion complexes, avec au moins deux alternances entre l'haplotype donneur et receveur. Ces alternances correspondent à des restaurations de l'haplotype receveur dans l'hétéroduplex, qui peuvent résulter de corrections ponctuelles des mésappariements de l'hétéroduplex (voir figure 7).

Nous nous sommes intéressés à deux paramètres permettant d'estimer les fréquences de

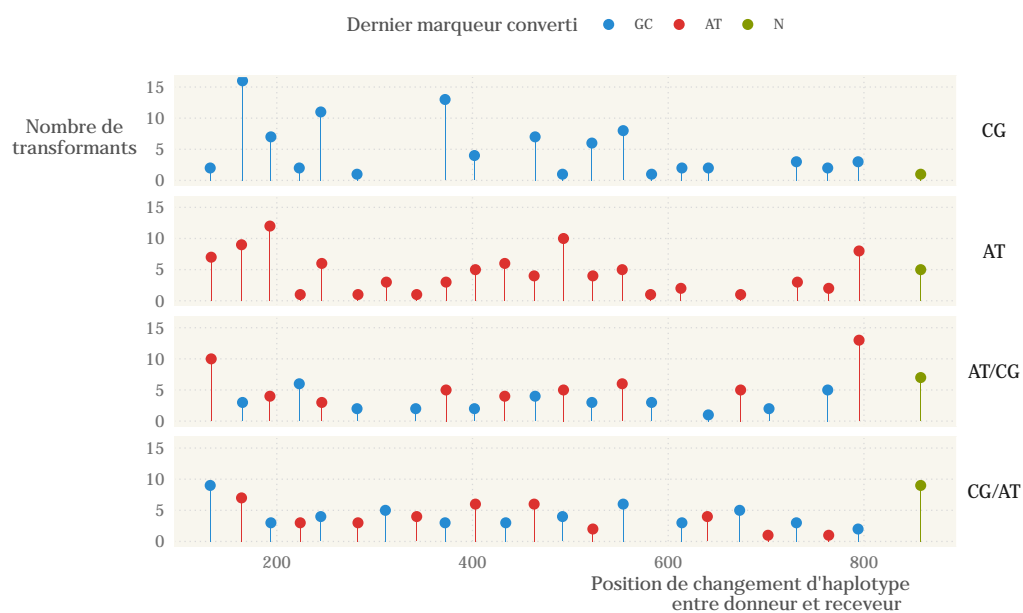


Figure 8 : **Distribution de la position du dernier marqueur converti**

Ce graphique représente en ordonnées le nombre de transformants dont le dernier marqueur converti est à la position représentée en abscisse. Les panneaux du graphique représentent les quatre constructions donneuses. La position du dernier marqueur converti indique la position du point de recombinaison. Les transformants qui ne montrent aucun marqueur converti sont indiqués par des *N*.

Table 2 : **Dénombrement des cas de restauration**

ADN synthétique donneur	Nucléotide restauré	
	AT	GC
AT	-	4
GC	0	-
AT/GC	3	2
GC/AT	1	4
Total	4	10

conversion en faveur de GC chez *A. baylyi* : la position du point de recombinaison, et la correction ponctuelle des mésappariements.

## 2.4 Comparaison de la longueur des régions converties

La longueur moyenne de région convertie par les donneurs GC et AT est de respectivement 446,6 et 395,6 pb ; l'écart entre les deux n'est pas significatif (test de Wilcoxon, probabilité critique = 0,31) (voir figure 8). De la même façon, la différence entre la longueur de région convertie par le donneur dAT/GC et celle par le donneur dGC/AT n'est pas significative (test de Wilcoxon, probabilité critique = 0,22). Globalement, le type de construction donneuse n'explique pas la variabilité de la longueur de région convertie (test de Kruskal-Wallis, probabilité critique = 0,10). Le type de donneur n'a pas d'influence sur la longueur de région convertie.

## 2.5 Restaurations de l'haplotype sauvage

Certains transformants montrent des régions de conversions qui alternent entre l'allèle sauvage receveur et l'allèle donneur. Ces alternances ponctuelles affectent de 1 à 3 marqueurs consécutifs (voir figure 7 et figure 12 en annexe 1.6 page iv). Nous avons confirmé qu'il s'agissait bien de restaurations de l'allèle sauvage de deux façons. 1) Expérimentalement, nous avons séquencé une sous-population de 31 clones issus d'un isolat séquencé en premier lieu. Tous montrent la même alternance au même site (voir figure 11 en annexe 1.5). 2) Analytiquement, le score de qualité moyen des sites montrant des restaurations de l'allèle sauvage permet de s'affranchir d'une possible erreur de séquençage : celle-ci se traduit généralement par un indice de qualité plus faible au site concerné. Le score de qualité moyen est de 49,36 aux sites restaurés, contre 52,79 aux sites non-restaurés. La différence entre les deux n'est pas significative (test de Wilcoxon, probabilité critique = 0,94). Les marqueurs correspondant à des restaurations de l'haplotype sauvage ne sont donc pas des erreurs de séquençage, et correspondent à un signal biologique.

Dans 10 des 14 cas de restaurations, la base sauvage restaurée est une base G ou C, contre 4 cas de restaurations d'une base A ou T. L'écart n'est pas significatif (test du  $\chi^2$ , probabilité critique = 0,11), mais indique un excès de restauration des bases G et C.



Table 3 : **Dénombrement des cas de conversion**

	dGC	dAT	dAT/GC	dGC/AT	Total
$N_{GC \rightarrow GC}$	-	750	489	422	1661
$N_{GC \rightarrow AT}$	-	1032	330	412	1774
$N_{AT \rightarrow AT}$	613	-	515	416	1544
$N_{AT \rightarrow GC}$	1060	-	375	404	1839

↓

	AT	GC
Marqueurs non convertis	1661	1544
Marqueurs convertis	1774	1839

## 2.6 Estimation des fréquences de conversion en faveur de GC

Sur l'ensemble des 6818 marqueurs, nous avons comparé le nombre de marqueurs AT qui ont été convertis en GC ( $N_{AT \rightarrow GC}$ ) avec le nombre de marqueurs GC qui ont été convertis en AT ( $N_{GC \rightarrow AT}$ ) (voir le tableau 3). Nous avons trouvé que la proportion de marqueurs AT convertis en GC (0,54 %) est significativement supérieure à la proportion de marqueurs GC convertis en AT (0,52 %) (test du  $\chi^2$ , probabilité critique = 0,03). Ce résultat est robuste au critère de qualité des marqueurs analysés ; l'analyse des 8561 marqueurs indique une probabilité critique de  $5 \times 10^{-3}$ .

Ce résultat important montre que les bases G et C ont une probabilité plus forte d'être transmises que les bases A et T au cours de la recombinaison à ce locus du génome d'*A. baylyi*.

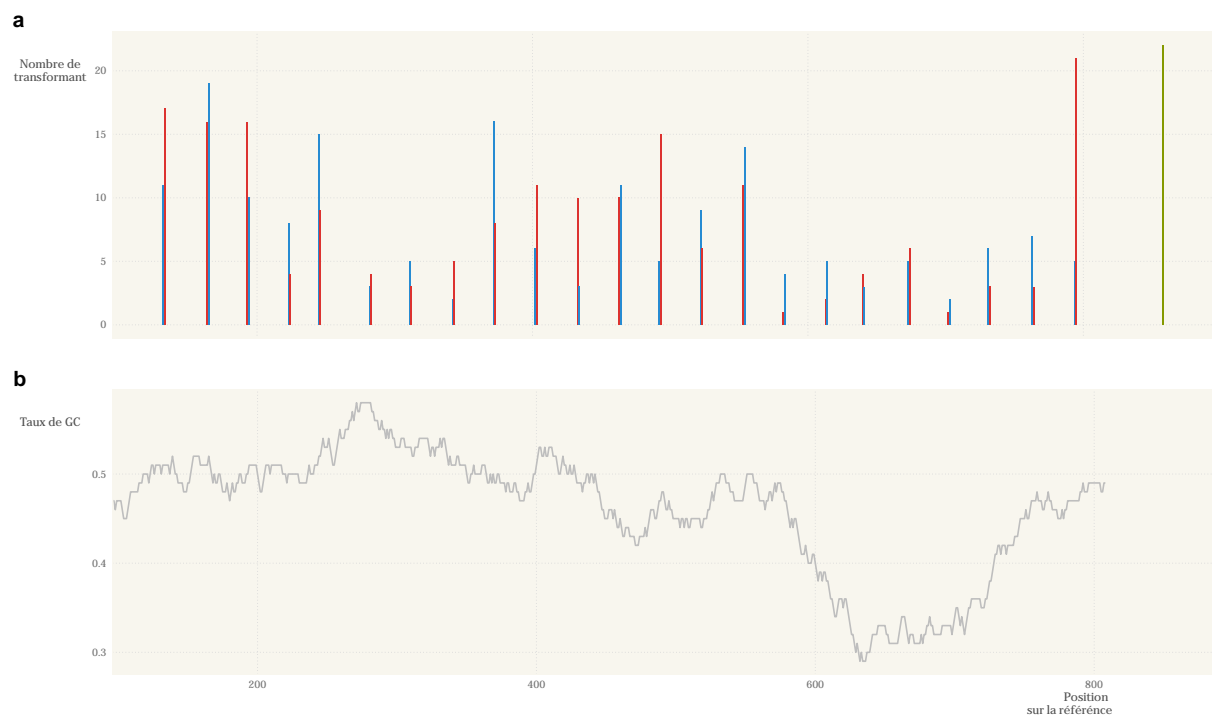


Figure 9 : **Comparaison du taux de GC moyen par fenêtre de 50 bp avec la distribution du dernier marqueur converti**

La figure a représente la distribution des derniers marqueurs convertis, indicateurs de la position du point de recombinaison. La figure b représente l'évolution du taux de GC par fenêtre de 50 pb au long de la séquence.

## 3 Discussions

Nous avons séquencé un grand nombre de clones d'*A. baylyi* recombinants à un locus choisi de façon à introduire la correction des mésappariements et à comparer la fréquence des événements de conversion conduisant à l'introduction de bases G ou C par rapport à ceux introduisant des bases A ou T.

### 3.1 Distribution des points de recombinaison

Le changement entre l'haplotype donneur et l'haplotype receveur se produit entre le dernier marqueur converti et le premier marqueur conservé. Cette information permet d'estimer la distribution du point de recombinaison et la longueur moyenne des régions converties. Les événements de recombinaison induits par les constructions donneuses sont des événements de type non-crossover : ils ne conduisent pas à l'échange des régions flanquantes<sup>1</sup>. La longueur moyenne des régions converties de  $393 \pm 228$  pb est comprise entre 100 et 1000 bp et correspond ainsi à celle observée pour les événements non-crossover chez l'Homme<sup>5</sup>. La distribution du point de recombinaison est assez différente de celle décrite par Yáñez-Cuna *et al.*<sup>20</sup> chez la bactérie *Rhizobium etli*. Ils observent que les recombinants changent d'haplotype pour la majorité dans les 200 premières paires de bases, et seuls 10 % des clones ont une région convertie de plus de 800 pb. Les différences entre leurs résultats et les nôtres peuvent potentiellement s'expliquer par les différences de modèles et de locus.

Quelle que soit la construction donneuse, la distribution du point de recombinaison suit globalement le même patron, les marqueurs entre 600 et 800 pb et autour de 350 pb sont moins souvent les derniers marqueurs convertis que les autres. Une hypothèse pour expliquer cette distribution serait que le taux de GC local conditionne la probabilité de rejet du brin d'ADN exogène. Un taux de GC plus faible diminue le nombre de liaisons hydrogène entre les brins et pourrait donc diminuer les probabilités d'appariement avec un brin d'ADN exogène. En effet, lors de la recherche par la protéine RecA d'une matrice homologue permettant de réparer la lésion, la reconnaissance est purement basée sur l'appariement Watson-Crick entre les brins<sup>21</sup>. La figure 9a montre un point froid de recombinaison dans la région entre 600 et 800 bp, environ 200 paires de bases après l'origine de l'hétéroduplex. La figure 9b montre l'évolution du taux de GC par fenêtre de 50 pb. Le point froid de recombinaison à 600 pb semble correspondre avec un taux de GC plus faible localement. En fait, la zone pauvre en GC à 600 pb correspond au taux



de GC moyen du génome d'*A. baylyi* de 40,3 %, mais le locus étudié a un taux de GC moyen de 46,8 %. Or ce locus est connu au laboratoire pour permettre de bonnes efficacités de transformations. Si on suppose qu'un taux de GC plus élevé entraîne une efficacité de recombinaison plus élevée, le taux de GC à ce locus supérieur à celui du génome pourrait ainsi expliquer ces efficacités de transformations. Cette supposition pourrait également expliquer les corrélations observées par Lassalle *et al.*<sup>8</sup> entre taux de GC et taux de recombinaison.

### 3.2 Des preuves de l'existence d'une conversion génique biaisée vers GC chez *A. baylyi*

Nous avons montré que la probabilité pour une base A ou T d'être convertie en G ou C est significativement supérieure à celle d'une base G ou C d'être convertie en A ou T. Cette observation correspond précisément aux prédictions du modèle gBGC. Quel pourrait être le mécanisme responsable de ce biais ?

Au sein des régions converties, 14 cas de restauration de l'allèle sauvage ont été détectés. Quels mécanismes peuvent les expliquer ? Ce ne sont probablement pas des mutations spontanées : bien que la recombinaison soit un processus mutagène en soi<sup>22,23</sup>, les restaurations sont retrouvées spécifiquement aux positions des marqueurs, et correspondent précisément à l'allèle sauvage. Une mutation spontanée pourrait introduire aléatoirement l'une des quatre bases. Chez les eucaryotes, des cas complexes de conversion ont déjà été décrits<sup>24,25</sup>. Chez la levure notamment, ces cas complexes ont été interprétés comme des basculements multiples entre le choix des matrices permettant la réparation des cassures<sup>26</sup>. Au cours de la recombinaison méiotique, les mésappariements sont corrigés préférentiellement en utilisant la chromatide homologue pour matrice, mais ces événements montrent que des alternances multiples entre le choix de la chromatide homologue et la chromatide sœur peuvent avoir lieu. Cependant, dans notre système, la seule matrice envisageable est le brin exogène. Il se pourrait également qu'une cellule en cours de réplication utilise le chromosome néo-synthétisé comme matrice alternative, mais, à notre connaissance, cette possibilité n'a encore jamais été documentée.

L'hypothèse la plus probable est que ces cas de restauration de l'allèle sauvage soient dus à l'action de la machinerie de correction des mésappariements de type BER (Base Excision Repair). Cette voie excise spécifiquement une base mésappariée en détectant les malformations qu'elle occasionne dans la structure de la double-hélice, puis la remplace par la base complémentaire à celle du brin opposé<sup>27</sup>. Chez les eucaryotes, la correction des mésappariements



dans les cellules en *mitose* est effectuée en partie par la voie du BER, qui est biaisée vers l'introduction des bases G et C chez les mammifères<sup>28,29</sup>. Ce biais est probablement dû à une adaptation pour compenser l'hypermutableté des cytosines méthylées, dont la déamination, spontanée en conditions physiologiques normales, conduit à l'introduction d'une thymine. Des enzymes dédiées, les thymines DNA-glycosylases, excisent spécifiquement ces bases. La voie du BER a donc été considérée comme l'un des moteurs possibles du gBGC chez les mammifères<sup>3</sup>, mais a été rejetée chez la levure<sup>4</sup>. Chez cette dernière, le gBGC est associé aux régions converties les plus longues et dans lesquelles tous les allèles sont convertis depuis le même haplotype donneur.

Le BER existe également chez les procaryotes. Lieb *et al.*<sup>30</sup> ont montré chez *E. coli* que les enzymes du VSP (Very Short Patch repair) réparaient un mésappariement dans un sens spécifique en fonction des bases au voisinage immédiat du mésappariement. Ce dernier point a également été démontré pour les enzymes du BER chez les eucaryotes<sup>29</sup>. Cette activité de réparation est spécifiquement liée au produit du gène *vsr*<sup>31,32</sup>. *A. baylyi* possède également certaines des enzymes impliquées dans les voies du VSP<sup>33</sup>. Si ce genre de mécanisme est à l'œuvre chez *A. baylyi*, les marqueurs AT et les marqueurs GC de nos constructions, qui sont adjacents, ne se situent pas rigoureusement dans le même contexte nucléotidique (voir figure 2). Pour palier à cet éventuel biais, il faudrait transformer à nouveau un clone ayant converti tous les marqueurs par la séquence sauvage : chaque évènement de conversion serait alors rigoureusement dans le même contexte.

Est-ce que le BER, s'il est actif chez *A. baylyi*, pourrait conduire à de la conversion biaisée vers GC ? Nous avons trouvé un excès de restauration des bases G et C par rapport aux cas de restauration des bases A et T, respectivement de 10 contre 4 cas. Cet écart n'est pas significatif. Si le rapport réel entre le nombre de cas de restaurations de l'allèle GC et le nombre de cas de restaurations de l'allèle AT est le même, il faudrait doubler la taille d'échantillon pour que le test soit significatif au seuil de  $5 \times 10^{-2}$ . Néanmoins, le résultat obtenu avec 6818 marqueurs est robuste à l'exclusion des marqueurs restaurés (probabilité critique =  $3 \times 10^{-2}$ ). Ces cas de restaurations de l'allèle sauvage au sein de la région convertie ne sont donc probablement pas le principal moteur du biais mesuré, bien qu'ils puissent y contribuer faiblement.

Le BER pourrait cependant contribuer de façon plus subtile à diminuer la probabilité de transmission des bases A et T. En effet, les premiers marqueurs conservés peuvent être la résultante de deux choses : 1) ils ne font pas partie de l'hétéroduplex, et ne sont pas soumis à de la conversion ou 2) ils font partie de l'hétéroduplex, mais l'allèle sauvage est restauré. Il n'est



donc pas exclu que l'hétéroduplex s'étende au delà du dernier marqueur converti et que le BER ou la machinerie de correction des mésappariements restaure l'allèle sauvage précisément sur ces marqueurs.

Le principal contributeur au biais mesuré semble donc être le degré d'extension de la région convertie. Cette extension est déterminée essentiellement par la portée de la résection induite par le MMR (Mis-Match Repair). En effet, lorsqu'une base mal appariée est détectée, elle est suivie par la résection d'une région de longueur variable et la resynthèse de la zone résectée en utilisant le brin intact pour matrice<sup>34</sup>. La portée de la résection détermine donc le nombre de marqueurs correspondant à l'haplotype du donneur qui sont intégrés. L'étude des régions converties chez des mutants d'*A. baylyi* délétés pour différentes fonctions clés de la correction des mésappariements devrait permettre de répondre formellement à la question de leur contribution au biais de conversion.

Une analyse globale des fréquences de conversion montre bien une sur-transmission des bases G et C. Ce résultat conforte l'extension du modèle gBGC aux procaryotes amorcée par Lassalle *et al.*<sup>8</sup>.

### 3.3 Conséquences évolutives de la conversion biaisée vers GC

L'écart entre la probabilité pour une base AT d'être convertie en GC (54 %) et la probabilité pour une base GC d'être convertie en AT (52 %) est faible. Un faible biais peut néanmoins avoir des conséquences importantes. Chez la levure, un mésappariement GC/AT a 50,62 % de chances d'être converti en GC, contre 49,38 % d'être converti en AT<sup>35,3</sup>. Mais de façon similaire à la sélection, l'impact du gBGC sur l'évolution des génomes dépend de son intensité par rapport à l'action de la dérive génétique. L'impact du gBGC à un locus donné dépend uniquement de la fréquence à laquelle le locus est impliqué dans un évènement de recombinaison, de la force du biais de conversion et de la taille de la population<sup>36,8,3,37</sup>. Ceci implique que même lorsque le biais est faible et la probabilité qu'un locus donné soit impliqué dans un évènement de recombinaison faible, dès lors que la taille de population est grande, l'allèle G ou C a une probabilité plus forte de se fixer. L'une des observations que ce modèle explique bien est le taux de GC très bas des bactéries endosymbiotiques, pour lesquelles le taux de recombinaison est effectivement nul<sup>38</sup>.

La conversion génique biaisée vers GC est l'un des candidats pour expliquer la mystérieuse force évolutive qui permet aux procaryotes de maintenir de tels taux de GC alors même que la mutation est universellement biaisée vers AT<sup>6,7</sup>.



# Conclusion

Nous avons pu au cours de cette étude valider l'approche expérimentale envisagée. L'utilisation d'une cassette de résistance pour sélectionner les clones et l'utilisation d'une ancre recombinogène ont permis d'obtenir un grand nombre de recombinants, pour toutes les constructions donneuses, et ce malgré les 3 % de divergence entre le locus cible et le gène de synthèse introduit. La fiabilité de la méthode de construction des plasmides donneurs a été éprouvée et devrait permettre de mettre au point d'autres constructions donneuses dans un temps raisonnable, à moindre coût et avec des résultats convaincants.

Nous avons séquencé 384 clones obtenus avec ces approches, dont l'étude a permis de décrire les zones de recombinaison, la distribution des points de recombinaison, la longueur des régions converties, les fréquences de restauration de l'haplotype sauvage et les fréquences de conversion des marqueurs. Une probabilité plus forte pour les marqueurs AT d'être convertis en GC que l'inverse a été démontrée, ce qui correspond précisément aux prédictions du modèle gBGC.

Ce résultat marque une deuxième étape importante dans l'extension de la théorie du biais de conversion génique aux procaryotes. Ce travail devra par la suite être étendu à d'autres locus du génome d'*A. baylyi*, plus proches et plus éloignés de l'origine de réplication. D'autres modèles bactériens sont également à l'étude : le modèle de la recombinaison homologue après transmission d'un plasmide porteur de sites variants par conjugaison chez *E. coli* et *Burkholderia ambifaria*. La confrontation des résultats obtenus pour ces trois modèles devrait permettre de quantifier formellement les fréquences de conversion en faveur de GC chez les procaryotes.

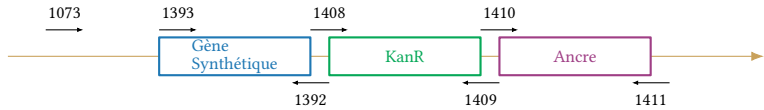
La démonstration d'une conversion génique biaisée vers GC chez les procaryotes serait importante, elle conduirait à repenser globalement les hypothèses qui attribuent au taux de GC un rôle adaptatif, et à intégrer la conversion génique dans l'ensemble des forces évolutives qui régissent la composition en base des génomes.

# Références

- [1] Chen, J.-M., D. N. Cooper, N. Chuzhanova, C. Férec, et G. P. Patrinos. 2007. Gene conversion : mechanisms, evolution and human disease. *Nat Rev Genet* 8 :762–775.
- [2] Pessia, E., A. Popa, S. Mousset, C. Rezvoy, L. Duret, et G. A. B. Marais. 2012. Evidence for Widespread GC-biased Gene Conversion in Eukaryotes. *Gen Biol Evol* 4 :675–682.
- [3] Duret, L. et N. Galtier. 2009. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annual Review of Genomics and Human Genetics* 10 :285–311.
- [4] Lesecque, Y., D. Mouchiroud, et L. Duret. 2013. GC-Biased Gene Conversion in Yeast Is Specifically Associated with Crossovers : Molecular Mechanisms and Evolutionary Significance. *Molecular Biology and Evolution* 30 :1409–1419.
- [5] Williams, A. L., G. Genovese, T. Dyer, N. Altemose, K. Truax, G. Jun, N. Patterson, S. R. Myers, J. E. Curran, R. Duggirala, J. Blangero, D. Reich, et M. Przeworski. 2015. Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *eLife* 4.
- [6] Hildebrand, F., A. Meyer, et A. Eyre-Walker. 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 6 :e1001107.
- [7] Hershberg, R. et D. A. Petrov. 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* .
- [8] Lassalle, F., S. Périan, T. Bataillon, X. Nesme, L. Duret, et V. Daubin. 2015. GC-Content Evolution in Bacterial Genomes : The Biased Gene Conversion Hypothesis Expands. *PLOS Genetics* 11 :e1004941.
- [9] Yahara, K., X. Didelot, K. A. Jolley, I. Kobayashi, M. C. J. Maiden, S. K. Sheppard, et D. Falush. 2016. The Landscape of Realized Homologous Recombination in Pathogenic Bacteria. *Mol Biol Evol* 33 :456–471.
- [10] Euzéby, J. P. 1997. List of Bacterial Names with Standing in Nomenclature : a folder available on the Internet. *International Journal of Systematic and Evolutionary Microbiology* 47 :590–592.
- [11] Chen, I. et D. Dubnau. 2004. DNA uptake during bacterial transformation. *Nat Rev Micro* 2 :241–249.
- [12] Vallet, D., E. Belda, A. Calteau, S. Cruveiller, S. Engelen, A. Lajus, F. L. Fèvre, C. Longin, D. Morinico, D. Roche, Z. Rouy, G. Salvignol, C. Scarpelli, A. A. T. Smith, M. Weiman, et C. Médigue. 2013. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucl. Acids Res.* 41 :D636–D647.
- [13] de Vries, J. et W. Wackernagel. 2002. Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proceedings of the National Academy of Sciences* 99 :2094–2099.
- [14] Meier, P. et W. Wackernagel. 2003. Mechanisms of homology-facilitated illegitimate recombination for foreign DNA acquisition in transformable *Pseudomonas stutzeri*. *Molecular Microbiology* 48 :1107–1118.
- [15] Sanger, F., S. Nicklen, et A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74 :5463–5467.
- [16] Ewing, B., L. Hillier, M. C. Wendt, et P. Green. 1998. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res.* 8 :175–185.
- [17] Edgar, R. C. 2004. MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 :1792–1797.
- [18] Cock, P. J. A., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, et M. J. L. d. Hoon. 2009. Biopython : freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25 :1422–1423.
- [19] R Core Team. 2015. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- [20] Yáñez-Cuna, F. O., M. Castellanos, et D. Romero. 2015. Biased gene conversion in *Rhizobium etli* is caused by preferential double strand breaks on one of the recombining homologs. *Journal of Bacteriology* p. JB.00768–15.
- [21] Lee, J. Y., T. Terakawa, Z. Qi, J. B. Steinfeld, S. Redding, Y. Kwon, W. A. Gaines, W. Zhao, P. Sung, et E. C. Greene. 2015. Base triplet stepping by the Rad51/RecA family of recombinases. *Science* 349 :977–981.
- [22] Rodgers, K. et M. McVey. 2016. Error-Prone Repair of DNA Double-Strand Breaks. *Journal of Cellular Physiology* 231 :15–24.
- [23] Hicks, W. M., M. Kim, et J. E. Haber. 2010. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329 :82–85.
- [24] Martini, E., V. Borde, M. Legendre, S. Audic, B. Regnault, G. Soubigou, B. Dujon, et B. Llorente. 2011. Genome-Wide Analysis of Heteroduplex DNA in Mismatch Repair-Deficient Yeast Cells Reveals Novel Properties of Meiotic Recombination Pathways. *PLoS Genet* 7.
- [25] Yeadon, P. J., J. P. Rasmussen, et D. E. A. Catchside. 2001. Recombination Events in *Neurospora crassa* May Cross a Translocation Breakpoint by a Template-Switching Mechanism. *Genetics* 159 :571–579.
- [26] Hoffmann, E. R. et R. H. Borts. 2005. Trans events associated with crossovers are revealed in the absence of mismatch repair genes in *Saccharomyces cerevisiae*. *Genetics* .
- [27] Krokan, H. E. et M. Bjørås. 2013. Base Excision Repair. *Cold Spring Harb Perspect Biol* 5 :a012583.
- [28] Brown, T. C. et J. Jiricny. 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* 50 :945–950.
- [29] Brown, T. C. et J. Jiricny. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell* 54 :705–711.
- [30] Lieb, M. 1985. Recombination in the  $\lambda$  repressor gene : evidence that very short patch (VSP) mismatch correction restores a specific sequence. *Mol Gen Genet* 199 :465–470.
- [31] Hennecke, F., H. Kolmar, K. Bründl, et H.-J. Fritz. 1991. The *vsr* gene product of *E. coli* K-12 is a strand- and sequence-specific DNA mismatch endonuclease. *Nature* 353 :776–778.
- [32] Lieb, M. 1983. Specific mismatch correction in bacteriophage lambda crosses by very short patch repair. *Molec Gen Genet* 191 :118–125.
- [33] Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, et M. Tanabe. 2016. KEGG as a reference resource for gene and protein annotation. *Nucl. Acids Res.* 44 :D457–D462.
- [34] Spies, M. et R. Fishel. 2015. Mismatch Repair during Homologous and Homologous Recombination. *Cold Spring Harb Perspect Biol* 7 :a022657.
- [35] Mancera, E., R. Bourgon, A. Brozzi, W. Huber, et L. M. Steinmetz. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454 :479–485.
- [36] Nagylaki, T. 1983. Evolution of a finite population under gene conversion. *PNAS* 80 :6278–6281.
- [37] Duret, L. et P. F. Arndt. 2008. The Impact of Recombination on Nucleotide Substitutions in the Human Genome. *PLoS Genet* 4 :e1000071.
- [38] Moran, N. A., J. P. McCutcheon, et A. Nakabachi. 2008. Genomics and Evolution of Heritable Bacterial Symbionts. *Ann Rev Genet* 42 :165–190.

# 1 Annexes

## 1.2 Amorces utilisées

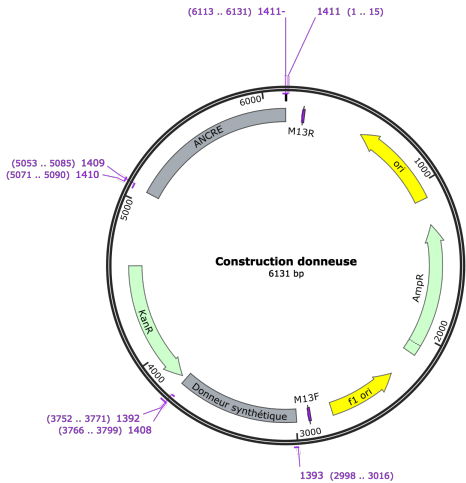


Cible	Amorce	Séquence	Tm (°C)
Génome Construction	1073	CAGGCTGACGTGATTGTTCA	56.6
	1392	AAGGTGGAAGAGAAGGAGGC	58.7
	1393	GCGAGGAGGAAAGCAAAGAG	58.2
	1408	CACCTTAATCACTAGTTAGACATCTAAATCTAGGTAC	61.50
	1409	GGTAAAGTCAGAGGAGAGGATGAGGAGGCAGATTG	68.66
	1410	TCCTCTGACTTTACCAACAAC	48.04
	1411	AGGCGGCCGCACTAGCTTTCTGAGGGGAACGATCA	71.62
Plasmide pGEM-T	M13R	GAGGAAACAGCTATGAC	47.8
	M13F	GTAAACGACGGCCAGT	53.9

Figure 10 : Liste des amorces utilisées

## 1.3 Conditions de PCR

## 1.4 Carte des constructions donneuses



## 1.5 Confirmations des restaurations des haplotypes sauvages

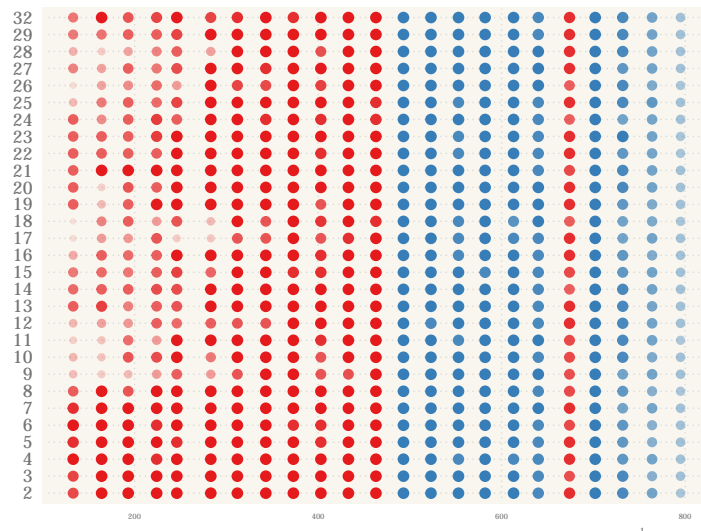


Figure 11 : **Confirmation d'une restauration d'haplotype sauvage**

Pour s'affranchir d'une possible erreur de séquençage sur les sites montrant des restaurations de l'haplotype sauvage dans la région convertie, nous avons séquencé la zone de recombinaison de 31 clones issus du clone séquencé en premier lieu. Nous avons également séquencé à nouveau la colonie "mère", de façon à comparer les sous-populations avec la population mère séquencée ; c'est le clone séquencé 32. Pour rappel, les points rouges correspondent à l'haplotype du receveur, les points bleus à l'haplotype du donneur. Les 31 clones séquencés montrent tous la même alternance au même marqueur que le clone 32. Nous avons ainsi confirmé que cette restauration de l'haplotype sauvage était bien due à une correction des mésappariements dans la population mère.

## 1.6 Traces de conversions

La figure 12 page 12 montre l'alignement de toutes les positions de marqueurs lorsque le donneur est GC, AT, GC/AT et AT/GC. L'interprétation des figures est détaillée dans la figure 7 en page 6.

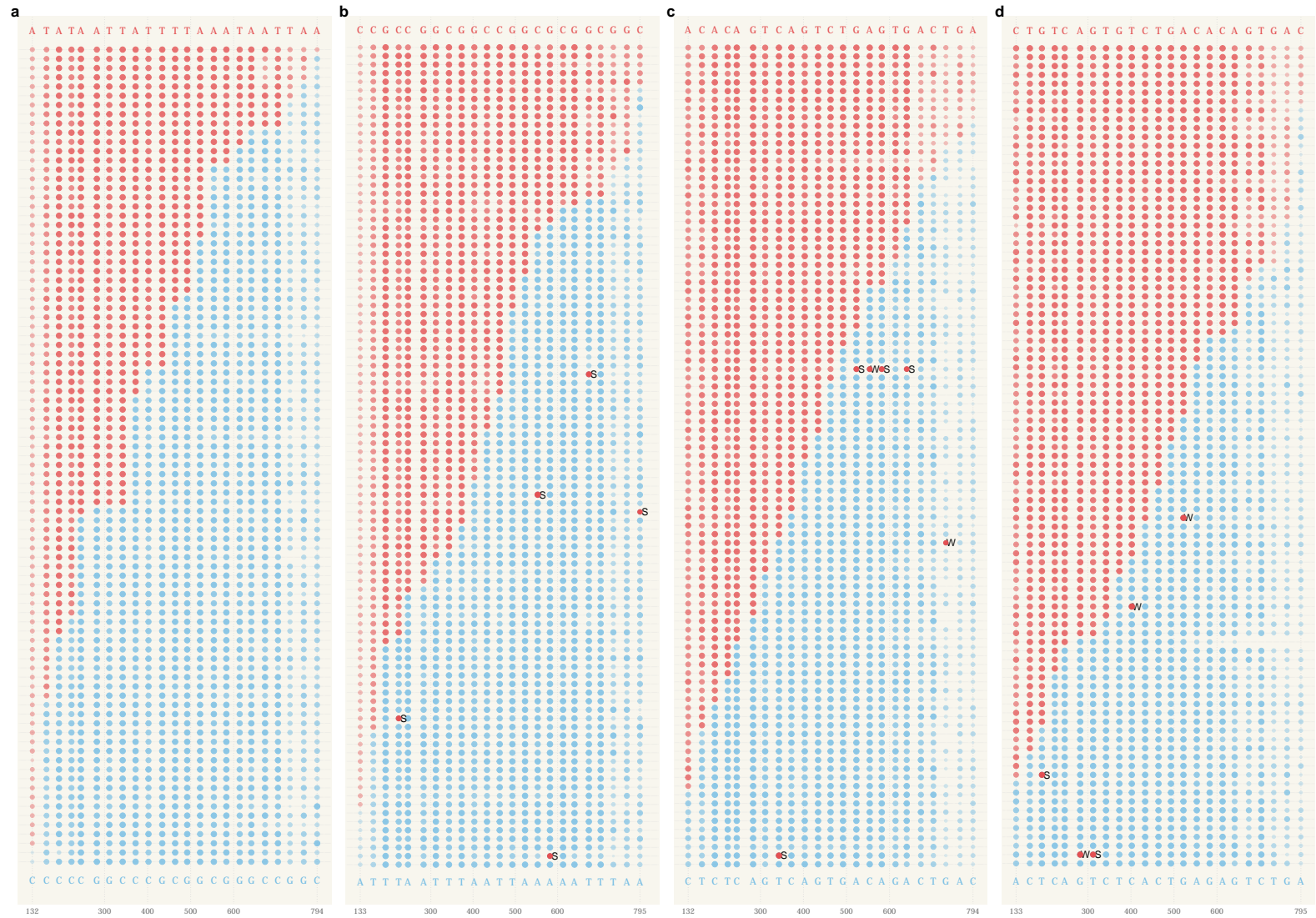


Figure 12 : Pellentesque dapibus suscipit ligula. Donec posuere augue in quam. Aliquam feugiat tellus ut neque. Nulla facilis, risus a rhoncus fermentum, tellus tellus lacinia purus, et dictum nunc justo sit amet elit.



## Étude expérimentale des fréquences de conversion en faveur de GC au cours de la recombinaison homologue chez *Acinetobacter baylyi*

La réparation de l'ADN joue un rôle important dans l'évolution de la composition en base des génomes. Chez les mammifères, la recombinaison homologue tend à augmenter le taux de GC des régions recombinantes par un processus appelé conversion génique biaisée (gBGC). L'hypothèse gBGC a été récemment étendue aux procaryotes. L'objectif de ce travail était de valider une approche expérimentale permettant de quantifier les fréquences de conversion en faveur de GC chez *Acinetobacter*. Nous avons transformé cette bactérie naturellement compétente par des gènes de synthèse introduisant des sites variants avec un locus génomique, sélectionné les recombinants et séquencé les produits de recombinaison. Nous avons pu démontrer la validité de l'approche expérimentale, en obtenant des fréquences de transformation de l'ordre de  $1 \times 10^{-5}$  et une moyenne de 3000 transformants par construction donneuse. Le séquençage de 384 recombinants a servi à décrire les régions recombinantes, la distribution des changements d'haplotype entre donneur et receveur, et à quantifier le nombre d'évènements de conversion dans le sens AT→GC. Les régions converties obtenues démontrent que les bases AT ont une probabilité plus forte d'être convertis en GC que l'inverse, une observation qui correspond précisément à l'hypothèse initiale. Ce résultat représente une seconde étape importante dans l'extension du biais de conversion génique vers GC aux procaryotes. Il vient appuyer la proposition du gBGC comme l'une des forces permettant aux procaryotes de maintenir des taux de GC élevés alors que la mutation est biaisée vers AT.

