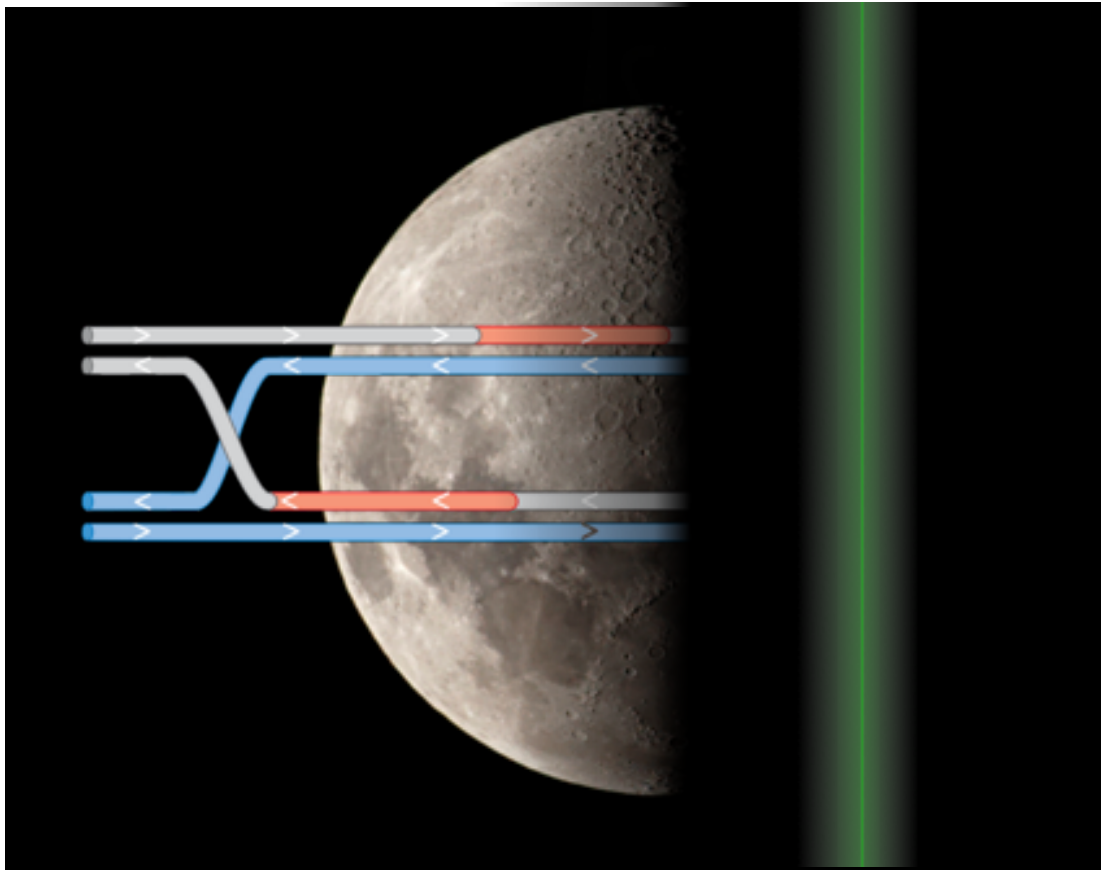

La conversion génique biaisée vers GC :

processus majeur d'évolution des génomes

Samuel BARRETO

2 Décembre 2015



Encadrement :

Franck Bertolla, UMR5557
Laurent Duret, UMR5558

UMR5557, Laboratoire d'écologie microbienne
Équipe 4 : Diversité et Adaptation des bactéries
phytopathogènes



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	2
1.1 Constructions des plasmides	2
1.2 Transformations d' <i>Acinetobacter baylyi</i>	4
1.3 Alignements	4
2 Résultats	5
2.1 Analyses visuelles des traces de conversion	5
1 Annexes	i
1.2 Amorces utilisées	i
1.3 Traces de conversions	i

Abbréviations

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

GC% Contenu en GC

Liste des points à traiter

insérer débuts de conclusions ici 2

Introduction

Le contenu en base est un paramètre clé de variation des séquences au sein des génomes et entre génomes. 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^{3,10,15}. Au cours de la réparation des cassures doubles brins par recombinaison homologue, les mésappariements locaux sont réparés par un mécanisme occasionnant de la *conversion génique*¹. Ce mécanisme est biaisé vers l'introduction préférentielle des bases C et G chez les mammifères et probablement chez un grand nombre d'eucaryotes¹¹. C'est un processus neutre : 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 C¹³.

En 2010, deux études simultanées^{8,7} démontrent que : 1) la *mutation* est universellement biaisée 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*⁹ et de Yahara *et al*¹⁶ ont montré que les zones recombinantes des génomes procaryotes ont un taux de GC plus élevé que les régions non-recombinantes, une observation compatible avec l'hypothèse du biais de conversion biaisé 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 modèle de transformation naturelle *Acinetobacter baylyi* ADP1^{*}. Les cellules d'*A.baylyi* en phase exponentielle de croissance sont naturellement compétentes à la transformation. Elles intègrent activement de l'ADN exogène

*. *A.baylyi* était appelée *A.calcoaceticus* avant 1995, mais la taxonomie du genre a fait l'objet d'une révision⁵

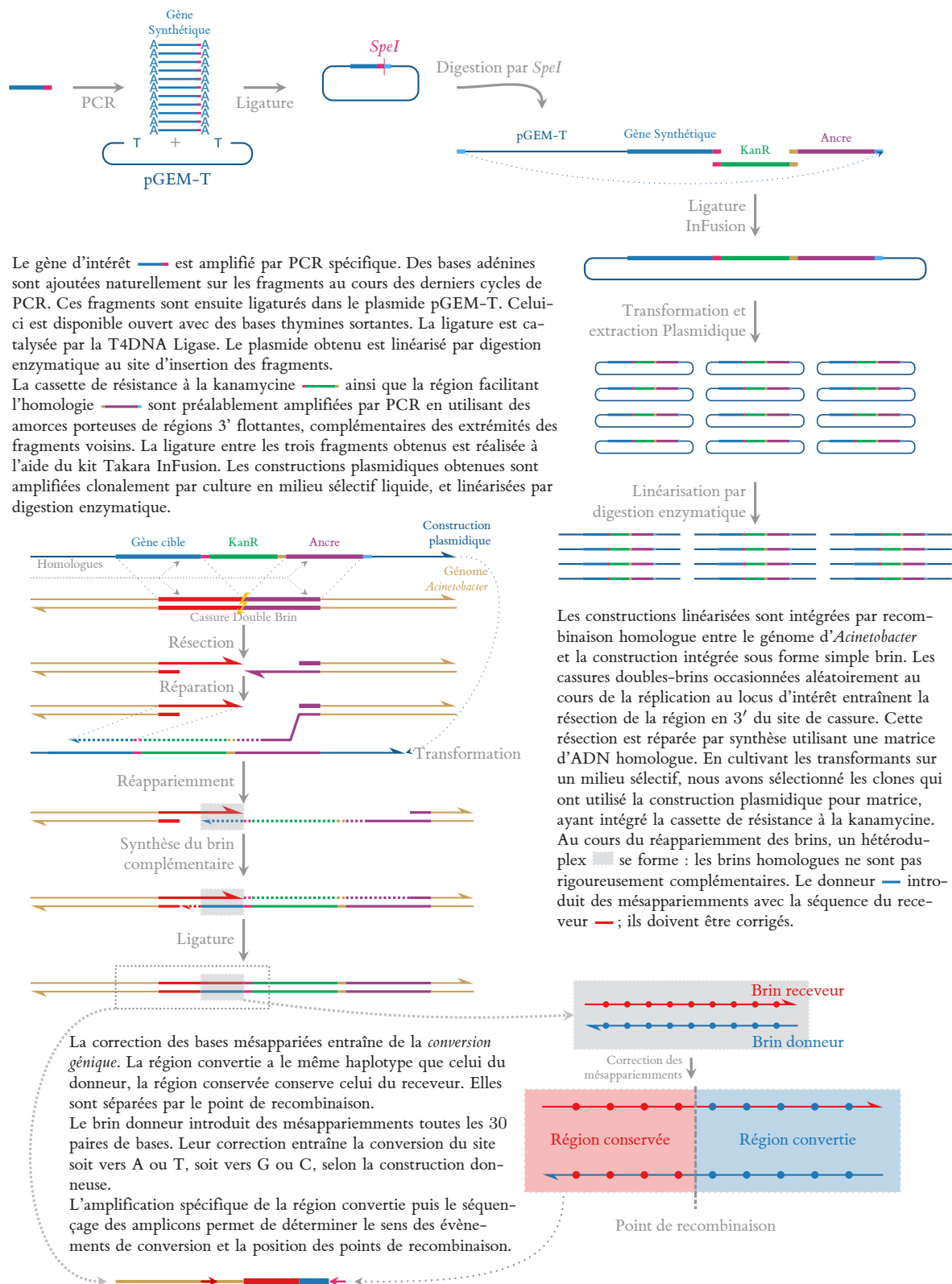


Figure 0.1 : **Constructions moléculaires et mécanismes de conversion génique au cours de la transformation naturelle**

sous forme simple brin par un mécanisme sous contrôle génétique. 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 — notre construction — et la séquence receveuse — le génome (voir figure 0.1). La réparation de ces mésappariements donne lieu à la conversion génique d'un brin par l'autre : l'information génétique portée par un brin est transmise unidirectionnellement du brin donneur au brin receveur.

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 0.1). Les constructions présentent des mésappariements avec la séquence receveuse. Ces marqueurs permettent 1) de localiser la position du point de recombinaison et 2) de déterminer la polarité de la conversion. À un site donné, une base peut être convertie ou restaurée lorsqu'elle conserve la base de la séquence receveuse. La réparation des mésappariements introduits nous renseigne sur les fréquences de conversion en faveur de GC chez *A.baylyi*.

insérer débuts de conclusions ici

1 Matériels & Méthodes

1.1 Constructions des plasmides

Nous avons utilisé deux approches. La première consiste à introduire exclusivement des bases GC ou AT. (voir figure). Sous l'hypothèse d'une conversion biaisée vers GC, les régions converties devraient être plus longues lorsqu'on introduit exclusivement des bases GC que lorsqu'on introduit des bases AT. Sur un locus du génome d'*Acinetobacter baylyi*, 23 dinucléotides séparés par trente bases ont été choisis. Ce sont tous des dinucléotides alternant AT et GC. Un gène de synthèse a été conçu pour introduire un mésappariement entre la base A ou T du génome et une base G ou C : cette construction sera appelée GC dans la suite de ce rapport. La construction AT a été conçue de façon à introduire un mésappariement entre les bases G ou C du génome et une base A ou T.

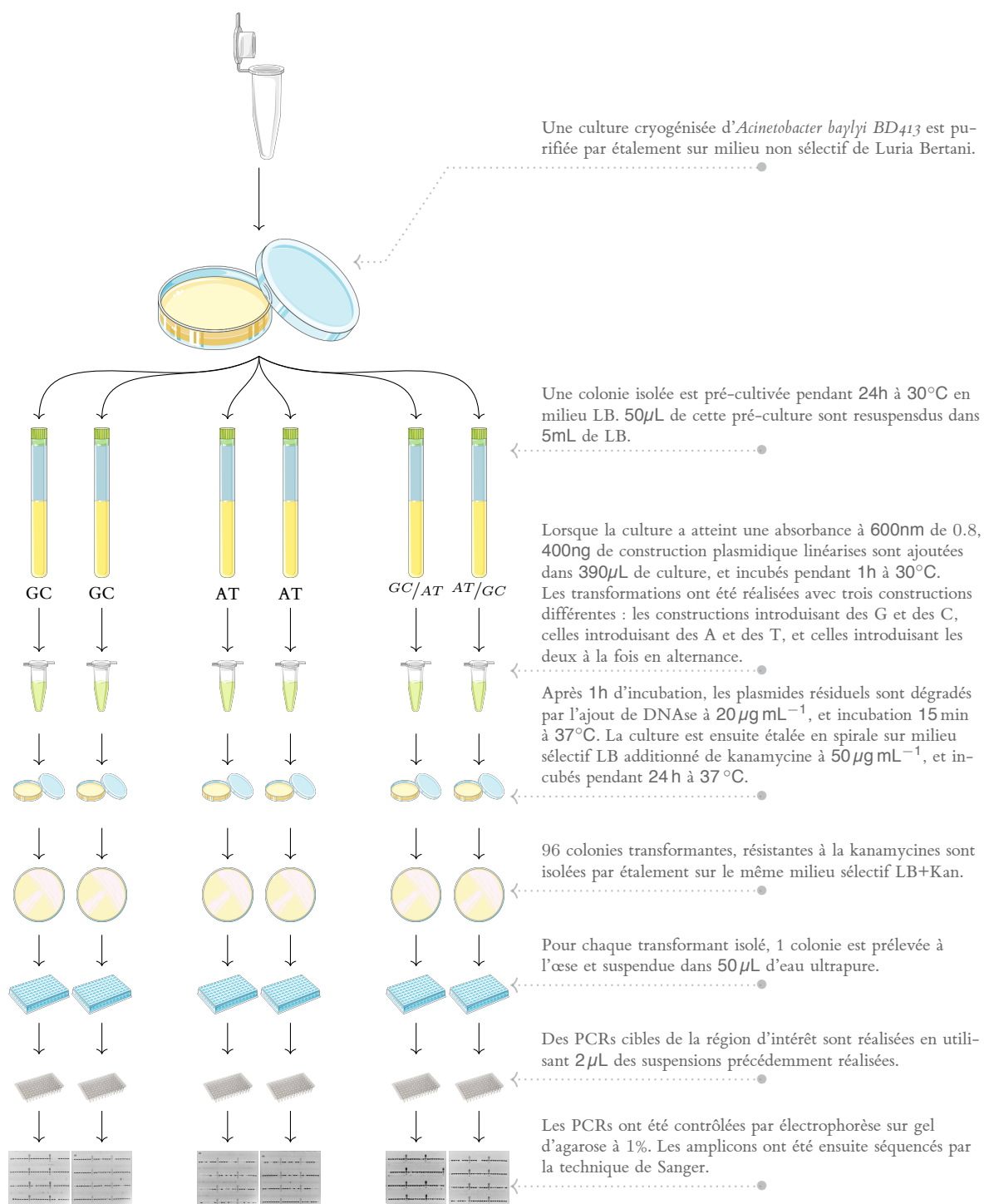


Figure 1.2 : **Protocole de transformation et d'obtention des amplicons des zones de re-combinaison.**

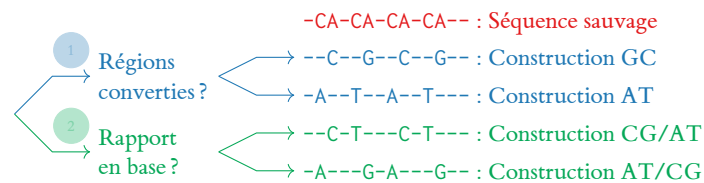


Figure 1.2 : **Schéma des deux approches et des constructions utilisées**

Des dinucléotides ont été choisis sur la séquence sauvage pour alterner AT et CG. La construction CG introduit uniquement des bases C et G, la construction AT des bases A et T, la construction AT/CG alterne AT et CG, et la construction CG/AT alterne CG et AT.

La seconde approche consiste à introduire à la fois des bases AT et CG. En l'absence de biais, le rapport en base ne devrait pas être affecté : on introduit autant de bases GC que de bases AT. Deux constructions ont donc été conçues sur la base des constructions exclusivement GC ou AT. La construction GC/AT alterne l'introduction d'une base GC avec celle d'une base AT. La construction AT/GC alterne l'introduction d'une base AT avec celle d'une base GC.

Une première partie du travail a consisté à construire le vecteur permettant d'introduire le gène synthétique porteur des mésappariements chez *Acinetobacter* par transformation naturelle. Les gènes synthétiques ont été synthétisés par quelqu'un. Les PCRs spécifiques du gène d'intérêt ont été réalisées. Les amplicons ainsi obtenus ont été insérés par ligature dans le plasmide pGEM-T. Le plasmide a été inséré dans la souche d'*E.coli* One Shot[®] TOP-10 (Invitrogen, Carlsbad, États-Unis), suivant le protocole du fabricant. Les cellules transformantes possédant l'insert ont été discriminées par un crible bleu / blanc sur gélose solide Luria-Bertani additionnée d'Ampicilline à $75 \mu\text{g mL}^{-1}$, de XGal et d'IPTG. Le sens d'insertion du fragment a été vérifié par PCR M13R-1392 (voir annexe la fameuse annexe). Les inserts des clones positifs ont été séquencés par séquençage Sanger (GATC Biotech, Constance, Allemagne). Les plasmides des clones validés par séquençage ont été extraits par le kit Nucleospin-Plasmid (Macherey-Nagel, Düren, Allemagne), et ouverts par digestion *SpeI*.

La cassette de résistance à la kanamycine *aphA3* est amplifiée par PCR avec les amorces 1408 et 1409. La région *ancree* facilitant la recombinaison est amplifiée avec les amorces 1410 et 1411. Les deux amplicons obtenus ont été ligaturés dans le plasmide pGEM-T porteur des constructions ouvert par *SpeI*. 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

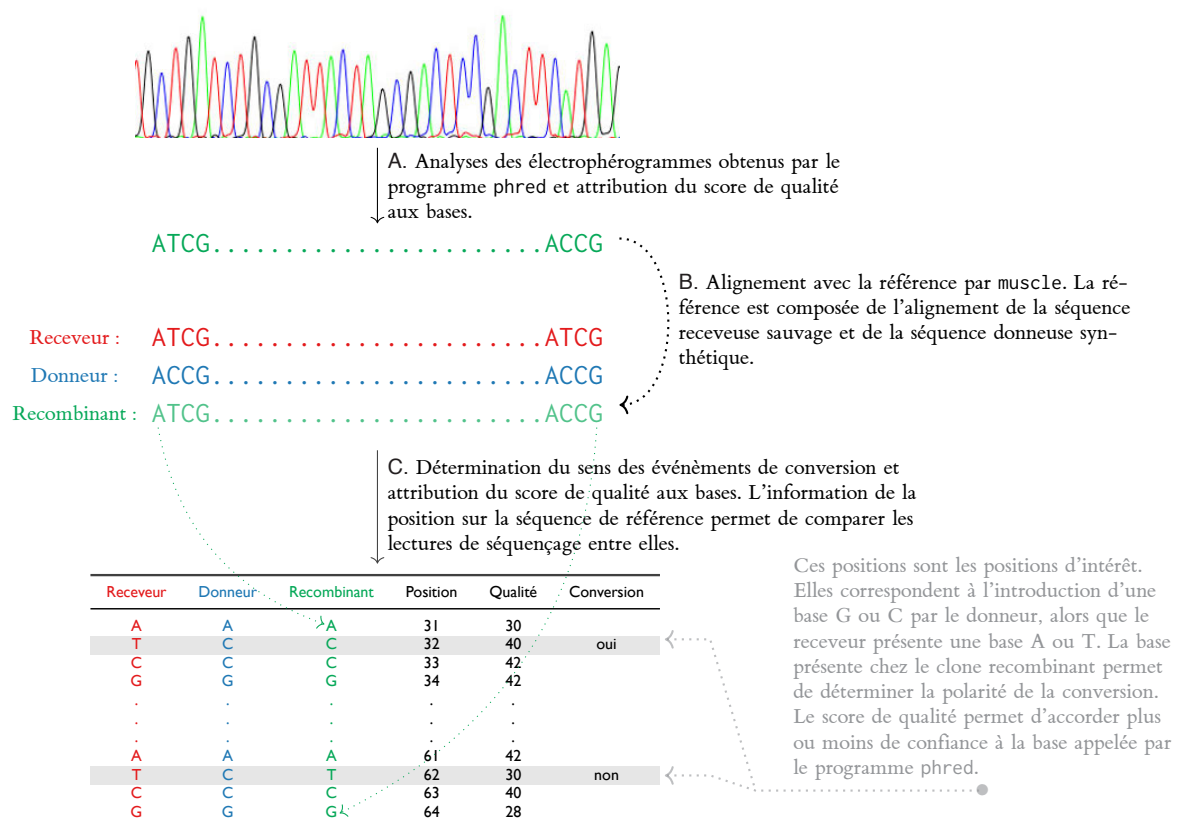


Figure 1.3 : Exemple d'analyse de la zone de recombinaison pour un clone transformant

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

kanamycine à $50 \mu\text{g mL}^{-1}$. Les transformants ont été confirmés par PCR spécifique de l'insert avec les amorces XXXX et YYYY.

1.2 Transformations d'*Acinetobacter baylyi*

1 μg de plasmide a été extrait et linéarisé par digestion *ScaI*. L'enzyme a été 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 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 incubés 24 h à 30°C . Les dénombrements ont été effectués via un compteur automatique Scan[®]1200 (InterScience).

96 clones isolés ont été purifiés par isolement sur milieu LB solide additionné de kanamycine à $50 \mu\text{mol mL}^{-1}$ et incubation 24 h à 30°C . Une colonie isolée par clone a été suspendue dans 50 μL d' H_2O ultra-pure. 2 μL de ces suspensions ont servi de matrice pour amplifier les régions recombinantes en PCR par la Taq polymérase haute fidélité Phusion (ThermoFischer, Waltham, États-Unis). L'amplification a été confirmée par migration sur gel d'agarose à 1%. Les amplicons ont été séquencés par la technique de Sanger¹⁴ (GATC Biotech).

1.3 Alignements

Les spectrogrammes de séquençage reçus au format propriétaire abi (Applied Biosystem, Foster City, États-Unis) ont été analysés par le programme phred⁶ et convertis en format universel FASTA (voir figure 1.3). Les séquences obtenues ont été alignées à la référence par muscle v3.8.31⁴. La référence en question correspond à l'alignement de la séquence sauvage et de la séquence du gène synthétique, respectivement *receveur* et *donneur* de l'évènement de recombinaison. Un script Python² a été développé pour analyser les alignements obtenus. Il infère les positions des SNPs d'intérêt dans l'alignement de référence, et détermine le génotype du clone séquencé. Les alignements par paire en colonne obtenus ont été analysés par R 3.2.3¹². 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>.

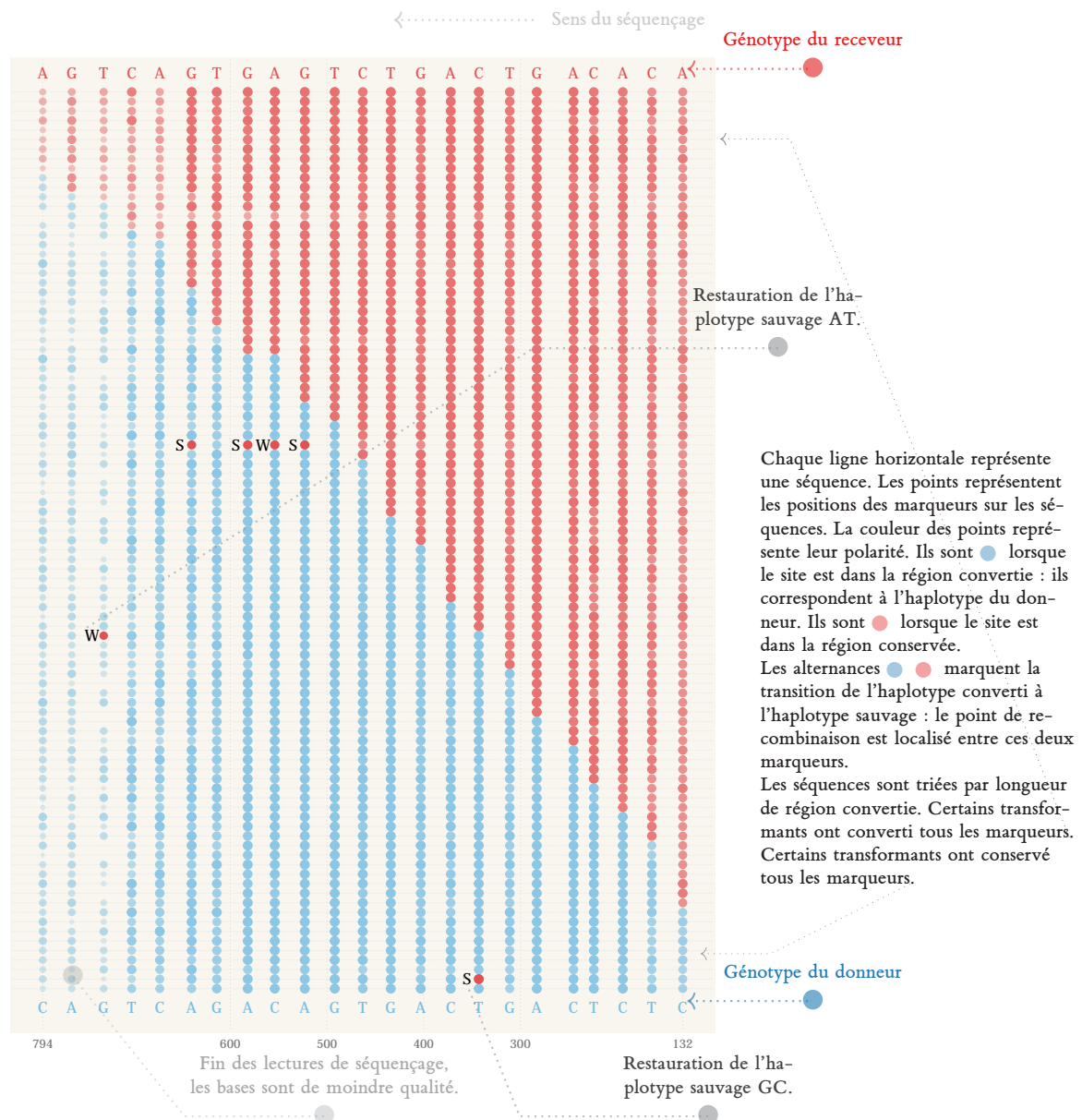


Figure 2.1 : Traces de conversion d'un locus génomique d'*Acinetobacter baylyi* par un gène synthétique donneur d'allèles GC et AT.

2 Résultats

2.1 Analyses visuelles des traces de conversion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget

enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

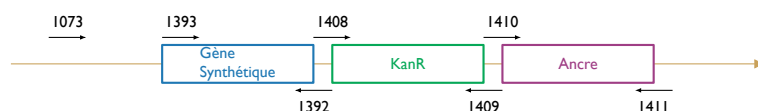
Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

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] 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.
- [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] Edgar, R. C. 2004. MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 :1792–1797.
- [5] 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.
- [6] 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.
- [7] Hershberg, R. et D. A. Petrov. 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* .
- [8] Hildebrand, F., A. Meyer, et A. Eyre-Walker. 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 6 :e1001107.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] R Core Team. 2015. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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 1.1 : Liste des amorces utilisées

1.3 Traces de conversions

