

Crédits photographiques

Photographie de l'image de couverture reproduite avec permissions depuis Wikimedia Commons, domaine public, accès libre.

https://commons.wikimedia.org/wiki/File:Dark_side_of_the_moon.jpg

Sommaire

Introduction	1
1 Matériels & Méthodes	3
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
2 Résultats	6
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 Distribution du dernier marqueur converti	8
2.6 Restaurations de l'haplotype sauvage	8
2.7 Mutations <i>de novo</i>	9
2.8 Estimation des fréquences de conversion en faveur de GC	9
3 Discussions	10
3.1 Distribution des points de recombinaison	10
3.1.1 Une influence du taux de GC local ?	10
3.1.2 Une influence des bases adjacentes ?	11
3.2 Restaurations de l'haplotype sauvage	11
3.3 Mutations <i>de novo</i> et recombinaison homologue	12
3.4 Comment augmenter la puissance du test ?	13
Conclusion	14
1 Annexes	ii
1.2 Amorces utilisées	ii
1.3 Conditions de PCR	ii
1.4 Carte des constructions donneuses	ii
1.5 Traces de conversions	ii
1.6 Confirmations des restaurations des haplotypes sauvages	ii

Abbréviations

BER Base Excision Repair

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

GC% Contenu en GC

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	9
10	Liste des amorces utilisées	ii
11	Confirmation des restaurations	iii
12	Régions converties	iv

Liste des tableaux

1	Fréquences de transformation	5
3	Dénombrement des cas de restauration	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 fonctionnel dans la survie de l'individu, ou est-ce qu'il est déterminé par des variations des patrons de mutations ou de réparation de l'ADN ? L'un des mécanismes essentiels de réparation de l'ADN est la recombinaison homologue, impliquée dans la réparation des cassures doubles brins. Au cours de la réparation de l'ADN par recombinaison homologue, 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*¹, 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 C¹³. Le biais de conversion génique vers GC (gBGC^a) tend à augmenter localement le taux de GC des régions fortement recombinantes^{3,10} ? 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^{6,4} 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.*⁹ et de Yahara *et al.*[?] ont montré que les zones recombinantes des génomes procaryotes ont un taux de GC plus élevé que les régions non-recombinantes. Elles également ont 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 est 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

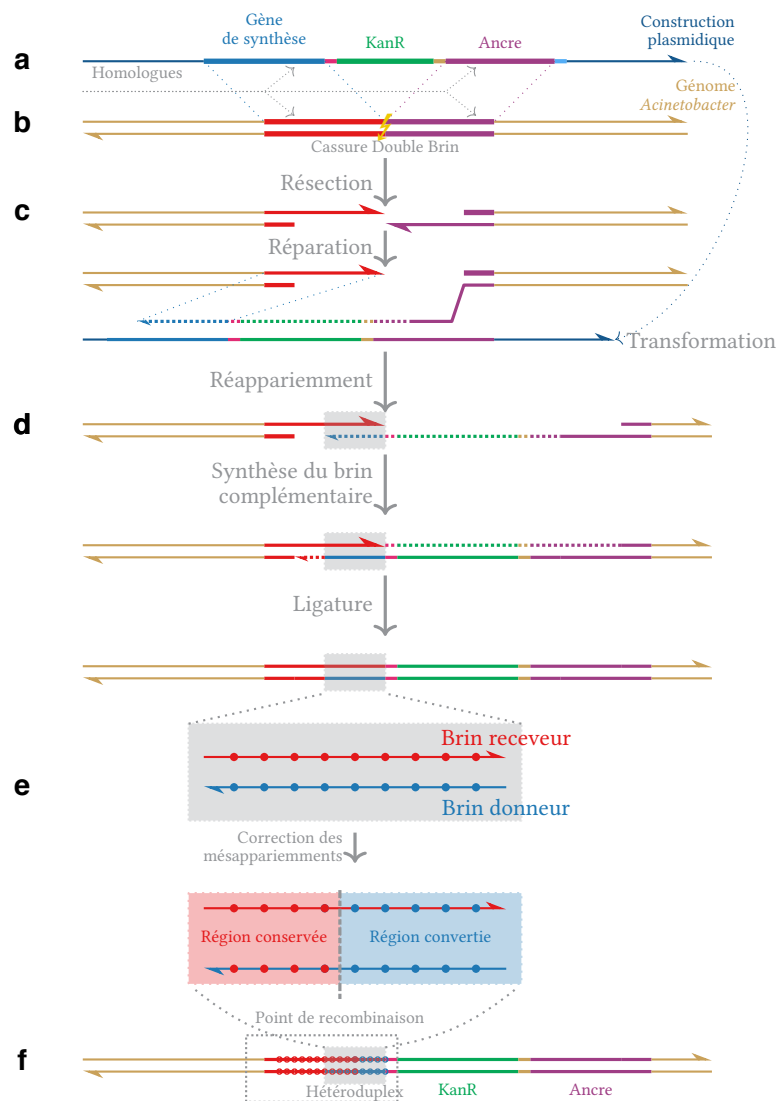


Figure 1 : **Mécanismes de conversion génique au cours de la transformation naturelle par les constructions plasmidiques**

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 un antibiotique (KanR) et une zone recombinogène (Ancre) qui permet d'augmenter la fréquence de transformation.

Lors d'une cassure double brin (b), la région en 3' subit une résection, qui est réparée sur la base d'une matrice d'ADN homologue (c). La transformation naturelle permet de réparer ces cassures en utilisant l'ADN exogène sous forme simple brin. 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^a pour deux raisons. 1) Le genre *Acinetobacter* montre une corrélation entre taux de GC et taux de recombinaison⁹ 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[?]. 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 2). Les mésappariements présents sur la séquence donneuse permettent 1) de localiser la position du point de recombinaison et 2) de déterminer la sens de la conversion. La réparation des mésappariements introduits nous renseigne sur 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. La taille d'échantillon n'est pour l'instant pas suffisamment élevée pour en tirer des conclusions significatives quant à l'existence du gBGC chez *A. baylyi*.

a. *A. baylyi* était appelée *A. calcoaceticus* avant 1995, mais la taxonomie du genre a fait l'objet d'une révision[?]

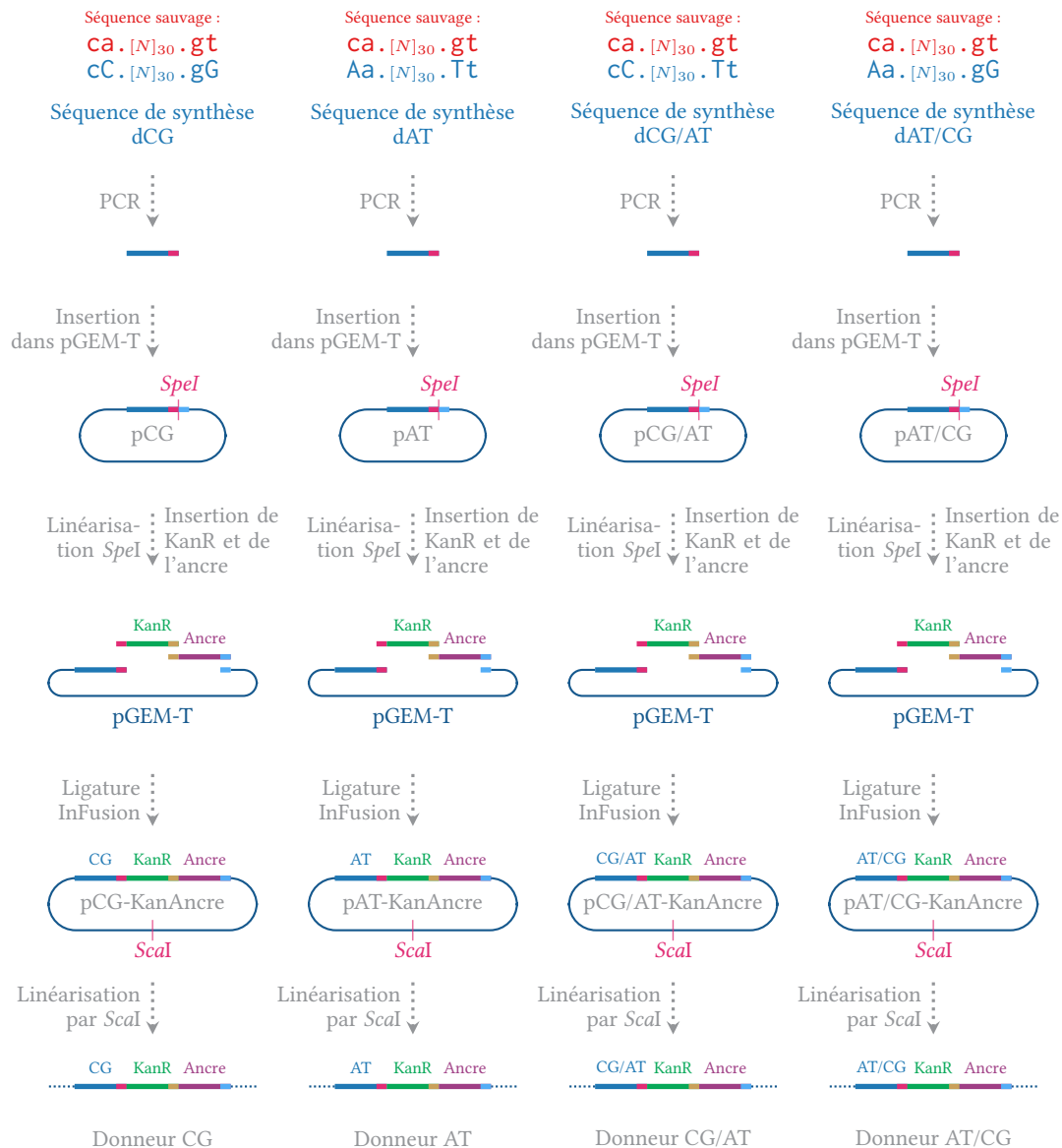


Figure 2 : **Constructions moléculaires et préparation de l'ADN recombinant**

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 digestions enzymatiques *SpeI* au niveau des sites d'insertion de la cassette de résistance à la kanamycine et de l'ancre. Ces fragments sont préalablement amplifiés par PCR en utilisant des amorces porteuses de régions 3' flottantes, complémentaires des extrémités des fragments voisins. La ligature 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 putativement une 3-oxoacyl-ACP reductase, impliqué dans la synthèse des acides gras¹⁴ 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'événements détectables.

Nous avons conçu 4 gènes de synthèse, nommés respectivement dCG^a, dAT, dCG/AT et dAT/CG (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 dCG/AT introduit d'abord un mésappariement GC | AT, puis un mésappariement AT | GC, le gène de synthèse dAT/CG 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 homologue à la séquence en 3' du locus d'intérêt et permet d'augmenter les fréquences de transformation^{2?}.

a. dCG pour donneur GC.

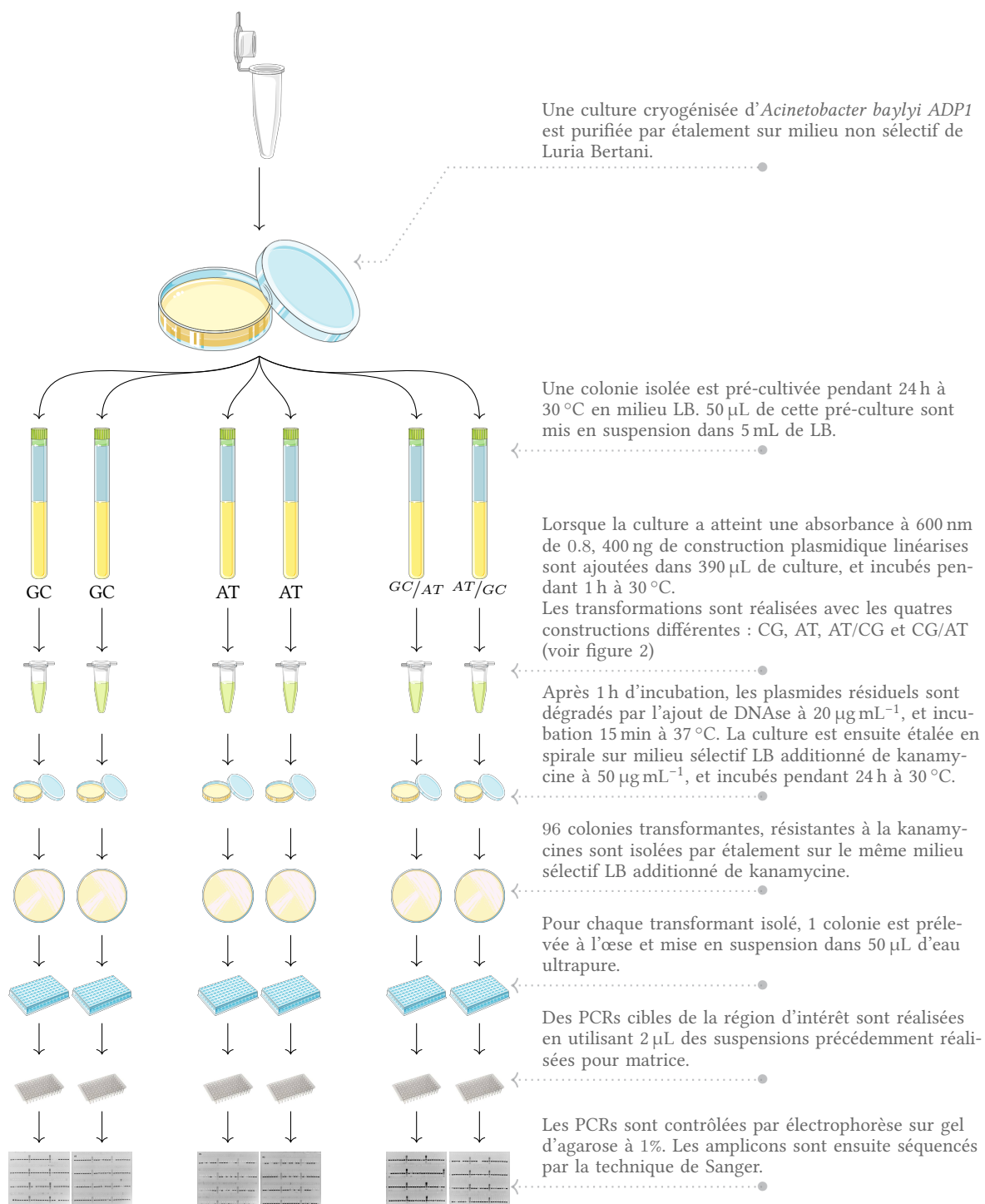


Figure 3 : **Protocole de transformation et d'obtention des amplicons des zones de recombinaison.**

Les gènes synthétiques ont d'abord été amplifiés par PCR ^a avec les amorces 1392 et 1393. Les plasmides obtenus ont été insérés dans la souche *E.coli* OneShot[®] 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 ligature a été réalisée simultanément par le kit InFusion (Takara Clontech, Saint Germain en Layes, France). Le produit de ligature 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 µg mL⁻¹ de kanamycine à 50 µg 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*

1 µg 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 µL 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 µmol 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 µg mL⁻¹ et incubées 24 h à 30 °C. Les dénombrements ont été effectués via un compteur automatique Scan[®]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 µmol mL⁻¹. Une colonie isolée par clone a été mise en suspension dans 50 µL d'H₂O ultra-pure. 2 µL de ces suspensions ont servi de matrice pour amplifier par PCR les régions recombinantes avec la Taq polymérase haute fidélité Phusion ^b (ThermoFischer, Waltham, États-

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

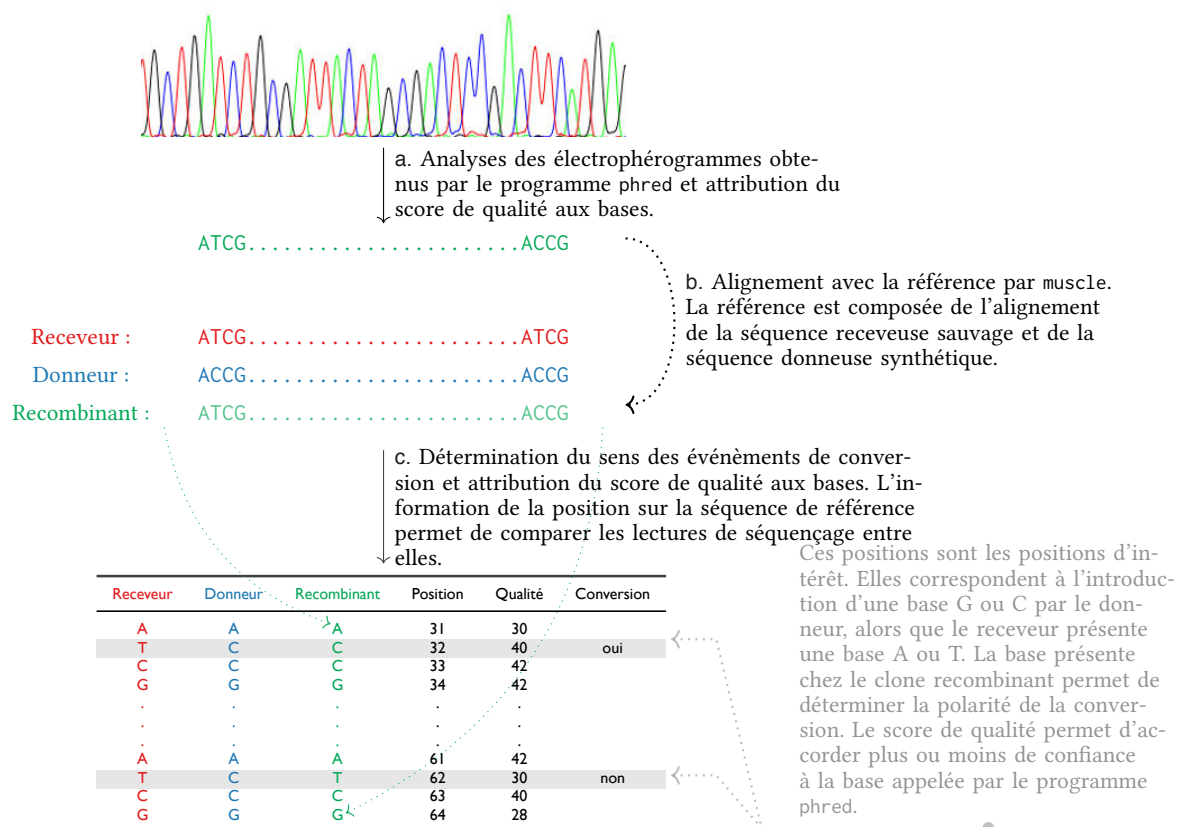


Figure 4 : **Exemple d'analyse de la zone de recombinaison pour un clone transformant**

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).

Unis). Les amplicons ont été vérifiés par migration sur gel d'agarose à 1% et séquencés par la technique de Sanger[?] (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[?] et convertis en format universel FASTA (voir figure 4). Les séquences obtenues ont été alignées aux références par muscle v3.8.31[?]. 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[?] 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[?]. Les programmes développés sont accessibles à l'adresse 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**

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

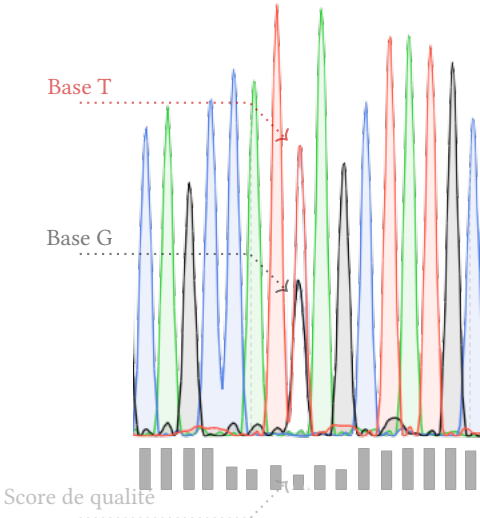


Figure 5 : **Exemple de marqueur montrant des traces de contaminations**

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.

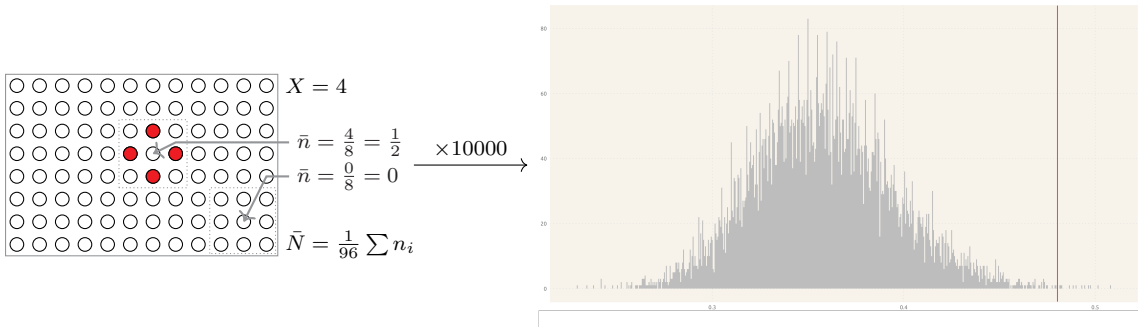


Figure 6 : **Des contaminations dues au hasard ?**

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 1×10^4 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 1×10^4 \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* 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 ci-contre, sont de l'ordre de 1×10^{-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

2 séquences trop courtes pour être alignées ont été rejetées. 9 autres séquences présentaient une qualité médiane significativement inférieure à celle des autres séquences et ont également été rejetées. Au total, 8561 positions variantes entre le donneur et le receveur ont été analysées. Globalement, la qualité des 873 séquences suit un profil attendu : les bases en début et fin de lecture sont d'une qualité inférieure à la qualité médiane de 51.

Cependant, 9,8% des marqueurs sont d'une qualité de séquençage inférieure à la qualité de la région environnante. C'est la marque de pics secondaires sur les électrophérogrammes (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. Pour y répondre, nous avons séquencé à nouveau 31 isolats issus d'un clone séquencé en premier

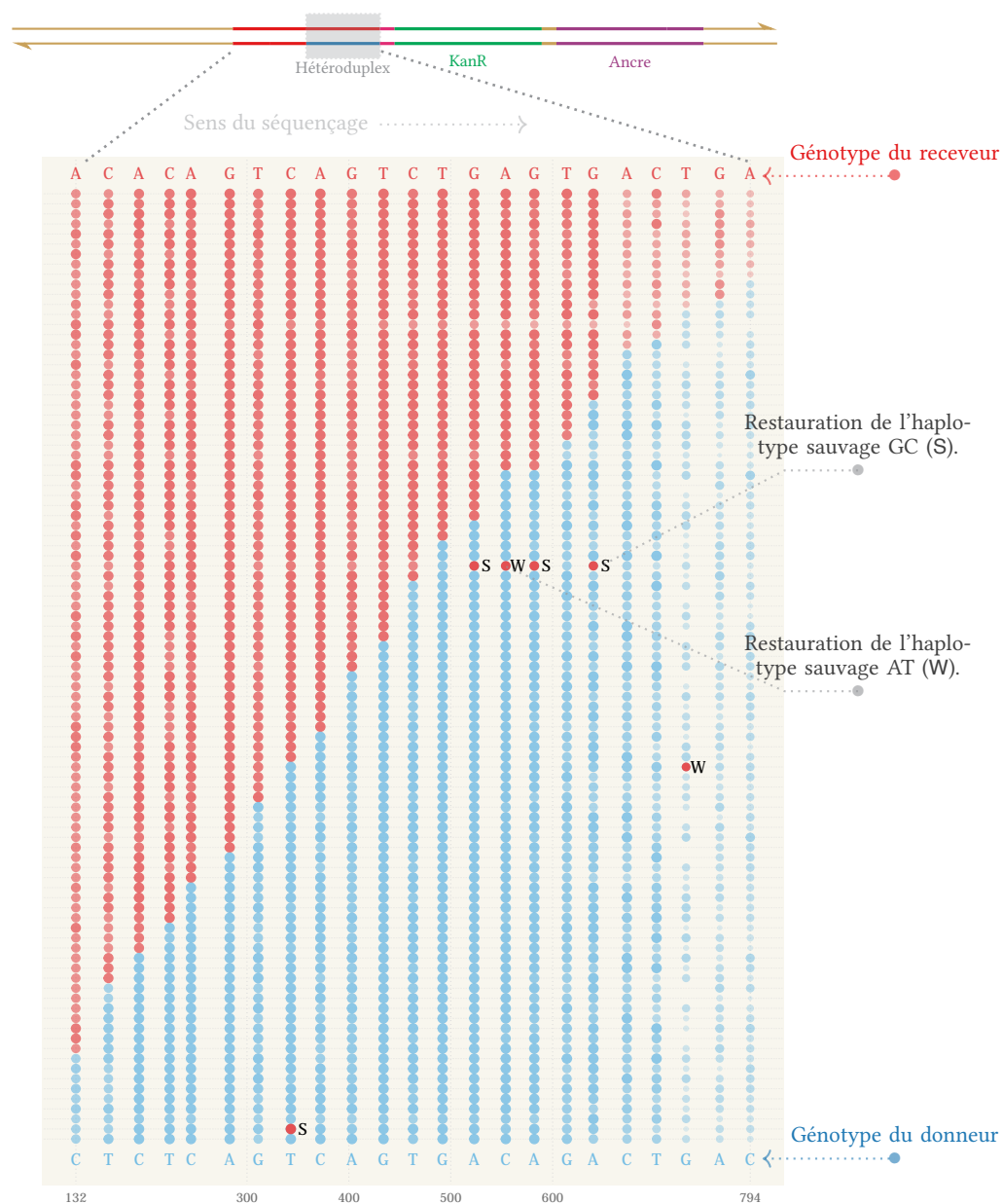


Figure 7 : **Zones de recombinaison entre un locus génomique d'*Acinetobacter baylyi* et un gène synthétique donneur d'allèles CG et AT.**

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.

lieu et qui montrait des pics secondaires. Aucun des sous-clones séquencé ne montrent l'allèle correspondant au pic secondaire (voir figure 11). De plus, nous avons adapté le protocole expérimental de façon à limiter l'apparition de ces artefacts. Nous avons supposé que des réminiscences de construction donneuse dans la suspension étalée aurait pu conduire à plusieurs événements de transformation consécutifs et donc à ces profils hétérogènes. Les suspensions ont donc été traitées par de la DNase, de façon à dégrader les plasmides résiduels. Enfin, pour s'affranchir de l'analyse de colonies hétérogènes, nous avons isolé chacun des transformants, et analysé un clone unique et purifié du transformant initial. La persistance de ces pics secondaires après adaptation du protocole suggère qu'ils ne reflètent pas le signal biologique envisagé.

2) S'il s'agit alors d'un artefact méthodologique, la population d'amplicon séquencée est hétérogène à la suite de contaminations entre les puits des plaques, qui ont pu avoir lieu au cours du séquençage ou au cours des PCRs. Dans ce cas, la répartition dans la plaque des puits dont la séquence montre des traces de contamination ne devrait pas être déterminée seulement par le hasard. Nous avons montré par simulation qu'elles ne l'étaient pas (voir figure 6). Les séquences montrant des pics secondaires sont plus souvent voisines avec une autre séquence affectée que si elles étaient réparties aléatoirement. En conséquence, nous avons décidé de n'étudier que les marqueurs dont la qualité est supérieure à 40 : cela correspond à la probabilité 1×10^{-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 CG et AT. Les zones de recombinaisons des clones transformés par les donneurs CG, AT et AT/CG sont détaillées en annexe 1.5. En moyenne, à notre locus cible, 393 ± 228 nucléotides ont été transférés et intégrés dans le génome. La position du dernier marqueur converti indique que le point de recombinaison (voir figure 1 et 7) 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 montrent des cas de conversion simple,

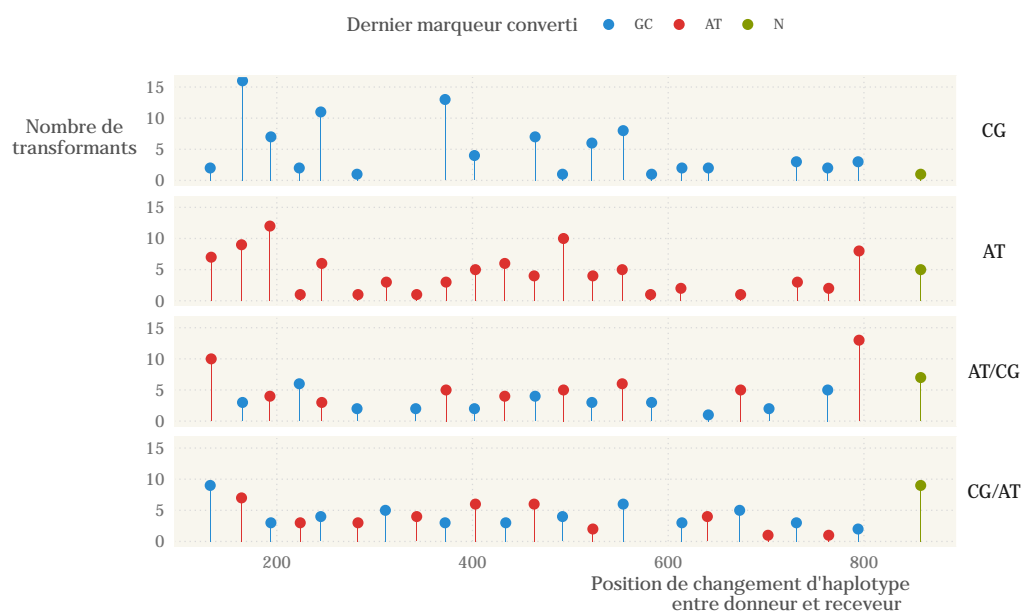


Figure 8 : **Distribution de la position du dernier marqueur converti**

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*.

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 sont des restaurations de l'haplotype receveur dans l'hétéroduplex.

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

Selon l'hypothèse gBGC, la région convertie devrait être plus longue lorsque la construction donneuse induit des réparations vers CG que lorsqu'elle induit des réparations vers AT. La différence entre la longueur moyenne de région convertie par les donneurs respectivement CG et AT n'est pas significative (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 AT/CG et celle par le donneur CG/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 Distribution du dernier marqueur converti

Les constructions alternant AT et CG permettent de déterminer si davantage de bases GC ont été introduites que de bases AT. Sur l'ensemble des marqueurs des recombinants transformés par pAT/CG et pCG/AT, 720 bases C ou G ont été introduites contre 677 bases A ou T. Cet écart n'est pas significatif (test du χ^2 , probabilité critique = 0,25).

2.6 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 1.5 page iv). Nous avons confirmé qu'il s'agissait

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

ADN synthétique donneur	Nucléotide Restauré	
	AT	CG
AT	-	4
CG	0	-
AT/CG	3	2
CG/AT	1	4
Total	4	10

bien de restaurations de l'allèle sauvage de deux façons. 1) Expérimentalement, nous avons séquencé une sous-population de 30 clones issus d'un isolat séquencé en premier lieu. Tous montrent la même alternance au même site (voir figure 11 en annexe 1.6). 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.

2.7 Mutations *de novo*

Nous avons détecté 20 cas de mutations *de novo*. Elles ne sont pas associées aux sites variants et sont d'une qualité moyenne supérieure à 41. Parmi ces 20 cas, 14 introduisent une base G ou C et 6 introduisent une base A ou T. L'écart n'est pas statistiquement significatif (test du χ^2 , probabilité critique = 0,07). Les 14 cas de mutations vers G ou C sont tous dans la région convertie, alors que les 6 cas de mutations vers A ou T sont tous dans la région conservée.

2.8 Estimation des fréquences de conversion en faveur de GC

Dans le but de comparer la quantité de bases A et T qui ont été converties en G et C ($N_{AT \rightarrow GC}$) à celle des bases G et C qui ont été converties en A et T ($N_{GC \rightarrow AT}$), nous avons échantillonné aléatoirement, par construction donneuse, 80 clones pour lesquels on dispose des 23 marqueurs. Cet échantillonnage permet de compenser les éventuelles différences de longueur moyenne de région convertie entre les constructions donneuses.

$N_{AT \rightarrow GC}$ est significativement supérieure à $N_{GC \rightarrow AT}$ (test du χ^2 , probabilité critique = 0,029 au premier tirage). Pour compenser un effet dû à la nature des transformants échantillonnés pendant ce tirage, nous avons répété la mesure 1×10^4 fois. Dans 99,98 % des cas, $N_{AT \rightarrow GC}$ est supérieure à $N_{GC \rightarrow AT}$. Le test est significatif au seuil de 0,05 dans 44,01 % des cas. Ce résultat indique que sur l'ensemble des transformants séquencés, un nombre plus élevé de bases AT ont été converties en GC que l'inverse, et que l'écart est significatif dans 44,01 % des cas.



Figure 9 : Taux de GC
TODO légènder

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, un processus qui est biaisé vers l'introduction des bases C et G chez certains eucaryotes.

3.1 Distribution des points de recombinaison

L'hypothèse gBGC prédit que la région convertie devrait être plus longue lorsque la conversion introduit des bases GC que lorsqu'elle introduit des bases AT. Nous n'avons pas détecté de différences significatives entre les deux constructions donneuses pCG/AT et pAT/CG. Cependant, quelle que soit la construction donneuse, la distribution du point de recombinaison suit globalement le même patron, avec une zone de plus faible intensité entre 600 et 800 pb et une autre autour de 350 pb. En fait, la distribution du point de recombinaison est assez différente de celle décrite par Yáñez-Cuna *et al.* [?] chez un modèle de *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 800pb. Les différences entre leurs résultats et les nôtres peuvent potentiellement s'expliquer par les différences de modèles et de locus. Nous avons en effet supposé que la composition en base du locus avait un impact sur l'efficacité locale de la recombinaison.

Une influence du taux de GC local ?

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 50pb. Le point froid de recombinaison à 600pb semble correspondre avec un taux de GC plus faible localement. Nous avons donc supposé que l'efficacité de la recombinaison, qui dépend pour partie de la probabilité de rejet du brin exogène et de la probabilité d'extension de l'hétéroduplex, pourrait être influencée par le taux de GC local. 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 pûrement basée sur l'appariement Watson-Crick entre les brins [?].

En fait, la zone pauvre en GC à 600pb 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.*⁹ entre taux de GC et taux de recombinaison.

Une influence des bases adjacentes ?

Lieb *et al.*[?] ont montré chez *E. coli* que les enzymes du VSP (Very Short Patch repair) ré-pairaient un mésappariement dans un sens spécifique en fonction des bases au voisinage immédiat du mésappariement. Cette activité de réparation est spécifiquement liée au produit du gène *vsr*^{??}. *A. baylyi* possède également certaines des enzymes impliquées dans les voies du VSP⁷.

Si ce genre de mécanisme est à l'œuvre chez *A. baylyi*, les marqueurs AT et les marqueurs GC de nos constructions ne sont 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 événement de conversion serait alors rigoureusement dans le même contexte.

3.2 Restaurations de l'haplotype sauvage

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 soi⁵, 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 décrits¹². Chez la levure, ces cas complexes ont été interprétés comme des basculements multiples entre le choix des matrices permettant la réparation des cassures[?].

Dans le contexte de la conversion génique biaisée vers GC, est-ce que ces cas de restaurations pourraient expliquer un biais ? Chez les eucaryotes, la correction des mésappariements dans

les cellules en mitose est effectuée en partie par la voie du 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 opposé⁸. Cette voie a été considérée comme l'un des moteurs possibles du gBGC chez les mammifères³, mais a été rejetée chez la levure¹⁰. 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.

Est-ce que le BER, s'il est actif chez les procaryotes, pourrait conduire à de la conversion biaisée vers GC ? L'étude des régions converties chez des mutants d'*A. baylyi* déléétés pour les fonctions clés de la correction des mésappariements par le BER devrait permettre de répondre à cette question.

Ces cas de restaurations de l'haplotype sauvage ajoutent une incertitude sur la nature des premiers marqueurs conservés (en rouge sur la figure 7). En effet, ces marqueurs 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. Dans tous les cas, la base séquencée correspond à l'allèle sauvage du marqueur. Ces cas peuvent perturber la mesure de la longueur de région convertie.

3.3 Mutations *de novo* et recombinaison homologue

Nous avons observé 20 cas de mutations spontanées, qui sont dans 70 % des cas localisés dans la région convertie, et introduisent des bases G et C. Ils peuvent être induits par deux phénomènes.

1) Ces mutations spontanées peuvent traduire des mutations survenues *ex-vivo*, au cours de la construction et/ou de l'amplification des plasmides donneurs chez *E. coli* (voir figure 2).

2) Elles peuvent également traduire des mutations survenues *in-vivo*, au cours de l'intégration de l'ADN dans la cellule ou de la réparation par recombinaison homologue du brin lésé. En effet, la recombinaison homologue a été démontrée à plusieurs reprises comme mutagène^{5??}, les polymérases impliquées tendent à introduire plus d'erreurs que les polymérases classiques[?]. Cependant, à notre connaissance, il n'a jamais été démontré que la mutation favorisait l'introduction des bases C et G au cours de la recombinaison homologue. Au contraire, la mutation est universellement biaisée vers l'introduction des bases A et T, chez les eucaryotes¹¹ comme chez

les procaryotes^{6,4}. Est-ce que le biais de conversion génique vers GC pourrait être influencé par un processus mutationnel spécifique à la recombinaison homologue ?

3.4 Comment augmenter la puissance du test ?

Pour détecter une différence de l'ordre de 30 paires de bases entre longueurs de régions converties correspondant à l'introduction d'un marqueur supplémentaire, en supposant qu'elles sont normalement distribuées, avec l'écart type observé dans nos résultats (227), avec une puissance de 0,8 et un niveau de confiance de 0,05, il faudrait une taille d'échantillon de 717 (test t) par groupe.

Pour détecter une différence significative entre le nombre de marqueurs où l'allèle sauvage respectivement AT ou GC est restauré, en supposant que l'effet observé soit moyennement fort ($\omega = 0,5$), avec un niveau de confiance de 0,05 et une puissance de 0,8, il faudrait observer 31 marqueurs. En présumant que les cas de restauration se produisent de façon indépendante entre chaque clones, étant donné que nous avons observé des cas de restaurations 4 % des recombinants, il faudrait séquencer 775 clones au total. Si l'effet est plus faible ($\omega = 0,3$), il faudrait observer 87 marqueurs, donc séquencer 2175 clones.

Cette taille d'échantillon n'est pas envisageable en utilisant les techniques actuelles : le séquençage par la technique de Sanger implique un coût prohibitif pour une telle taille d'échantillon. Nous avons étudié différents procédés permettant de séquencer les régions converties à haut débit. Le séquençage à haut débit permettrait de réduire les coûts de séquençage. Il nécessite néanmoins de prendre en compte les facteurs suivants. 1) La taille des lectures est de l'ordre de 2×300 paires de bases par lecture, soit un total de 600pb, inférieure à la taille du locus considéré. 2) Lors de la construction de la librairie de séquence, les amplicons sont mélangés les uns aux autres. Il faut pouvoir les discriminer les uns des autres facilement, de façon à individualiser les événements de recombinaison. 3) La divergence très faible des amplicons peut introduire des artefacts lors de la capture des clusters par la caméra de séquençage.

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 et les fréquences de restauration de l'haplotype sauvage. Ces informations sont capitales pour mettre au point un plan expérimental permettant de confirmer ou non l'existence du biais de conversion vers GC chez les procaryotes. Les données obtenues à ce jour ne permettent pas de conclure sur ce point. Il semble cependant que les cas de restaurations montrent une tendance à la restauration préférentielle des bases C et G. De plus, sur l'ensemble des recombinants obtenus, davantage de bases AT ont été converties en GC que l'inverse.

Ce travail devra 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* *NOMDESPECE* ?. 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.

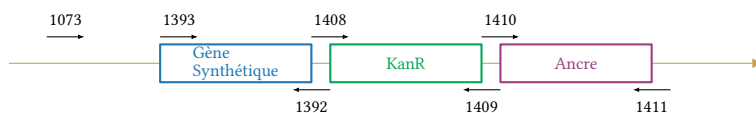
Références

- [1] **Arbeithuber, B., A. J. Betancourt, T. Ebner, et I. Tiemann-Boege.** 2015. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *PNAS* 112 :2109–2114.
- [2] **Chen, I. et D. Dubnau.** 2004. DNA uptake during bacterial transformation. *Nat Rev Micro* 2 :241–249.
- [3] **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.
- [4] **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.
- [5] **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.
- [6] **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.
- [7] **Edgar, R. C.** 2004. MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 :1792–1797.
- [8] **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.
- [9] **Ewing, B., L. Hillier, M. C. Wendl, et P. Green.** 1998. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res.* 8 :175–185.
- [10] **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.
- [11] **Hershberg, R. et D. A. Petrov.** 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* .
- [12] **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.
- [13] **Hildebrand, F., A. Meyer, et A. Eyre-Walker.** 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 6 :e1001107.
- [14] **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* .
- [15] **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.
- [16] **Krokan, H. E. et M. Bjørås.** 2013. Base Excision Repair. *Cold Spring Harb Perspect Biol* 5 :a012583.
- [17] **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.
- [18] **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.
- [19] **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.
- [20] **Lieb, M.** 1983. Specific mismatch correction in bacteriophage lambda crosses by very short patch repair. *Molec Gen Genet* 191 :118–125.
- [21] **Lieb, M.** 1985. Recombination in the λ repressor gene : evidence that very short patch (VSP) mismatch correction restores a specific sequence. *Mol Gen Genet* 199 :465–470.
- [22] **Lynch, M.** 2010. Rate, molecular spectrum, and consequences of human mutation. *PNAS* 107 :961–968.
- [23] **Malkova, A. et J. E. Haber.** 2012. Mutations Arising During Repair of Chromosome Breaks. *Annual Review of Genetics* 46 :455–473.
- [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] **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.
- [26] **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.

- [27] Pomerantz, R. T., M. F. Goodman, et M. E. O'Donnell. 2013. DNA polymerases are error-prone at RecA-mediated recombination intermediates. *Cell Cycle* 12 :2558–2563.
- [28] R Core Team. 2015. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- [29] Rodgers, K. et M. McVey. 2016. Error-Prone Repair of DNA Double-Strand Breaks. *Journal of Cellular Physiology* 231 :15–24.
- [30] 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.
- [31] Vallenet, D., E. Belda, A. Calteau, S. Cruveiller, S. Engelen, A. Lajus, F. L. Fèvre, C. Longin, D. Mornico, D. Roche, Z. Rouy, G. Salvignol, C. Scarpelli, A. A. T. Smith, M. Weiman, et C. Médiague. 2013. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucl. Acids Res.* 41 :D636–D647.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.

1 Annexes

1.2 Amorces utilisées



Cible	Amorce	Séquence	T _m (°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	GTAAAACGACGGCCAGT	53.9

Figure 10 : Liste des amorces utilisées

1.3 Conditions de PCR

1.4 Carte des constructions donneuses

1.5 Traces de conversions

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

1.6 Confirmations des restaurations des haplotypes sauvages

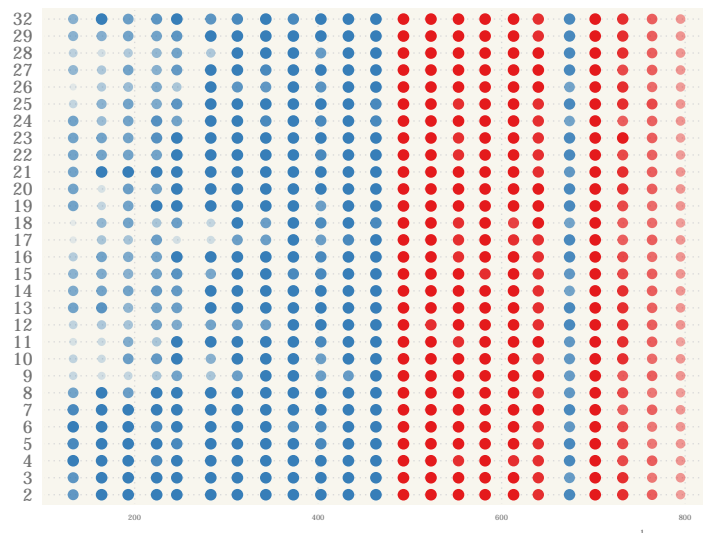


Figure 11 : **Confirmation d'une restauration d'haplotype sauvage**

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. 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.

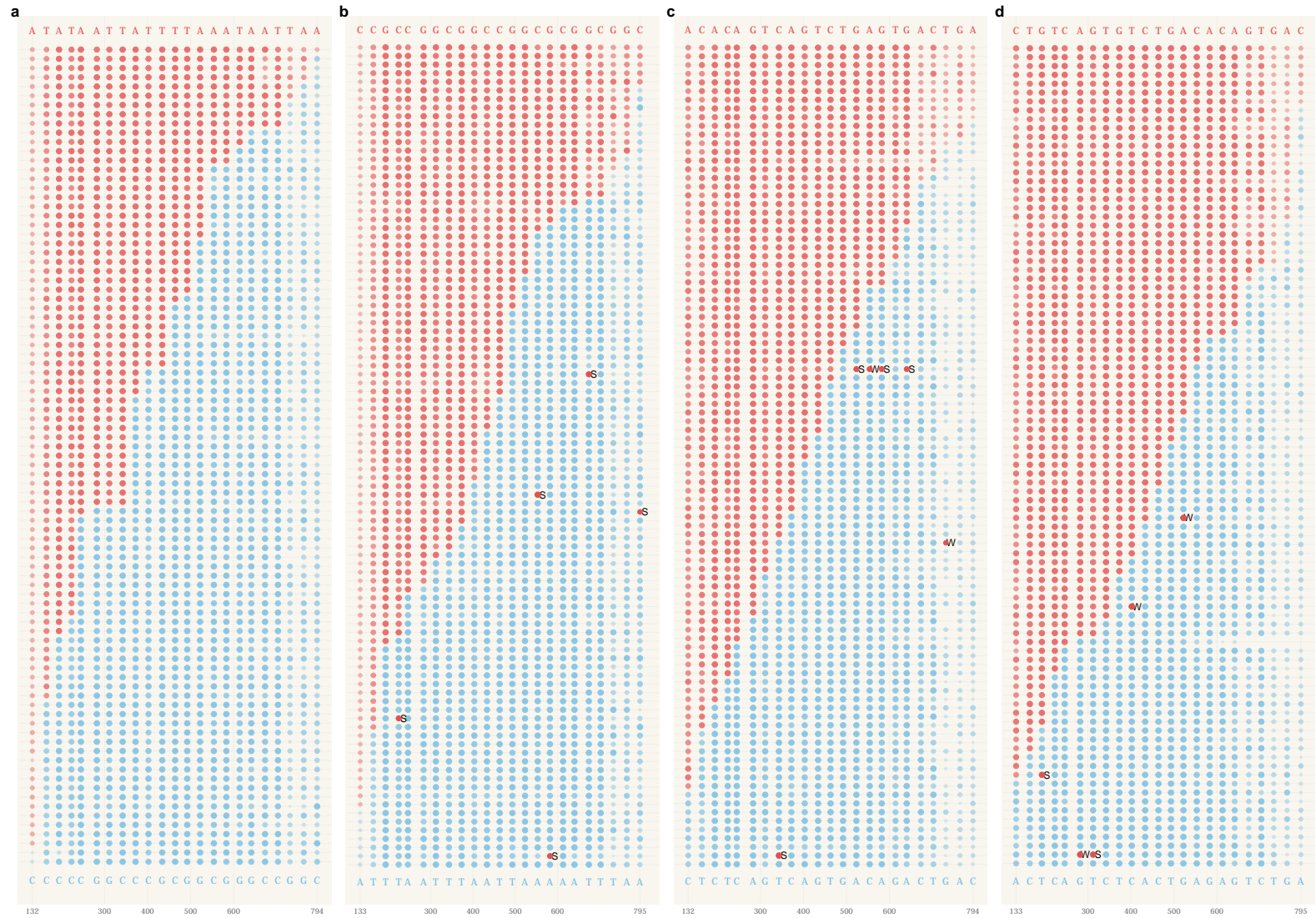


Figure 12 : Pellentesque dapibus suscipit ligula. Donec posuere augue in quam. Aliquam feugiat tellus ut neque. Nulla facilisis, risus a rhoncus fermentum, tellus tellus lacinia purus, et dictum nunc justo sit amet elit.

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 1×10^{-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 montrent que d'avantage de bases A ou T ont été converties en G ou C que l'inverse, une observation compatible avec notre hypothèse. La taille d'échantillon n'est toutefois pas suffisante pour que les comparaisons soient significatives. Ces résultats suggèrent que la conversion génique biaisée pourrait être à l'œuvre chez les procaryotes, mais le signal obtenu n'est pour l'instant pas suffisamment fort pour en tirer des conclusions.

