



### *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](https://commons.wikimedia.org/wiki/File:Dark_side_of_the_moon.jpg)

# 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 Analyses visuelles des zones de recombinaison . . . . .	6
2.2 Analyses de la position des points de recombinaison . . . . .	6
2.3 Analyses des restaurations de l'haplotype sauvage . . . . .	7
<b>3 Discussions</b>	<b>8</b>
<b>Conclusion</b>	<b>9</b>
<b>1 Annexes</b>	<b>ii</b>
1.2 Amorce utilisées . . . . .	ii
1.3 Carte des constructions donneuses . . . . .	ii
1.4 Traces de conversions . . . . .	ii
1.5 Confirmations des restaurations des haplotypes sauvages . . . . .	ii

# Abbreviations

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

**GC%** Contenu en GC

## 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	Zones de recombinaison détaillée . . . . .	5
6	Distribution de la position du dernier marqueur . . . . .	6
7	Liste des amorces utilisées . . . . .	ii
8	Zones de recombinaison des clones transformés par la construction AT . . . . .	iii
9	Zones de recombinaisons des clones transformés par la construction GC . . . . .	iv
10	Zones de recombinaisons des clones transformés par la construction AT/GC . . . . .	v
11	Confirmation des restaurations . . . . .	vi

## Liste des tableaux

1	Dénombrements des derniers marqueurs avant le point de recombinaison . . . . .	6
2	Dénombrement des cas de restauration . . . . .	7





# Introduction

Chez les procaryotes, le taux de guanine et cytosine (GC%) varie de 16,5 % à 75 % dans les génomes séquencés à ce jour. Une telle amplitude soulève plusieurs questions. Celles qui nous intéressent concernent les mécanismes responsables de ces variations : est-ce qu'ils sont associés à la réplication de l'ADN, à sa réparation ou à une pression de sélection sur l'usage du code génétique ? Au cours des quinze dernières années, il a été démontré que la recombinaison homologue tend à augmenter localement le taux de GC dans les régions fortement recombinantes des génomes eucaryotes<sup>5,13,20</sup>. Ce mécanisme est essentiel pour la cellule : il permet de réparer les lésions de l'ADN. Au cours de la recombinaison, les brins associés au sein d'un hétéroduplex peuvent présenter des mésappariements. Leur correction entraîne de la *conversion génique*<sup>2</sup> : c'est une transmission à sens unique de l'information portée par un brin vers l'autre. Ce mécanisme est biaisé vers l'introduction préférentielle des bases C et G chez les mammifères et probablement chez un grand nombre d'eucaryotes<sup>15</sup>. C'est un processus non adaptatif : il impacte les régions codantes et non-codantes. Il peut s'opposer à l'action de la sélection en augmentant la probabilité de fixation des allèles G et C<sup>17,9</sup>.

En 2010, deux études simultanées<sup>11,10</sup> démontrent que : 1) le patron de *mutation* est universellement biaisé vers A et T chez les procaryotes, et 2) 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.* avancent que le GC% est en soi un trait soumis à la sélection, et rejettent l'hypothèse d'un biais de conversion génique biaisé vers GC chez les procaryotes. Cependant, les analyses récentes de Lassalle *et al.*<sup>12</sup> et de Yahara *et al.*<sup>21</sup> ont montré que les zones recombinantes des génomes procaryotes ont un taux de GC plus élevé que les régions non-recombinantes, une observation qui correspond précisément aux prédictions du biais de conversion vers GC (gBGC). Elles montrent également que la fixation préférentielle des allèles GC va à l'encontre de la fixation des allèles optimaux des codons, une signature caractéristique de l'action d'un processus non adaptatif tel que le gBGC. Ces observations sont compatibles avec l'hypothèse d'une conversion génique biaisée vers GC chez les procaryotes.

L'objectif de ce travail était d'obtenir une estimation expérimentale des fréquences de conversion vers GC chez la bactérie naturellement transformable *Acinetobacter baylyi* ADP1\*. En effet, cette bactérie développe un stade de compétence pendant la phase exponentielle de

---

\*. *A. baylyi* était appelée *A. calcoaceticus* avant 1995, mais la taxonomie du genre a fait l'objet d'une révision<sup>7</sup>

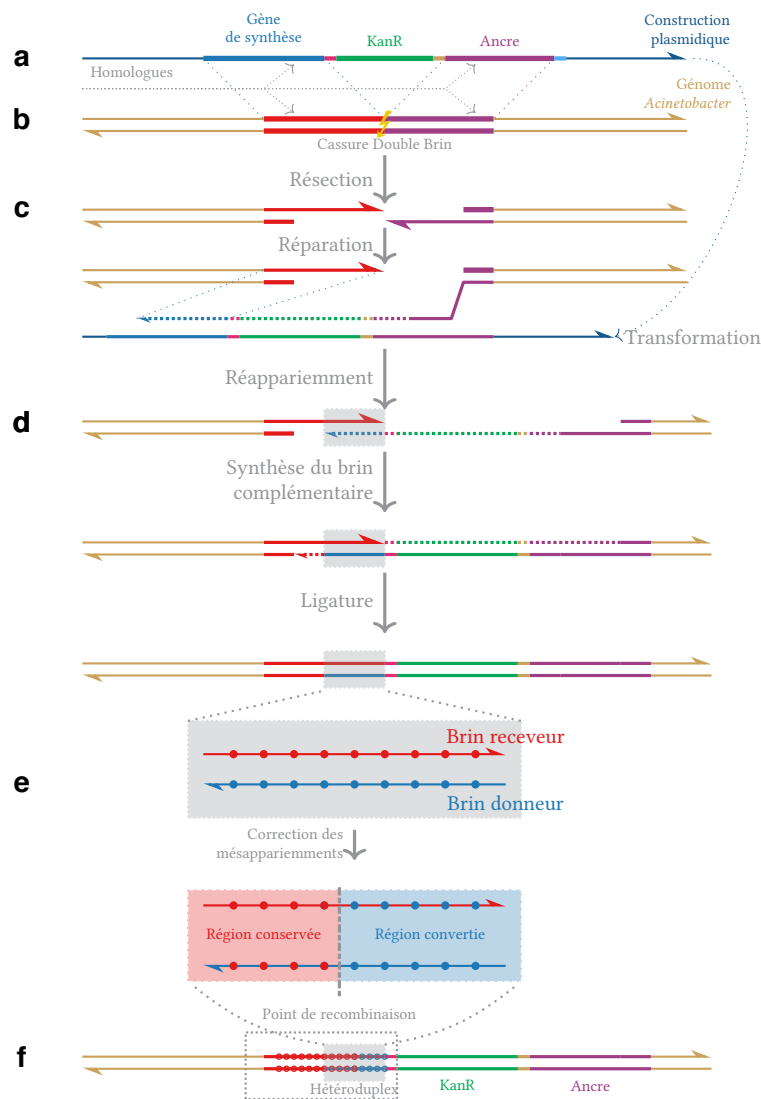


Figure 1 : **Mécanismes de conversion génique au cours de la transformation naturelle par les constructions plasmidiques**

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 100% homologue qui permet d'augmenter la fréquence de transformation.

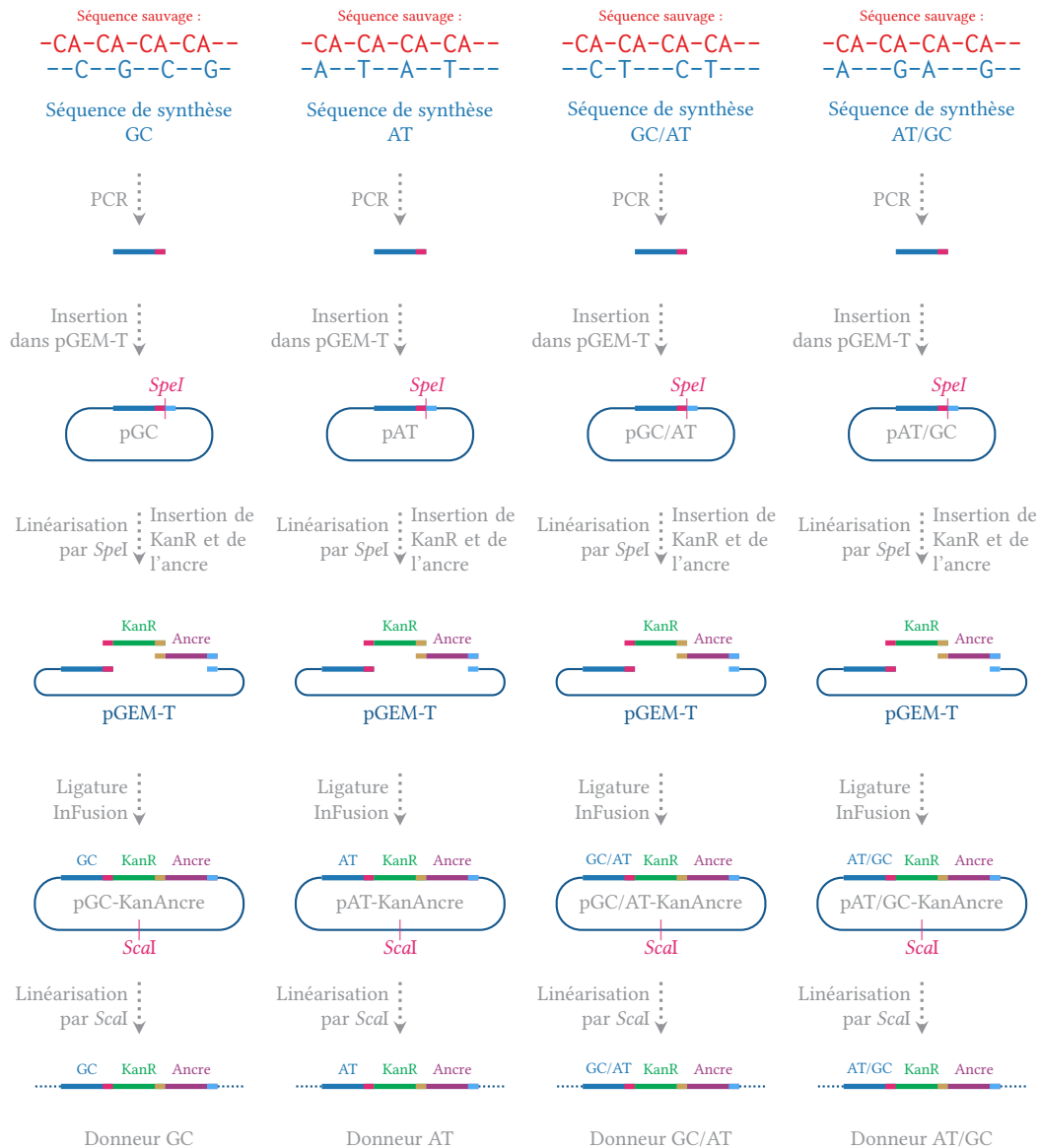
Lors d'une cassure double brin (b), la région en 3' subit une résection, qui est réparée sur la base d'une matrice d'ADN homologue (c). La transformation naturelle permet de réparer ces cassures en utilisant l'ADN exogène sous forme simple brin. Le réappariement (d) avec le brin parental conduit à la 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.



croissance, et intègre de l'ADN exogène sous forme simple brin<sup>1</sup> 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*.

insérer débuts de conclusions ici



**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 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 permettant d'introduire ces mésappariements. Nous avons choisi 23 sites sur un locus neutre du génome, séparés par 30 paires de bases\*. Ce locus est connu au laboratoire pour permettre de fortes efficacités de transformations. Il est localisé à la position 47184 du génome de référence. Il code putativement une 3-oxoacyl-ACP reductase, impliqué dans la synthèse des acides gras<sup>19</sup>.

Les sites choisis sont des dinucléotides alternant une base GC avec une base AT. Le gène de synthèse GC introduit des mésappariements GC aux sites AT (GC | AT), le gène de synthèse AT introduit des mésappariements AT aux sites GC (AT | GC) (voir figure 2). Pour s'affranchir d'un possible effet dû à l'écart de GC% entre la séquence synthétique 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 CG/AT introduit d'abord un mésappariement GC | AT, puis un mésappariement AT | GC, le gène de synthèse AT/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<sup>4,14</sup>.

Les gènes synthétiques ont d'abord été amplifiés par PCR avec les amorces 1392 et 1393 (voir annexe 1.2), puis ligaturés dans le plasmide pGEM-T. Les plasmides obtenus ont été insérés dans la souche *E.coli* OneShot® TOP-10 (Invitrogen, Carlsbad, États-Unis), suivant le protocole

---

\*. Ils introduisent 3% de divergence avec la séquence sauvage. Un plus grand nombre de marqueurs diminue les fréquences de recombinaison ; un nombre inférieur diminue la qualité du signal.

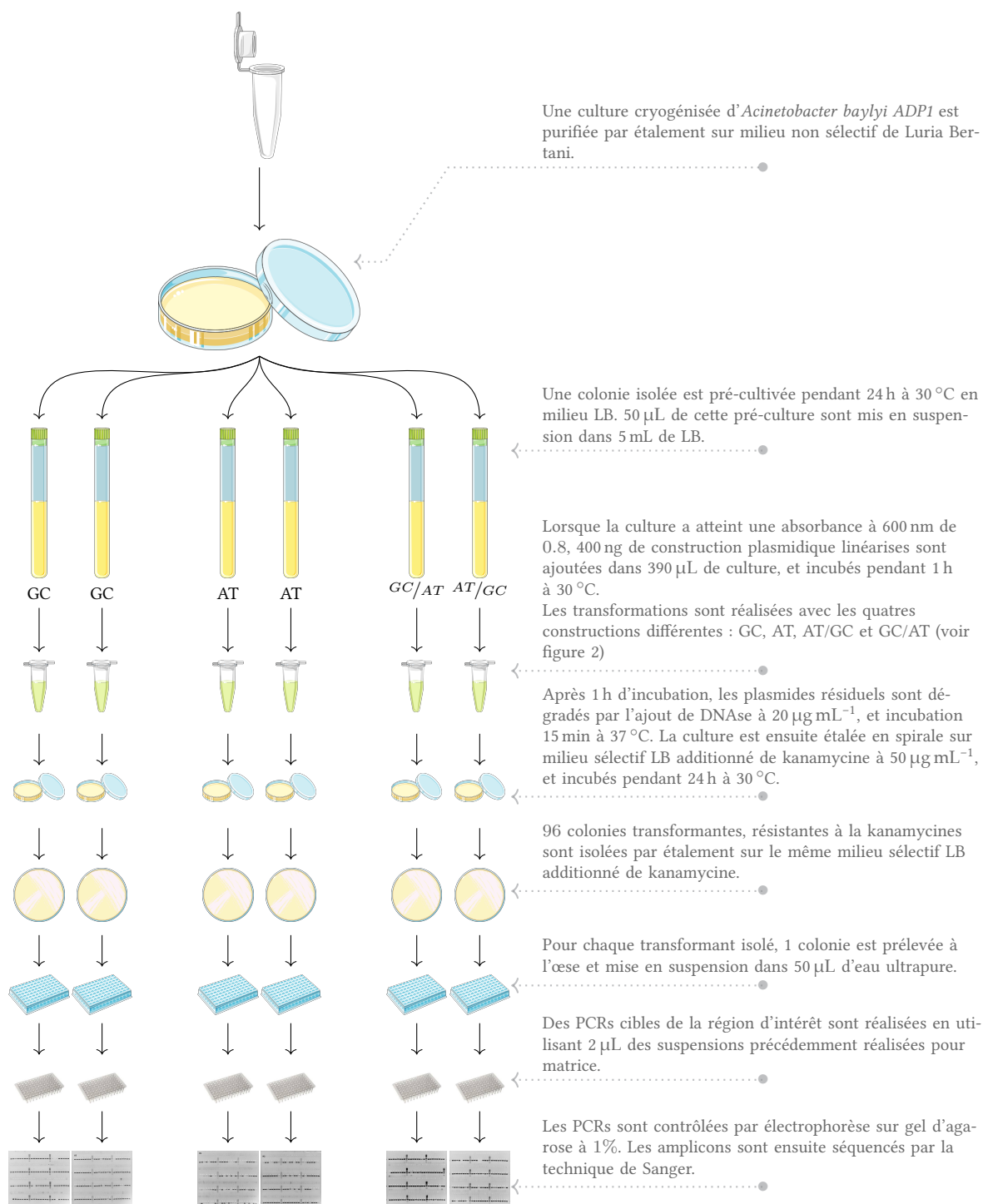


Figure 3 : **Protocole de transformation et d'obtention des amplicons des zones de recombinaison.**

du fabriquant. Le sens d'insertion du gène synthétique dans le plasmide a été vérifié par PCR avec les amorces M13R et 1392 (voir annexe 1.2), et l'absence de mutation a été vérifié 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  $75 \mu\text{g mL}^{-1}$  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 (voir annexe 1.2).

### 1.3 Transformations d'*Acinetobacter baylyi*

1  $\mu\text{g}$  de plasmide a été extrait et linéarisé par l'enzyme *Scal* (ThermoFisher). L'enzyme a été ensuite inactivée par incubation 10 min à  $70^\circ\text{C}$ . 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  $28^\circ\text{C}$  en présence de 200 ng de plasmide linéarisé. Les suspensions ont été ensuite incubées 15 min à  $37^\circ\text{C}$  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  $30^\circ\text{C}$ . Les dénombrements ont été effectués via un compteur automatique Scan<sup>®</sup> 1200 (InterScience).

96 clones recombinants ont été isolés et incubés 24 h à  $30^\circ\text{C}$  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 (ThermoFischer, Waltham, États-Unis). Les amplicons ont été vérifiés par migration sur gel d'agarose à 1% et séquencés par la technique de Sanger<sup>18</sup> (GATC Biotech).

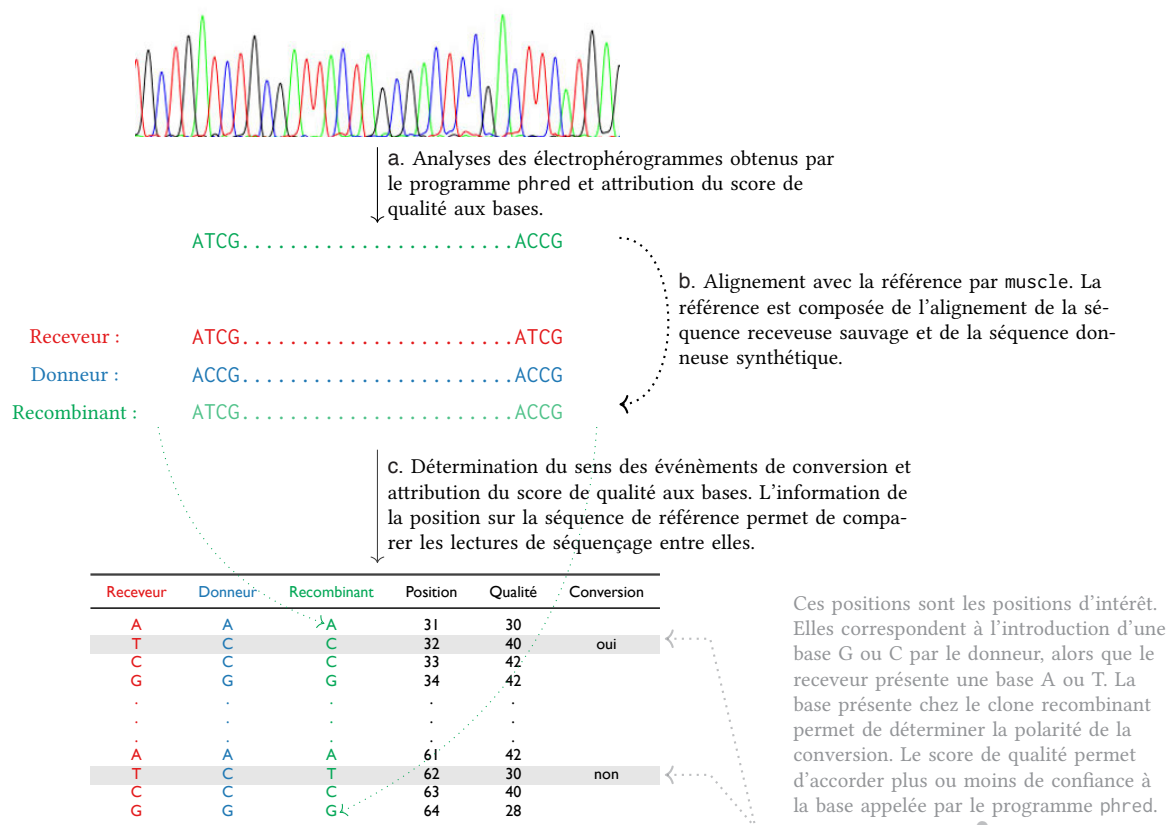


Figure 4 : **Exemple d'analyse de la zone de recombinaison pour un clone transformant**

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).

## 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>8</sup> et convertis en format universel FASTA (voir figure 4). Les séquences obtenues ont été alignées aux références par *muscle* v3.8.31<sup>6</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>3</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>16</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>.

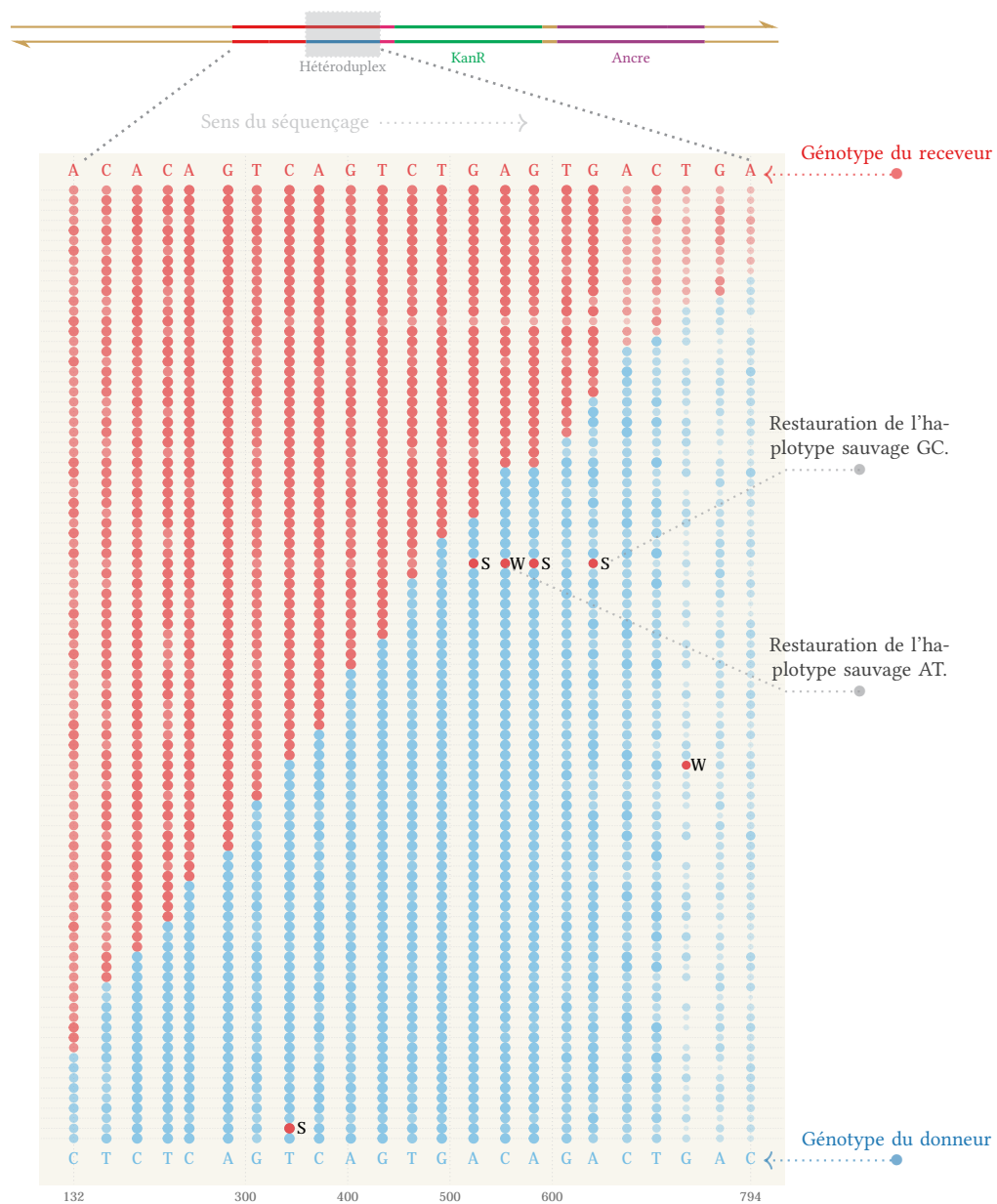


Figure 5 : **Zones de recombinaison entre un locus génomique d'*Acinetobacter baylyi* et un gène synthétique donneur d'allèles GC et AT.**

Chaque ligne horizontale représente une séquence. Les points représentent les positions des marqueurs sur les séquences. L'intensité et le diamètre des points représentent le score de qualité 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 est dans la région conservée. Les séquences sont triées par longueur de région convertie. Les alternances rouge / bleu marquent la transition de l'haplotype converti à l'haplotype sauvage : le point de recombinaison est localisé entre ces deux marqueurs.



## 2 Résultats

### 2.1 Analyses visuelles des zones de recombinaison

Nous avons transformé une suspension d'*Acinetobacter* par des constructions dont l'intégration dans le génome par recombinaison homologe entraîne la réparation des mésappariements, un mécanisme qui est biaisé vers l'introduction des bases GC chez les eucaryotes. Pour chaque construction, nous avons représenté graphiquement les zones de recombinaison (voire figure 5). L'origine du graphique correspond au début de la région convertie. Les 96 séquences obtenues sont triées par longueur de région convertie. La figure 5 représente le détail des zones de recombinaison obtenues en transformant avec une construction alternant CG et AT. Les zones de recombinaisons des clones transformés par les donneurs GC, AT et AT/GC sont détaillées en annexe 1.4.

### 2.2 Analyses de la position des points de recombinaison

Selon l'hypothèse gBGC, la région convertie devrait être plus longue lorsque la construction donneuse induit des réparations vers GC que lorsqu'elle induit des réparations vers AT. La différence entre la longueur moyenne de région convertie par les donneurs GC et la longueur moyenne de région convertie par des donneurs AT n'est pas significative (test de Wilcoxon, probabilité critique = 0,31) (voir figure 6). De la même façon, la différence entre la longueur de région convertie par le donneur AT/GC et celle par le donneur GC/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.

Les constructions alternant AT et GC permettent de déterminer si le point de recombinaison se situe plus souvent après un marqueur introduisant une conversion vers GC qu'après un marqueur introduisant une conversion vers AT (voir table 1). Le dernier marqueur converti est AT dans 92 transformants ; il est GC dans 83 transformants. Cet écart n'est pas significatif (voir table 1).

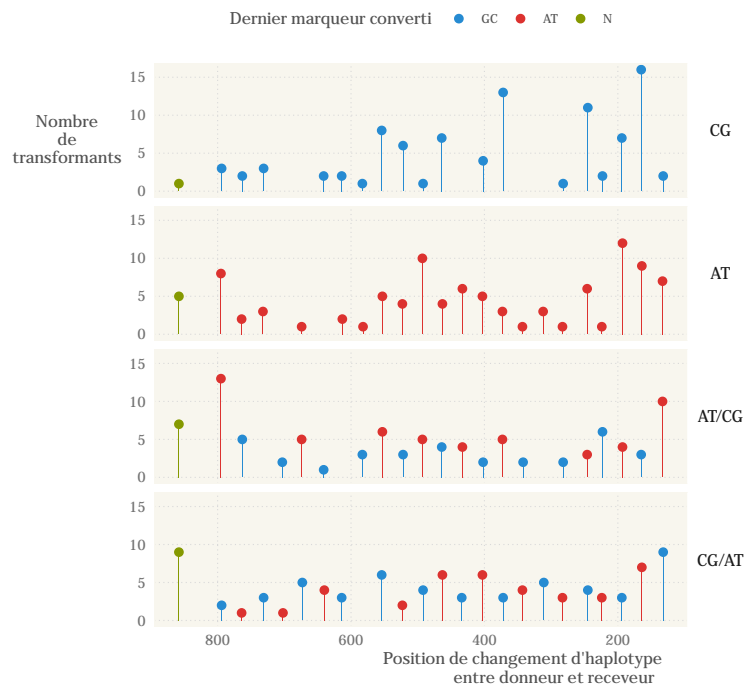


Figure 6 : **Distribution de la position du dernier marqueur converti**

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 pas de région converties sont indiqués par des *N*. La distribution du dernier marqueur converti est très variable.

Dernier marqueur converti	Donneur	Total
AT	AT/CG	55
	CG/AT	37
		92
CG	AT/CG	33
	CG/AT	50
		83

Table 1 : **Dénombrements des derniers marqueurs avant le point de recombinaison.**

Nous avons compté le nombre de séquences dont le dernier marqueur dans la région convertie est un marqueur AT ou GC. Selon l'hypothèse gBGC, ce nombre devrait être plus élevé aux positions GC qu'aux positions AT. Contrairement à l'hypothèse, il y a plus de transformants dont le dernier marqueur converti est AT que de transformants dont le dernier marqueur converti est GC. Cet écart n'est pas statistiquement significatif (test du  $\chi^2$  d'indépendance, probabilité critique = 0.49).

## 2.3 Analyses des 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 5). 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 clones issues 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.

Sur les 14 cas de restaurations de l'haplotype sauvage, 4 sont des restaurations de l'allèle AT, 10 sont des restaurations de l'allèle GC. Cet écart n'est pas significatif (voir la table 2).

Allèle restauré	Donneur	Total
CG	AT	4
	AT/CG	2
	CG/AT	4
		10
AT	CG	0
	AT/CG	3
	CG/AT	1
		4

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

Certains transformants montrent des restaurations ponctuelles de l'allèle sauvage dans la région convertie. Ces restaurations peuvent être des restaurations de l'allèle AT ou de l'allèle CG. Sur les 14 cas observés, 10 sont des restaurations de l'allèle GC, 4 de l'allèle AT. L'écart n'est pas statistiquement significatif (test du  $\chi^2$  d'indépendance, probabilité critique = 0,11).

# 3 Discussions



## Conclusion

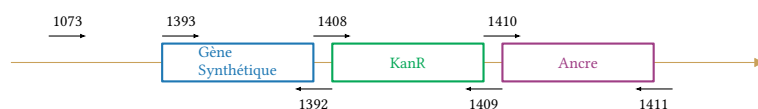
# Références

- [1] **Chen, I. et D. Dubnau.** 2004. DNA uptake during bacterial transformation. *Nat Rev Micro* 2 :241–249.
- [2] **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.
- [3] **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.
- [4] **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.
- [5] **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.
- [6] **Edgar, R. C.** 2004. MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 :1792–1797.
- [7] **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.
- [8] **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.
- [9] **Galtier, N., L. Duret, S. Glémin, et V. Ranwez.** 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends in Genetics* 25 :1–5.
- [10] **Hershberg, R. et D. A. Petrov.** 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* .
- [11] **Hildebrand, F., A. Meyer, et A. Eyre-Walker.** 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 6 :e1001107.
- [12] **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.
- [13] **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.
- [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] **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.
- [16] **R Core Team.** 2015. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- [17] **Ratnakumar, A., S. Mousset, S. Glemin, J. Berglund, N. Galtier, L. Duret, et M. T. Webster.** 2010. Detecting positive selection within genomes : the problem of biased gene conversion. *Philosophical Transactions of the Royal Society B : Biological Sciences* 365 :2571–2580.
- [18] **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.
- [19] **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. Médigue.** 2013. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucl. Acids Res.* 41 :D636–D647.
- [20] **Williams, A. L., G. Genovese, T. Dyer, N. Altamose, 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.
- [21] **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.



# 1 Annexes

## 1.2 Amorces utilisées



Cible	Amorce	Séquence	Tm
Génome Construction	1073		
	1392		
	1393		
	1408	CACCTTAATCACTAGTTAGACATCTAAATCTAGGTAC	61.50
	1409	GGTAAAGTCAGAGGAGAGGATGAGGAGGCAGATTG	68.66
	1410	TCCTCTGACTTTACCAACAAC	48.04
	1411	AGGCGGCCGCACTAGCTTTCTGAGGGGAACGATCA	71.62
Plasmide pGEM-T	M13R		
	M13F		

Figure 7 : Liste des amorces utilisées

## 1.3 Carte des constructions donneuses

## 1.4 Traces de conversions

## 1.5 Confirmations des restaurations des haplotypes sauvages

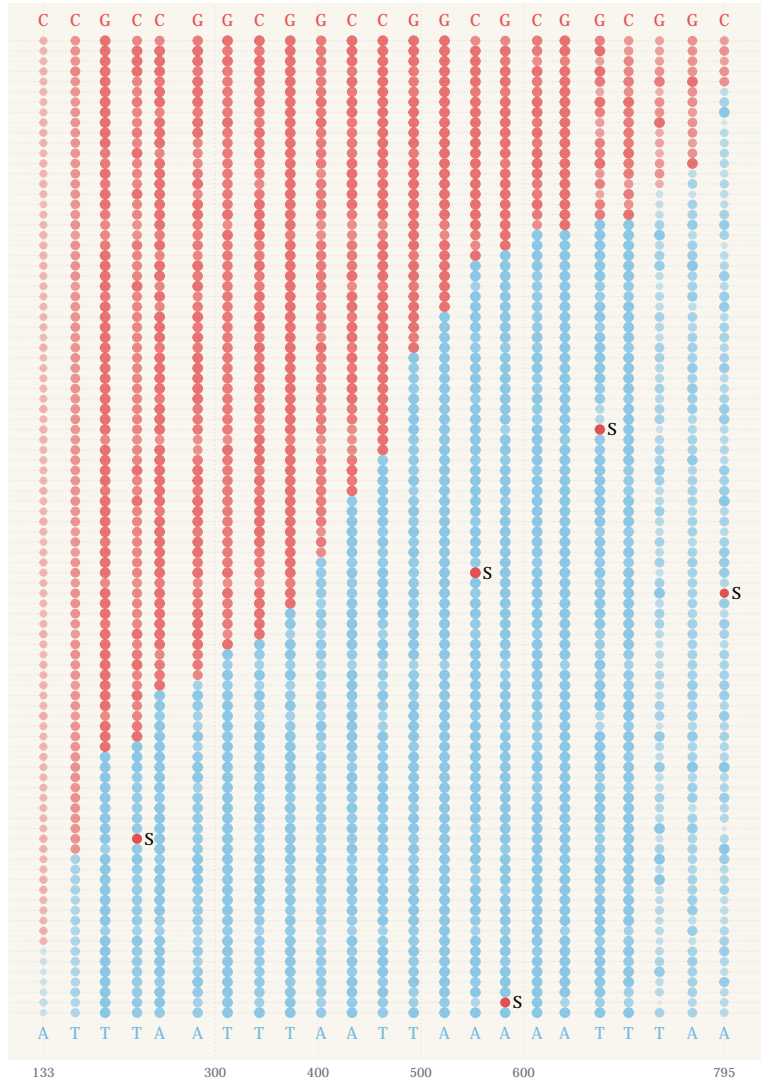


Figure 8 : **Zones de recombinaisons des clones transformés par la construction AT.**

Lorsqu'on introduit uniquement des polymorphismes AT, on observe quatre cas indépendants de restauration du génotype sauvage, et un cinquième pour une base de qualité inférieure.

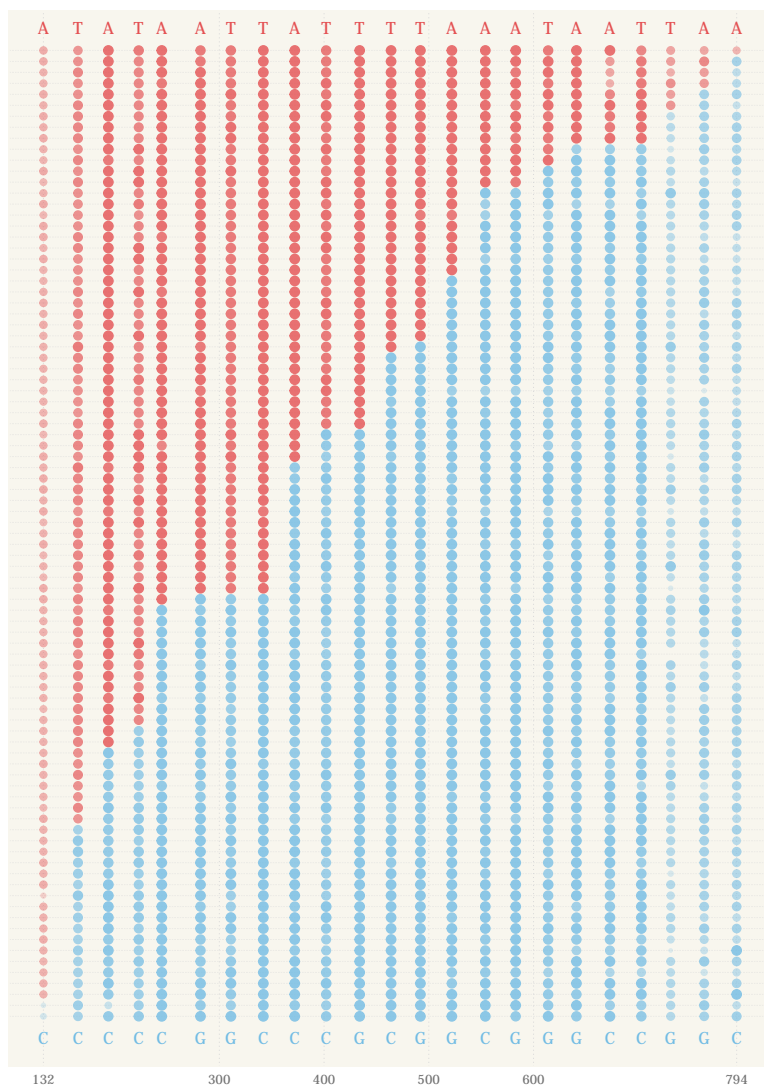


Figure 9 : **Zones de recombinaisons des clones transformés par la construction GC.**  
 Lorsqu'on introduit uniquement des polymorphismes GC, on n'observe pas de cas de res-  
 tauration du génotype sauvage.

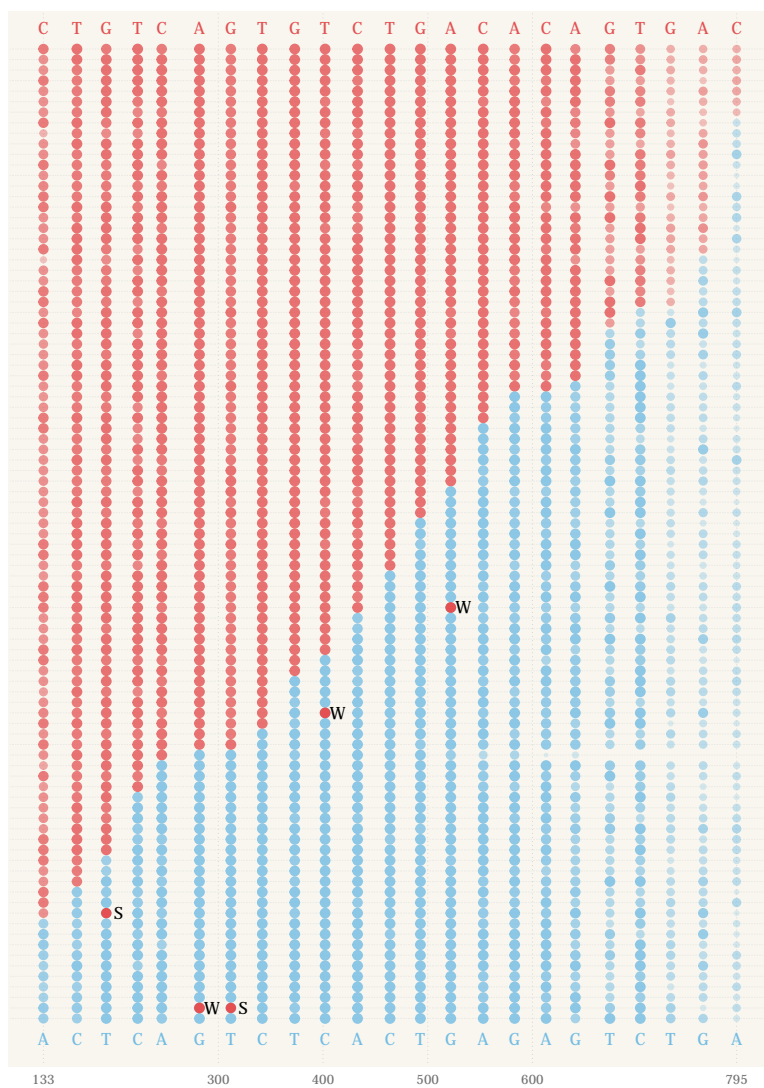


Figure 10

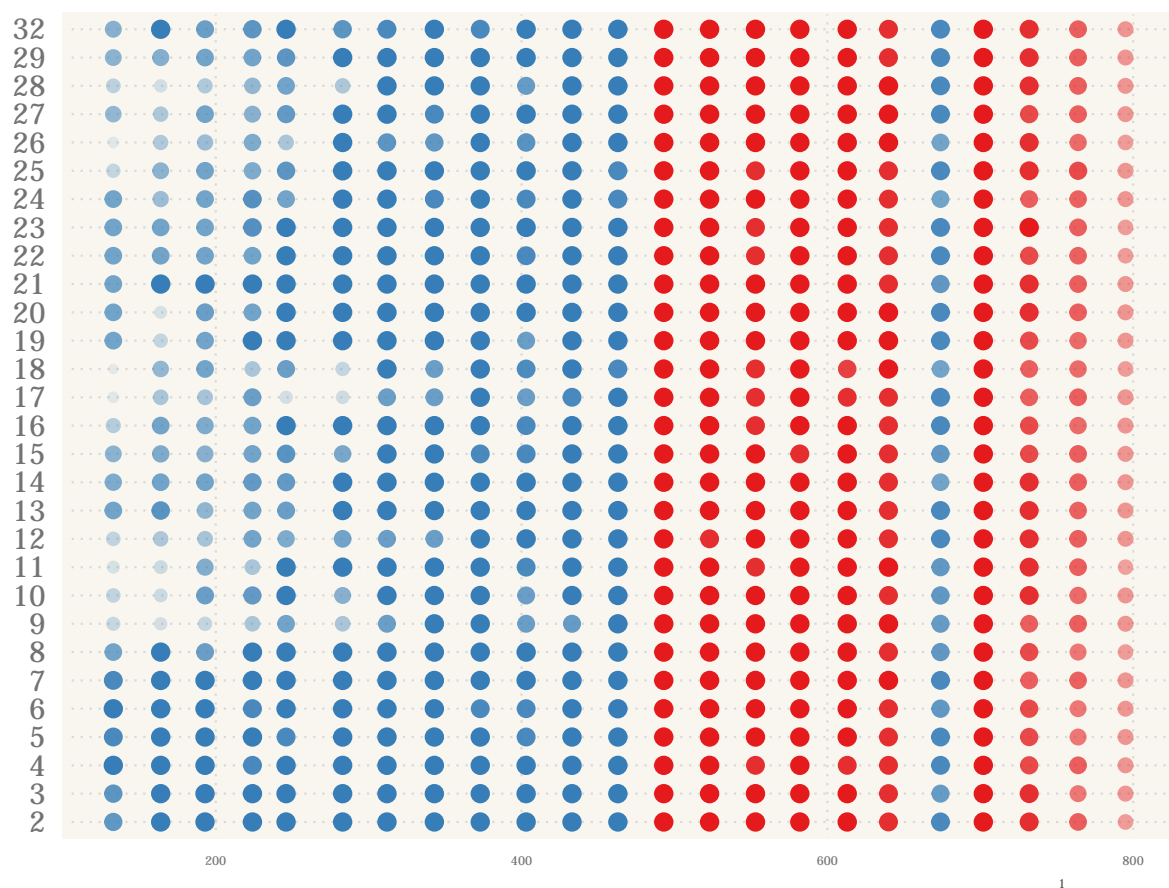


Figure 11 : **Confirmation d'une restauration d'haplotype sauvage**

Pour s'affranchir d'une possible erreur de séquençage, 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.

Les 31 clones séquencés montrent tous la même alternance au même marqueur. Nous avons ainsi confirmé que cette restauration de l'haplotype sauvage était bien due à une correction des mésappariements dans la population mère.

