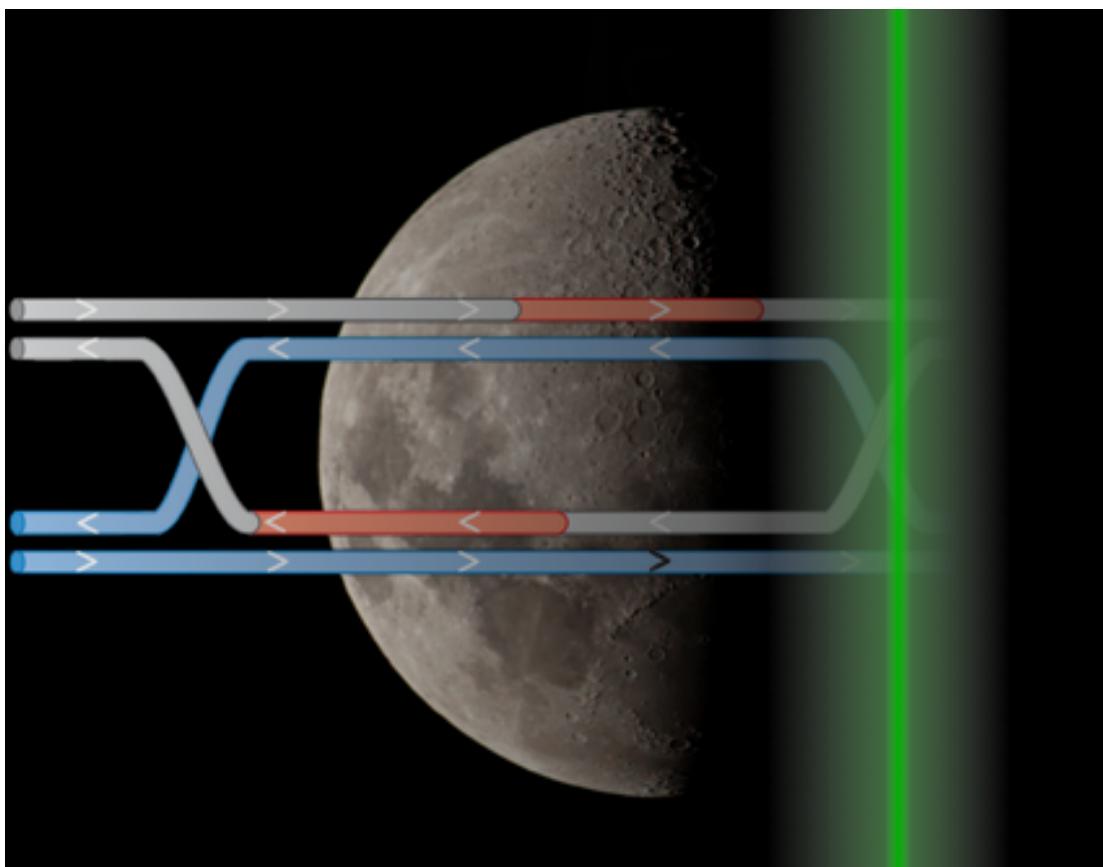


La conversion génique biaisée vers GC : processus majeur d'évolution des génomes

Samuel BARRETO

2 Décembre 2015



Encadrement :

Franck Bertolla, UMR5557

Laurent Duret, UMR5558



Crédits photographiques

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

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

Sommaire

1 Qu'est-ce que la conversion génique ?	2
1.1 Recombinaison homologue	2
1.1.1 Mécanismes de recombinaison homologue	2
1.1.2 La recombinaison méiotique : étape clé de la méiose	3
1.1.3 La recombinaison chez les procaryotes	4
1.2 Conversion génique	4
1.3 Conversion génique biaisée vers GC	5
2 Quelles hypothèses pour expliquer le gBGC ?	6
Des propriétés inhérentes à la réparation ?	6
Compenser la mutation ?	7
3 Quelles sont les conséquences du gBGC ?	7
3.1 Le gBGC structure le contenu en GC	7
3.1.1 Le contenu GC des génomes mammifères et la théorie des isochores	8
3.1.2 Un gBGC procaryote ?	8
3.2 Le gBGC interfère avec la sélection	9
3.3 Le gBGC fausse les tests de sélection	10
Le gBGC fausse les tests de sélection	10

Abbréviations

BER Base Excision Repair

DSB Double strand break : cassures doubles brins

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

GC% Contenu en GC

MMR Mismatch repair : machinerie de réparation des mésappariements

NER Nucleotide Excision Repair

Introduction

La recombinaison homologue est un processus majeur d'évolution des génomes. C'est un processus très conservé, des procaryotes aux eucaryotes⁷. C'est le mécanisme clé de réparation des lésions de l'ADN. Elles sont réparées sur la base d'une matrice d'ADN homologue, donnant lieu à la formation transitoire d'un hétéroduplex. Un hétéroduplex est un appariement entre deux brins homologues, qui ne sont pas nécessairement rigoureusement complémentaires. Chez les eucaryotes, la matrice est typiquement la région homologue correspondante, sur la chromatide sœur. Les crossover méiotiques sont générés sous contrôle génétique et impliquent une étape obligatoire de recombinaison²⁴. Chez les procaryotes, cette matrice peut être acquise par transfert horizontal. La recombinaison est le moyen principal de transfert de gène entre procaryotes.

Au voisinage de la région lésée — puis réparée — , des échanges d'information génétique de petite échelle existent¹². Ils sont dûs à la correction des erreurs d'appariement dans l'hétéroduplex. Ce sont des événements de conversion génique. Ils donnent lieu à des transmissions non-mendéliennes d'allèles : l'information génétique est transmise d'un brin vers l'autre, de façon unidirectionnelle. Cette conversion est biaisée dès lors que l'un des allèles a une plus forte probabilité de transmission, à l'échelle de la population. Chez les eucaryotes, de nombreuses observations montrent que la conversion est biaisée vers les bases G ou C^{28;24;10}. Les allèles G ou C sont plus souvent transmis que les allèles A ou T. Ce biais de conversion génique en faveur de GC — gBGC — a un impact fort sur l'évolution des génomes : il conditionne la structuration en contenu GC des génomes¹⁰ ; il interfère avec la sélection naturelle¹⁶ ; et, d'un point de vue pratique, il brouille les tests de sélection naturelle³⁰.

Lorsque le biais est observé, le contenu en GC est anormalement élevé dans les régions fortement recombinantes¹⁰. Ce mécanisme permet d'expliquer en partie les variations du GC%, au sein d'un génome et entre génomes. L'explication de ces variations suscite un débat entre les partisans des hypothèses adaptatives, et ceux des hypothèses neutres. Les premiers veulent que le contenu en GC soit un trait sélectionné : il contribue en soi à la valeur adaptative de l'individu¹⁸. Les seconds avancent qu'il est également déterminé par des processus de dérive génétique et de mutation. Le gBGC est un autre processus neutre contribuant à déterminer le GC%. En fait, c'est la contribution relative des trois processus à la détermination du taux de GC qui fait débat.

Le biais peut également permettre de fixer des mutations délétères, si celles-ci sont portées par les allèles G ou C du gène. Il contrecarre l'action de la sélection naturelle^{16;15}. Enfin, l'action du biais

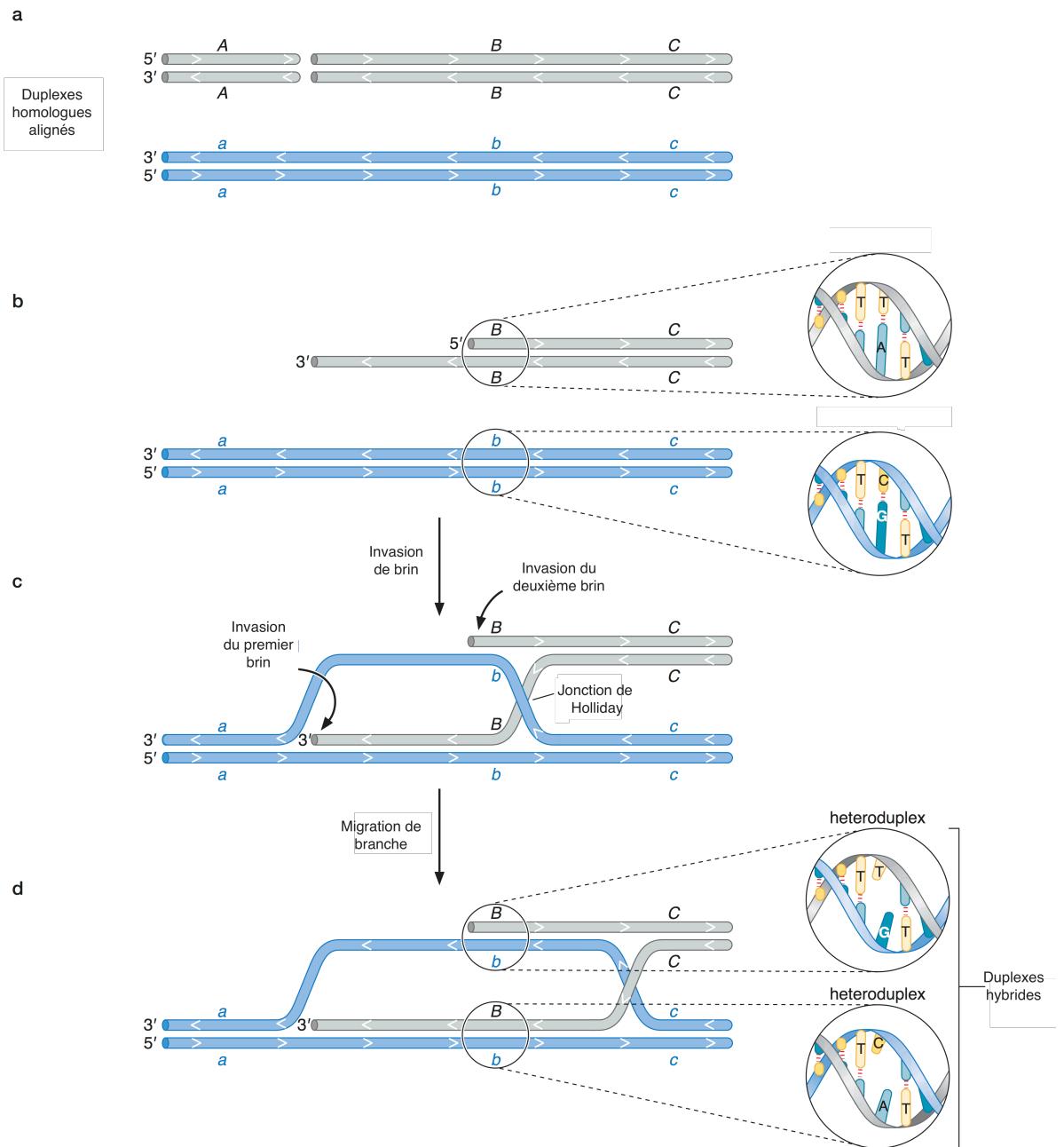


FIGURE 1 – Le modèle classique de formation d'un hétéroduplex par invasion de brin. Après la formation d'une cassure double brin en a, les extrémités 3' sont exposées par résection, en b. Le premier brin gris porteur de l'allèle B envahit le brin bleu porteur de l'allèle b, formant une jonction de Holliday, en c. Localement, l'hétéroduplex ainsi formé montre des mésappariements, en d.

Adapté de Molecular Biology Of The Gene, Watson, 2012.

laisse des traces génomiques équivalentes à celles de la sélection naturelle : les tests classiques de détection confondent l'action de la sélection avec celle du gBGC³⁰.

Récemment, l'hypothèse du biais vers GC s'est étendue aux procaryotes²¹. Des observations montrent que les régions fortement recombinantes ont un contenu en GC significativement plus élevé. L'extension de cette hypothèse aux procaryotes représente une avancée majeure : elle pourrait permettre d'expliquer l'évolution du contenu en GC des populations bactériennes.

Nous verrons donc ce qu'est la conversion génique biaisée et les mécanismes qui la sous-tendent, dans une première partie ; dans une deuxième, les hypothèses permettant d'expliquer son existence ; et dans une dernière, ses conséquences évolutives.

1 Qu'est-ce que la conversion génique ?

La conversion génique est un mécanisme qui intervient au cours de la réparation de l'ADN. Lors d'une cassure double brin — DSB, Double Strand Break — , la lésion est réparée par recombinaison homologue. Les mésappariements induits peuvent être corrigés, donnant lieu à une transmission d'allèle non-réiproque⁶.

1.1 Recombinaison homologue

La réPLICATION de l'ADN est une étape clé du cycle cellulaire. Elle implique la formation de fourches de réPLICATIONS. Les lésions subies par l'ADN en cours de réPLICATION empêchent le déplacement de ces fourches, et notamment les cassures double brins. Ces cassures sont extrêmement cytotoxiques : elles causent l'arrêt de la réPLICATION, la perte de chromosomes, sont mutagènes et conduisent à la mort cellulaire. La bonne conduite de la réPLICATION nécessite de réparer ces cassures, de façon non-mutagène. La recombinaison homologue est le moyen préférentiel de réparation. Compte tenu de son importance dans le cycle cellulaire, les mécanismes qui la sous-tendent sont ubiquitaires et extrêmement conservés, des phages aux mammifères⁷.

Mécanismes de recombinaison homologue

Les cassures doubles brins, ou DSB, sont générées par des agents mutagènes ou sous contrôle génétique, notamment lors de la méiose. Après la formation des DSB, les extrémités 5' de la lésion

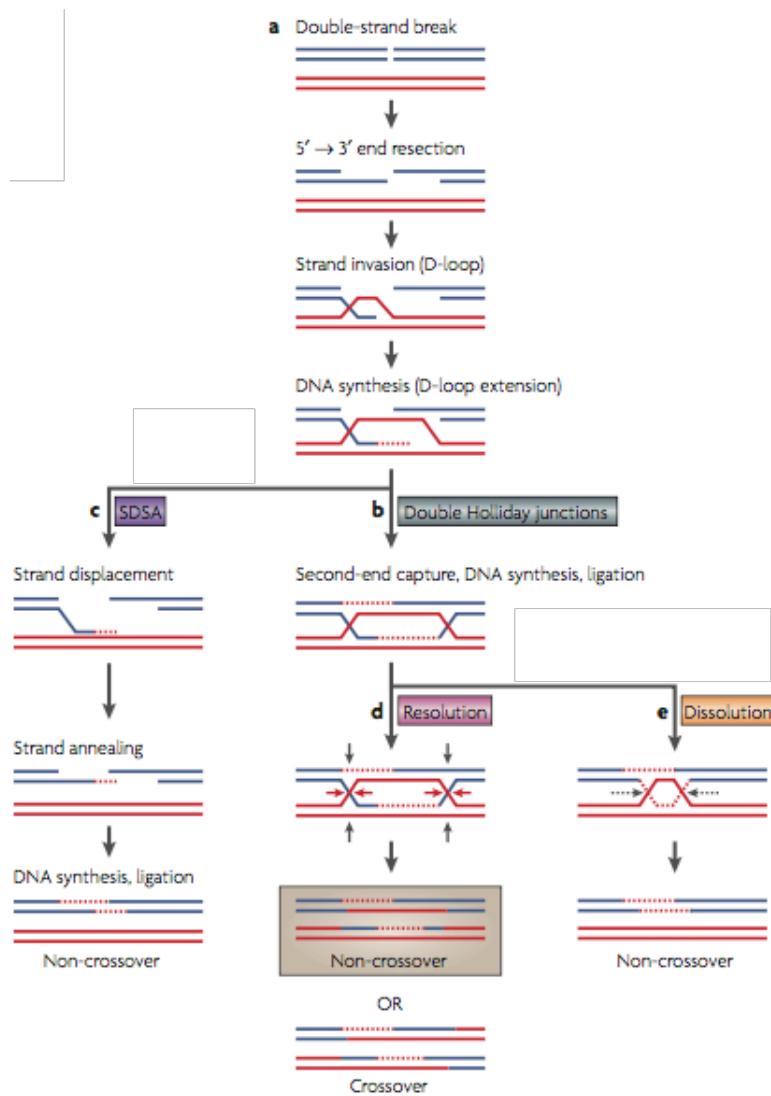


FIGURE 2 – Le modèle classique de réparation des cassures doubles brins par recombinaison homologue Les cassures doubles brins sont suivies d'une résection dans le sens $5' \rightarrow 3'$. L'un des brins ainsi exposé cherche ensuite activement une séquence homologue. Au cours de l'invasion de brin, une boucle D se forme, ainsi qu'une jonction de Holliday. La synthèse d'ADN a lieu en utilisant la séquence intacte comme matrice. La résolution de ces structures peut passer par différentes voies. En b, le brin réparé doit être apparié à la l'extrémité du brin originel : c'est la *second-end capture*. Selon le mode de clivage des résolvases, le produit obtenu est non-crossover ou crossover (d). Chez les eucaryotes, la *dissolution* est une autre voie de résolutions des doubles jonctions de Hollidays (e). En c, la voie SDSA, Synthesis-Dependent Strand-Annealing, implique une étape de dénaturation, puis de ré-appariement du brin envahisseur avec l'autre extrémité 3' de la cassure. La synthèse se poursuit et est suivie d'une étape de ligation. Dans tous les cas, des hétéroduplex sont formés, dès lors que les séquences appariées ne sont pas rigoureusement identiques.

Tiré de Chen et al., 2007⁵.

subissent une résection, délétion locale d'une section d'ADN. Chez *E.coli*, le complexe RecBCD est responsable de cette digestion. Il a une activité d'hélicase : il dissocie les deux brins d'ADN ; et de nucléase : il digère localement les brins dissociés⁹.

Le brin à l'extrémité 3' envahit les brins complémentaires de la molécule intacte. Ce brin recherche ensuite activement les zones d'homologie. Lorsqu'une zone est déterminée, les brins sont échangés. L'ensemble de ce processus est catalysé par la nucléo-protéine RecA⁶ ou ses homologues. C'est l'acteur moléculaire clé de la recombinaison homologue.

L'échange des brins entraîne la formation d'un hétéroduplex (voir Figure 1). L'intégrité de l'hétéroduplex est maintenue par une structure appelée jonction de Holliday. Des mésappariements peuvent exister entre les brins de l'hétéroduplex : deux brins homologues ne sont pas systématiquement complémentaires. Leur correction peut donner lieu à une conversion génique.

La zone de résection est ensuite comblée par la synthèse d'ADN en utilisant le brin homologue comme matrice. Enfin, les intermédiaires de recombinaisons sont résolus par des résolvases qui clivent les jonctions de Holliday. La résolution des intermédiaires de recombinaisons peut donner des produits dits crossovers ou non-crossovers, entraînant respectivement l'échange des régions flanquantes ou non²⁴.

La réparation des cassures est la fonction principale et première de la machinerie de recombinaison homologue. Cependant, les mécanismes en jeu sont le lieu d'un brassage génétique, aussi bien lors de la méiose eucaryote que lors des transferts de gène procaryotes³¹.

La recombinaison méiotique : étape clé de la méiose

Chez les eucaryotes, la méiose implique la formation de DSB, par les enzymes Spo11, sous contrôle génétique rigoureux. Ils sont réparés par recombinaison homologue⁴. Cependant, la distribution des sites de coupure est variable : il existe des hotspots de cassure, et donc de recombinaison. Par opposition, les coldspots sont des régions moins soumises que d'autres aux cassures.

La réparation des DSB par recombinaison homologue est requise pour l'appariement et la ségrégation des chromosomes homologues au cours de la méiose. Selon le mode de clivage des jonctions de Holliday par les résolvases, des crossovers se forment entre les chromosomes parentaux. Ces crossovers entraînent le brassage des allèles, un processus bénéfique sur le plan évolutif³³. En effet,

il casse les liaisons entre allèles : la sélection élimine alors plus efficacement les variants délétères et promeut les variants bénéfiques²⁷.

La recombinaison chez les procaryotes

Étant donné la taille des populations bactériennes et les temps évolutifs en jeu, la recombinaison a un impact majeur sur l'évolution procaryote⁸. C'est le moteur des transferts de gène. Ceux-ci sont médiés soit par des vecteurs, les plasmides ou les phages, soit par un état de compétence naturelle, *via* l'acquisition passive ou active d'ADN exogène. La principale fonction de la recombinaison homologue semble être la réparation des lésions de l'ADN¹⁴. L'acquisition de matériel génétique exogène est un effet secondaire des mécanismes de réparation de l'ADN. Cet effet secondaire est bénéfique sur le plan évolutif dès lors que le matériel acquis apporte un avantage sélectif à l'individu*.

Après la résolution des intermédiaires de recombinaison, des mésappariements peuvent exister entre les différents brins. Leur correction entraîne une conversion génique.

1.2 Conversion génique

La conversion génique est l'échange non réciproque d'information génétique. C'est une transmission non-mendéienne : l'un des allèles a une plus forte probabilité d'être transmis que l'autre⁵.

Considérons le cas de la transmission de l'allèle A et de son homologue α , au cours de la méiose. Après la méiose, le génotype attendu est AA $\alpha\alpha$. Un évènement de conversion de gène peut conduire à des génotypes de type A $\alpha\alpha\alpha$ ou AAA α .

Au cours de la réparation des DSB, la conversion peut subvenir de deux façons. i) L'allèle A est proche du site d'initiation de la cassure. Il fait partie de la résection, l'allèle α est copié vers le brin réparé. A $\alpha\alpha\alpha$ est le génotype obtenu. ii) L'intermédiaire de recombinaison présente un polymorphisme A α sur l'un des hétéroduplex. La machinerie de réparation des mésappariements — MMR — les prend en charge. α est alors converti en A, ou réciproquement.

Chez *E.coli*, la détection des mésappariements est effectuée par les dimères des enzymes MutS. Les mésappariements sont reconnus par la distorsion qu'ils causent à la structure de l'ADN. Les enzymes MutL et MutH sont alors recrutées. Une cassure est introduite dans l'un des brins, suivie par une

*. ou s'il manipule le comportement de reproduction de l'hôte en faveur de sa dissémination...

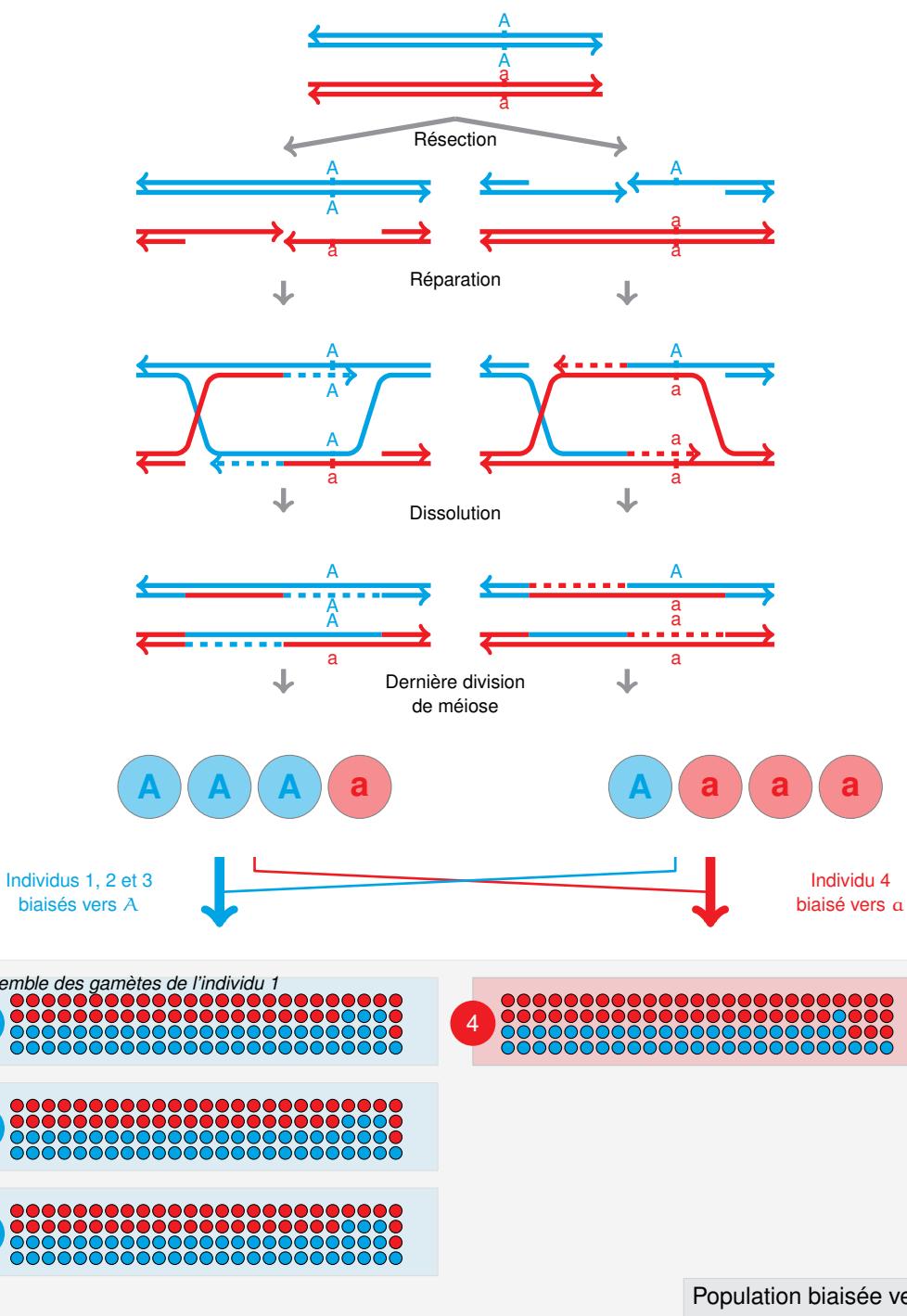


FIGURE 3 – La conversion génique à ses différentes échelles. À un locus A, dont les allèles sont A et a, la conversion génique entraîne soit la conversion de a par A, à gauche, soit l'inverse, à droite. À l'échelle d'un évènement de méiose, les gamètes obtenus sont soit AAAa, soit Aaaa. À l'échelle de l'*ensemble de la gamétogénèse* d'un individu, la conversion n'est biaisée que si AAAa est plus souvent obtenu qu'Aaaa. À l'échelle de la *population*, la conversion n'est biaisée vers A que si les individus dont la conversion est biaisée vers A sont plus fréquents.

résection souvent supérieure à 1 kb à proximité de la cassure. Une ADN polymérase utilise ensuite le brin intact pour synthétiser la région complémentaire. Les eucaryotes possèdent des protéines aux fonctions homologues, appelées respectivement MSH et MLH pour MutS Homologs et MutL Homologs. Ce sont des composants de la voie NER, Nucleotide Excision Repair.

Au cours de la recombinaison, le système de MMR est la voie préférentielle de correction des mésappariements dans l'hétéroduplex. Néanmoins, la voie BER, Base Excision Repair, est une alternative à ce système.

Elle entraîne l'excision de l'une des bases du mésappariement, puis son remplacement par la base complémentaire à l'autre. Les ADN glycosylases excisent les bases avec une spécificité de substrat : chaque base A, T, C ou G a une ADN glycosylase correspondante et spécifique.

Dans tous les cas, le génotype de la région — ou de la base — digérée est converti par celui du brin intact. Le transfert a lieu entre séquences homologues, qu'elles soient sur des chromatides sœurs, sur le même chromosome ou sur des chromosomes différents⁵.

En théorie, la conversion $a \leftrightarrow A$ a lieu avec la même fréquence que celle de la conversion $A \leftrightarrow a$. Cependant, dès lors qu'un allèle est plus souvent converti que l'autre, à l'échelle de la population, la conversion génique est biaisée (voire Figure 3). Chez les eucaryotes, de nombreuses observations montrent que les mésappariements GA, GT, CA ou CT sont plus fréquemment corrigés en GC qu'en AT¹².

1.3 Conversion génique biaisée vers GC

Mancera *et al*²⁴ ont génotypé l'ensemble des quatre haplotypes — les tétrades — résultants des produits de méiose de 46 levures, à haute résolution. Ils montrent qu'1% du génome de chaque produit de méiose est soumis à de la conversion génique. Ces régions montrent une transmission biaisée en faveur des allèles G ou C. Ils sont transmis avec une probabilité 1.3% plus élevée qu'attendu sous l'hypothèse d'une transmission mendélienne²⁴. Ce biais, bien que faible, peut affecter très fortement la probabilité de fixation des allèles GC dès lors que la taille de la population est grande²⁶.

Chez la levure, le gBGC est associé spécifiquement aux produits de recombinaisons entraînant des crossovers²². Il est également associé aux événements de conversion simple — par opposition aux

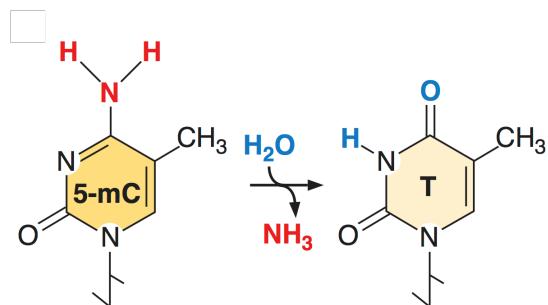


FIGURE 4 – La déamination spontanée des méthyl-cytosines. La perte du groupement amine d'une cytosine méthylée génère une base naturelle de l'ADN : la thymine. Cette perte peut avoir lieu dans des conditions physiologiques normales. La réPLICATION d'une telle erreur conduit à l'introduction d'un A sur le brin opposé, au lieu du G attendu. Le mécanisme de Base Excision Repair excise préférentiellement les thymines chez les vertébrés, probablement pour compenser la déamination spontanée des cytosines méthylées.

Adapté de Molecular Biology Of The Gene, Watson, 2012.

événements complexes. Lors d'un évènement de conversion simple, le même brin est le donneur de la conversion sur l'ensemble de la région convertie. Lors d'un évènement complexe, les deux brins de l'hétéroduplex peuvent être donneur.

Les causes moléculaires de l'existence d'un tel mécanisme suscitent beaucoup d'interrogations. Différentes hypothèses ont été avancées : elles font l'objet de la partie suivante.

2 Quelles hypothèses pour expliquer le gBGC ?

Les mécanismes précis responsables du biais de conversion génique vers GC sont encore inconnus à ce jour. Parmi les hypothèses avancées, on distinguera ici les mécanismes moléculaires potentiellement responsables d'un tel biais, des raisons d'être évolutives de la conversion biaisée.

Des propriétés inhérentes à la machinerie de réparation ?

La machinerie de réparation pourrait présenter dans sa structure un biais en faveur de la transmission des allèles G ou C, au cours de la conversion génique. Chez l'Homme, la réparation des mésappariements dans les cellules en mitose est fortement biaisé vers G ou C. Les ADN glycosylases de la voie BER ciblent spécifiquement les bases thymines, probablement pour compenser l'hypermutableté des cytosines³ (*cf* Figure 4). Si la voie BER est active au cours de la réparation des erreurs d'appariement de la recombinaison, le biais observé pourrait être dû à l'activité d'un mécanisme spécifiquement destiné à la réPLICATION mitotique.

Chez la levure, l'hypothèse de l'intervention du BER a été exclue²². En effet, étant donné la courte portée du BER, les traces de conversion obtenues devraient être complexes, avec une alternance du génotype non-biaisé et biaisé sur de courtes échelles. Pourtant, le biais de conversion n'est observé que dans les traces simples : le brin donneur est le même sur l'ensemble de la région convertie.

Deux modèles alternatifs ont été proposés²² : i) le modèle de rejet de brin, et ii) le modèle du MMR biaisé. Le modèle de rejet de brin intervient au moment de la recherche d'homologie par le complexe RecA-ADN simple brin : si un brin riche en AT est moins souvent rejeté que son homologue riche en GC, la conversion a plus souvent lieu du brin riche en GC vers le brin riche en AT. Ce qui causerait une sur-transmission de GC.

Le modèle du MMR biaisé dépend du choix de brin matrice pour la réparation des mésappariements. Sur la Figure ??, MutH introduit une cassure sur le brin porteur de l'allèle G. Si au contraire, la cassure est plus souvent introduite sur les brins porteurs des allèles A ou T, G ou C est plus souvent transmis.

Un processus sélectionné pour compenser la mutation ?

L'intérêt évolutif d'un tel mécanisme est de second ordre : la mutation est universellement biaisée vers AT^{23;17}. Le gBGC pourrait avoir été sélectionné pour contrecarrer les effets de ce biais mutationnel^{25;2}. Autrement dit, le gBGC permettrait de *guérir* les mutations vers AT par recombinaison homologue. Il est également possible que le gBGC soit dû à des mécanismes de réparation mitotiques, dont l'action biaisée vers GC est conservée au cours de la recombinaison homologue¹².

3 Quelles sont les conséquences du gBGC ?

Puisqu'elle augmente la probabilité de fixation des allèles G ou C, la conversion biaisée joue un rôle important dans la structuration du contenu en GC des génomes. La conversion biaisée n'est pas en soi liée à la sélection naturelle. Elle affecte cependant la fixation d'allèles d'une façon similaire à la sélection²⁶. Elle a donc deux conséquences directes et indirectes : elle interfère avec la sélection et confond les tests de sélection naturelle.

3.1 Le gBGC structure le contenu en GC

Les bases C et G sont liées par trois liaisons hydrogènes : elles sont plus stables que les liaisons doubles entre A et T. Certains pensent qu'en soi, le taux de GC est un trait adaptatif : à l'échelle du génome, un contenu en GC supérieur en augmenterait la stabilité. Ce modèle rencontre néanmoins de nombreuses difficultés, chez les eucaryotes comme les procaryotes. La conversion biaisée vers GC a été proposée comme modèle alternatif expliquant les variations de GC%, au sein d'un génome et entre génomes.

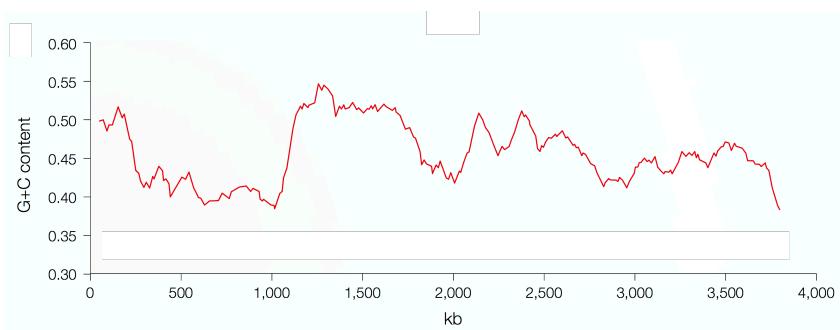


FIGURE 5 – Les isochores : des variations de GC% à grande échelle. Est représentée ici la distribution du taux de GC sur le chromosome humain 6. Des régions relativement homogènes en taux de GC se distinguent. Leur distribution est très variable sur une échelle de 4Mb. Le contenu en GC est corrélé à un grand nombre d'autres facteurs, tels que la densité de gène, le taux de transcription ou encore la vitesse de réPLICATION.

Tiré de Eyre-Walker & Hurst, 2001¹³

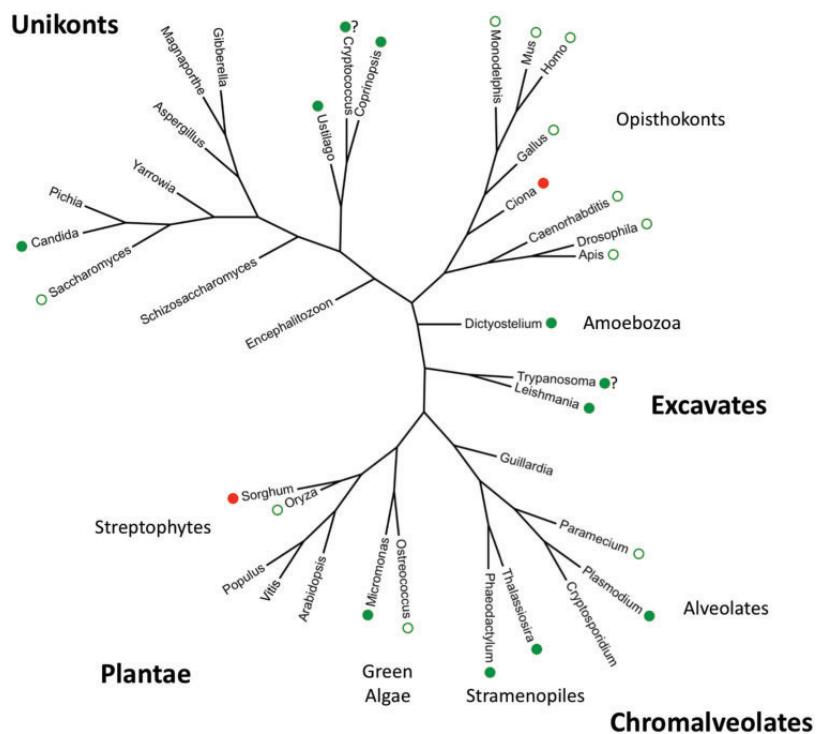


FIGURE 6 – Un gBGC universel ? Corrélation entre le taux de recombinaison et le contenu en GC chez les eucaryotes. Parmi 36 espèces tirées des groupes eucaryotes majeurs, Pessia et collaborateurs ont cherché à déterminer la relation entre GC% et taux de recombinaison. Les ● et ○ indiquent une corrélation positive entre le taux de GC et le taux de recombinaison local. Les deux ● indiquent les corrélations négatives non-compatibles avec l'hypothèse gBGC. Cette étude semble montrer que le gBGC est un mécanisme universel chez les eucaryotes.

Tiré de Pessia et al., 2012²⁸.

Le contenu GC des génomes mammifères et la théorie des isochores

Les mammifères montrent des variations intragénomiques de grande échelle en taux de GC¹³ (>100kb). Ces régions relativement homogènes en taux de GC ont été baptisées isochores*. Leur origine fait débat : est-ce un trait sélectionné ou une conséquence évolutive des patrons de mutations ?

Le modèle sélectionniste se heurte au fait que les variations du GC affectent les sites fonctionnels comme neutres. En fait, l'évolution des isochores résulte de l'accumulation de mutations. Il faudrait donc un avantage sélectif significatif à l'acquisition d'une mutation ponctuelle vers G ou C, dans un isochrome de plus de 100kb.

Le gBGC a été proposé pour expliquer l'apparition et le maintien des isochores riches en GC¹¹. Un argument fort de l'hypothèse gBGC est que les zones fortement recombinantes ont un GC% supérieur. C'est le cas chez l'Homme^{10;1}. L'apparition et la disparition successive de points chauds de recombinaison explique la succession des épisodes de gBGC : il conditionne le contenu en GC local, permettant d'expliquer la structuration des isochores riches en GC.

La taille des chromosomes a un impact fort sur le GC% : le taux de recombinaison à l'échelle de la Mb est fortement corrélé à la taille du chromosome, chez le poulet et l'Homme²⁰. Autrement dit, les grands chromosomes recombinent peu, les petits beaucoup. Comme attendu sous l'hypothèse gBGC, chez l'opossum, les petits chromosomes ont un taux de GC plus élevé que les grands.

Cette corrélation entre le taux de recombinaison et le contenu en GC local a également été observée dans la plupart des taxons eucaryotes (voir Figure 6).

Ainsi, chez les mammifères, le contenu en GC est déterminé par la recombinaison : elle augmente la probabilité de fixation des mutations vers GC. Elle a pour impact de structurer localement le GC% au gré des épisodes de points chauds de recombinaisons. De nombreuses preuves indirectes attestent de l'existence du gBGC chez les eucaryotes. Qu'en est-il chez les procaryotes ?

Un gBGC procaryote ?

Le taux moyen de GC chez les procaryotes est extrêmement diversifié : il varie de 14 à 75% selon les espèces. Certains y voient une adaptation aux conditions environnementales. En effet,

*. Voir Figure 5

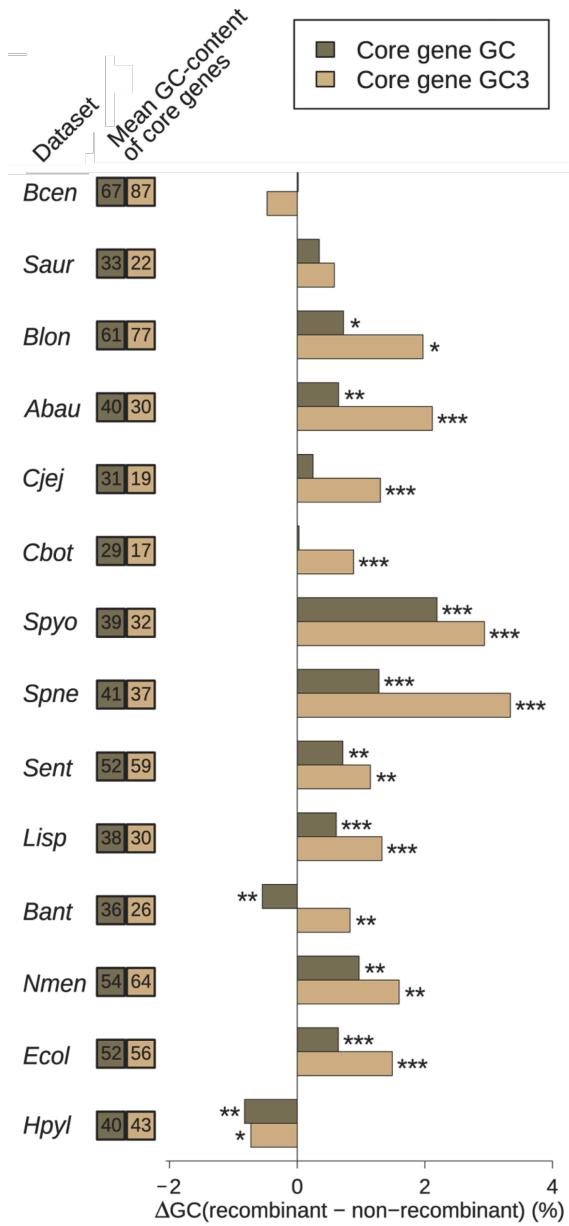


FIGURE 7 – Le gBGC chez les procaryotes ? Effet de la recombinaison sur le contenu en GC du *core genome*. La différence entre le contenu en GC des gènes recombinants et des gènes non-recombinants est mesurée sur l'ensemble de la séquence codante (■) et sur la troisième position de codon uniquement (■). La troisième position est moins soumise à la sélection : les mutations peuvent être synonymes. Le taux de GC des gènes recombinants est significativement supérieur à celui des non-recombinants, *a fortiori* lorsqu'on considère les positions les moins contraintes par la sélection.

Tiré de Lassalle et al., 2015²¹.

la température de croissance optimale est corrélée avec le taux de GC par exemple. Cependant, ces effets environnementaux sont faibles, et les pressions de sélection associées mystérieuses. Le modèle classique considère que le GC% est essentiellement déterminé par la mutation, qui est biaisée vers AT^{17;32}.

Récemment, il a été démontré que les gènes recombinants ont un taux de GC supérieur aux non-recombinants²¹, chez 21 espèces bactériennes. La troisième position des codons est d'autant plus affectée qu'elle est moins soumise à la sélection. Le code génétique étant redondant, une mutation en troisième position ne change pas nécessairement l'acide aminé : la mutation est synonyme. L'excès de substitutions AT → GC en troisième position peut être dû à la pression de sélection sur les autres positions, qui tend à conserver la fonction de la protéine. Les régions intergéniques flanquantes des gènes recombinants ont également un GC% supérieur à celles des régions flanquantes des gènes non-recombinants. C'est un patron attendu sous l'hypothèse gBGC. Cette corrélation entre taux de recombinaison et contenu en GC est similaire quantitativement à celle observée chez l'Humain²¹.

Le gBGC explique donc en partie la structuration en GC des génomes mammifères, eucaryotes et probablement procaryote. Il augmente la probabilité de fixation des allèles G ou C : il peut même s'opposer à la sélection naturelle si cette fixation est faiblement délétère.

3.2 Le gBGC interfère avec la sélection

La conversion génique affecte la probabilité de fixation d'un allèle de façon similaire à la sélection²⁶. Si un allèle faiblement délétère est porté par une substitution AT → GC, le gBGC peut entraîner sa fixation dans la population. À l'inverse, il peut empêcher la fixation d'une mutation GC → AT. Il contrecarre les effets de la sélection naturelle.

Galtier et collaborateurs ont analysé la séquence des protéines de primates qui montrent un taux d'évolution plus rapide depuis la divergence avec les macaques¹⁶. Cette accélération de la vitesse de substitution dans les séquences codantes peut *a priori* être due à un changement de fonction adaptatif. Cependant, les séquences analysées montrent un excès significatif de mutations AT → GC. De plus, ces mutations sont significativement plus souvent non-synonymes : elles changent l'acide aminé. En clair, des régions auparavant conservées ont subi un ou plusieurs épisodes de gBGC, qui ont entraîné deux choses : i) le GC% local a augmenté, et ii) la fonction des régions a changé (*cf* Figure 8).

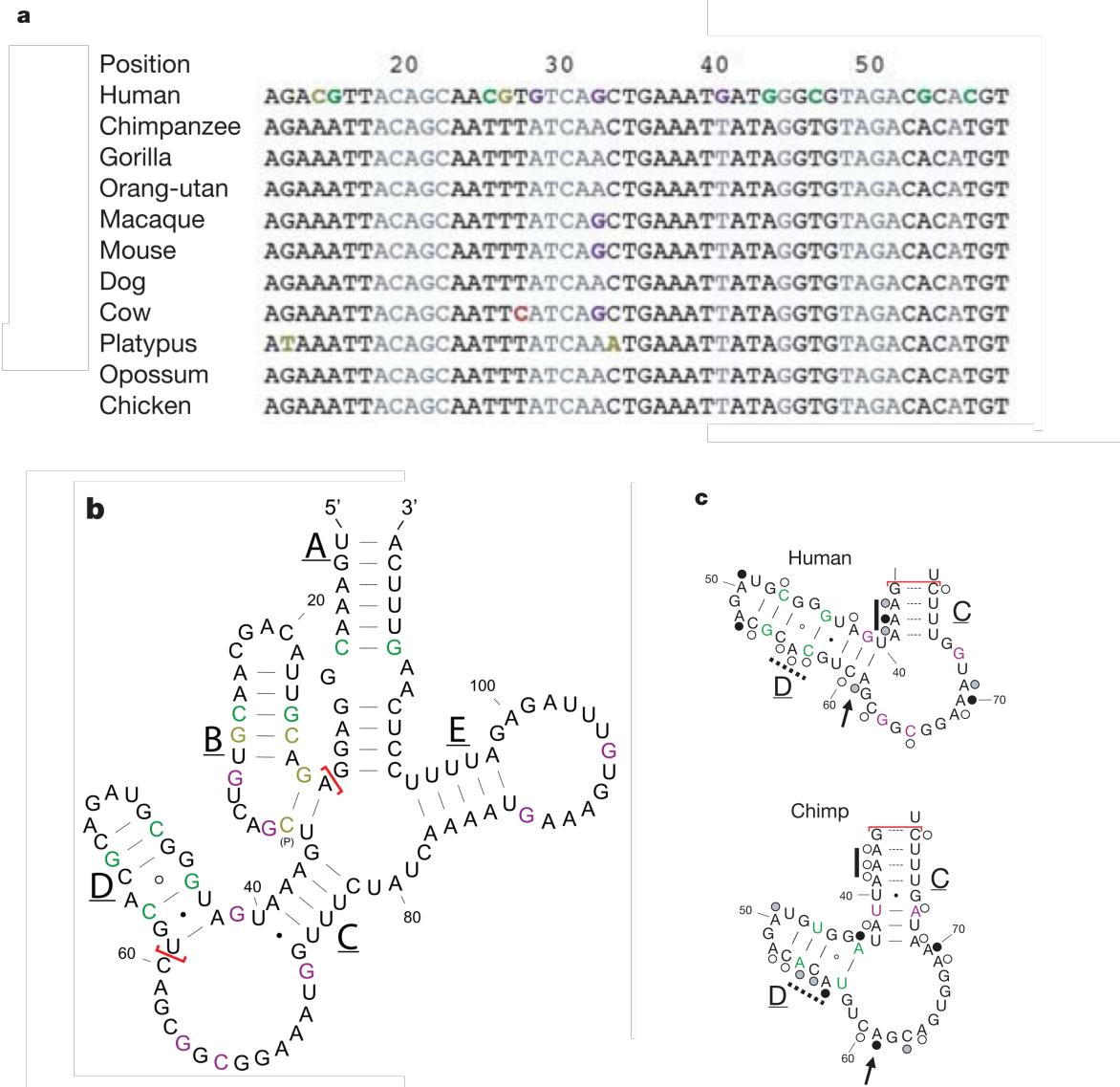


FIGURE 8 – Le gBGC a-t-il interféré avec la sélection pour le développement des régions corticales humaines ? Pollard et collaborateurs ont analysé les régions non-codantes dont la vitesse d'évolution a augmenté uniquement dans la lignée humaine : les HAR, Human Accelerated Regions. Parmi ces régions, la région HAR1 est particulièrement intéressante : elle comprend un gène codant pour un ARN régulateur exprimé au cours du développement des régions corticales. De façon surprenante, les 18 mutations spécifiques à l'Homme sont de type AT → GC (en a). Ces changements ont une influence sur la structure de l'ARN, représenté en b : l'hélice D est plus longue chez l'Homme que chez le Chimpanzé (en c). Autrement dit, le biais vers GC aurait pu influencer la divergence entre l'Homme et le Chimpanzé. Un ARN non-codant extrêmement conservé du Poulet au Chimpanzé a brutalement accéléré dans la lignée humaine. La sélection et le gBGC ont pu agir de concert pour altérer la structure de cet ARN.

Adapté de Pollard et al., 2006²⁹.

3.3 Le gBGC fausse les tests de sélection

Les tests de sélection reposent classiquement sur deux principes généraux¹⁹ : ce qui évolue lentement est fonctionnel, ce qui évolue vite est adaptatif, par rapport à la vitesse d'évolution aux sites neutres. L'approche privilégiée pour détecter les régions influencées par la sélection est de comparer les génomes, puis d'identifier les régions qui évoluent rapidement, sur une branche particulière de l'arbre phylogénétique obtenu³⁰.

Le gBGC a cependant une empreinte sur les séquences similaire à celle de la sélection dirigée : il peut brouiller les tests de sélection. L'action de la sélection naturelle est alors confondue avec celle d'un processus neutre voire mal-adaptatif. Typiquement, le ratio d_N/d_S , qui résume le rapport entre le taux de mutations non-synonymes d_N et synonymes d_S , est supérieur à 1 lorsque la protéine est sous sélection positive, en faveur du changement de fonction. Ratnakumar et collaborateurs ont estimé qu'environ 20% des régions avec un ratio d_N/d_S élevé pourraient avoir été sous l'influence du gBGC.

Le gBGC brouille les traces de la sélection naturelle.

Conclusion

La conversion génique biaisée vers GC est un mécanisme potentiellement universel, qui affecte localement le taux de GC des régions recombinantes. Bien que généralement faible, le biais peut avoir un impact fort sur la fixation d'un allèle, si la taille efficace de la population est grande. Il contribue à structurer le taux de GC des génomes, peut contrecarrer la sélection, et pose des problèmes de détection de celle-ci. De nombreuses interrogations restent en suspens. Les mécanismes qui le sous-tendent sont encore mystérieux, de même que sa raison d'être évolutive. L'extension récente de l'hypothèse gBGC aux procaryotes reste à confirmer expérimentalement. Elle constitue néanmoins un terrain d'exploration nouveau, qui pourrait permettre d'étudier plus avant la machinerie moléculaire responsable d'un phénomène *a priori* anodin, mais dont les conséquences sont nombreuses.

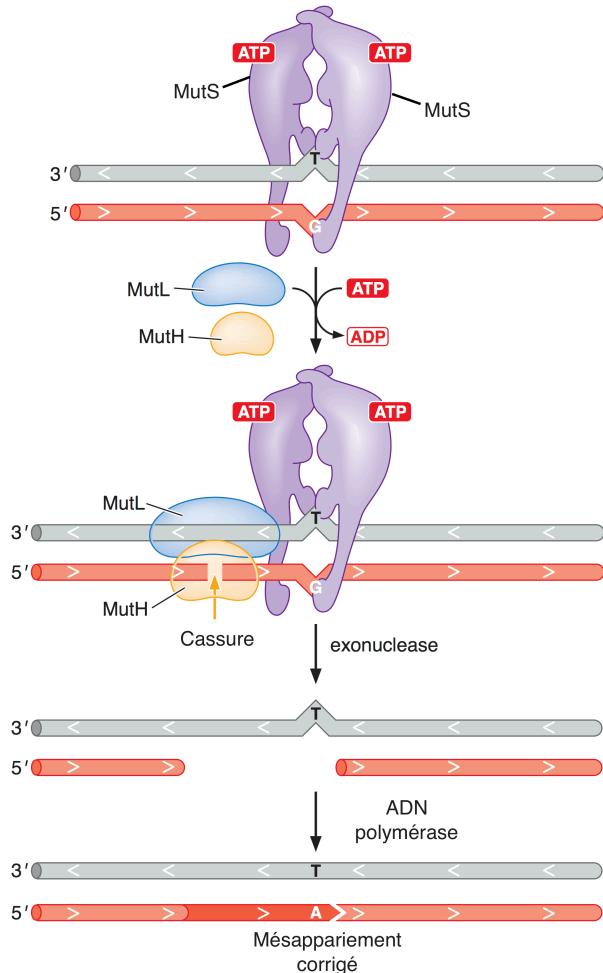
Références

- [1] **Berglund, J., K. S. Pollard, and M. T. Webster.** 2009. Hotspots of Biased Nucleotide Substitutions in Human Genes. *PLoS Biol* **7**:e1000026.
- [2] **Birdsell, J. A.** 2002. Integrating Genomics, Bioinformatics, and Classical Genetics to Study the Effects of Recombination on Genome Evolution. *Mol Biol Evol* **19**:1181–1197.
- [3] **Brown, T. C. and J. Jiricny.** 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* **50**:945–950.
- [4] **Chapman, J. R., M. R. G. Taylor, and S. J. Boulton.** 2012. Playing the End Game : DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell* **47**:497–510.
- [5] **Chen, J.-M., D. N. Cooper, N. Chuzhanova, C. Férec, and G. P. Patrinos.** 2007. Gene conversion : mechanisms, evolution and human disease. *Nat Rev Genet* **8**:762–775.
- [6] **Chen, Z., H. Yang, and N. P. Pavletich.** 2008. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* **453**:489–484.
- [7] **Cromie, G. A., J. C. Connelly, and D. R. F. Leach.** 2001. Recombination at Double-Strand Breaks and DNA Ends : Conserved Mechanisms from Phage to Humans. *Molecular Cell* **8**:1163–1174.
- [8] **Didelot, X. and M. C. Maiden.** 2010. Impact of recombination on bacterial evolution. *Trends Microbiol* **18**:315–322.
- [9] **Dillingham, M. S. and S. C. Kowalczykowski.** 2008. RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. *Microbiol Mol Biol Rev* **72**:642–671.
- [10] **Duret, L. and P. F. Arndt.** 2008. The Impact of Recombination on Nucleotide Substitutions in the Human Genome. *PLoS Genet* **4**:e1000071.
- [11] **Duret, L., A. Eyre-Walker, and N. Galtier.** 2006. A new perspective on isochore evolution. *Gene* **385**:71–74.
- [12] **Duret, L. and N. Galtier.** 2009. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annual Review of Genomics and Human Genetics* **10**:285–311.
- [13] **Eyre-Walker, A. and L. D. Hurst.** 2001. The evolution of isochores. *Nat Rev Genet* **2**:549–555.
- [14] **Fall, S., A. Mercier, F. Bertolla, A. Calteau, L. Gueguen, G. Perrière, T. M. Vogel, and P. Simonet.** 2007. Horizontal Gene Transfer Regulation in Bacteria as a “Spandrel” of DNA Repair Mechanisms. *PLoS ONE* **2**:e1055.
- [15] **Galtier, N. and L. Duret.** 2007. Adaptation or biased gene conversion ? Extending the null hypothesis of molecular evolution. *Trends in Genetics* **23**:273–277.
- [16] **Galtier, N., L. Duret, S. Gléménin, and V. Ranwez.** 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends in Genetics* **25**:1–5.
- [17] **Hershberg, R. and D. A. Petrov.** 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* .

- [18] **Hildebrand, F., A. Meyer, and A. Eyre-Walker.** 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* **6** :e1001107.
- [19] **Hurst, L. D.** 2009. Genetics and the understanding of selection. *Nat Rev Genet* **10** :83–93.
- [20] **Kaback, D. B., D. Barber, J. Mahon, J. Lamb, and J. You.** 1999. Chromosome size-dependent control of meiotic reciprocal recombination in *Saccharomyces cerevisiae* : the role of crossover interference. *Genetics* **152** :1475–1486.
- [21] **Lassalle, F., S. Périan, T. Bataillon, X. Nesme, L. Duret, and V. Daubin.** 2015. GC-Content Evolution in Bacterial Genomes : The Biased Gene Conversion Hypothesis Expands. *PLOS Genetics* **11** :e1004941.
- [22] **Lesecque, Y., D. Mouchiroud, and 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.
- [23] **Lynch, M.** 2010. Rate, molecular spectrum, and consequences of human mutation. *PNAS* **107** :961–968.
- [24] **Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz.** 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454** :479–485.
- [25] **Marais, G.** 2003. Biased gene conversion : implications for genome and sex evolution. *Trends in Genetics* **19** :330–338.
- [26] **Nagylaki, T.** 1983. Evolution of a finite population under gene conversion. *Proceedings of the National Academy of Sciences* **80** :6278–6281.
- [27] **Otto, S. P. and T. Lenormand.** 2002. Resolving the paradox of sex and recombination. *Nat Rev Genet* **3** :252–261.
- [28] **Pessia, E., A. Popa, S. Mousset, C. Rezvoy, L. Duret, and G. A. B. Marais.** 2012. Evidence for Widespread GC-biased Gene Conversion in Eukaryotes. *Genome Biology and Evolution* **4** :675–682.
- [29] **Pollard, K. S., S. R. Salama, N. Lambert, M.-A. Lambot, S. Coppens, J. S. Pedersen, S. Katzman, B. King, C. Onodera, A. Siepel, A. D. Kern, C. Dehay, H. Igel, M. Ares, P. Vanderhaeghen, and D. Haussler.** 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443** :167–172.
- [30] **Ratnakumar, A., S. Mousset, S. Glemin, J. Berglund, N. Galtier, L. Duret, and 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.
- [31] **Redfield, R. J.** 2001. Do bacteria have sex ? *Nat Rev Genet* **2** :634–639.
- [32] **Sueoka, N.** 1988. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **85** :2653–2657.
- [33] **Webster, M. T. and L. D. Hurst.** 2012. Direct and indirect consequences of meiotic recombination : implications for genome evolution. *Trends in Genetics* **28** :101–109.

Annexes

Correction des mésappariements : le modèle E.coli



La réparation des mésappariements par le complexe MutSLH. Le dimère de MutS reconnaît l'anomalie structurelle de l'ADN causé par un mésappariement. Il recrute alors l'enzyme MutL, qui à son tour recrute l'endonucléase MutH. Celle-ci introduit une cassure sur l'un des brins. Elle est suivie d'une résection par une exonucléase, puis par la synthèse *via* une ADN polymérase, sur la base du brin intact.

Adapté de Molecular Biology of The Gene, Watson, 2012.

