

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Microbiologie - Virologie**

Arrêté ministériel : 7 août 2006

Présentée par

Larissa CAUDWELL

Thèse dirigée par **Dominique SCHNEIDER** et
co-encadrée par **Olivier TENAILLON**

préparée au sein du **Laboratoire Adaptation et Pathogénie des
Microorganismes** (CNRS UMR5163)
dans l'**École Doctorale Chimie et Sciences du Vivant**

Dynamiques théorique et expérimentale des taux de mutations

Thèse soutenue publiquement le **23 octobre 2015**,
devant le jury composé de :

M. Eduardo ROCHA

DR, Paris, Rapporteur.

M. Guillaume BESLON

PU, Lyon, Rapporteur.

M. Jean-Luc RAVANAT

DR, Grenoble, Président.

M. Ivan MATIC

DR, Paris, Examinateur.

M. Dominique SCHNEIDER

PU, Grenoble, Directeur de thèse.

M. Olivier TENAILLON

DR, Paris, Co-encadrant de thèse.



*A mon père,
qui m'a plongée dans la marmite de la science
quand j'étais petite.*

Remerciements

Au travers de ce manuscrit s'achèvent trois années de labeur sur un sujet qui m'est cher : l'évolution. J'ai été heureuse de faire mes recherches sur une expérience unique au monde. Tout cela n'aurait pas été possible sans un grand nombre de personnes que je souhaite remercier ici.

J'aimerais d'abord remercier mon directeur de thèse, Dominique Schneider, qui m'a accueillie dans son laboratoire il y a plus de trois ans. Travailler avec vue sur les sommets de Belledonne a été un luxe incroyable après avoir vécu dans la capitale pendant plusieurs années. Dominique a été un excellent encadrant qui m'a donnée à la fois la liberté dont j'avais besoin pour travailler, tout en m'apportant soutien et encadrement lorsque c'était nécessaire, malgré ses nombreuses obligations. Il a toujours été là pour répondre à mes questionnements scientifiques ou pratiques. Tout ce travail n'aurait pas été possible sans son aide.

Je remercie aussi mon co-encadrant Olivier Tenaillon. Son aide m'a été précieuse pour commencer la bioinformatique. Son regard extérieur sur mes travaux a aussi permis bien des avancées. Merci d'avoir répondu à mes questions, même si j'ai mis du temps à comprendre des tes réponses !

Merci aussi à Fabien Pierrel et Ivan Matic, membres de mon comité de suivi de thèse, dont les conseils m'ont été précieux pour faire avancer mon projet sans m'engager dans des fausses pistes. A Eduardo Rocha et Guillaume Beslon, merci d'avoir accepté d'être mes rapporteurs et de relire ma thèse. J'espère que la lecture n'a pas été trop fastidieuse. A Jean-Luc Ravanat, merci d'avoir accepté de collaborer avec nous sur ce projet. Il semblerait que cela ne soit que le début.

Il me faut ensuite remercier mes collègues, sans qui mon passage au laboratoire aurait été bien triste. A tous ceux que je quitte d'abord : Thomas et Joël, les piliers de l'équipe, merci pour vos encouragements et votre expertise. Ludovic, Fabien, Patricia, Corinne, nous nous serons juste croisés, mais votre arrivée a égayé l'équipe. A la partie hospitalière de notre équipe, Max, Sandrine, Yvan, Claire, merci pour votre assistance, en particulier Yvan, pour m'avoir fourni les antibiotiques les plus improbables pour mes expériences. A l'équipe des midis, Jessika, Otmene, Jérémy, Mikael, Laurie-Anne, (maman) Antonia et parfois Thibaut. Je me souviendrai des parties endiablées de *Bang!* ou *Six qui prend*, et des *cheese day* qui viennent couronner la fin de semaine. Jérémy, des oeuvres d'art que tu as laissées sur ma paillasse et de ton humour absurde, je garde un souvenir impérissable (ou pas).

A ceux qui m'ont précédée : Sébastien W. pour avoir initié ce projet et m'avoir formée en Master 2. A Delphine A., ma collègue stagiaire de M2. A Maxime L., Jessica P. et Colin R. qui ont accompagné le début de ma vie de thésarde et m'ont aidée à me familiariser à ce laboratoire. A Amandine G., qui a fait tourner le laboratoire et géré les commandes pendant une bonne partie de ma thèse, merci pour ton aide.

Une partie de mes travaux n'aurait pas été possible sans Bastien T. et Delphine A. qui m'ont donné accès à la plateforme de cytométrie, ainsi que Séverine S. et Christiane V., qui ont coulé les milliers de boîtes nécessaires à mes expériences. Merci pour votre assistance.

A Marion S. et Aleksandra B-G., je vous remercie de m'avoir initiée au journalisme scien-

tifique. La pratique régulière de l'écriture m'a été d'un grand soutien dans la rédaction de ma thèse et je garde de cette expérience un goût prononcé pour la vulgarisation scientifique.

Enfin, je remercie mes proches pour leur soutien et leur affection : mes parents qui m'ont encouragée à faire ce que j'aimais et m'ont permis de faire ces longues études ; Tobit, avec qui j'ai partagé les hauts et les bas de la thèse et qui a été un soutien sans faille ; et enfin, mes nombreux amis thésards, eux aussi. Ceux de la prépa, Thomas et Caroline, ceux de l'ENS Cachan, tous les BGB, en particulier Lucas et Claire , mais aussi dans les autres sections ; ceux du cours Pasteur, Caroline, Fanny, Grégoire G. et enfin ceux que j'ai rencontrés à Grenoble au cours des diverses formations, comme Grégoire R.. Tous mes encouragements à ceux qui n'ont pas encore fini !

Table des matières

I	Introduction	9
1	Les mutations, un des principaux moteurs de l'évolution	10
1.1	Altération de l'ADN et génération de diversité	10
1.2	Protection et réparation de l'ADN pour la stabilité du génome	12
1.3	Populations hypermutatrices	17
2	La dynamique des taux de mutation : entre stabilité et adaptabilité	22
2.1	Adaptation vs. Stabilité	22
2.2	Des approches analytiques ou informatiques	26
2.3	Des approches expérimentales	28
3	Biais de mutation, points chauds mutationnels	32
3.1	Spectres et points chauds mutationnels	32
3.2	Biais de répartition topologique	33
3.3	Biais de sequence	34
4	Modèle expérimental d'évolution à long terme	37
4.1	Description du modèle	37
4.2	Six populations hypermutatrices	39
4.3	Objectifs de la thèse et adéquation du modèle d'evolution pour l'étude des taux de mutation et des biais de mutation	41
II	Evolution des taux de mutations au cours du temps	45
1	Préambule	46
2	Article 1	49

2.1	Abstract	50
2.2	Introduction	50
2.3	Results	53
2.4	Discussion	58
2.5	Conclusion	61
2.6	Materials and Methods	61
	Bibliography	64
2.7	Figure Legends	69
2.8	Tables	71
2.9	Figures	73
3	Supplementary Material	78

III Biais de mutations dans les populations hypermutatrices de l'expérience d'évo-

	lution à long terme	83
1	Préambule	84
2	Article 2	87
2.1	Abstract (up to 250 words)	88
2.2	Introduction	89
2.3	Results	93
2.4	Discussion	97
2.5	Materials & Methods	101
	Bibliography	102
2.6	Figure Legends	106
2.7	Tables	107
2.8	Figures	109
2.9	Supplementary Information	112

IV Discussion

1	Considérations évolutives : évolution de l'évolution	127
1.1	Dynamique au cours du temps	127
1.2	Dynamique spatiale, influence du contexte	128

1.3	Modulation des taux de mutation et évolution de l'évolution	130
2	Implications cliniques	131
3	Perspectives	133
3.1	Mécanistiques	133
3.2	Evolutives	134
	Bibliographie	136

Chapitre I

Introduction

1 Les mutations, un des principaux moteurs de l'évolution

1.1 Altération de l'ADN et génération de diversité

Dans l'*Origine des espèces* (1859), Charles Darwin présente l'évolution comme un processus sélectif plutôt que directif. Ce processus de sélection naturelle implique au préalable la production d'une certaine diversité.

J'ai donné le nom de sélection naturelle ou de persistance du plus apte à cette conservation des différences et des variations individuelles favorables et à cette élimination des variations nuisibles.

Tandis que la variation phénotypique et son héritabilité sont observées et étudiées depuis longtemps, l'origine de cette variation n'est connue que depuis les avancées notables de la génétique au début du XXème siècle (travaux de Gregor Mendel 1865 ; Thomas Morgan 1911), et la découverte de l'ADN dans les années 1950 (Avery et al. 1944 ; Watson et Crick 1953). Une des sources majeures de variabilité est constituée par les mutations, qui sont des altérations de la séquence nucléotidique de l'ADN. Comme toute molécule, l'ADN est soumis à la dégradation et aux altérations chimiques, qui conduisent inévitablement à la modification de sa séquence nucléotidique et donc à l'apparition de mutations.

Certaines mutations surviennent suite à des stress environnementaux, comme l'exposition aux UV ou à des agents mutagènes, ou encore à l'oxydation, alors que d'autres résultent de processus intracellulaires influençant l'ADN, par exemple au travers de modifications épigénétiques, d'interactions avec des protéines ou métabolites, ou suite à des erreurs de réplication de l'ADN. En effet, lors de sa réplication et en l'absence de correction par les systèmes de réparation de l'ADN, une erreur est faite en moyenne toutes les 10 000 paires de bases.

Modifications chimiques Les modifications de l'ADN sont décrites dans la littérature de façon extensive (Friedberg et al. 1995). Elles diffèrent par le type de modification chimique induit sur l'ADN. Les dimères de pyrimidines résultent de liaisons se formant entre les cycles carbonés de deux résidus pyrimidines voisins, suite à l'activation par les UV. L'alkylation fait suite à l'exposition à des agents alkylants, qui rajoutent un groupement méthyle aux bases nucléotidiques de

l'ADN. D'autres agents chimiques peuvent établir des liaisons entre les bases appariées des deux brins d'ADN, créant un pont inter-brin. Mais parmi les modifications spontanées, les dommages oxydatifs et la désamination des bases nucléotidiques sont les plus fréquents (Friedberg et al. 1995).

Les dommages oxydatifs engendrent des modifications au niveau des nucléotides, mais peuvent aussi générer des cassures simple brin. En effet, en conditions réductrices, les molécules de dioxygène forment des espèces réactives de l'oxygène ou ROS (reactive oxygen species). Elles se réduisent en ions superoxyde O_2^- , qui peuvent ensuite réagir pour former du peroxyde d'hydrogène H_2O_2 . Le peroxyde d'hydrogène peut alors subir une réaction de Fenton, produisant ainsi des ions HO. très réactifs. En leur présence, les bases nucléotidiques peuvent s'oxyder. Le dGTP en particulier, quand il est oxydé, forme du 8-oxo-dGTP, qui a la propriété de pouvoir s'apparier avec un dATP ou un dCTP avec la même probabilité, ce qui en fait un des agents mutagènes les plus puissants dans la cellule. L'intégration de 8-oxo-dGTP dans l'ADN génère des mutations AT→CG. Les ROS peuvent aussi s'attaquer aux liaisons phosphatidyl-di-esther des nucléotides dans l'ADN, provoquant des cassures simple brin.

La désamination des bases est un remplacement du groupement amine par un hydroxyle sur leur cycle benzénique. La désamination de la cytosine produit de l'uracile, mais cette réaction est encore plus fréquente lorsque la cytosine est méthylée, suite à un marquage épigénétique par la cellule. Il en résulte une mutation C→T très fréquente au niveau des sites de méthylation de l'ADN. Les sites de méthylation de l'ADN sont donc des points chauds de mutation (voir partie 1.3).

Autres sources de variabilité Outre ces modifications chimiques, certaines mutations surviennent également suite à une erreur lors de la réplication de l'ADN par les ADN polymérases. Il peut s'agir de l'incorporation d'une mauvaise base ou même d'un glissement des polymérases notamment au niveau de répétitions de mononucléotides ou d'une série de bases nucléotidiques, ce qui peut alors provoquer l'insertion ou la délétion d'une ou plusieurs paires de bases.

Même si les mutations représentent une source majeure de diversité, d'autres sources de variabilité génétique existent. Alors que les mutations correspondent à des modifications à petite échelle (sur un ou plusieurs nucléotides), d'autres modifications impliquent des régions de taille

bien plus importante. Ainsi, les réarrangements chromosomiques, par exemple par recombinaison homologue entre des séquences répétées, permettent la délétion, la duplication, l'amplification, l'inversion ou la translocation de gènes, parfois impliquant de grandes parties du chromosome (Feher et al. 2012 ; Raeside et al. 2014 ; Darmon et al. 2014). Ces réarrangements chromosomiques impactent non seulement sur la structure mais également sur l'expression des génomes. Ils ont été caractérisés chez de nombreuses bactéries et se produisent généralement par recombinaison entre séquences répétées (copies d'ADN ribosomique, séquences d'insertion IS, . . . , Darmon et al. 2014). En plus de ces réarrangements, les transferts horizontaux représentent également une source importante de variation chez les microorganismes (Denamur et al. 2000 ; Syvanen 2012). Ils permettent notamment l'acquisition d'ADN à partir d'une source extérieure, et donc rendent possible l'échange ou l'acquisition de nouveaux gènes, ainsi que le rétablissement d'allèles fonctionnels lorsque la copie d'origine est mutée. Les événements de transferts horizontaux de gènes, par transformation, conjugaison ou transduction, permettent donc potentiellement d'acquérir de nouvelles fonctions de façon extrêmement rapide et efficace.

L'ensemble de ces modifications permettent donc des variations dans les génomes et donc leur évolution suite à différentes pressions de sélection. Cependant, ces mutations peuvent bien évidemment être également néfastes en modifiant des gènes nécessaires à la survie des organismes vivants. Face à toutes ces modifications, les cellules disposent d'une machinerie complexe leur permettant de maintenir leur information génétique au cours du temps. Ces systèmes, permettant de corriger mais introduisant également des erreurs, entraînent ainsi un équilibre entre l'évolvabilité, c'est-à-dire la capacité de produire des innovations phénotypiques, et la stabilité des génomes des organismes vivants.

1.2 Protection et réparation de l'ADN pour la stabilité du génome

La survie des organismes implique nécessairement le maintien de l'intégrité de leur information génétique. De fortes modifications de cette information peuvent nuire au fonctionnement cellulaire et donc à la survie. Malgré une stabilité supérieure à l'ARN, l'ADN reste une molécule sujette aux mutations, d'autant plus qu'il s'agit d'une molécule de grande taille, plus grande que la cellule bactérienne elle-même. Ainsi, divers systèmes ont été développés au cours de l'évolution pour

protéger et réparer le génome. En général, les mésappariements et les modifications de l'ADN à l'origine des mutations provoquent une torsion anormale de la double hélice. Cette torsion est détectée et permet la réparation par les enzymes cellulaires. La conservation de l'information génétique se joue à plusieurs niveaux : de la fidélité de la réplication à la réparation, en passant par la protection de l'ADN. Je décrirai surtout ces mécanismes présents chez les Bactéries, car celles-ci constituent mon modèle d'étude.

Correction des erreurs de réplication Un premier niveau de conservation est réalisé au niveau de la réplication de l'ADN, à travers la fidélité contrôlée par des enzymes de réparation. Lors de la polymérisation, les ADN polymérases produisent environ un mésappariement toutes les 10 000 à 100 000 paires de bases. Avec un tel taux d'erreur, la réplication d'un génome d'*Escherichia coli* contiendrait donc plus de 40 erreurs. En réalité, les ADN polymérases réplcatives possèdent une fonction de relecture qui permet d'augmenter la fidélité de réplication d'un facteur 1000 à 10 000. Cette fonction de relecture détecte et élimine certains mésappariements. Elle est réalisée par la sous-unité ϵ de l'ADN polymérase III codée par le gène *dnaQ* (Fijałowska et al. 2012).

Malgré cette activité de relecture, des mésappariements peuvent subsister. En l'absence de réparation, ils engendrent des mutations au cours du cycle de réplication suivant. Ces mésappariements sont réparés par le système MMR (Methyl-directed Mismatch Repair), un système dépendant de la méthylation de l'ADN (Figure 1). En effet, l'ADN des bactéries est méthylé. La méthylation de l'ADN intervient après la réplication. Au moment précis de la réplication de l'ADN, l'ADN est héli-méthylé, avec seul le brin matrice portant ces modifications. Il est ainsi possible de distinguer le brin matrice du brin néo-synthétisé, ce qui permet de cibler les réparations selon le brin matrice. Le système MMR repose essentiellement sur 3 protéines : MutH, MutS et MutL. La protéine MutS détecte le mésappariement, puis interagit avec la protéine MutL qui permet le recrutement de la protéine MutH au sein du complexe et son activation. La protéine MutH reconnaît la méthylation de l'ADN. Elle discrimine donc le brin matrice du brin néo-synthétisé et clive ce dernier au niveau du site de méthylation (Figure 1). Ceci provoque le recrutement d'une hélicase et d'exonucléases permettant l'élimination du mésappariement. Il est ensuite corrigé par l'ADN polymérase III (Wolfram et al. 1995 ; Li 2008).

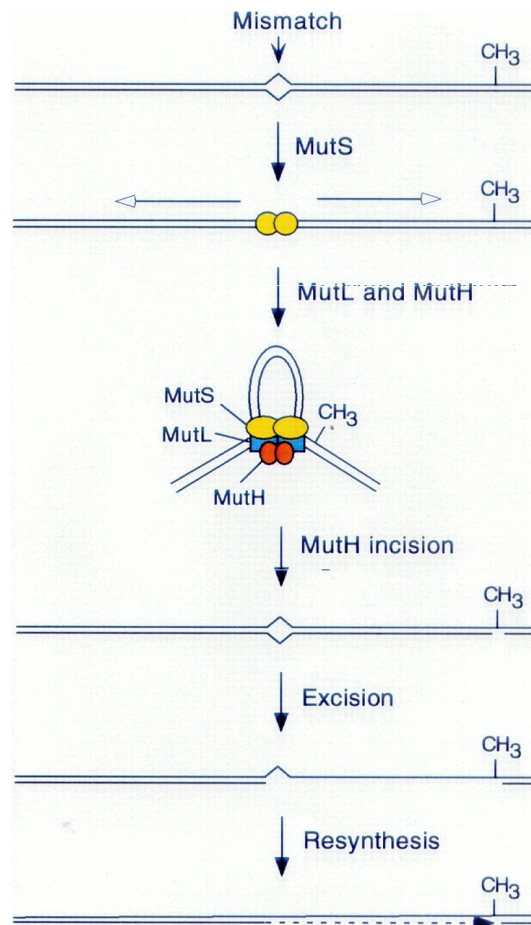


FIGURE 1 : Système de réparation des mésappariements MMR (Harfe et Jinks-Robertson 2000).

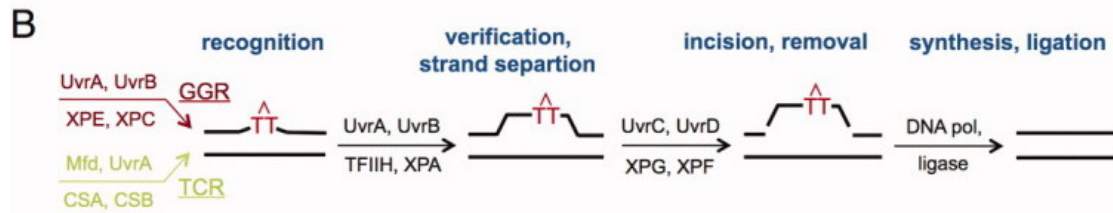


FIGURE 2 : Réparation des dimères de pyrimidines par le système d'excision de nucléotides (Yang 2011) Deux voies de reconnaissance des mutations aboutissent à ce processus, la reconnaissance globale du génome (GGR) et la reconnaissance couplée à la transcription (TCR).

Réparation des altérations Les autres types de dommages à l'ADN sont pris en charge par des systèmes différents. Ainsi, les dommages aux UV, comme les dimères de pyrimidines, sont pris en charge par un système d'excision de nucléotides impliquant les protéines UvrA, UvrB et UvrC, qui détectent, recrutent et éliminent la mutation respectivement, comme dans le système MMR (Figure 2, Kisker et al. 2013). La lésion est ensuite corrigée par l'action jointe de l'hélicase UvrD, de l'ADN polymérase I et d'une ligase. Les modifications des bases nucléotidiques, à l'exception de l'alkylation, créent peu de distorsion dans l'ADN contrairement aux autres mutations. Elles sont corrigées par le système d'excision de base. Ce système fait intervenir des glycosylases, qui détectent et enlèvent la base modifiée, créant un site abasique. Ce site abasique est ensuite enlevé et comblé par une AP-endonucléase (AP pour APurique ou APyrimidique) et une ADN polymérase (Krokan et Bjørås 2013). L'alkylation et les cassures simple brin sont réparées par un système de réparation direct.

La prévention La prévention des mutations peut aussi se jouer en amont de la réplication. Un facteur essentiel se trouve dans le maintien du pool de nucléotides. En effet, le dGTP possède un fort potentiel d'oxydation en 8-oxo-dGTP. Une fois incorporé dans l'ADN, il s'apparie également avec un dATP ou dCTP, provoquant des mutations AT→CG. Avant d'être incorporé, il peut être détoxifié et hydrolysé en 8-oxo-dGMP (oG) par la protéine MutT. Si toutefois, ce nucléotide modifié s'intègre dans l'ADN, il est alors pris en charge par deux glycosylases, MutM et MutY. MutM enlève la base oxydée dans les mésappariements oG:C et MutY enlève la base du A dans les mésappariements A:oG. Lorsque le A fait partie du brin matrice, l'intervention de la protéine MutY est alors mutagène, créant des mutations AT→CG (Figure 3, Maki et Sekiguchi

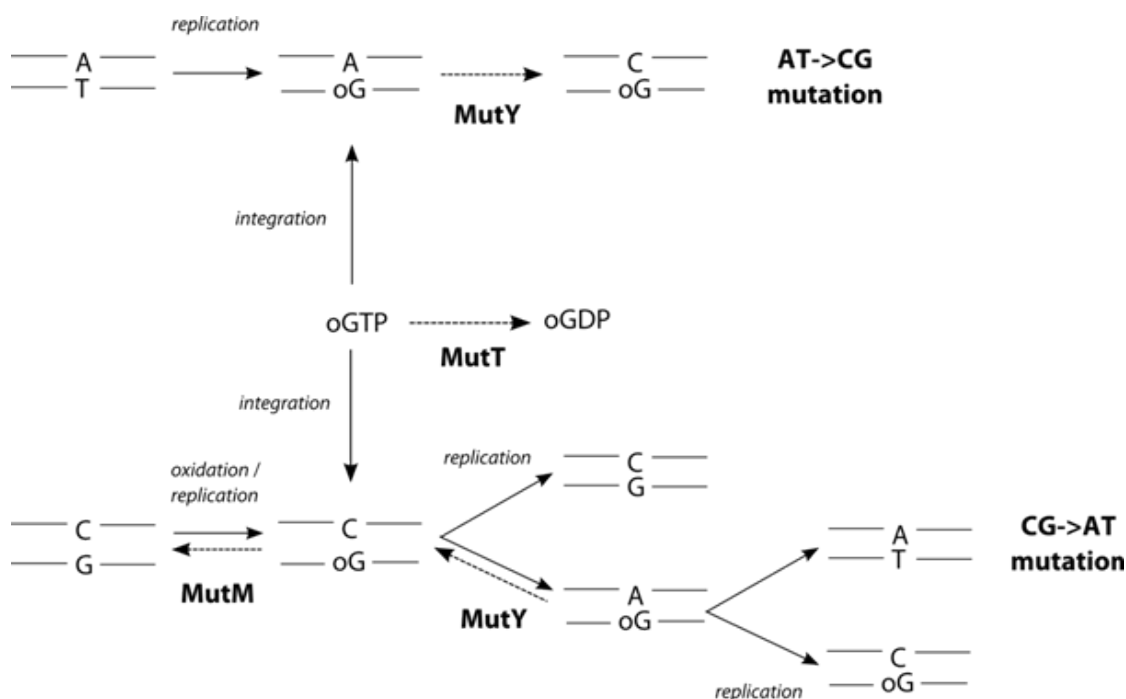


FIGURE 3 : Prévention et réparation des mutations liées au 8-oxo-dGTP. Les protéines sont indiquées en gras, et les processus provoquant les mutations en italique. Les mutations résultant sont indiquées en gras sur le côté droit.

1992). Ainsi, l'absence de la protéine MutT se traduit par une forte augmentation des mutations AT→CG. MutY étant en partie responsable de ces mutations, une l'absence de MutY peut diminuer en partie ces mutations, et compenser le phénotype *mutT*⁻.

En plus de ces systèmes de réparation, l'ADN est protégé par diverses protéines. La protéine SSB (Single-Strand Binding protein) lie l'ADN simple brin, très sensible aux lésions, lors de la réplication et de la transcription (Meyer et Laine 1990). D'autres protéines (H-NS, HU) protègent l'ADN et participent à la régulation de son expression via sa compaction (Dorman 2009). Enfin, certaines protéines comme Dps protègent l'ADN lors de la phase stationnaire en formant des structures cristallines (Zeth 2012).

Les systèmes de réparation de l'ADN et de fidélité de réplication permettent de limiter l'apparition de mutations. L'activité de ces enzymes concerne uniquement des mutations qui peuvent apparaître spontanément (Figure 4) : les mutations CG→GC sont très rares et de ce fait, sont très mal reconnues par le système MMR. L'absence des systèmes de réparation révèle donc

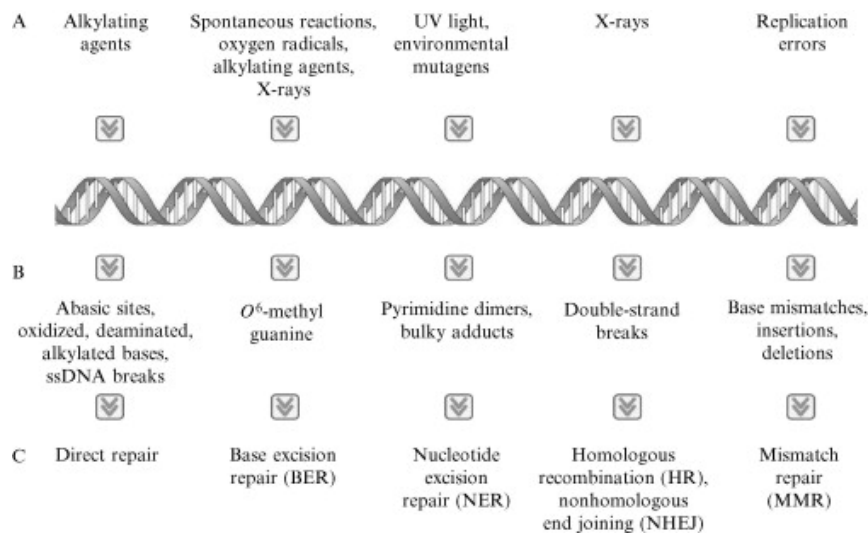


FIGURE 4 : Types de mutations et système de réparation associé (Joseph et al. 2006)

quelles mutations apparaissent spontanément, mais provoque également l'augmentation des taux de mutation.

1.3 Populations hypermutatrices

Le taux de mutation correspond au nombre de mutations par paire de base et par génération. Grâce à la fidélité de la réplication et aux systèmes de réparation de l'ADN, ce taux est d'environ 10^{-9} mutation/génération chez *E. coli* (Wielgoss et al. 2011) et semble constant dans le phylum bactérien (Drake 1991). Pourtant, des mutants présentant des taux de mutation plus élevés, dits hypermutateurs, ont depuis longtemps été observés en laboratoire (Treffers et al. 1954). Cette mise en évidence a soulevé la question de savoir si ce phénotype était propre aux conditions de laboratoire ou s'il pouvait également être retrouvé dans la nature. En effet, la majorité des échantillons observés dans l'environnement présentent un taux de mutation faible. Lors des premiers recensements, les génotypes hypermutateurs représentaient une forte minorité des échantillons : 1/408 (Gross et Siegel 1981), 2/300 (Piechocki et Fleischmann 1984), ou encore 9/268 (Leclerc et al. 1996). Ce n'est que plus tard qu'ils ont été isolés en plus forte proportion, grâce à l'utilisation d'autres techniques (test de papillation) (14%, Matic et al. 1997), plus particulièrement dans des échantillons cliniques (Denamur et al. 2002 ; Oliver et al. 2000). Enfin, encore plus récem-

ment, le phénotype mutateur peut être sélectionné ou émerger lors d'expériences d'évolution en laboratoire et même chez des patients (Sniegowski et al. 1997; Maharjan et al. 2012; Marvig et al. 2013). Quoique majoritairement bas, le taux de mutation est donc variable. Le phénotype hypermutateur résulte de mutations dans les gènes impliqués dans les systèmes de réparation de l'ADN ou dans les facteurs de fidélité de la réplication (Table 1). Ainsi, le premier phénotype hypermutateur observé par Treffers (1954) résulte d'une mutation dans le gène *mutT*, impliqué dans l'élimination du 8-oxo-dGTP (Yanofsky et al. 1966; Bhatnagar et al. 1990).

La dynamique des taux de mutation représente une question évolutive importante car les mutations sont une des sources de variabilité et donc les substrats de la sélection naturelle. Cette dynamique dépend en effet des tensions existant entre les deux forces contradictoires que sont l'adaptabilité et la stabilité.

TABLE 1: Gènes dont les mutations sont à l'origine de phénotypes hypermutateurs (Miller 1998).

Locus	Type de mutation	Fonction
<i>Stress oxydatif</i>		
<i>mutY</i>	G:C → T:A	Glycosylase, enlève le A des mésappariements G:A and 8-oxodG:A
<i>mutM</i>	G:C → T:A	Glycosylase, enlève le 8-oxodG du mésappariement 8-oxodG:C
<i>mutT</i>	A:T → C:G	Hydrolyse le 8-oxodG pour empêcher son incorporation dans l'ADN et la formation de mésappariements 8-oxodG:A
<i>sodA, sodB</i>	substitutions de base	Superoxide dismutase, élimine les radicaux superoxide (ROS)
<i>oxyR</i>	principalement A:T → T:A	Régulateur positif des gènes de réparation des dommages oxydatifs
<i>nth, nei</i>	G:C → A:T	Glycosylases et AP-lyases, enlèvent les pyrimidine oxydées
<i>Excision de bases endommagées ou altérées</i>		
<i>ung</i>	G:C → A:T	Uracil-DNA glycosylase, enlève l'uracile de l'ADN double brin
<i>xth, nfo</i>	Substitutions de bases	exo III et endo IV, enlèvent et réparent les sites AP
<i>usr</i>	G:C → A:T	Endonuclease, promeut l'élimination des T mésappariés avec G dans certains contextes nucléotidiques
<i>ogt, ada</i>	G:C → A:T	Methyltransferases, enlèvent le groupement méthyle de la O ⁶ -methylguanine
<i>Mismatch repair</i>		

Locus	Type de mutation	Fonction
<i>dam</i>	G:C → A:T, A:T → G:C, décalages du cadre de lecture	Méthylase, agit sur les adénines dans les séquences 5'-GATC-3'
<i>mutH</i>	G:C → A:T, A:T → G:C, décalages du cadre de lecture	Endonucléase impliquée dans la réparation dirigée par la méthylation (MMR)
<i>mutS</i>	G:C → A:T, A:T → G:C, décalages du cadre de lecture	Protéine qui reconnaît et lit les mésappariements
<i>mutL</i>	G:C → A:T, A:T → G:C, décalages du cadre de lecture	Collabore avec les protéines <i>MutS</i> et <i>MutH</i>
<i>uvrD</i>	G:C → A:T, A:T → G:C, décalages du cadre de lecture	Hélicase
<i>Erreurs de répllication</i>		
<i>mutD (dnaQ)</i>	Substitutions de bases et décalage du cadre de lecture	Sous-unité (ϵ) de la pol III, activité de relecture
<i>polA</i>	Décalages du cadre de lecture, délétions	pol I, impliquée dans la synthèse d'ADN réparatrice
<i>dnaE</i>	Substitutions de bases et décalages du cadre de lecture	sous unité α de l'ADN pol III ; sélection de bases et activité de relecture
<i>Processus de recombinaison</i>		
<i>recG</i>	Décalages du cadre de lecture (hors division)	Hélicase impliquée dans la migration des jonctions de Holliday
<i>topB</i>	Délétions	Topoisomérase III

Locus	Type de mutation	Fonction
<i>hns</i> (<i>bglY</i>)	Délétions	affecte la topologie du chromosome
<i>sbcA</i>	Inversions au niveau des micro-homologies	active <i>recE</i>
<i>sbcB</i>	Délétions	Altère l'exonucléase I
<i>sbcC</i>	Inversions au niveau des micro-homologies	Clive les structures secondaires qui se forment au niveau des fourches de répllication
<i>Autre</i>		
<i>mutA</i> , <i>mutC</i>	A:T→ T:A, G:C→ T:A, parfois A:T→ C:G	Glycine l'ARNt de la glycine s'insère sur les codons asparagine. Peut créer des polymérases mutatrices (ϵ)
<i>miaA</i>	G:C→ T:A	Transférase impliquée dans la modification des ARNt

2 La dynamique des taux de mutation : entre stabilité et adaptabilité

2.1 Adaptation vs. Stabilité

Les mutations ont un caractère ambivalent : elles permettent de créer une diversité nécessaire pour l'évolution, mais elles peuvent aussi nuire au fonctionnement de la cellule en interrompant ou modifiant des gènes et leurs séquences régulatrices. Ainsi, les mutations contribuent à l'évolvabilité, la capacité à évoluer, mais diminuent la stabilité du génome. Lorsqu'elles améliorent l'adaptation à l'environnement, elles sont dites bénéfiques. Elles sont en revanche délétères lorsqu'elles diminuent la valeur sélective, c'est-à-dire la capacité à produire des descendants adaptés à leur environnement, et létales si elles affectent directement la viabilité de l'individu qui les porte. Les mutations peuvent aussi être neutres quand elles n'ont pas d'effet détectable sur un caractère phénotypique donné.

Il est de façon générale bien plus simple de détériorer un système fonctionnel que de l'améliorer en particulier car les systèmes biologiques résultent de milliards d'années d'évolution : les mutations ont donc en majorité des effets délétères car elles se produisent également de façon aléatoire (Fisher 1930). De plus, les systèmes de réparation de l'ADN et les processus de fidélité de la réplication sont très répandus et conservés au cours de l'évolution. Ils permettent le maintien du taux de mutation à un niveau bas, ce qui contribue au maintien de l'intégrité du génome et donc de sa stabilité. Cependant, ces systèmes de réparation ont un coût énergétique et peuvent ralentir la reproduction (la croissance) en ralentissant la réplication de l'ADN (Kimura 1967). Leur absence peut donc avoir un avantage sélectif direct. De plus, elles ont aussi un effet indirect : en effet, leur absence accroît la fréquence de mutations et augmente donc les chances d'acquérir des mutations bénéfiques. Il s'agit dans ce cas d'une sélection de second ordre. Ces mutations n'ont alors pas d'effet direct sur la valeur sélective, mais un effet indirect via les mutations qu'elles engendrent (Tenaillon 2001). Bien évidemment, ces processus sont à mettre en parallèle avec la charge génétique imposée par des taux de mutation élevés qui augmentent également la proportion de mutations délétères.

Autostop Un phénotype hypermutateur résulte d'une mutation dans un gène impliqué dans la réparation de l'ADN ou la fidélité de la réplication. Avec un taux de mutation plus élevé, un génotype hypermutateur possède ainsi une probabilité plus importante de produire des mutations bénéfiques, et présente donc une évolvabilité accrue. Lorsque l'allèle mutateur et la mutation bénéfique sont sur le même chromosome, seule une recombinaison peut briser le lien physique qui existe entre ces deux mutations (Leigh 1970; Johnson 1999; Tenaillon et al. 2001). En l'absence de recombinaison, les deux mutations restent liées et la sélection s'appliquant sur une mutation, s'applique également sur l'autre. Ainsi, l'avantage sélectif procuré par la mutation bénéfique profite aussi à l'allèle mutateur. Ce dernier peut alors se répandre dans la population, voire s'y fixer avec la mutation bénéfique : c'est le phénomène d'entraînement génique ou auto-stop (Figure 5, Maynard-Smith et Haigh 1974). Ceci est particulièrement le cas chez les bactéries qui possèdent en grande majorité un seul chromosome.

L'émergence de l'allèle mutateur par auto-stop est fortement dépendante du taux de mutations bénéfiques, délétères et létales, ainsi que de la taille de la population (Taddei et al. 1997; Tenaillon et al. 1999). Le taux de mutations peut être élevé si les individus ne sont pas encore adaptés à leur environnement ou si l'environnement est fluctuant. Si le taux de mutation bénéfique est suffisamment élevé, plusieurs mutations bénéfiques peuvent survivre à la dérive génétique et augmenter en fréquence de façon quasi déterministique. A l'inverse, si le taux de mutation létale est trop élevé, la population hypermutatrice reste en faible fréquence. Le taux de mutations bénéfique est d'autant plus élevé si l'environnement fluctue, car le potentiel de mutations bénéfique est entretenu au cours du temps, favorisant ainsi l'émergence de mutateurs (Johnson 1999; Pal et al. 2007). Suite à la production de mutations bénéfiques et en l'absence de recombinaison, seul un mutant ou une combinaison de mutants apparus successivement peut envahir la population. Il y a donc compétition entre ces mutations lors d'un processus appelé interférence clonale (Gerrish et Lenski 1998). Cette interférence ralentit l'émergence des allèles hypermutateurs. De plus, ces derniers étant initialement rares, la population non mutatrice peut aussi produire des mutations avantageuses. Cependant, les génotypes hypermutateurs étant plus à même de générer une succession de mutations avantageuses, ils ont tendance à sortir vainqueurs de cette interférence si celle-ci est soutenue et donc à devenir majoritaires au sein de la population.

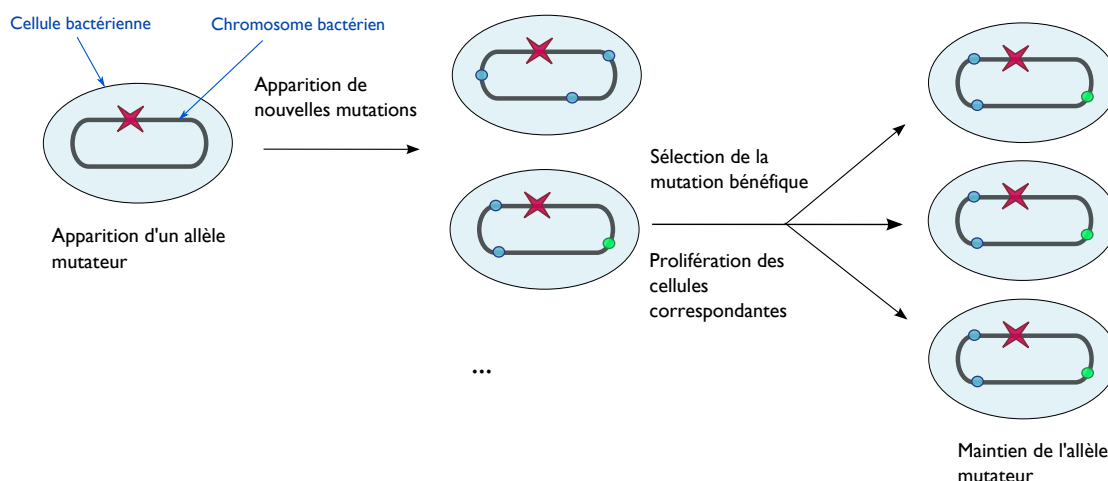


FIGURE 5 : Auto-stop ou entraînement génique. Les cercles verts représentent des mutations bénéfiques et les bleus des mutations neutres ou délétères.

Fardeau mutationnel Un allèle mutateur peut émerger dans la population suite à son entraînement avec une mutation bénéfique. Cependant, il est également affecté par les mutations délétères qu'il produit. De plus, ces dernières étant majoritaires, elles seront en excès. Un génotype hypermutateur produit et accumule donc rapidement des mutations qui lui sont néfastes. A l'équilibre mutation-sélection, la fréquence des mutations délétères n'est pas nulle. En effet, même si une mutation délétère est contre-sélectionnée, elle a des chances non négligeables de réapparaître si le taux de mutation est élevé. A l'équilibre, une mutation délétère d'effet s sur la valeur sélective et apparaissant avec un taux μ a une fréquence égale à $\frac{\mu}{s}$. La diminution de la valeur sélective résultant de la récurrence des mutations délétères est appelée *fardeau mutationnel*. Dans le cas précédent, la valeur sélective moyenne est de $1 * (1 - \frac{\mu}{s}) + \frac{\mu}{s} * (1 - s)$ et est donc égale à $(1 - \mu)$. Le *fardeau mutationnel* est donc égal au taux de mutation lui-même et ne dépend pas du coefficient de sélection des mutations délétères. C'est le principe d'Haldane-Muller. Par conséquent, plus le taux de mutation est élevé, plus la valeur sélective à l'équilibre de la population diminue.

Coût des mutations délétères et accumulation par auto-stop Les mutations délétères peuvent avoir des effets plus ou moins forts sur la valeur sélective. Les mutations létales ont un coût immédiat, elles annulent la viabilité des individus. Lorsque les mutations délétères ont un

effet fort, l'équilibre mutation-sélection se met en place rapidement pour la population hypermutatrice : sa valeur sélective est de $1 - \mu$, comme décrit précédemment. Lorsque les mutations délétères ont un effet plus faible, celles-ci peuvent s'accumuler au cours du temps, et il s'installe alors une dynamique de mutation-sélection accompagnant la sélection de mutations bénéfiques et l'élimination des mutations délétères. Les mutations délétères de faible effet peuvent être maintenues dans la population si elles sont compensées par des mutations bénéfiques. Le phénomène d'auto-stop peut aussi bénéficier aux mutations délétères de faible effet : lorsqu'une mutation bénéfique compense la mutation délétère et augmente la valeur sélective totale, la mutation délétère est alors entraînée par la mutation bénéfique

Faible sélection : cliquet de Muller Lorsque la sélection est faible, les mutations délétères, fréquentes dans une population hypermutatrice, peuvent s'accumuler jusqu'au point où tous les clones présents possèdent au moins une mutation délétère. Aucun descendant de cette population ne peut alors posséder un nombre inférieur de mutations délétères. En l'absence de recombinaison, ce processus peut continuer, entraînant l'accumulation de mutations délétères, et donc la réduction progressive de la valeur sélective de la population. Cette théorie a été formulée pour la première fois par Hermann Joseph Muller (1964) et est plus connue sous le nom de cliquet de Muller. Elle constitue un argument fort pour l'apparition de la reproduction sexuée, qui permet de briser les liaisons physiques entre mutations par recombinaison.

Echelle de temps de la sélection : court terme, et effet sur les taux de mutation

L'évolution des taux de mutation est donc soumise à deux forces opposées, que sont l'adaptabilité et la stabilité, qui se jouent sur deux échelles de temps différentes, à court et long terme. Un clone hypermutateur peut être avantagé à court terme grâce à la production et à l'entraînement par une mutation bénéfique. Mais sur le long terme, sa valeur sélective est plus faible qu'un clone ayant la même mutation bénéfique mais un taux de mutation plus bas, car il produit et accumule rapidement des mutations délétères. Ainsi, sous l'influence de l'adaptation et de la sélection, les taux de mutation sont susceptibles de varier au cours du temps. De nombreuses études, dont la présente thèse, se sont focalisées sur cette dynamique, tant au niveau théorique qu'expérimental.

2.2 Des approches analytiques ou informatiques

Les premières études sur l'évolution du taux de mutations précèdent la connaissance de la nature des mutations et du séquençage de l'ADN (Haldane 1927 ; Fisher 1930). En effet, dans ces modèles théoriques, il suffit de considérer l'effet des mutations sur la valeur sélective et leurs interactions. Plusieurs paramètres peuvent être pris en compte : la taille de la population (finie ou infinie), le poids respectif des mutations bénéfiques ou délétères, la force et le rythme de la sélection, voire le paysage adaptatif (Clune 2008). Certains modèles s'intéressent à l'émergence des hypermutateurs tandis que d'autres recherchent un taux de mutation stable au niveau évolutif, c'est-à-dire qu'une population possédant un tel taux de mutation ne peut pas être envahie par un clone ayant un taux de mutation différent (Maynard-Smith 1982).

Les modèles analytiques les plus simples ne considèrent que les mutations délétères. Il est alors prouvé que le taux de mutation évolue vers des valeurs minimales (Kimura 1967 ; Lieberman et Feldman 1986). Cette valeur minimale des taux de mutation serait déterminée par la dérive génétique (Lynch 2011). Néanmoins, lorsque les mutations bénéfiques ont un poids plus important ou que l'environnement fluctue, le taux de mutation à l'équilibre n'est pas nul (Leigh 1970). Dans son modèle, Johnson (1999) estime que le taux de mutation stable au niveau évolutif est égal à la variation de l'environnement. D'autres modèles (Kondrashov 1995 ; Dawson 1998) prennent en compte un avantage sélectif direct lié à l'allèle mutateur. Il diminuerait en effet le coût et le temps de la réplication, ce qui résulte en un taux de mutation à l'équilibre non nul. Cependant, ces modèles supposent que les mutateurs restent en faible proportion dans la population, ce qui n'est pas toujours le cas, comme observé dans les populations hypermutatrices.

Modèle d'invasion Les génotypes hypermutateurs peuvent augmenter en fréquence suite à la production de mutations bénéfiques. La dynamique d'invasion des mutateurs et leur effet sur la valeur sélective moyenne de la population a été simulée (Taddei et al. 1997). Dans une population infinie, les mutateurs envahissent la population rapidement, participant à l'augmentation rapide de la valeur sélective de la population, puis disparaissent progressivement une fois que le pic adaptatif est atteint, par réversion du phénotype hypermutateur. Dans les populations de taille finie, l'invasion des mutateurs est plus stochastique et ne survient que dans une minorité des simulations. Lorsque les mutateurs parviennent à émerger dans la population, la valeur sélective

augmente rapidement. Cependant, une fois le pic adaptatif dépassé, la valeur sélective décroît progressivement, dû à l'accumulation de mutations délétères, contrairement aux populations de taille infinie. En effet, la population étant de taille finie, les chances de réversion sont plus faibles.

Stochasticité de l'invasion : probabilité de mutation et dérive La stochasticité de l'invasion dans une population finie peut s'expliquer par la faible fréquence des clones hypermutateurs. Lorsqu'un génotype hypermutateur apparaît dans une population, il est en faible nombre et atteint un équilibre mutation/sélection/dérive associé à son fardeau mutationnel élevé. Après cette phase d'équilibre, les mutateurs sont toujours minoritaires tant qu'ils ne produisent pas de mutation bénéfique. De par ce faible nombre, les chances pour qu'une mutation bénéfique apparaisse dans la sous-population mutatrice restent faibles, malgré le taux de mutation plus élevé.

Phase de sélection par entraînement génique : comment émerger ? Les génotypes hypermutateurs peuvent émerger dans la population grâce à la production de mutations bénéfiques et l'entraînement génique (sélection de second ordre). Cependant, des mutations bénéfiques différentes peuvent apparaître dans différents clones de la population et à des temps proches. Il y a alors une concurrence entre les clones porteurs de ces différentes mutations, ce qui ralentit leur émergence : c'est le phénomène d'interférence clonale. Si deux mutations bénéfiques apparaissent dans un contexte génétique non mutateur et hypermutateur respectivement, l'interférence clonale ralentit alors l'émergence des génotypes hypermutateurs.

La sélection par entraînement génique d'un allèle mutateur dépend de la taille de la population, de la force du mutateur, et du nombre de mutations à acquérir pour l'adaptation (Tenailon et al. 1999). En effet, si le génotype mutateur entraîne une augmentation trop forte du taux de mutation, il produit un nombre trop important de mutations délétères pour pouvoir émerger dans la population. A l'inverse, s'il entraîne une augmentation trop faible du taux de mutation, les chances de produire des mutations bénéfiques sont minces, surtout si le mutateur est en très faible nombre dans la population. De même, plus il y a de mutations bénéfiques possibles, plus les génotypes mutateurs sont avantagés. En effet, un premier entraînement génique avec une mutation bénéfique va augmenter la taille de la sous-population mutatrice, qui a alors plus de chances de produire une mutation bénéfique, provoquant des événements successifs d'auto-stop.

Grâce à cette boucle de rétroaction positive, les mutateurs peuvent émerger dans la population. Ainsi, différents paramètres influent sur l'émergence des génotypes hypermutateurs.

Devenir après invasion André et Godelle (2006) ont étudié le taux de mutation à l'équilibre en considérant un modèle dans lequel les mutateurs pouvaient, contrairement aux études précédentes, envahir la population et donc atteindre une fréquence de 1 (les simulations autorisaient cela). Dans ce modèle, chaque balayage sélectif constitue un goulot d'étranglement qui réduit fortement le polymorphisme présent dans la population. Les taux de mutation qui en résultent ne sont pas stables au niveau évolutif car l'invasion par un génotype hypermutateur est toujours possible. Dans leur modèle, Gerrish et al. (2007) poussent cette observation encore plus loin, et décrivent que l'entraînement génique entre allèle mutateur et mutation bénéfique peut aboutir à une augmentation incontrôlée des taux de mutation, provoquant l'extinction de la population. Ceci résulterait de la vue à court terme du processus de l'évolution au cours de laquelle les individus avec la meilleure valeur sélective sont sélectionnés en dépit des mutations délétères qu'ils vont produire par la suite. À terme, l'avantage associé à la production de mutations bénéfiques est de plus en plus rapidement contré par la production massive de mutations délétères, jusqu'à ce que la production de ces dernières devienne insurmontable et qu'elles s'accumulent de façon irrémédiable (cliquet de Muller), provoquant l'effondrement de la population. Ces considérations restent extrêmes et négligent vraisemblablement d'autres paramètres comme l'épuisement des mutations bénéfiques.

Ainsi, les modèles théoriques concluent qu'il est possible pour un génotype hypermutateur d'envahir une population. La phase de sélection dépend de la taille de la population et de la force de l'allèle mutateur. Après invasion, le taux de mutation tend à diminuer vers une valeur intermédiaire, dans un équilibre instable. Les vérifications expérimentales de ces modèles théoriques n'arrivent que plus tardivement, mais elles se développeront grâce au séquençage de génomes de souches évoluées au laboratoire.

2.3 Des approches expérimentales

Au niveau expérimental, la variation du taux de mutations a été observée dès 1950. En 1954, Treffers décrit un variant d'*E. coli* K-12 pouvant développer rapidement une résistance à la

streptomycine. Il démontre que ce variant a en fait un taux de mutation plus élevé que la souche initiale. Cette souche comportait en fait une mutation dans le gène *mutT* (Bhatnagar et al. 1990). Ces variants, appelés alors hypermutateurs, ont un avantage sélectif fort sur leurs homologues non mutateurs.

Dynamique d'invasion : stochasticité et avantage indirect Lorsque deux populations d'*E. coli*, l'une non mutatrice et l'autre mutatrice, sont co-cultivées, la population mutatrice prend rapidement le dessus (Gibson et al. 1970 ; Cox et Gibson 1974). Chao et Cox (1983) se sont intéressés à la dynamique et aux conditions favorisant l'invasion des hypermutateurs. Ils ont mis en compétition une souche d'*E. coli* avec un variant mutateur *mutT* dans un chemostat sur une centaine de générations, avec différentes fréquences de départ, et ont suivi l'évolution de chaque sous-population au cours du temps. L'invasion de la population non mutatrice par les hypermutateurs n'arrive qu'après une durée d'environ 60 générations et seulement si la fréquence initiale des mutateurs est supérieure à un certain seuil (7×10^{-5}). Cette phase précédant l'invasion des mutateurs correspond en fait au temps nécessaire à l'acquisition et à l'émergence d'une mutation bénéfique. De plus, lorsque les mutateurs sont en faible proportion, leur probabilité d'acquisition de mutations bénéfiques est contre-balançée par leur rareté. A l'inverse, lorsque la souche non mutatrice est en forte proportion, la probabilité d'apparition d'une mutation bénéfique dans cette population augmente. Ainsi, le temps nécessaire et la dépendance à la fréquence de l'avantage sélectif indiquent que les mutateurs bénéficient des mutations bénéfiques qu'ils produisent et émergent par entraînement génique. De même, Gentile et al. (2011) ont réalisé des compétitions entre mutateurs de différentes forces, en utilisant des simple- ou double-mutants dans des gènes de réparation de l'ADN (*mutL* ou *mutL/dnaQ*). Le double mutant, qui présente un taux de mutation 40 fois plus élevé que le simple mutant, est alors avantagé lorsque la population est petite (fréquence-dépendance) ou la pression sélective forte, car il favorise grandement la probabilité d'apparition de mutations bénéfiques. D'autres travaux chez la levure ont confirmé l'influence de la taille de la population sur ce phénomène d'auto-stop mutationnel. Une population de grande taille permet l'apparition de mutations dans les génotypes non mutateurs, et il existe alors une interférence clonale entre les mutateurs et non mutateurs ayant des mutations bénéfiques (Raynes et al. 2012). Ainsi, ces expériences confirment que les hypermutateurs

peuvent émerger dans une population suite à un entraînement génique. Cependant, l'impact de la fréquence-dépendance se pose. En effet, lorsqu'un hypermutateur apparaît dans une population suite à une mutation, il représente un individu dans toute la population. De son apparition à sa fixation, un hypermutateur est donc nécessairement soumis à la dérive génétique au cours de laquelle il peut disparaître, comme décrit dans les modèles théoriques (Taddei et al. 1997; Tenaillon et al. 1999).

Emergence suite à une mutation : dérive et interférence clonale, deux obstacles à l'invasion Pendant une expérience de co-évolution de *Pseudomonas fluorescens* avec le phage SBW25Φ2, une population sur quatre développe un phénotype hypermutateur (Pal et al. 2007). Dans ces conditions, la pression évolutive est très forte et les génotypes hypermutateurs sont avanta-gés. Cette minorité de populations hypermutatrices malgré les conditions stringentes illustre la compétition intense pour produire des mutations bénéfiques. Les mutateurs peuvent l'emporter, mais ce n'est pas une nécessité. Dans une autre expérience d'évolution (expérience d'évolution à long terme, voir partie IV), Shaver et al. (2002) ont suivi l'émergence de génotypes hypermutateurs dans trois populations différentes. De leur apparition à leur fixation, il s'écoule plus de 1000 générations. Ce temps est bien inférieur au temps nécessaire à une fixation par dérive génétique, confirmant que ces génotypes hypermutateurs ont été sélectionnés. Cependant, cette longue période de co-existence entre génotypes mutateurs et non mutateurs illustre l'interférence clonale entre ces deux sous-populations. En effet, les deux sous-populations mutatrices et non mutatrices ont augmenté leur valeur sélective au cours du temps. Elles ont ainsi pu interférer avec l'évolution de l'autre, ralentissant l'émergence du phénotype hypermutateur. Si la sous-population hypermutatrice survit à la phase de dérive génétique, puis à l'interférence clonale, elle peut alors se fixer dans la population. Or lorsque l'adaptation est terminée, il n'y a plus d'avantage à acquérir de nouvelles mutations, et le phénotype hypermutateur devrait être désavantagé.

Déclin du taux de mutation après adaptation Cette question du déclin du phénotype hypermutateur a été étudiée en 1984 par Tröbner et Piechocki. Après 2200 générations dans un chémostat, des mutants *mutT* ont un taux de mutation réduit. De même, des levures adaptées depuis des années à leur milieu (souches de laboratoires adaptées à leur milieu de culture) ont

été modifiées génétiquement pour obtenir un phénotype mutateur, puis cultivées pendant près de 6700 générations (mutation dans *msh2*, McDonald et al. 2012). Sur les 8 populations hypermutatrices, 4 avaient un taux de mutation réduit. Ainsi, le phénotype hypermutateur est avantage à court terme, mais est néfaste à long terme. Ces conclusions ont été confirmées *in vivo* grâce à des expériences de colonisation intestinale chez la souris (Giraud et al. 2001). Les génotypes hypermutateurs ont un avantage sélectif pendant la colonisation. Par contre, leur taux de transmission à un autre individu est réduit, diminuant ainsi le bénéfice à long terme. Contrairement aux modèles de Gerrish et al. (2007), le taux de mutation tend donc à décroître avec le temps et non à augmenter.

Depuis la connaissance des allèles mutateurs, de nombreuses expériences ont été réalisées pour mieux comprendre l'apparition et l'émergence des hypermutateurs. Un nombre beaucoup moins important d'expériences se sont focalisées sur leur devenir au long terme et aux mécanismes impliqués dans le déclin des taux de mutation au cours du temps. C'est dans ce cadre que s'inscrit cette thèse.

L'accumulation des mutations au cours du temps a aussi permis d'observer que toutes les mutations ne sont pas équivalentes, et qu'elles peuvent se concentrer dans certaines régions du génome. Ces biais de mutations, bien que surprenants, n'ont pu être étudiés que récemment suite à la démocratisation du séquençage de génomes entiers.

3 Biais de mutation, points chauds mutationnels

3.1 Spectres et points chauds mutationnels

En 1961, Benzer soulevait la question suivante : les mutations sont-elles équitablement réparties sur le génome ? Le génome étant une structure organisée, il est légitime de penser que la répartition des mutations peut être hétérogène. En effet, certains segments du génome sont exprimés et d'autres non. Or, durant l'expression, l'ADN se trouve de façon transitoire dans un état simple brin qui est plus vulnérable. D'autre part, le génome comporte des régions plus ou moins compactées ou structurées (Valens et al. 2004). Enfin, certains motifs sont plus ou moins mutagènes, compliquant l'action de la réplicase ou facilitant des altérations chimiques. Par exemple, la répétition de nucléotides peut entraîner un glissement de la polymérase, tandis que la méthylation de la cytosine favorise sa déamination. En utilisant la région *rII* du bactériophage T4, Benzer mit en évidence l'existence de points chauds mutationnels, où les mutations étaient concentrées. Plus tard, une expérience utilisant le gène *lacI*, codant le répresseur de l'opéron lactose, comme cible de mutations confirma l'existence de ces points chauds, impliquant les cytosines méthylées (Coulondre et al. 1978).

Avec l'émergence du séquençage haut-débit, l'accès à l'identification des mutations est devenu plus simple et plus exhaustif. Ainsi, un autre biais a été observé au niveau des spectres de mutations. Il existe 2 types de transitions (mutation purine vers purine, ou pyrimidine vers pyrimidine) contre 4 types de transversions (pyrimidine vers purine ou inversement). Le taux théorique de transitions par rapport à celui des transversions, aussi appelé Ts/Tv , est donc théoriquement de 0,5. Or, dans toutes les espèces étudiées, il existe une prédominance de transitions par rapport aux transversions, à l'exception de *Cenorhabditis elegans* (Lynch et al. 2008 ; Lynch 2010 ; Hershberg et Petrov 2010). De même, les mutations $GC \rightarrow AT$ sont majoritaires, tandis que les mutations $GC \rightarrow CG$ sont très rares. Ce biais presque universel semble lié au processus de mutation lui-même et non à la sélection naturelle. En effet, dans ces études la pression sélective a été minimisée, soit en réalisant une expérience d'accumulation de mutations (Lynch 2008), soit en choisissant des organismes ayant évolué sous une pression sélective faible (Hershberg et Petrov 2010).

Ainsi, toutes les mutations ne sont pas équivalentes ni réparties aléatoirement dans le génome.

La nature et l'origine de ces biais suscite un intérêt grandissant car ils représentent une force directrice de l'évolution mais également une façon d'appréhender plus en détail les mécanismes d'action des systèmes de réparation de l'ADN. Il est possible de distinguer deux types de biais : un biais de séquence suggérant l'influence des bases voisines sur l'apparition d'une mutation, et un biais topologique selon lequel différentes régions du chromosome n'ont pas la même mutabilité.

3.2 Biais de répartition topologique

Pourquoi s'attendre à voir des biais topologiques ? Le génome est structuré. Il comporte deux régions qui se distinguent par leur sens de réplication, les réplichores, séparés par l'origine et la région terminus de réplication. La réplication détermine aussi une différence entre les brins sens et non-sens, l'un d'entre eux étant répliqué via la synthèse de fragments d'Okazaki et l'autre non. A travers ces différences de réplifications, les mutations peuvent apparaître de façon différenciée entre les deux brins et faire varier leur composition (Fijalkowska et al. 1998 ; Rocha et al. 2006). De même, le génome comporte des régions plus ou moins structurées, les macrodomaines (Valens et al. 2004). Cette structuration pourrait empêcher, ou au contraire, favoriser les mutations. De plus, certaines régions sont exprimées et d'autres non. Par conséquent, il est possible d'imaginer que toutes les régions du chromosome ne soient pas exposées de façon équivalente aux mutations, et ces dernières pourraient donc se produire de manière différentielle le long du chromosome et selon les brins.

Observations de biais de brin d'ADN En examinant la composition des génomes d'*E. coli*, de *Bacillus subtilis* et d'*Haemophilus influenzae*, Lobry (1996) découvre un fort biais d'utilisation des nucléotides C ou G symétrique par rapport à l'origine de réplication (un phénomène appelé GC-skew). En effet, du côté gauche de l'origine une majorité de C par rapport à G est détectée tandis que la partie droite comporte une majorité de G. Ce biais de composition résulterait d'un biais de mutation plutôt que de la sélection, car il est plus fort dans les séquences intergéniques (Lobry 1996). De plus, la symétrie par rapport à l'origine suggère que ce biais est lié aux mécanismes de réplication, en particulier à la synthèse des fragments d'Okazaki. Des biais dans la composition des brins ont par la suite été observés dans 21 espèces bactériennes (Rocha et al. 2001). Un tel biais est encore observé dans une expérience d'accumulation de mutations chez

un mutant *mutL* d'*E. coli* dans lequel les transitions AT→GC sont favorisées (Lee et al. 2012). Dans le réplicore droit, les résidus A sont mutés à une fréquence deux fois plus élevée que les résidus T sur le brin supérieur. L'inverse est observé dans le réplicore droit. Les résidus T ont donc une probabilité plus élevée d'être mutés sur le brin matrice que sur le brin retard. De même, les transitions GC→AT affectent plus les résidus G sur le brin matrice.

Motif en vaguelette, pas de correspondance avec les macrodomaines En plus de cette symétrie selon le réplicore, les mutations semblent s'accumuler selon un motif en vaguelette détecté lors d'une expérience d'accumulation de mutations (Foster et al. 2013). Lorsque les mutations de cette expérience sont analysées en considérant des intervalles de 100kpb environ, leur nombre augmente et diminue successivement le long du chromosome. Des observations semblables ont été faites chez *B. subtilis* (Sung et al. 2015) et *Pseudomonas fluorescens* (Long et al. 2015). Il existe donc des régions du chromosome plus sujettes à mutations que d'autres. De plus, les mutations sont présentes autour de l'origine de réplication, tandis qu'elles sont plus denses au niveau des régions terminales (Foster et al. 2013). Cependant, les variations de densité de mutations ne correspondent pas aux macrodomaines du chromosome et semblent plutôt liées au processus de réplication (Foster et al. 2013; Niki et al. 2000). De même, chez la levure *Saccharomyces cerevisiae*, aucun biais de répartition de mutation n'a été détecté entre chromosomes (Zhu et al. 2014). Cependant, cette étude exclut les points concentrant les mutations.

Ainsi, il existe un biais de répartition des mutations le long du génome, entre les brins directs et indirects, ainsi qu'entre les régions correspondant à l'origine et au terminus de réplication Ori et Ter, respectivement. Les biais de mutations peuvent aussi être observés à plus petite échelle, avec une forte influence des nucléotides avoisinants.

3.3 Biais de sequence

Observation de biais de séquence Suite à l'observation de biais de répartition des mutations, l'influence des bases voisines a été analysée. En étudiant les transitions AT→GC au niveau d'un codon de la région *rII* du phage T4, Koch a mis en évidence une différence d'un facteur 5 à 10 entre la mutagenèse des codons UAG et UAA (Koch 1971). Ces biais sont fréquents. Ainsi, la cytosine est 10 fois plus sujette à mutation lorsqu'elle est suivie d'une guanine (Hess et al.

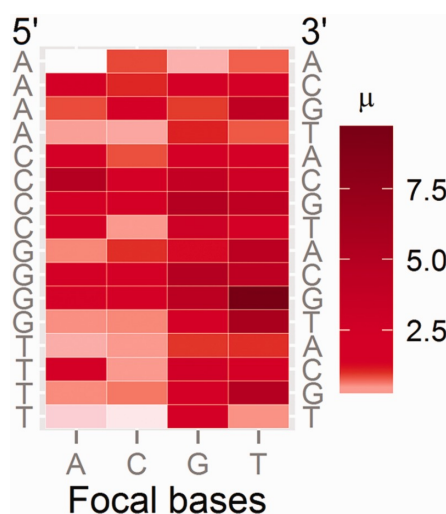


FIGURE 6 : Taux de mutation en fonction du contexte chez *P. fluorescens* au cours d'une expérience d'accumulation de mutations (Long et al. 2015). Dans cette expérience, le triplet le moins touché est AAA, tandis que le plus touché est GTG, avec une augmentation du taux de mutation de 67 entre ces deux triplets.

1994); les mutations de ces dimères CpG sont à l'origine d'un tiers des maladies héréditaires chez l'humain (Cooper et Krawczak 1990; Duret 2009). De plus, une forte mutabilité de la cytosine dans les triplets CCG et TCG a été observée chez *S. cerevisiae* (Zhu et al. 2014). Chez *P. fluorescens*, le triplet GTG, avec le résidu T muté, est le plus touché tandis que AAA, avec le résidu A muté, est le moins touché, avec une différence d'un facteur 67 entre les deux (Figure 6, Long et al. 2015). Cette forte dépendance au contexte nucléotidique est reporté chez *B. subtilis* où les mutations des triplets CAA, TTG, GGC et ATG sont sur-représentées en l'absence de MMR (Sung et al. 2015).

Simulations confirmant le biais de mutations Des simulations ont confirmé le rôle de ces biais dans l'évolution (Hwang et Green 2004; Baele et al. 2010). La première étude a modélisé l'évolution de séquences non exprimées chez les mammifères en permettant un taux de mutation différent en fonction du contexte. Les taux de mutations obtenus sont bien différents d'un contexte à l'autre. Les mutations peuvent ainsi être classées en 14 catégories dans lesquelles les taux de mutations évoluent de façon similaire entre les différentes branches de la phylogénie, certaines s'accumulant de façon constante au cours du temps et entre les différentes branches. De la même manière, les auteurs observent des biais de brin, les transitions de pyrimidines étant plus

fréquentes sur le brin sens. La seconde étude réalise une modélisation similaire chez les primates, mais cette fois en observant un biais du deuxième ordre, avec le rôle des nucléotides se trouvant deux positions avant ou après le résidu muté.

Origine de ces biais Les causes de ces biais mutationnels sont variées : modification de nucléotides favorisant les mutations, erreurs de réplication ou mauvaise reconnaissance par les systèmes de réparation de l'ADN. Un facteur majeur de biais de séquence est attribué à la méthylation des cytosines. En effet, les résidus cytosines méthylés se désaminent spontanément en thymine (voir partie I). Chez *E. coli*, les sites CC*AGG, où C* est un résidu cytosine méthylé, sont des points chauds de mutations. Ce biais n'est pas observé dans la souche B d'*E. coli* qui ne présente pas de méthylation au niveau de cette séquence (Coulondre et al. 1978). Il arrive parfois que les mésappariements soient mal reconnus par les systèmes de réparation et cela peut être influencé par les bases voisines de la mutation. Les liaisons C:G sont plus fortes que les liaisons A:T. Par conséquent, un mésappariement dans un environnement riche en C et G ne provoquera qu'une faible déformation de la double hélice et sera donc peu reconnu. Ainsi, les séquences riches en C:G, comme les séquences répétées de C ou de G, sont des points chauds de mutation. Cette influence de l'environnement est vue plus généralement chez *E. coli*, où la réparation des mésappariements produisant des transversions (A:G ou C:T) dépend de leur position (Su et al. 1988 ; Parker et al. 1992 ; Joseph et al. 2006).

Les études des biais de mutations à l'échelle génomique sont très récentes. Elles sont soit basées sur la phylogénie d'espèces plus ou moins éloignées, soit sur des expériences d'accumulation de mutations, menées sur quelques centaines de générations. Aucune étude exhaustive sur une expérience d'évolution à long terme n'a encore été faite. Grâce au modèle expérimental utilisé dans cette thèse, l'expérience d'évolution la plus longue en cours actuellement dans le monde, nous avons accès à un grand nombre de mutations ainsi qu'à plusieurs populations présentant des phénotypes mutateurs. Il est donc possible d'étudier de manière extensive les biais contextuels et topologiques des mutations. Ce travail est présenté dans le troisième chapitre de cette thèse, qui vise à mieux comprendre les mécanismes et la directionnalité de l'évolution. Tout comme l'étude de la dynamique des taux de mutations, elle s'appuie sur le modèle expérimental d'évolution à long terme développé par Richard Lenski depuis 1988.

4 Modèle expérimental d'évolution à long terme

4.1 Description du modèle

L'évolution expérimentale permet d'aborder l'évolution en contrôlant les paramètres. Parmi les expériences d'évolution à long terme, la plus longue à ce jour a été initiée par Richard Lenski en 1988 (Figure 7, Lenski et al. 1991). A partir d'un ancêtre *E. coli* B REL606 Ara⁻, incapable de consommer l'arabinose, et de son révertant métabolisant l'arabinose (Ara⁺) REL607, 12 populations ont été initiées en parallèle. Ainsi, 6 populations sont Ara⁻ et numérotées de 1 à 6 (Ara-1 à Ara-6) et 6 autres sont Ara⁺ (Ara+1 à Ara+6). Ces populations sont cultivées dans 10 ml de milieu minimum (Davis minimal DM, Carlton et Brown 1981) contenant 25 mg/l de glucose. Les populations sont propagées par repiquages quotidiens de 100 μ l de culture dans 10 ml de milieu frais (dilution 1/100). La limitation en glucose dans le milieu résulte en une croissance en trois phases : une phase de latence (qui tend à disparaître au cours du temps évolutif), une phase de prolifération exponentielle au cours de laquelle le glucose est consommé, et une phase stationnaire quand la ressource nutritive est épuisée. Il s'agit donc d'un milieu saisonnier et limité. Environ 7 générations (6,64 précisément) sont réalisées en 24 h, si bien que les 12 populations évoluent maintenant dans ce milieu depuis plus de 60 000 générations.

Tout au long de l'expérience, des échantillons ont été prélevés et congelés à -80°C toutes les 500 à 1000 générations. L'ensemble de ces échantillons constituent donc une banque d'archives fossiles revivifiables, permettant de revenir dans le passé. Il est ainsi possible de suivre l'évolution de chaque population, voire de rejouer son évolution. De plus, les phénotypes Ara⁺ et Ara⁻ permettent de mettre en compétition ces populations entre elles et avec leur ancêtre, ce qui permet de suivre l'évolution de leur valeur sélective au cours du temps. Depuis l'avènement des technologies de séquençage de génomes, les génomes de clones isolés à différents temps dans ces populations (et de mélanges de populations) ont été séquencés et il est donc possible de retracer la totalité de leur histoire mutationnelle.

L'évolution de ces 12 populations dans ces conditions se caractérise par un parallélisme phénotypique important lors de l'adaptation de ces populations (Figure 8) : le taux de croissance a augmenté de façon similaire dans l'ensemble des 12 populations, les cellules qui les composent sont de plus grande taille et leur valeur sélective a augmenté également par des trajectoires

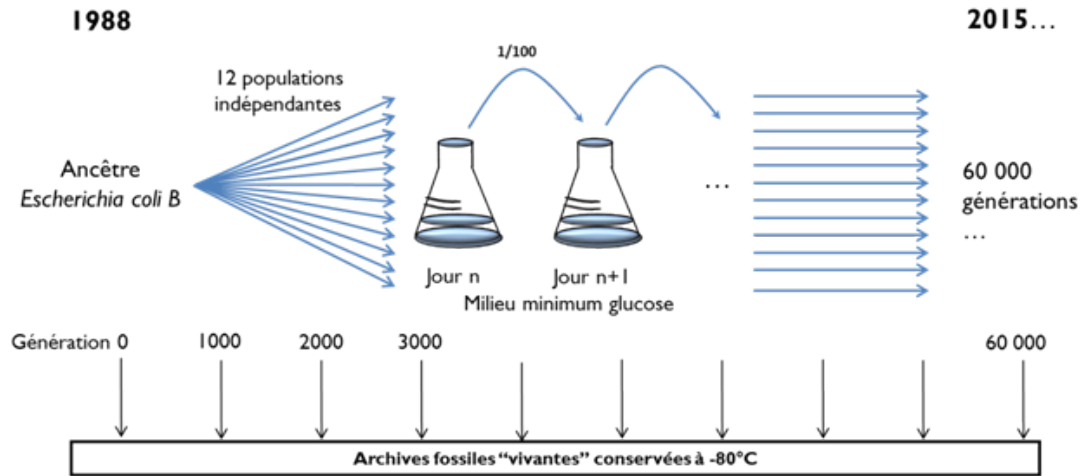


FIGURE 7 : Expérience d'évolution à long terme (Lenski et al. 1991). A partir d'un ancêtre *E. coli* B, 12 populations ont été initiées. Elles sont maintenues par repiquages quotidiens dans un milieu minimum glucose. A raison d'environ 7 générations par jour, l'expérience représente aujourd'hui plus de 60 000 générations.

similaires (Lenski et Travisano 1994; Philippe et al. 2007; Wiser et al. 2013). La moitié des populations ont développé un phénotype hypermutateur. L'augmentation de la valeur sélective est hyperbolique : elle croît fortement au début de l'expérience, au cours des 2000 premières générations, puis ralentit (Lenski et Travisano 1994). Elle est due à l'accumulation de mutations bénéfiques au cours du temps. Ces mutations bénéfiques affectent des gènes de différentes fonctions : *rbs* et *malT* entraînant la perte du métabolisme du ribose et du maltose, *pbpA* impliqué dans la synthèse de la paroi cellulaire, *spoT* impliqué dans la réponse stringente aux carences nutritionnelles, *topA* et *fis* impliqués dans le contrôle de la superhélicité de l'ADN (Cooper et al. 2001; Pelosi et al. 2006; Crozat et al. 2005; Barrick et al. 2009, Philippe et al. 2009; Crozat 2010). Il est intéressant de noter que près de la moitié des mutations affectent des gènes codant des régulateurs de l'expression des gènes (Philippe et al. 2007; Hindré et al. 2012). De plus, les mêmes gènes sont affectés par des mutations dans une majorité des 12 populations, révélant ainsi un fort parallélisme génétique qui pourrait ainsi expliquer le parallélisme phénotypique observé. Les mutations sont cependant différentes au sein des différentes populations, avec une forte divergence allélique. Enfin, les 12 populations présentent de nombreux réarrangements chromosomiques (délétions, duplications, inversions), certaines régions étant affectées de façon répétée

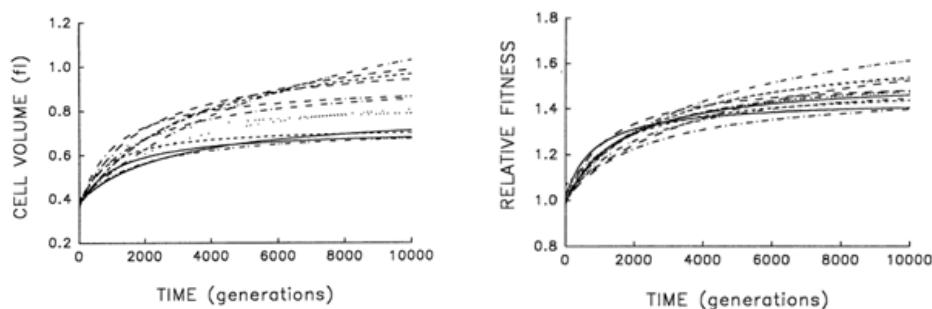


FIGURE 8 : Evolution de la taille cellulaire et de la valeur sélective dans les 12 populations de l'expérience d'évolution à long terme (Lenski et Travisano 1994). La taille cellulaire et la valeur sélective augmentent de façon similaire dans chaque population, illustrant un fort parallélisme au cours de l'évolution.

(Raeside et al. 2014).

Outre ce parallélisme, les populations ont aussi développé des caractéristiques qui leurs sont propres. Par exemple, l'émergence d'une co-existence au long terme de deux écotypes distincts a été mise en évidence dans la population Ara-2 dès 6500 générations, ces deux écotypes se maintenant par la suite (Rozen et Lenski 2000 ; Rozen et al. 2005, 2009 ; Le Gac et al. 2012 ; Plucain et al. 2014). La population Ara+3 a acquis la capacité de métaboliser le citrate présent dans le milieu après plus de 30 000 générations d'évolution et l'acquisition de plusieurs mutations (Blount et al. 2008, 2012).

4.2 Six populations hypermutatrices

Parmi les 12 populations, 6 ont développé un phénotype hypermutateur. Le taux de mutation des 12 populations a été mesuré après 10 000 générations par des tests de fluctuation en utilisant les résistances à l'acide nalidixique et au bactériophage T5 (Sniegowski et al. 1997). Trois populations présentent des taux de mutations supérieurs aux autres : Ara-2, Ara-4 et Ara+3 (Figure 9). Une copie sauvage des gènes du système MMR a permis de compléter le phénotype hypermutateur, ce qui a permis de montrer que la population Ara-2 est hypermutatrice suite à une mutation du gène *mutL* et que les populations Ara+3 et Ara-4 ont une mutation dans les gènes *mutS* et *uvrD*, respectivement (Sniegowski et al. 1997). Par la suite, le phénotype hypermutateur a été détecté dans 3 autres populations.

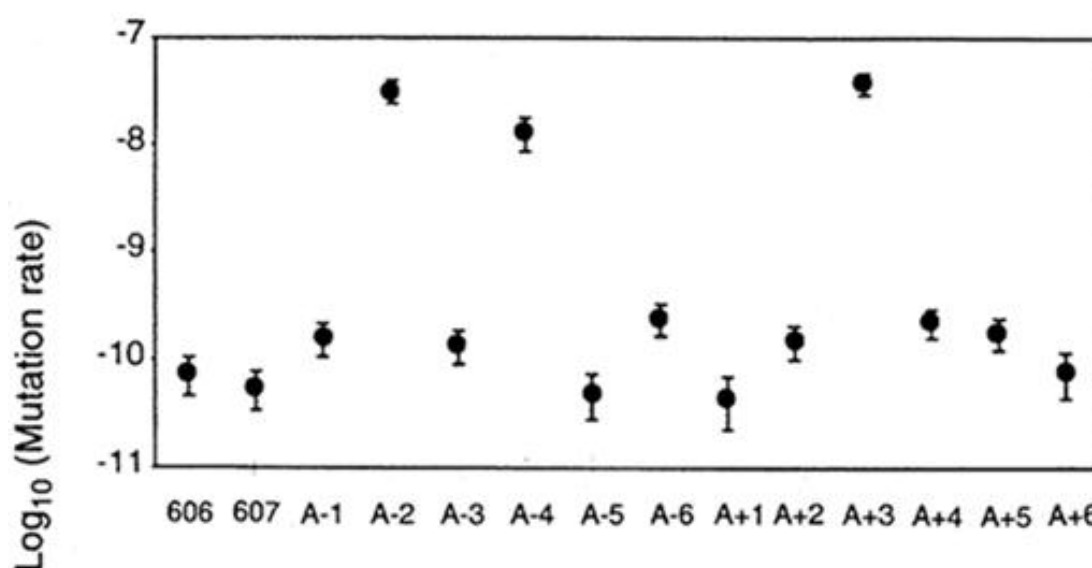


FIGURE 9 : Taux de mutation des 12 populations de l'expérience d'évolution à long terme après 10 000 générations – environ 4 ans (Sniegowski et al. 1997) mesuré par un test de fluctuation. Trois populations ont un taux de mutation supérieur aux autres : Ara-2, Ara-4 et Ara+3.

La population Ara-3, qui a acquis la capacité d'utiliser le citrate comme source de carbone, a développé un phénotype hypermutateur très tardivement, après 30 000 générations. Il résulte d'une mutation dans le gène *mutS* (Blount et al. 2012). Ainsi, 4 populations sur 12 ont un phénotype hypermutateur résultant d'une mutation dans un gène du système MMR. Cependant, aucune de ces populations n'a été étudiée de façon détaillée pour ce caractère hypermutateur. Deux autres populations sont hypermutatrices suite à des mutations dans un autre système mutagène, lié cette fois au 8-oxo-dGTP. Dans la population Ara+1, le phénotype hypermutateur n'émerge qu'entre 25 000 et 27 000 générations suite à une délétion d'un nucléotide dans le gène *mutT*. Enfin, la population Ara+6 est une population détectée comme hypermutatrice grâce au séquençage de génomes. Elle a fait l'objet d'une partie de mon travail de thèse et sera donc détaillée dans la partie Résultats.

Parmi les 6 populations mutatrices, seule Ara+1 a été étudiée pour son aspect mutateur (Wielgoss et al. 2013). Pour la première fois, la dynamique des taux de mutation a pu être étudiée au long terme, notamment suite à l'émergence d'un phénotype hypermutateur. Le taux de mutation de cette population augmente d'un facteur 150 entre 20 000 et 27 000 générations, puis diminue de 55 à 65% après 40 000 générations. Les mécanismes moléculaires à l'origine des

variations de taux de mutation ont été identifiés. L'augmentation du taux de mutation est liée à une mutation du gène *mutT* et se traduit par une augmentation de la proportion des mutations AT→CG. Par la suite, la compensation du taux de mutation élevé se produit au moins deux fois de façon indépendante dans la population, à chaque fois suite à une mutation du gène *mutY*. Les deux mutations de *mutY* sont différentes. Ce parallélisme dans l'évolution pourrait indiquer un avantage sélectif d'un taux de mutation réduit. En effet, le fardeau mutationnel associé aux génotypes *mutT* a été estimé à 0,013, tandis que celui des génotypes compensés grâce aux mutations *mutY* était de 0,0073 et 0,0093, respectivement. Ainsi, le fardeau mutationnel généré par l'augmentation des taux de mutation aurait favorisé la sélection de mutations compensatrices entraînant une diminution des taux de mutation lorsque la population est adaptée à son environnement.

L'augmentation des taux de mutation dans ces populations hypermutatrices s'accompagne d'une adaptation plus rapide que dans les autres populations (Figure 10, Wiser et al. 2013).

Dans cette thèse, je m'intéresse à ces populations hypermutatrices et plus particulièrement à Ara+6. Elle constitue mon modèle pour l'étude de l'évolution des taux de mutation à long terme.

4.3 Objectifs de la thèse et adéquation du modèle d'évolution pour l'étude des taux de mutation et des biais de mutation

Le modèle expérimental d'évolution à long terme choisi pour cette thèse présente de grands avantages. D'une part, il donne accès à l'ancêtre et à l'ensemble des temps évolutifs auxquels des échantillons ont été prélevés et conservés (soit toutes les 500 à 1000 générations). Ces derniers sont revivifiables, et il est donc possible de mesurer la valeur sélective de ces populations, voire de rejouer l'évolution. Les clones conservés sont aussi modifiables génétiquement. Les mutations identifiées peuvent ainsi être soit remplacées par les allèles ancestraux correspondants soit introduites dans le chromosome ancestral par recombinaison, pour créer des souches isogéniques et mesurer ainsi l'effet de ces mutations, seules ou en combinaison, sur différents phénotypes.

D'autre part, nous disposons également de séquences génomiques. La moitié des populations ayant développé un phénotype hypermutateur, le nombre de mutations recensées est élevé (50 000

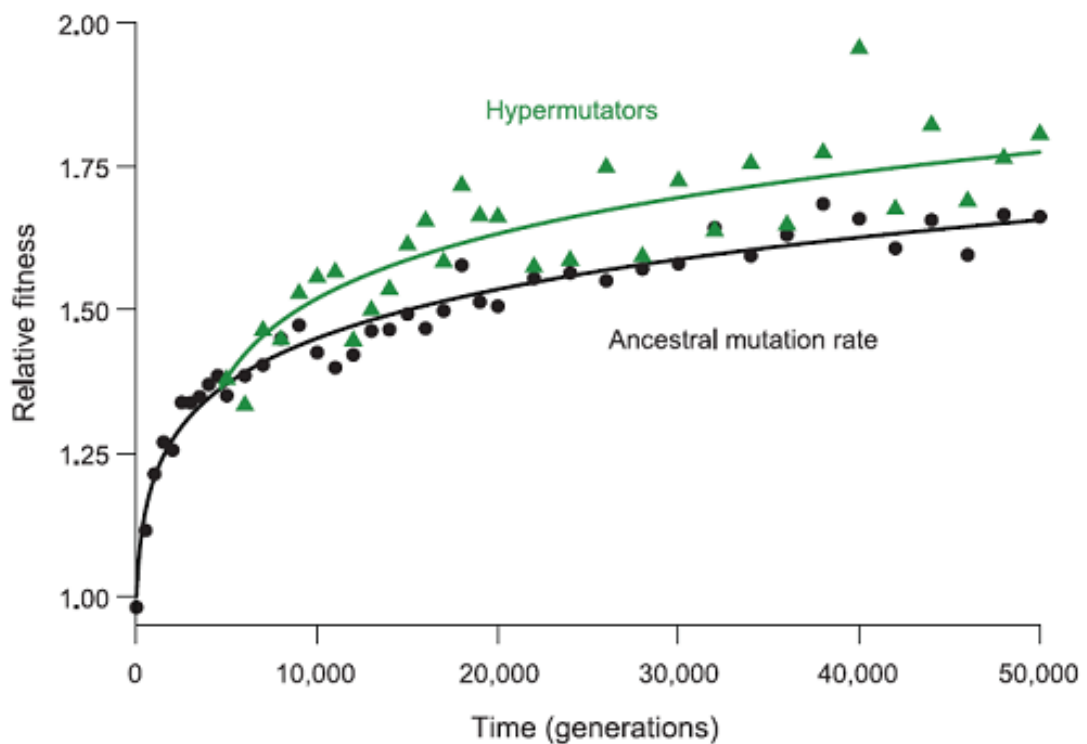


FIGURE 10 : Effet du phénotype hypermutateur sur l'évolution de la valeur sélective (Wiser et al. 2013). Les valeurs sélectives de chaque population ont été mesurées puis moyennées en fonction de leur phénotype mutateur ou sauvage. Les hypermutateurs s'adaptent plus vite que leurs homologues non mutateurs.

au total), et constitue ainsi une banque de données considérable. De plus, ces mutations sont informatives sur l'adaptation des douze populations au cours du temps. Nous pouvons observer les parallélismes et les divergences entre populations, en particulier au niveau des biais de mutations, et ceci en fonction du type de mutateur étudié.

Un autre avantage majeur à ce modèle expérimental est sa durée. Cette expérience d'évolution *in vitro* est la plus longue jamais réalisée, avec plus de 60 000 générations à partir d'un ancêtre commun. Cette durée a permis d'accumuler les mutations, mais aussi des données sur l'adaptation au cours du temps. Malgré la simplicité du milieu utilisé, il a été possible d'observer une grande variété de phénomènes, comme la diversification et la co-existence à long terme, la modification des réseaux métaboliques et de régulation, et les mécanismes d'adaptation de façon générale.

Bien que très étudiée, cette expérience comporte encore de nombreux aspects non explorés. Dans cette thèse, je m'intéresse à la dynamique des taux de mutation et à la nature des mutations. D'une part, j'ai mesuré les taux de mutation dans une des populations hypermutatrices sur le long terme, Ara+6. En effet, j'ai pu montrer que cette population a évolué vers un phénotype mutateur dès 5000 générations, contrairement à la population Ara-1, étudiée précédemment (Wielgoss et al. 2013), dont le phénotype mutateur ne survient qu'après 27 000 générations. Grâce à la durée de l'expérience d'évolution, il a donc été possible de voir évoluer la balance adaptation/stabilité au cours du temps, résultant en une variation du taux de mutation et ce, en particulier pendant une période d'adaptation.

D'autre part, je me suis intéressée à la distribution des mutations dans le génome et aux biais éventuels. Pour cela, j'ai utilisé l'ensemble des mutations recensées dans les douze populations. Suite au phénotype hypermutateur de la moitié des populations, cette banque de données contient plus de 50 000 mutations. Ces mutations ont un spectre particulier en fonction du type de mutateurs (transversions AT→CG pour les mutateurs associés à des mutations du gène *mutT*, transitions pour les mutateurs associés aux mutations du système MMR). Grâce à cette exhaustivité, il m'est possible d'avoir un signal statistique suffisant pour étudier la distribution des mutations dans le génome, et même de voir des différences éventuelles entre les différents types de mutateurs.

Dans la première partie, j'évoquerai mon travail expérimental sur l'évolution des taux de mutation dans la population Ara+6 au travers d'un manuscrit soumis à la revue *Molecular*

Biology and Evolution. Puis la deuxième partie rendra compte de mon travail théorique sur les biais de mutation dans l'expérience d'évolution à long terme, sous forme d'un manuscrit qui sera prochainement soumis à la revue *Molecular Biology and Evolution*.

Chapitre II

Evolution des taux de mutations au cours du temps

1 Préambule

Les taux de mutations résultent d'un équilibre entre stabilité et évolvabilité. Or ces deux forces agissent à deux échelles de temps différentes. Tout d'abord, les allèles hypermutateurs peuvent émerger grâce à leur entraînement génétique (auto-stop) avec une mutation bénéfique (qu'ils génèrent à une fréquence plus élevée). Ainsi, l'évolvabilité peut être sélectionnée indirectement (sélection de second ordre) sur des échelles de temps courtes. Cependant, ces hypermutateurs produisent également continuellement des mutations ayant des effets délétères, nuisant à la stabilité du génome. Ainsi, un fardeau mutationnel s'accumule avec le temps et la sélection pour la stabilité se produit ainsi à des échelles de temps plus longues. Par le passé, cette dynamique a surtout été étudiée de façon théorique, et les études expérimentales se sont focalisées sur la dynamique à court terme des taux de mutations, avec l'émergence de populations hypermutatrices (à quelques exceptions près, notamment dans notre laboratoire). Dans cette partie, nous étudions la dynamique évolutive des taux de mutation de façon expérimentale et au long terme (50 000 générations) dans une des douze populations de l'expérience d'évolution au long terme chez *E. coli*, la population Ara+6. Le caractère hypermutateur de cette population a été révélé par le biais de séquences génomiques (Wielgoss et al. 2011). Les résultats obtenus sont synthétisés dans un manuscrit qui fera l'objet d'une soumission au journal *Molecular Biology and Evolution*.

Dans un premier temps, nous nous sommes intéressés aux bases génétiques du phénotype hypermutateur. Les séquences de génomes de clones isolés dans cette population nous ont permis de mettre en évidence deux mutations : une première dans le gène *mutT*, qui participe à l'élimination du nucléotide mutagène 8-oxo-dGTP (Figure 1 de l'article), et une autre dans le gène *dnaE*, codant la sous-unité α de l'ADN-polymérase III qui est un facteur important dans la fidélité de la réplication de l'ADN. Les mutations trouvées dans les génomes sont majoritairement des transversions AT->CG, qui sont caractéristiques d'un défaut d'activité du gène *mutT* (Table 1 de l'article). De plus, nous avons isolé dans la population à 4000 générations un clone évolué possédant uniquement la mutation du gène *mutT*. Nous avons procédé à un échange allélique pour le gène *mutT* entre la souche ancestrale et un clone évolué, remplaçant l'allèle évolué par sa version ancestrale et réciproquement. Ainsi, nous avons obtenu les clones suivants : un clone ancestral avec l'allèle évolué de *mutT* et l'allèle ancestral de *dnaE*, (*AncmutT^{evo}*) et un clone

évolué avec la version ancestrale du gène *mutT* et l'allèle évolué pour *dnaE* (10K-B*mutT*^{anc}). En comparant leurs taux de mutation avec les clones ancestraux et évolués, ainsi qu'avec le clone évolué isolé à 4000 générations et possédant l'allèle évolué de *mutT* (Figure 3 de l'article), nous avons pu mettre en évidence que le phénotype hypermutateur résultait essentiellement de la mutation du gène *mutT*. Le clone ancestral avec l'allèle évolué de *mutT* et le clone évolué isolé à 4000 générations et possédant l'allèle évolué de *mutT* possèdent toutefois un taux de mutation plus élevé que le clone évolué isolé à 10 000 générations et possédant les allèles évolués des deux gènes *mutT* et *dnaE*. Ces résultats suggèrent donc que la mutation *dnaE* permet une compensation du phénotype hypermutateur.

Après avoir identifié les bases moléculaires du phénotype hypermutateur, nous nous sommes intéressés à l'évolution des taux de mutations au cours du temps, à la fois aux niveaux génomique et phénotypique (Figure 4 de l'article). Au niveau génomique, les variations sont peu visibles. Ces données ont donc été complétées par des modèles dont le plus vraisemblable comporte 5 taux de mutations différents, suggérant une très forte dynamique de ces taux de mutation : une première forte augmentation suivie de deux diminutions successives le long de la branche principale de l'arbre phylogénomique, et deux diminutions sur les branches terminales (Figure 5 de l'article). Ce modèle est conforté par les mesures expérimentales du taux de mutation, au moins pour les deux branches terminales de l'arbre phylogénomique et la deuxième diminution le long de la branche principale entre 30 000 et 40 000 générations. Ainsi, plusieurs événements de compensation se sont produits au cours de l'évolution de cette population.

Nous avons étudié les bases moléculaires de ces compensations dans les clones évolués isolés à 20 000 et 40 000 générations, le gène candidat le plus probable étant *mutY*, voire *mutM*, car ils appartiennent tous les deux au système de réparation des lésions 8-oxo-G (Figure 1 de l'article). Deux mutations du gène *mutY* ont en effet été trouvées dans les clones évolués 20K-A et 40K-A, le taux de mutation du clone 20K-A pouvant être à nouveau augmenté par l'expression d'une copie non mutée du gène *mutY* clonée sur un plasmide (voir Supplementary Data de l'article). Cependant, aucune mutation de *mutY* ou *mutM* n'a été retrouvée dans les clones 20K-B et 40K-B, ayant pourtant tous deux un taux de mutation réduit. D'autres voies ont été envisagées, comme la réduction du taux d'oxydation ou l'excrétion de dérivés oxydés, mais aucune ne semble être impliquée. La compensation du phénotype hypermutateur dans ces clones résulte donc d'un

mécanisme inconnu pour l'instant.

Les résultats de ces travaux apportent une meilleure compréhension de l'évolution de génotypes hypermutateurs à long terme, qui, avant cette étude, n'avait jamais été suivie sur une aussi longue période (Raynes et Sniegowski 2014, Wielgoss et al. 2013, McDonald et al. 2013). Nous avons observé une diminution des taux de mutations, initialement prévisible avec l'augmentation du fardeau mutationnel. Cependant, cette diminution n'est pas aussi directe qu'attendue, contrairement à ce qui avait été observé dans une autre des douze populations de l'expérience d'évolution à long terme, la population Ara-1 (Wielgoss et al. 2013). Ici, la diminution des taux de mutation s'est produite en plusieurs étapes, par une succession d'essais/erreurs, avec différentes mutations du gène *mutY* dont certaines n'ont pas été fixées au sein de la population. Nous avons également pu mettre en évidence l'existence de mécanismes alternatifs à ceux décrits dans la littérature pour compenser le phénotype hypermutateur associé à une mutation du gène *mutT*. Ces mécanismes pourront être étudiés plus en détail afin de mettre en évidence de nouveaux mécanismes participant à la fidélité de la réplication de l'ADN et au maintien du génome.

2 Article 1

Article

Discoveries

Long-Term Dynamics of Mutation Rates in a Hypermutator Population of *Escherichia coli*

Larissa Caudwell,^{1,2} Olivier Tenaillon,^{3,4} Sebastien Wielgoss,⁵ Sylvain Caillat,^{6,7} Jean-Luc Ravanat,^{6,7} Richard E. Lenski,^{8,9} and Dominique Schneider^{*,1,2}

¹Univ. Grenoble Alpes, Laboratoire Adaptation et Pathogénie des Microorganismes (LAPM), F-38000 Grenoble, France

²Centre National de Recherche Scientifique (CNRS), LAPM, F-38000 Grenoble, France

³IAME, Unité Mixte de Recherche (UMR) 1137, INSERM, F-75018 Paris, France

⁴IAME, UMR 1137, Université Paris Diderot, Sorbonne Paris Cité, F-75018 Paris, France

⁵Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland.

⁶Univ. Grenoble Alpes, Laboratoire Lésions des Acides Nucléiques, CEA, INAC-SCIB, Grenoble, France

⁷CEA, INAC-SCIB, Grenoble, France

⁸Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, 48824, USA.

⁹BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824, USA

***Corresponding author:** E-mail: dominique.schneider@ujf-grenoble.fr

2.1 Abstract

Mutations are the ultimate source of variation in evolution. However, while they contribute to adaptation, they may also disrupt functional genes and decrease genome stability. Mutation rates therefore result from a tension between stability and evolvability. Hypermutator genotypes have been observed in both clinical and natural isolates. They result from mutations in genes involved in DNA-repair systems and theory has shown that they arise in populations by hitchhiking with beneficial mutations. However, it is also expected that mutators will lose advantage when populations are adapted to their environment, owing to the genetic load of deleterious mutation accumulation. The evolution of hypermutators has been widely studied theoretically. Experimentally, most studies have focused on the rise of mutators, but only few investigated the fate of hypermutators. In this paper, we monitored by genetic, molecular and genomic approaches an *Escherichia coli* hypermutator population in a long-term evolution experiment over a 50,000-generation time period. The hypermutator phenotype emerged in the population before 5,000 generations and the mutation rate started to decrease after 20,000 generations. In contrast to previous observations in other populations, the mutation rate did not decrease in a single event, but by successive steps suggesting that it is a selectable trait that can be fine-tuned over long evolutionary time during fast and slower adaptive periods.

2.2 Introduction

Microorganisms have colonized almost all ecological niches on Earth, including environments with diverse temperature, salt, O₂, pH and osmotic levels. Moreover, the last decades have witnessed bacterial adaptation to a large set of human selective pressures, such as antibiotics, pollutants as well as diverse controlled laboratory conditions. Therefore, the evolutionary success of bacteria is not only related to their billions of years of existence, but also to their enormous capacity to adapt to new environments, in other terms to their evolvability. Mutations being the ultimate source of genetic variability, mutation rates and their evolution constitute an important factor contributing to bacterial evolvability.

Mutations arise from DNA damage or replication errors when not repaired. They may have various effects on organismal fitness. In a given environment, most mutations are either neutral,

with no measurable effect on fitness, or deleterious thereby reducing fitness (Fisher 1930). Yet in all environments, a small proportion of mutations are beneficial leading to higher fitness and allowing adaptation (Eyre-Walker and Keightley 2007). Mutations are therefore the double-edged sword of evolution: they can alter or improve the organism. The mutation rate of a given organism, *i.e.* the number of mutations per cell and generation (or time), results from a tension between evolvability and stability. Understanding how the mutation rate evolves stands as a keystone to investigate evolutionary mechanisms and trajectories (Raynes and Sniegowski 2014).

In order to preserve their genome integrity, living organisms have developed proofreading enzymes and DNA repair systems thereby reducing mutational events (Friedberg et al. 1995; Miller and Perrino 1996). The mutation rate has been observed to be maintained at a constant level in DNA-based microorganisms (Drake 1991). However, selection, in particular in fluctuating environments, can favor the emergence of high mutation rate alleles in genes involved in DNA-repair or replication fidelity resulting in hypermutator strains. The hypermutator state is characterized by an increase of 10- to 100-fold of the mutation rate. Bacterial hypermutators have been found in both environmental and clinical settings (Tröbner and Piechocki 1984; LeClerc et al. 1996; Matic et al. 1997; Oliver et al. 2000; Denamur et al. 2002). Invasion of hypermutators has also been observed when bacteria are propagated during laboratory evolution experiments (Sniegowski et al. 1997; Giraud et al. 2001; Notley-McRobb et al. 2002; Wielgoss et al. 2013).

In the absence of recombination, mutations remain fixed to the genetic background in which they occurred. Consequently, hypermutator alleles can invade the population by hitchhiking with the beneficial mutations they have produced (Chao and Cox 1983; Tenaillon et al. 2000), thereby opening a period of increased mutation rates that may allow faster accumulation of beneficial mutations (Tenaillon et al. 1999; Travis and Travis 2002; Wiser 2013). However, the number of deleterious mutations, which are known to be predominant (Sturtevant 1937), also increases leading to a higher mutational load. This tension between adaptation and genetic load therefore raises questions about the evolutionary dynamics of mutation rates, especially as a function of the environmental conditions.

Theoretical models commonly predict that mutation rate cannot be readily optimized (André and Godelle 2006). However, when the supply of beneficial mutations of substantial fitness effect

is large, all models agree that selection favors an increase in mutation rate, even if the deleterious mutations are predominant. This is related to the short time associated with overproduction of beneficial mutations compared to the longer term cost of the increased accumulation of deleterious mutations (Sniegowski and Gerrish 2010; Desai and Fisher 2011). However, once the organisms are well adapted to their environment, models predict that a mutation rate decrease should be favored (Taddei et al. 1997; Orr 2000; André and Godelle 2006; Clune et al. 2008). Models have nevertheless mostly focused on the initial rise of mutator alleles, and barely on the dynamics of mutation rate decay. Evidences of mutation rate fluctuation exist in nature. Indeed, signatures of past inactivation and re-functionalization through horizontal gene transfer of mismatch repair genes have been detected in their nucleotide sequences (Denamur et al. 2000). However, in asexual populations propagated in the laboratory, such gene transfer cannot occur. Yet hypermutator selection followed by compensatory events has been observed (Tröbner and Piechocki 1984; McDonald et al. 2012; Wielgoss et al. 2013). In most cases, compensation was moderate and the dynamics of mutation rate decay was not well documented.

In a previous study, we documented the rise and compensation of mutation rate in one population of the long term evolution experiment (LTEE). During the LTEE, twelve populations were initiated from a common ancestor strain of *Escherichia coli* B in 1988 and have been propagated ever since in a glucose-limited minimal medium by daily 1:100 dilution transfers, representing more than 60,000 generations of evolution (Lenski et al. 1991; Wiser et al. 2013). Samples have been collected and frozen at regular time intervals providing a revivable fossil record. Six of the twelve populations evolved a hypermutator phenotype, including four early during the experiment well before 10,000 generations and two much later, after 20,000 generations. We previously investigated one population, called Ara-1, for its mutation rate at the genomic level (Wielgoss et al. 2013). The Ara-1 population evolved a hypermutator phenotype after 20,000 generations owing to the substitution of a mutation in *mutT* encoding an enzyme that purges the intracellular oxidized guanine nucleotides (8-oxo-dGTP). The increase in mutation rate was ~100-fold. After the emergence of this hypermutator, at least two independent lineages with an antimutator phenotype swept to fixation in less than 10,000 generations. Each of these two antimutator lineages had a two-fold reduced mutation rate owing to mutations in *mutY* encoding a DNA repair glycosylase. Genomic analyses showed that these antimutators were selected for

their reduced genetic load at an evolutionary time where the bacterial population was already well adapted to its environment.

In this previous study, only one population was investigated, and it was characterized by both the late emergence of the hypermutator genotype and a single mechanism of compensation that left the overall mutation rate fairly high (10-fold). Moreover, no isogenic strains, except for the DNA repair genes, were constructed in that previous study. Here, we investigated the dynamics of mutation rates in another population of the LTEE, called Ara+6, that evolved a hypermutator phenotype much earlier during evolution, before 6000 generations. Moreover, we used both genomic and genetic approaches. Compared to the Ara-1 population, an even higher dynamics of mutation rates was detected, with mutation rate changes occurring even after more than 50,000 generations of evolution. The early occurrence of a hypermutator allele resulted therefore in a higher and more complex dynamics of mutation rates.

2.3 Results

Ara+6 is a *mutT* Hypermutator Population We sequenced the genome of 24 evolved clones that were sampled from population Ara+6 at different time points during evolution. While less than 5 mutations per genome were detected in the evolved clones up to 5000 generations, more than 250 mutations were found in the 6000-generation clones (Table 1). Moreover, synonymous changes were detected only in the evolved clones sampled from 6000 generations. Almost 99% of all mutations were AT→CG changes. Therefore, the population Ara+6 evolved a hypermutator phenotype before 6000 generations with a highly skewed mutational spectrum for AT→CG mutations that are diagnostic of a defect in *mutT*.

We indeed found a frameshift in *mutT* in all evolved clones sampled from generation 6000 onward that resulted in a truncated protein of 80 aminoacids instead of 129. The *mutT* gene encodes an 8-oxo-dGTP diphosphatase which prevents the integration of the mutagenic 8-oxo-dGTP into DNA (Fig. 1; Maki and Sekiguchi 1992). The 8-oxo-dGTP, an oxidized form of dGTP, can pair equally well with either A or G, thereby potentially leading to AT->GC mutations. We also found mutations in other genes involved in the 8-oxo7,8-dihydroguanine (8-oxoG) repair pathway (Fig. 2). First, a mutation in *dnaE* (leading to the M1006L change) which encodes the

α -subunit of DNA polymerase III was found in all clones having the *mutT* mutation. Interactions between mutations in *mutT* and *dnaE* have been reported previously with some *dnaE* alleles being able to either compensate the *mutT* mutator phenotype or further increase the mutation rate (Schaaper 1996; Yamada et al. 2012). Second, after 20,000 generations, three different mutations were observed in *mutY* (leading to the Y323*, Y82D, and C208W changes) and one in *mutM* (leading to the N139H change) encoding the MutY and MutM enzymes, respectively (Fig. 1, Michaels et al. 1992; Fowler et al. 2003). Whereas MutT hydrolyzes the modified nucleotide, MutY removes the mispaired A in front of the integrated 8-oxoG and MutM is part of the base-excision repair pathway removing the 8-oxoguanine from DNA. It has been shown that high mutation rates associated with a *mutT* mutation can be compensated by a mutation in *mutY* (Fowler et al. 2003), while a triple mutant with changes in *mutT*, *mutY* and *mutM* experienced a shift of mutational pattern from AT→CG to CG→AT with no reduction of the overall mutation rates.

Involvement of *mutT* and *dnaE* in the Hypermutator Phenotype All evolved clones that we studied had either none or both the *mutT* and *dnaE* mutations. We used three different approaches to discriminate the effect of each of these two mutations on the mutation rate. First, we screened 96, 93 and 75 evolved clones sampled after 4000, 5000 and 6000 generations, respectively, for the presence of each of the two mutations to investigate their order of appearance. No mutation was found in the 96 clones sampled at 4000 generations and in 60 clones sampled at 5000 generations, while both mutations were detected in both the remaining 33 clones from 5000 generations and all 75 clones from 6000 generations. Therefore, both mutations invaded the population between 4000 and 6000 generations and we were not able to discriminate their order of appearance.

Second, despite the absence of detection of hypermutator clones at 4000 generations, we hypothesized that a minor hypermutator lineage may already exist at that time that will harbor only one of the two *mutT* and *dnaE* mutations. We therefore tried to enrich this minority. Hypermutators produce more mutations and thus develop a higher rate of resistance to antibiotics like fosfomycin and rifampicin that can be acquired through point mutations in the *fosA*, *fosB* or *fosX* genes and *rpoB* gene, respectively. To enrich for mutators, we selected from the mixed

population sampled from Ara+6 at 4000 generations clones that were resistant to both fosfomycin and rifampicin. We then screened these clones by PCR for their *mutT* and *dnaE* alleles. Of 24 antibiotic-resistant clones, one had the evolved *mutT* allele alone while retaining the *dnaE* ancestral allele, suggesting that the evolved *mutT* allele occurred before the evolved *dnaE* allele. This clone was named 4K17.

Third, we tried to construct isogenic strains except for the *mutT* and/or *dnaE* alleles. We both moved the *mutT* evolved allele in the ancestral background (*AncmutT^{evo}*) and replaced the *mutT* evolved allele by its ancestral counterpart in the 10K-B evolved clone (*10K-BmutT^{anc}*). We were however unable to perform any allelic exchange for *dnaE* which is an essential gene and therefore less amenable to genetic manipulation than *mutT*. We then performed Luria-Delbrück fluctuation tests to compare the mutation rates of the above isolated and constructed clones (fig. 3). The *AncmutT^{evo}* isogenic construct displayed a hypermutator phenotype, with a mutation rate 100-fold higher than the ancestral strain. Conversely, the *10K-BmutT^{anc}* strain displayed a reduced mutation rate, comparable to that of the ancestor. These results strongly suggest that the *mutT* mutation is responsible for the first rise of mutation rates in the Ara+6 population, resulting in the hypermutator phenotype. Moreover, two lines of evidence suggest that an early and fast compensation of mutation rate occurred. First, the mutation rate of the evolved clone 10K-B is 4-fold lower than that of *AncmutT^{evo}*. Second, the 4K17 clone sampled at 4000 generations and harboring only the *mutT* evolved allele had also a higher mutation rate compared to 10K-B which has both the *mutT* and *dnaE* mutations. Therefore, after the first rise of the mutation rate, associated with the *mutT* mutation, a small and fast 4-fold compensatory reduction occurred. The *dnaE* mutation is likely contributing to that decay.

After the substitution of the two *mutT* and *dnaE* evolved alleles, the phylogenetic tree revealed that the mutation rate remained high over 50,000 generations and that additional mutations in the 8-oxoguanine repair pathway occurred (Fig. 1). We therefore investigated the dynamics of mutation rates in the Ara+6 population over time at both the genomic and phenotypic levels.

Dynamics of Genomic Mutation Rates over Time The genomic mutation rate was computed for each of the evolved mutator clones (Table 1 and Fig. 4A) by the ratio between the

difference in the number of synonymous mutations found in genomes from two consecutive time points, and the generational time difference between these genomes. This computation assumes that the rate was constant since the last common ancestor of the sampled clones. The genomic mutation rate confirmed the hypermutator state from generation 5000. Later on, the evolved clone 20K-A, and the two sampled at 30K, 40K and 50K revealed lower genomic mutation rates, with a further reduction for 40K-B and 50K-B. On average, an overall decrease in mutation rates may have occurred between 20,000 and 30,000 generations.

To rigorously quantify these differences, we compared the likelihood of models with different mutation rates. We used only AT→GC synonymous mutations and the phylogeny (Fig. 2) to compute the likelihood of different changes in mutation rates. All branching times were optimized in a Monte Carlo Metropolis Hasting Markov Chain, in which the various mutation rates were computed according to branch length and mutation number (supplementary table S1, Supplementary Material online). We rejected a one-rate model of log likelihood = -121.8 and showed that a 5-rate model was much better with a mean log likelihood of -69.65. This 5-rate model assumed an initial genome wide AT→GC synonymous mutation rate of 8.1 per 1000 generations (95% confidence interval 7.5-9.0). It decreased to 3.4 (95% confidence interval 3.4-4.2) in the 20K-A clone, to 4.2 (95% confidence interval 3.8-4.9) in 30K-A, 30K-B and 40K-A, to 1.5 (95% confidence interval 1.2-1.8) in 40K-B, and to 1.0 (95% confidence interval 0.8-1.2) in 40K-A and 50K-B (fig. 5). The reduced mutation rates in the evolved clones 20K-A, 40K-A, 50K-A and 50K-B may be explained by the presence of *mutY* mutations (fig. 2) which is further confirmed by an increase of the *mutY* pattern of mutations (CG→AT, table 1). However, the reduced mutation rates in clones 30K-A and -B could not be explained by obvious genomic changes. Intriguingly, the model does not support a mutation rate increase in the 50K-B evolved clone that had a *mutM* mutation in addition to *mutT* and *mutY* (the model marginally supports an actual decrease log likelihood = -67.7), suggesting that these mutations may not alter the protein functions.

Dynamics of Experimental Mutation Rates over Time The genomic estimation of mutation rates provides a history of mutation rate evolution on the phylogeny of the sampled clone. However, we may have missed mutation rate changes first at the population level as we sampled

only two clones per time point and second at the individual clone level if mutator alleles were substituted at a time point very close to the sampling time. We therefore also experimentally measured mutation rates by fluctuation tests, using both the mixed samples (fig. 4B) and the individual clones (fig. 4C) collected from population Ara+6 over time. As for the genomic estimation, a global reduction was detected after the initial rise of mutation rates, yet differences were also observed.

First, while both genomic and experimental measures were consistent on reduced mutation rates in 20K-A that had a *mutY* mutation, the mutation rate of 20K-B was also lower by fluctuation tests compared to the genomic estimate. Therefore, a compensation event of unknown nature may have happened just before 20,000 generations. Second, whereas the model predicts a mutation rate reduction at 30,000 generations, which is supported by the fluctuation tests using mixed populations, both evolved clones 30K-A and -B showed a mutation rate as high as the hypermutator clone 10K-B. While mutations rates of individual clones and phylogenetic analysis suggested the recurrent emergence of evolved clones with lower mutation rates, the population-wide mutation rate suggested otherwise with a high dynamics of mutation rates including consecutive ups and downs. Hence, decreases were observed at both 20,000 and 40,000 generations, followed by higher mutation rates at 30,000 and 50,000 generations (Fig. 4B).

Mechanisms of Mutation Rate Compensation Compensation of high mutation rates associated with alteration of *mutT* has been shown to target either DNA polymerase III encoded by *dnaE* to increase its fidelity or MutY encoded by *mutY* which removes an adenine mispaired with an 8-oxoguanine (Fowler and Shaper 1997). When the removed A is located on the matrix strand, MutY activity is mutagenic, producing AT→CG mutations. Inactivating *mutY*, when *mutT* is impaired, thereby reduces mutation rates.

All evolved clones bearing a *mutY* mutation (20K-A, 40K-A, 50K-A and 50K-B) had a high proportion of CG→AT mutations, which is diagnostic of *mutY* alterations (supplementary table S1, Supplementary Material online). This trait is specific to these clones. We overexpressed a functional *mutY* gene in 20K-A, which has a Y323* *mutY* mutation. This overexpression resulted in an increase of the mutation rate, thereby confirming that the *mutY* mutation was responsible for the compensation (supplementary methods and Fig. S1, Supplementary Material

online). No compensation was observed in other clones due to fluctuating growth during the test, which is quite common in late evolved clones since they are specialized to the glucose evolution environment. Moreover, growth difficulties have already been reported in population Ara+6 (Wiser et al. 2013).

The genome sequences of the evolved clones revealed that mutations in *mutY* were not the only genetic way to compensate the *mutT* hypermutator phenotype (Fig. 2). Indeed, both the 20K-B and 40K-B evolved clones had reduced mutation rates without carrying any mutation in either *mutY* or *mutM*. It has been recently shown that another way of compensating high *mutT* mutation rates is to decrease oxidative stress in the cells by overexpressing efflux pumps (Guelfo et al. 2010). Therefore, we measured the level of oxidative stress by flow cytometry in the evolved clones 40K-A and 40K-B, bearing or not a *mutY* mutation respectively, and in the hypermutator 30K-A and 30K-B clones as references (supplementary methods and Fig. S2, Supplementary Material online). We used the dihydro-rhodamine 123 fluorescent dye that becomes fluorescent when oxidized. The 40K evolved clones did not show lower fluorescence than the 30K hypermutator clones, precluding overexpression of efflux pumps as a compensatory mechanism for the decreased mutation rates observed in 40K-A and 40K-B. Further study will be needed to identify and characterize the compensation mechanisms in the two evolved clones 20K-B and 40K-B.

2.4 Discussion

The Ara+6 population of the LTEE evolved a hypermutator phenotype very early during evolution, thereby providing a unique opportunity to study long-term dynamics in mutation rates both at the genomic and experimental levels. We observed a very high dynamics of mutation rates, despite long-term evolution (>50,000 generations) in the same medium. First, a hypermutator clone emerged before 4000 generations, after a mutation in the *mutT* gene, which was characterized by a 50-fold increase in the mutation rate compared to the ancestor. This high mutation rate quickly decreased. Indeed, at 5000 generations, the *mutT* mutation was found to be linked with a mutation in *dnaE*, which presumably contributed to this early mutation rate compensation. At 6000 generations, this clone swept into the population. In another population,

called Ara+1, a compensation was also observed within a few thousands of generations after the fixation of a *mutT* hypermutator (Wielgoss et al. 2013). Yet, in Ara+1, compensation occurred through the inactivation of *mutY*. In Ara+6, *mutY* mutations occurred more than 15,000 to 30,000 generations after the *mutT* mutator invaded the population. The first compensation event by *dnaE* may have been sufficient to limit the genetic load during the early stages of the evolution experiment.

After this first compensation, at least four additional ones occurred over evolutionary time, which significantly reduced mutation rates. Two compensation events involved *mutY* mutations in the evolved clones 20K-A and 40K-A, reminiscent to what was found in Ara+1. The compensation that occurred in 40K-B involved an unknown mechanism. Indeed, we found no mutation in *mutY* and *mutM* (Fowler and Shaaper 1997) and efflux pumps seemed not to be involved since no reduction of oxidation was observed (Guelfo et al. 2012). Another compensation event has also occurred in 20K-B and has to be investigated. In contrast to *mutY* compensations in which the mutational spectrum changes from AT→CG to CG→AT, no biases were observed in the mutational spectrum of these two clones. By simulations, we inferred mutation rate changes at 30,000 generations. In the genome sequences of the evolved clones 30K-A and 30K-B, we found mutations in *dnaQ*, *mutS* and *mutL*, three genes that are known to be involved in the control of mutation rates (Li 2008). These mutations were substituted in all later clones. These mutations may provide a potential way for mutation rate compensation without shifting the mutational bias. More detailed investigation is however needed to identify the various compensation pathways for *mutT* mutations. Finally, it is interesting to note that these late compensation events, which reduced the mutation rate up to 5-fold, only happened more than 15,000 generations after the *mutT* inactivation, suggesting that the decreased mutation rate strategy did not provide a high enough beneficial effect or that it did not succeed to hitchhike to fixation with some other beneficial mutations. Yet, the absence of change in mutational bias for some compensation events may argue for a true selection for decreased mutation rates. Indeed, when *mutY* compensation events are selected for, two mechanisms may be involved, including counter-selection of high mutation rates to decrease the genetic load and selection for increased mutation rates on another type of mutations, both hypotheses being not mutually exclusive. Here, the presence of clones with reduced mutation rates without any change in mutational bias suggests a strong selective

pressure to decrease the genetic load.

At the entire population level, we found that the mutation rate was not only decreasing after the first rise in mutation rate but rather revealed waves of fluctuations. Two decreases at 20,000 and 40,000 generations separated by two increases at 30,000 and 50,000 generations were detected. Mutation rates eventually reached an intermediate value at 50,000 generations, which is 7 fold lower than the initial mutation rate of the *mutT* hypermutator but 6 fold higher than the ancestral one. These waves contrasted with the Monte Carlo simulation that determined at least 4 mutation rate decays. (It may even have changed more often since we did not have access to all extinct branches). Indeed, the phylogenetic tree, based on 2 evolved clones at each time point, gives no information on the prevalence of members of the different branches present at a given time point and possibly differing in their mutation rates. The non-linear decrease of mutation rates in a hypermutator population challenges previous models stating that they would drop and stabilize to the lowest possible level (Leigh 1970; Johnson 1999). If that case, the compensated clones would have swept to fixation and settled into the population as early as 20,000 generations. Instead, population-wide increases in mutation rates at 30,000 and 50,000 generations respectively, suggest that mutators outcompete these compensated clones, presumably as they produced beneficial mutations at a faster rate than the compensated clones.

Nevertheless, despite these fluctuations, the mutation rate decreased progressively in the population. This decrease has to be paralleled with the decreased in the rate of adaptation observed through time. The fitness trajectories measured in the long-term evolution experiment show that adaptation slows down very quickly after 2500 generations (Lenski and Travisano 1994; Barrick et al. 2009), although the fitness trajectory of population Ara+6 itself could not be determined (Wiser et al. 2013). Mutator populations also tended to adapt faster than non-mutator populations. Therefore, we may legitimately speculate that the Ara+6 population adapted very quickly to its stable environment. So, along with the exhaustion of beneficial mutations of large effects, the mutation rate slowly decreases. However, since some adaptive mutations are still possible, mutator can still sweep into the population. Therefore, it seems that the decrease of mutation rate is not straightforward but happens by successive trials and errors. Such a dynamic was not observed in models because either the distribution of beneficial mutations was kept constant or only a small subset of large effect mutations was considered. In

the former case (Gerrish et al. 2007; Söderberg et al. 2011), the benefit of high mutation rate was constantly high, and consequently, the mutation rate could increase and reach intolerable values despite antimutator mutations. In the latter case, mutator selection was favored until all mutations were fixed and then only a decrease in mutation rate was selected for (Taddei et al. 1997).

2.5 Conclusion

In the hypermutator population Ara+6, the mutation rate decreased by consecutive steps through a trial and error process. Several known and unknown molecular mechanisms were involved in the reduction and allowed a strain inactivated for *mutT* to have its mutation rate reduced by more than 16-fold. These steps presumably occurred to reduce the genetic load imposed by a high mutation rate that becomes too strong when the flux of beneficial mutations of high effect becomes limiting, *i.e.* when the population is highly adapted to its environment. This dynamics suggests that mutation rate, and consequently evolvability, can be fine-tuned as a continuous trait, according to the needs for adaptation. Even if we could investigate the mutation rate dynamics over 50,000 generations of evolution, it still remains to be determined whether a balance will ever be reached over even longer time frames.

2.6 Materials and Methods

Long-Term Evolution Experiment and Strains Twelve populations were initiated by a common ancestor *Escherichia coli* B REL606 and propagated by 1/100 serial daily transfers in a glucose-limited medium 25mg/l, DM25; Lenski et al. 1991). Mixed population samples, from which individual evolved clones can be collected, have been frozen at 500-generation intervals. All strains used in this study were derived from one of the twelve populations of the long-term evolution experiment, called Ara+6 (table 1). We used mixed population samples from 4000, 5000, 6000, 10,000, 20,000, 30,000, 40,000 and 50,000 generations. Two evolved clones were sampled at different time points and named after the time at which they were sampled. For example, 20K-A is an evolved clone that was sampled at 20,000 generations. The *E. coli* strain JM109 (Yanisch-Perron et al. 1985) was used for cloning experiments.

Plasmids, Primers and Media The pKO₃ suicide-plasmid was used for allelic exchange (Link et al. 1997). The pQE-80L plasmid (Qiagen) was used for overexpression of the MutY protein. The pCRII-Topo plasmid (Invitrogen) was used for cloning experiments. Primers used in this study are listed in table 2. Strains were grown in LB and DM supplemented with different concentrations of glucose (25mg/l for DM25, 250mg/l for DM250, 1g/l for DM1000, 2g/l for DM2000, and 4g/l for DM4000). Strains were plated on solid DM4000 medium supplemented with 100g/l agar. When necessary, rifampicin (DM-Rif) and ampicillin were added at 100mg/l each, chloramphenicol at 30mg/l, and fosfomycin at 50mg/l.

Isogenic Strain Construction and PCR-Screening The Anc *mutT*^{Tevo} and 10K-B*mutT*^{anc} strains were reconstructed by allelic exchange using the pKO₃ plasmid (Link et al. 1997). Briefly, the *mutT* ancestral and evolved alleles were PCR-amplified as ~1000bp products containing ~500bp of adjacent sequences on each side of the mutation position. The PCR products were first cloned into pCRII-Topo and then sub-cloned into pKO₃ digested by BamHI and NotI. After electro-transformation of the relevant strain (ancestor or 10K-B) with the appropriate plasmid, chloramphenicol-resistant cells with a copy of this non-replicative plasmid integrated into the chromosome were selected at 43°C. Subsequent plating on 5% sucrose-LB plates without NaCl (Blomfield et al. 1991) allowed the selection of plasmid loss, because pKO₃ carries *sacB*, which renders cells sensitive to killing by sucrose. After selection, clones were screened by ARMS-PCR (Amplification refractory mutation system PCR; Newton et al. 1989) to score for the presence or absence of the *mutT* mutation. Positive clones were then checked by PCR and sequencing for both *mutT* and *dnaE*. The same PCR screening was also applied to isolate the 4K17 clone at 4000 generations (see Results).

Fluctuation Tests Fluctuation tests were performed at least three times independently for each strain or population sample as described at <http://barricklab.org/twiki/bin/view/Lab>. Thirty replicates for each strain were grown each in 10ml DM2000 during 24h at 37°C with shaking at 120rpm. A total of 24 were then plated onto selective DM-Rif plates while the 6 others were diluted 10⁶-fold and plated on DM4000 to compute the total cell number. One batch containing Anc*mutT*^{evol}, 4K17 and 10K-B was grown in DM1000 to reduce density on

selective plates. Mutation rates were then estimated using the Fluctuation Analysis Calculator (FALCOR, Hall et al. 2009) with the Ma-Sandri-Sarkar Likelihood Estimation method (Ma et al. 1992).

Selection of an Early Mutator A sample of the population at 4000 generations was first plated on fosfomycin-containing LB plates and incubated overnight at 37°C. Resistant clones were then streaked on two LB plates containing each either rifampicin or fosfomycin and incubated at 37°C for 24h. A total of 24 clones which grew on both plates were conserved as glycerol suspensions at -80°C. Their *mutT* and *dnaE* allele were screened by ARMS-PCR. One clone was found to have an evolved *mutT* allele and an ancestral *dnaE* allele. This clone was named 4K17 and its genotype for both *mutT* and *dnaE* was confirmed by sequencing.

Genome sequencing Genomes of the sampled clones (Table 1) were sequenced on the Illumina Genome Analyzer platform at the Centre National de Séquençage, Genoscope, with one lane of single-end 36-bp reads per genome. Reads were compared to the ancestral genome to identify mutations, using the BRESEQ pipeline (Barrick et al. 2009).

Estimation of Genomic Mutation Rates and Phylogenetic Tree Genomic mutation rates were calculated as described previously (Wielgoss et al. 2013). We computed the mutation rate between two consecutive time points as the number of mutations divided by time and normalized by the total number of sites. Only synonymous events corresponding to the *mutT* mutational spectrum (AT→CG) were considered.

The phylogenetic tree was constructed with the maximum parsimony method on the basis of the presence or absence of mutations using the MEGA software (Tamura et al. 2013).

Estimation of Genomic Mutation Rate along the Phylogeny Using the topology of the phylogeny and the number of AT→CG synonymous mutations, we optimized the branch length and mutation rates of the tree with a Monte Carlo Markov Chain with Metropolis Hasting sampling (MCMCMS). Each sequenced clone was assigned its sampling age in generations. Then, some initial set of branch length compatible with the sampling age of the clones used to start the MCMCMS process. A mutation rate over a set of branches sharing the same mutation rate

was computed as the ratio of the total number of mutations found on these branches to their total length. The likelihood was derived for each branch using a Poisson law. At each cycle of the MCMCMS, 1 to 5 of the branch lengths were modified with a Gaussian deviate and the new likelihood of the data computed on the tree. The new tree was accepted and replaced the previous one if its likelihood Lk_new was higher than the previous one Lk_ref or if a random number r , ranging from 0 to 1, was lower than the ratio Lk_new / Lk_ref . A total number of 5000 cycles of burning during which the tree reached a stable likelihood were discarded. The parameters of the tree, sampled every 100 over 95,000 cycles, were used to compute the posterior probability of each branch length and mutation rate.

Bibliography

- André J, Godelle B. 2006. The evolution of mutation rate in finite asexual populations. *Genetics*, 172: 611–626.
- Barrick J, Yu D, Yoon S, Jeong H, Oh T, Schneider D, Lenski R, Kim J. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461: 1243–1247.
- Blomfield I, Vaughn V, Rest R, Eisenstein B. 1991. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis* *sacB* gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol*, 5: 1447–1457.
- Chao L, Cox E. 1983. Competition between high and low mutating strains of *Escherichia coli*. *Evolution*, 37: 125–134.
- Clune J, Misevic D, Ofria C, Lenski R, Elena S, R S. 2008. Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. *PLoS Comput Biol*, 4: e1000187.
- Denamur E, Lecointre G, Darlu P, Tenaillon O, Acquaviva C, Sayada C, Sunjevaric I, Rothstein R, Elion J, Taddei F, Radman M, Matic I. 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell*, 103: 711–721.

- Denamur E, Bonacorsi S, Giraud A, Duriez P, Hilali F, Amorin C, Bingen E, Andremont A, Picard B, Taddei F, Matic I. 2002. High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *J Bacteriol*, 184: 605–609.
- Desai M, Fisher D. 2011. The balance between mutators and nonmutators in asexual populations. *Genetics*, 188: 977–1014.
- Drake J 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A*, 88: 7160–7164.
- Eyre-Walker A, Keightley P. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet*, 8: 610–6018.
- Fisher R 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- Fowler R, White S, Koyama C, Moore S, Dunn R, Schaaper R. 2003. Interactions among the *Escherichia coli* *mutT*, *mutM*, and *mutY* damage prevention pathways. *DNA Repair*, 2: 159–173.
- Fowler RG, Schaaper RM. 1997. The role of the *mutT* gene of *Escherichia coli* in maintaining replication fidelity. *FEMS Microbiol Rev*, 21: 43–54.
- Friedberg E, Walker G, Siede W. 1995. *DNA Repair and Mutagenesis*. ASM Press, Washington, DC.
- Gerrish P, Colato A, Perelson A, Sniegowski P. 2007. Complete genetic linkage can subvert natural selection. *Proc Natl Acad Sci U S A*, 104: 6266–6271.
- Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, Taddei F. 2001. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science*, 291: 2606–2608.
- Guelfo J, Rodríguez-Rojas A, Matic I, Blázquez J. 2010. A MATE-family efflux pump rescues the *Escherichia coli* 8-oxoguanine-repair-deficient mutator phenotype and protects against H₂O₂ killing. *PLoS Genet*, 6: e1000931.

- Hall B, Ma C, Liang P, Singh K. 2009. Fluctuation AnaLysis CalculatOR (FALCOR): a web tool for the determination of mutation rate using Luria-Delbrück fluctuation analysis. *Bioinformatics*, 25: 1564–1565.
- Johnson T 1999. The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *Proc Biol Sci*, 266: 2389–2397.
- LeClerc J, Li B, Payne W, Cebula T. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science*, 274: 1208–1211.
- Leigh E 1970. Natural selection and mutability. *Am Nat*, 104: 301–305.
- Lenski R, Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci U S A*, 91: 6808–6814.
- Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat*, 138: 1315–1341.
- Link A, Phillips D, Church G. 1997. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: Application to open reading frame characterization. *J Bacteriol*, 179: 6228–6237.
- Ma W, Sandri G.V. H, Sarkar S. 1992. Analysis of the Luria-Delbrück distribution using discrete convolution powers. *J Appl Probab*, 29: 255–267.
- Maki H, Sekiguchi M. 1992. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature*, 355: 273–275.
- Matic I, Radman M, Taddei F, Picard B, Doit C, Bingen E, Denamur E, Elion J. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science*, 277: 1833–1834.
- McDonald M, Hsieh Y, Yu Y, Chang S, Leu J. 2012. The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. *Curr Biol*, 22: 1235–1240.
- Michaels M, Tchou J, Grollman A, Miller J. 1992. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry*, 31: 10964–10968.

- Miller H, Perrino F. 1996. Kinetic mechanism of the 3'→5' proofreading exonuclease of dna polymerase III. analysis by steady state and pre-steady state methods. *Biochemistry*, 35: 12919–12925.
- Newton C, Graham A, Heptinstall L, Powell S, Summers C, Kalsheker N, Smith J, Markham A. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*, 17: 2503–2516.
- Notley-McRobb L, Seeto S, Ferenci T. 2002. Enrichment and elimination of *mutY* mutators in *Escherichia coli* populations. *Genetics*, 162: 1055–1062.
- Oliver A, Cantón R, Campo P, Baquero F, Blázquez J. 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science*, 288: 1251–1254.
- Orr H 2000. The rate of adaptation in asexuals. *Genetics*, 155: 961–968.
- Raynes Y, Sniegowski P. 2014. Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity*, 113: 375–380.
- Schaaper R 1996. Suppressors of *Escherichia coli mutT*: antimutators for dna replication errors. *Mutat Res*, 350: 17–23.
- Sniegowski P, Gerrish P. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philos Trans R Soc Lond B Biol Sci*, 365: 1255–1263.
- Sniegowski P, Gerrish P, Lenski R. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*, 387: 703–705.
- Söderberg R, Berg O. 2011. Kick-starting the ratchet: the fate of mutators in an asexual population. *Genetics*, 187: 1129–1137.
- Sturtevant A 1937. Essays on evolution. I. On the effects of selection on mutation rate. *Q Rev Biol*, 12: 464–467.
- Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutator alleles in adaptive evolution. *Nature*, 387: 700–702.

- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*, 30: 2725–2729.
- Tenaillon O, Toupance B, Le Nagard H, Taddei F, Godelle B. 1999. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics*, 152: 485–493.
- Tenaillon O, Le Nagard H, Godelle B, Taddei F. 2000. Mutators and sex in bacteria: conflict between adaptive strategies. *Proc Natl Acad Sci U S A*, 97: 10465–10470.
- Travis J, Travis E. 2002. Mutator dynamics in fluctuating environments. *Proc Biol Sci.*, 269: 591–597.
- Tröbner W, Piechocki R. 1984. Selection against hypermutability in *Escherichia coli* during long term evolution. *Mol Gen Genet*, 198: 177–178.
- Wielgoss S, Barrick J, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Médigue C, Lenski R, Schneider D. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3 (Bethesda)*, 1: 183–186.
- Wielgoss S, Barrick J, Tenaillon O, Wiser M, Dittmar W, Cruveiller S, Chane-Woon Ming B, Médigue C, Lenski R, Schneider D. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A*, 110: 222–227.
- Wiser M, Ribeck N, Lenski R. 2013. Long-term dynamics of adaptation in asexual populations. *Science*, 342: 1364–1367.
- Yamada M, Shimizu M, Katafuchi A, Grúz P, Fujii S, Usui Y, Fuchs R, Nohmi T. 2012. *Escherichia coli* DNA polymerase III is responsible for the high level of spontaneous mutations in *mutT* strains. *Mol Microbiol*, 86: 1364–1375.
- Yanisch-Perron C, Vieira J, Messing J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33: 103–119.

2.7 Figure Legends

Fig. 1. The MutT, MutY and MutM DNA repair pathway and its role in the control of 8-oxoG. MutT hydrolyzes the oxidized form of dGTP, 8-oxo-dGTP (denoted as oGTP) and prevents its incorporation into DNA. However, when it is incorporated or formed by direct oxidation (denoted as oG), MutM excises 8-oxoG that is then replaced by G base. If cell division occurs before 8-oxoG is removed by MutM an adenine base could be inserted opposite 8-oxoG and MutY is able to excise A pairs with 8-oxoG to reduce the possibility to induce a GC to TA mutation. When A was part of the initial sequence, i.e. following incorporation of 8-oxodGTP instead of dTTP during DNA synthesis that mostly occur when MutT is mutated, the activity of MutY is mutagenic.

Fig. 2. Phylogenetic tree for evolved clones sampled from the Ara+6 population, rooted by using the ancestor. The tree was built using generational time as a scale and the branching was based on maximum parsimony. Colored branches indicate changes in mutation rates (see Fig. 5). The *mutT*, *dnaE* and *mutM* evolved alleles are indicated in yellow, pink and orange, respectively. Three different *mutY* alleles are shown using three different blue colors. The scale for 5,000 generations is shown.

Fig. 3. Effect of the *mutT* and *dnaE* mutations on mutation rates. Mutation rates were determined by fluctuation tests with rifampicin for the ancestor, the two evolved clones 4K17 and 10K-B, and the two constructed strains Anc*mutT*^{evo} and 10K-B*mutT*^{anc}. The *mutT* and *dnaE* alleles were checked by PCR and are indicated below each strain. Error bars show 95% confidence intervals.

Fig. 4. Dynamics of mutation rates in the Ara+6 population over time. (A) The mutation rates inferred from the genome sequences of evolved clones sampled over time. Only AT→CG synonymous mutations, typical of a *mutT* genetic background, were considered. The ancestral mutation rate was calculated as in Wielgoss et al. (2011) and is represented by the dashed line. (B and C) The mutation rates toward rifampicin resistance (Rif^R) inferred from fluctuation tests using mixed population samples (B) and individual A and B evolved clones (C) sampled over time. Evolved A and B clones (Table 1) were randomly sampled in the population at different time points. Therefore, there is no linkage between each A or B clone from successive time

points. Triangles and circles represent the mutation rates of the A and B clones, respectively, while the diamond represents the one of the ancestor. Error bars show 95% confidence intervals.

Fig. 5. Computed dynamics of mutation rates. Mutation rates were determined by a Monte Carlo Markov chain simulation. The graph shows only the 5-rate model, represented by different colors that correspond to those used for the phylogenetic tree (Fig. 2). The estimated times of mutation rate decreases are $\sim 22,600$ and $39,100$ generations. The transition times between the different mutation rates on terminal branches are unknown.

2.8 Tables

Table 1: Mutational Spectrum and Genotype for *mutT* and *dnaE* of Evolved Clones Sampled from Ara+6 over Time.

Clone Name	Non-Synonymous			Synonymous			Genotype	
	AT→CG	CG→AT	Other	AT→CG	CG→AT	Other	<i>mutT</i>	<i>dnaE</i>
0.5K-A	0	0	1	0	0	0	Anc	Anc
0.5K-B	0	1	1	0	0	0	Anc	Anc
1K-A	1	1	1	0	0	0	Anc	Anc
1K-B	0	1	2	0	0	0	Anc	Anc
1.5K-A	1	2	1	0	0	0	Anc	Anc
1.5K-B	0	2	1	0	0	0	Anc	Anc
2K-A	1	1	1	0	0	0	Anc	Anc
2K-B	2	1	1	0	0	0	Anc	Anc
5K-A	1	1	2	0	0	0	Anc	Anc
5K-B	1	1	2	0	0	0	Anc	Anc
6K-A	269	1	2	47	0	1	Evol	Evol
6K-B	284	1	2	43	0	1	Evol	Evol
10K-A	514	1	2	72	0	1	Evol	Evol
10K-B	519	1	2	80	0	2	Evol	Evol
15K-A	749	1	3	113	1	1	Evol	Evol
15K-B	752	2	3	118	1	1	Evol	Evol
20K-A	939	8	4	140	1	1	Evol	Evol
20K-B	1032	1	4	162	1	1	Evol	Evol
30K-A	1391	1	4	217	1	1	Evol	Evol
30K-B	1434	1	5	211	1	1	Evol	Evol
40K-A	1633	1	7	269	4	1	Evol	Evol
40K-B	1529	1	5	229	2	1	Evol	Evol
50K-A	1909	5	11	295	3	1	Evol	Evol
50K-B	1714	15	9	279	15	1	Evol	Evol

NOTE.- The number of mutations for each mutational category is given for each clone. Anc and Evol indicate the ancestral and evolved alleles of *mutT* and *dnaE*, respectively. Only one evolved allele was found for each gene. The sequences reported in this paper have been deposited in the European Nucleotide Archive (ENA) Sequence Read Archive (accession no XXXX).

Table 2: Primers and Plasmids Used

Primer	Sequence	Restriction Site	Experiment
mutT_fw	TGCCTGAAGAGGTTGAGGAG	-	Allelic exchange, sequencing
mutT_rev	GATGACTGGAGCGAAGAACC	-	Allelic exchange, sequencing
dnaE_fw	CCGGTCTTTACCATTTCAC	-	Allelic exchange, sequencing
dnaE_rev	TAACCGCAGTCAGAGAATCG	-	Allelic exchange, sequencing
dnaE_anc	GGGGCTCGTTGTTGCCGCGCGGGTCA	-	ARMS-PCR
dnaE_evo	GGGCTCGTTGTTGCCGCGCGGGTCC	-	ARMS-PCR
dnaE_rev	AACGCAACCGCGCGCGTGCATCCGCC	-	ARMS-PCR
mutT_anc_sh	AGAAGTCGGGATTACCCCTC	-	ARMS-PCR
mutT_evo_sh	AGAAGTCGGGATTACCCCTA	-	ARMS-PCR
mutT_rev	GCCTTATCCGACCTACAAAAAC	-	ARMS-PCR
mutY_fw	GGCGCACAAAACACTACGGATA	-	Sequencing
mutY_rv	TGACCTTCTGCTTCACGTTG	-	Sequencing
mutY-pQE80F	CTATGGATCCATGCAAGCGTCGCA	BamHI	Expression
mutY-pQE80R	TATTAAGCTTCTAAACCGGCGCGC	HindIII	Expression
Plasmid		Cut by	Experiment
pQE-80L	-	BamHI/BclI	Overexpression
pKO ₃	-	SmaI/NotI	Allelic exchange
pCRII-Topo	-	BamHI/NotI	Cloning

2.9 Figures

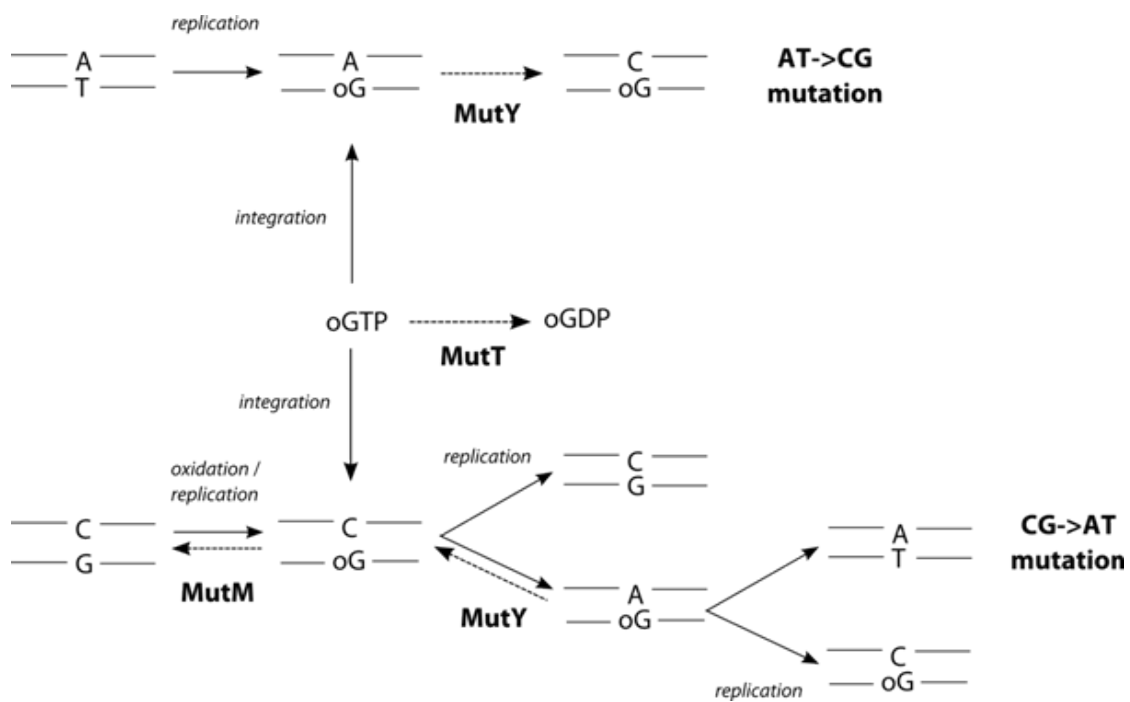


Fig 1.

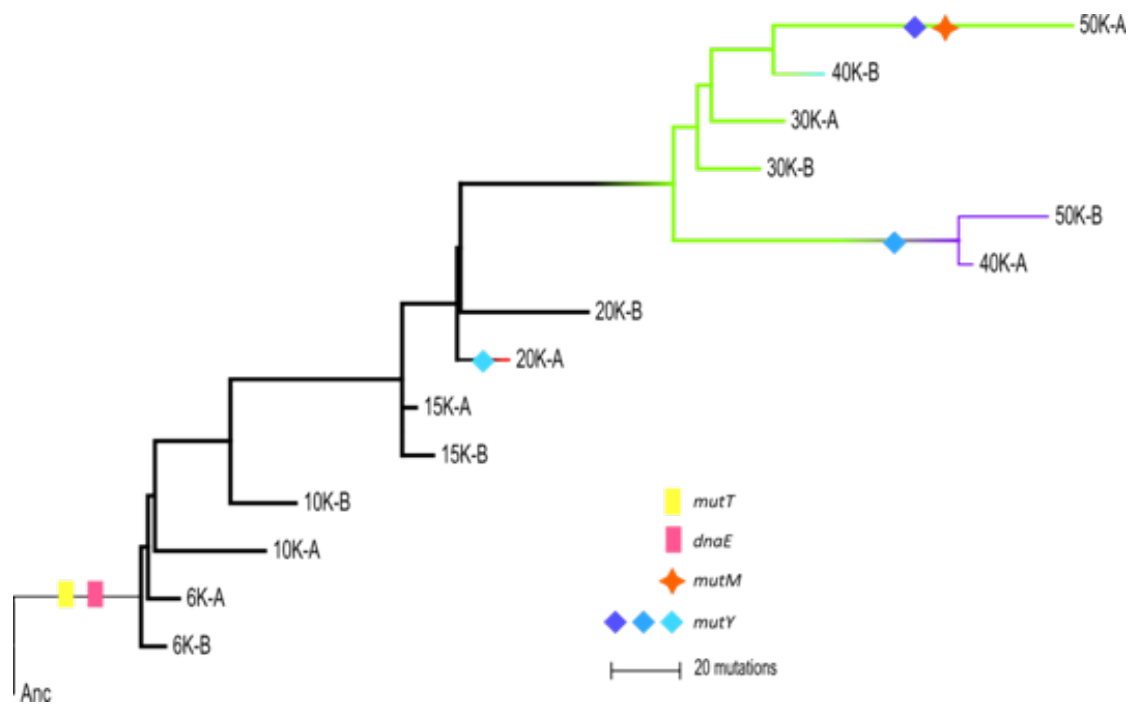


Fig. 2

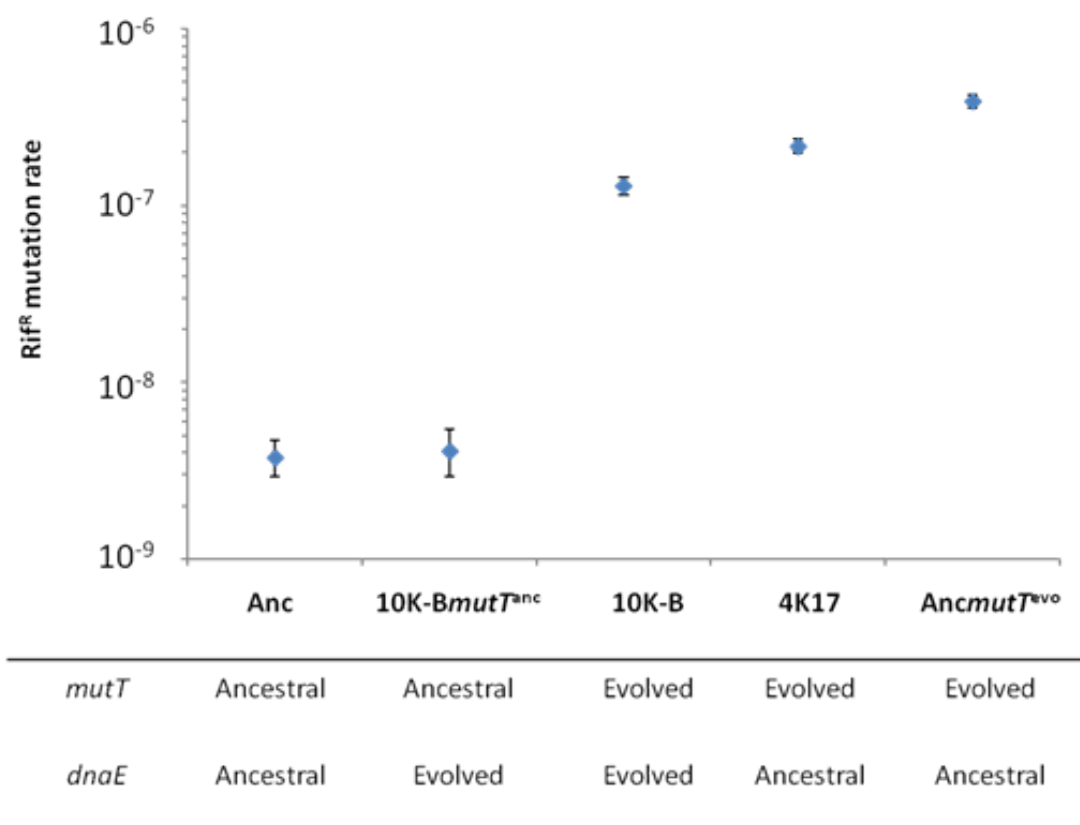
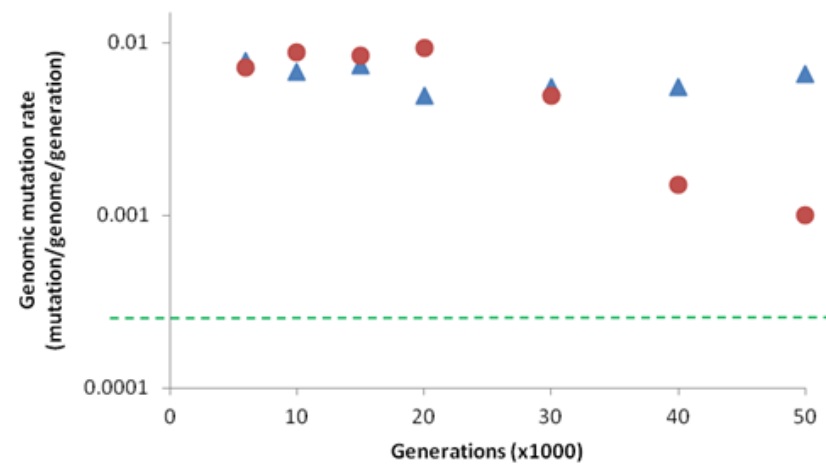
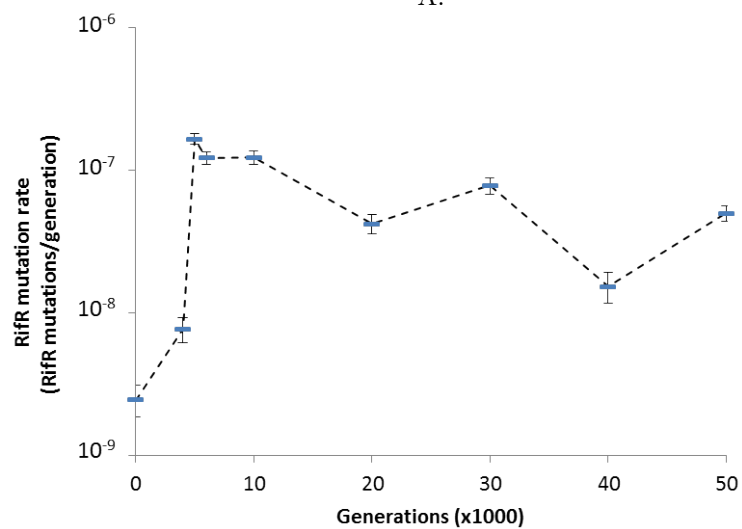


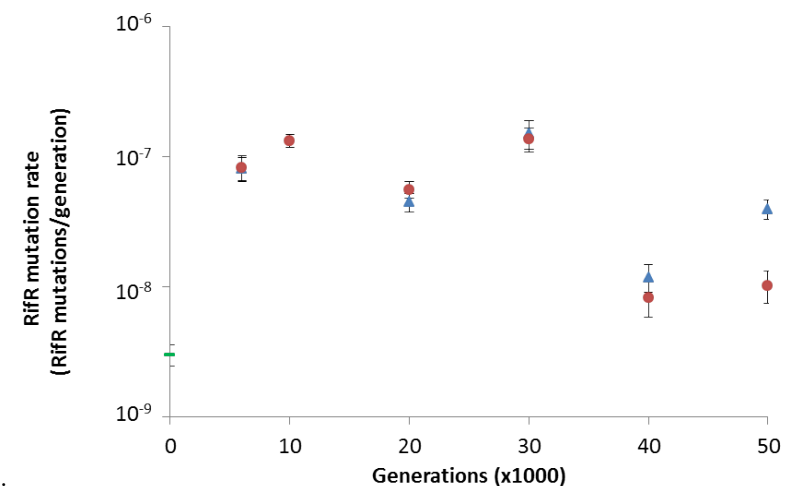
Fig. 3



A.



B.



C.

Fig. 4

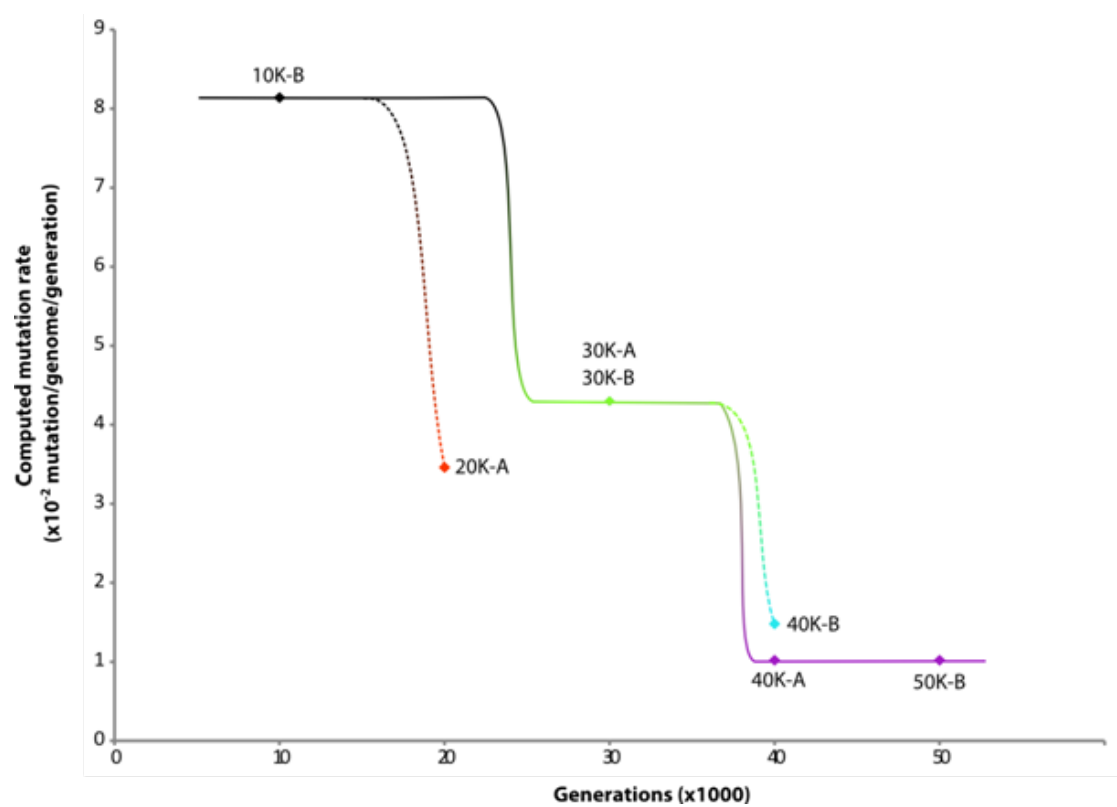


Fig. 5

3 Supplementary Material

Long-Term Dynamics of Mutation Rates in a Hypermutator Population of *Escherichia coli*

Larissa Caudwell, Olivier Tenaillon, Sebastien Wielgoss, Sylvain Caillat, Jean-Luc Ravanat,
Richard E. Lenski, and Dominique Schneider

Table S1. Computation of Mutation Rates in Ara+6 using Monte Carlo Metropolis Hasting Markov Chain simulation.

Fig. S1. Effect of overexpression of *mutY* and *mutM* on mutation rates.

Fig. S2. Oxidative stress measurements by flow cytometry.

Supplementary Methods:

Overexpression of *mutY* and *mutM*, and Spotting Test

Flow Cytometry

Table S1. Computation of Mutation Rates in Ara+6 using Monte Carlo Metropolis Hasting Markov Chain simulation.

Initial Rate	20K-A Branch	40K-B Branch	40K-A, 50K-B	Later Rate	Log Likelihood	AIC
5.26 (4.94-5.53)					-121.9	245.8
5.34 (5.04-5.61)	3.13 (2.59-3.93)				-120.74	245.48
5.42 (5.15-5.69)		1.59 (1.33-1.97)			-110.75	225.5
5.95 (5.63-6.30)			0.97 (0.74-1.15)		-93.59	191.18
7.62 (6.96-8.31)				3.56 (3.30-3.92)	-97.46	198.92
6.04 (5.73-6.39)	3.30 (2.79-3.92)		0.96 (0.72-1.15)		-91.25	188.5
6.12 (5.79-6.45)		1.06 (0.88-1.18)	1.50 (1.31-1.78)		-83.31	172.62
7.56 (6.91-8.83)			1.05 (0.86-1.17)	4.06 (3.57-4.64)	-83.58	173.16
6.26 (5.97-6.59)	3.36 (2.82-4.12)	1.63 (1.37-2.11)	0.92 (0.69-1.13)		-78.86	165.72
7.54 (6.95-8.33)		1.49 (1.22-1.91)	1.03 (0.81-1.17)	4.40 (3.88-4.93)	-73.81	155.62
8.14 (7.40-8.88)	3.46 (2.98-4.18)	1.48 (1.17-1.97)	1.02 (0.76-1.16)	4.30 (3.73-4.89)	-69.22	148.44

NOTE. - One to five different mutation rates were computed. Values in bold indicate the lowest AIC among all the simulations with the same number of parameters.

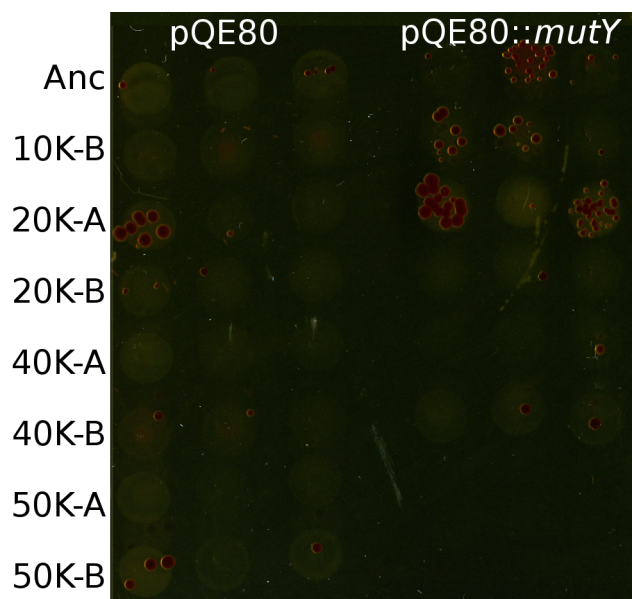


FIG. S1. Effect of overexpression of *mutY* and *mutM* on mutation rates. A spotting test was performed after introduction of plasmids overexpressing each of the two genes into the indicated clones. The pQE80 plasmid without any insert was used as a control.

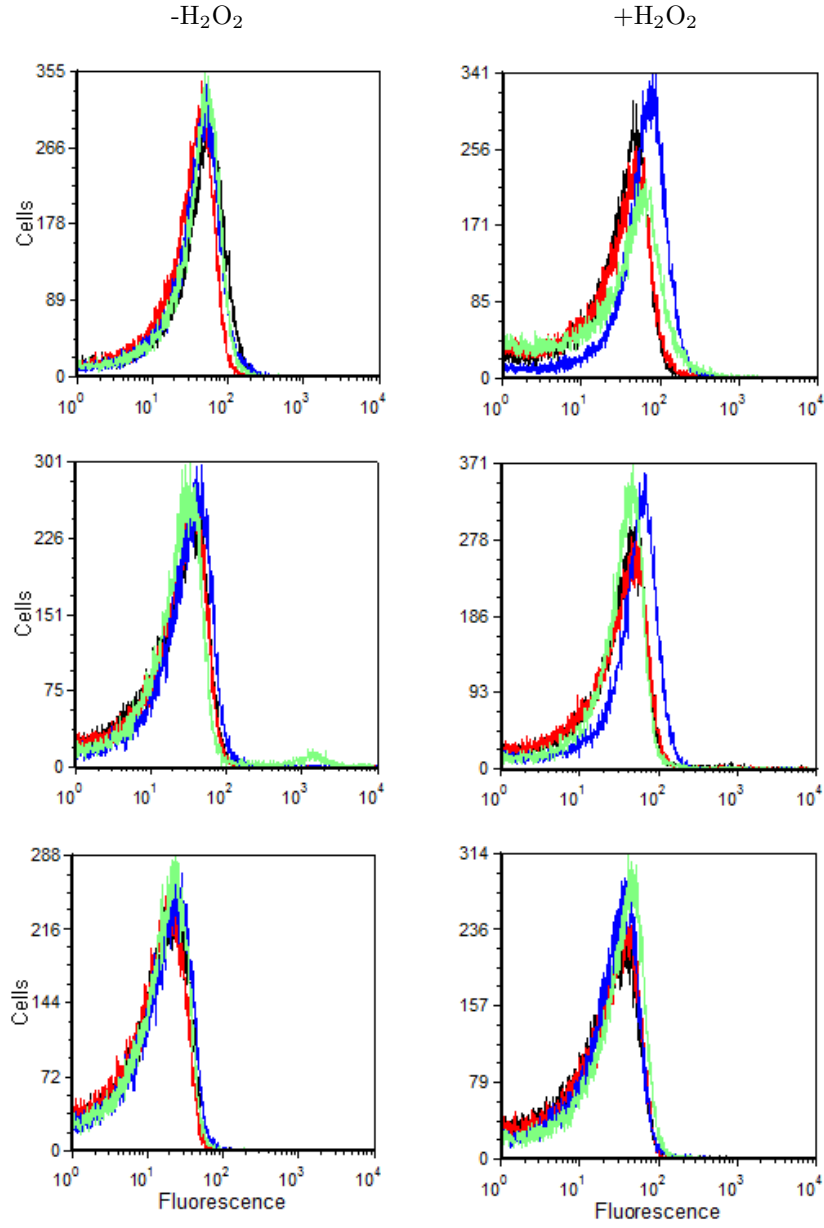


FIG. S2. Oxidative stress measurements by flow cytometry. Oxidation levels within cells were measured using the dihydro-rhodamine 123 fluorescent dye. Cells were treated (right) or not (left) with H_2O_2 before staining. Four evolved clones were analyzed: 30K-A (black), 30K-B (red), 40K-A (blue), and 40K-B (green). All experiments were done in triplicate (from top to bottom).

Supplementary Methods

Overexpression of *mutY*, and Spotting Test Overexpression was achieved by cloning each of the two *mutY* and *mutM* genes in the pQE80L expression plasmid digested by the restriction enzymes HindIII and BamHI. The *mutY* gene was PCR-amplified using the primer pairs mutY-pQE80F and -R. (The restriction sites HindIII and BamHI were inserted into the primers and are underlined). The PCR product, after sequencing to verify its nucleotide sequences, was then inserted into pQE80. Each of the two plasmids (pQE80 and pQE80::*mutY*) were then introduced by electro-transformation in the relevant clones (supplementary Fig. S1, Supplementary Material online).

All recipient cells were then tested for their mutation rates using a spotting test. After an overnight pre-culture in LB medium at 37°C with shaking at 200rpm, strains were grown in LB+ampicillin (100 mg/l) with or without 0.02mM IPTG during 32h at 37°C with shaking at 120rpm in 1ml 96-well plates. Cultures were then centrifuged and their supernatant removed to concentrate the cells. Drops of 10 μ l of each culture were deposited on DM4000+ampicillin and DM4000+rifampicin+ampicillin agar plates, and incubated at 37°C for 48h. The hypermutator phenotype was characterized by a high number of colonies within a spot on agar plates (supplementary Fig. S1, Supplementary Material online). All cultures were performed in triplicates.

Flow Cytometry After an overnight pre-culture in LB at 37°C with shaking at 120rpm, cells were acclimated in DM25 during 24h, then diluted at 1/100 in DM250, and incubated for 4h at 37°C with shaking at 120rpm. Two 2-ml aliquot of each culture were sampled and divided in 2, with 2ml being incubated with 5mM H₂O₂ for 30min and 2ml with nothing added. Samples were then washed and diluted in 5ml PBS. Cells were stained with 1x SYBR Green (Life Technologies) and 10 μ M dihydro-rhodamine 123 (Life Technologies).

Chapitre III

Biais de mutations dans les populations hypermutatrices de l'expérience d'évolution à long terme

1 Préambule

Bien qu'étant un phénomène aléatoire, toutes les mutations ne sont pas distribuées uniformément le long du génome et les changements de nucléotides ne sont pas tous dans les mêmes proportions. L'étude des biais de mutation a considérablement augmenté ces dernières années grâce au séquençage de génomes à haut débit. Divers organismes modèles ont été étudiés, tels qu'*E. coli* (Lee et al. 2012; Foster et al. 2013), *Saccharomyces cerevisiae* (Zhu et al. 2014), *Bacillus subtilis* (Sung et al. 2015) ou *Pseudomonas fluorescens* (Long et al. 2015). Ces études ont utilisé des expériences d'accumulation de mutations afin de réduire l'effet de la sélection. Les biais de mutations résultent soit des mécanismes chimiques provoquant les mutations, soit des processus de réplication ou réparation de l'ADN (Frank et Lobry 1999). Dans cette partie, nous étudions les biais de mutations dans les six populations hypermutatrices de l'expérience d'évolution à long terme (Ara-2, Ara-3, Ara-4, Ara+3, Ara-1 et Ara+6). En effet, le phénotype hypermutateur nous permet de disposer d'une banque de données considérable pour étudier la distribution des mutations dans le génome. En effet, nous avons accès à plus de 50 000 mutations, dont environ 15 000 événements distincts. Nous disposons aussi de deux types de phénotypes mutateurs : un premier, dans les populations Ara-2, Ara-3, Ara-4 et Ara+3, lié à un déficit en enzymes du système de réparation des mésappariements (*mutS*, *mutL*, ou *uvrD*, dit MMR-), et un second, dans les populations Ara-1 et Ara+6, dans le système de prévention du 8-oxo-dGTP (*mutT*-). Ainsi, nous pouvons analyser, par des approches de bioinformatique, les biais de mutations à plusieurs niveaux : au niveau des changements de nucléotides, de leur distribution topologique et de leur entourage nucléotidique.

Chez *E. coli* comme chez de nombreux organismes, les changements de type transition sont majoritaires par rapport aux transversions (Lee et al. 2012). Ce biais est d'autant plus marqué dans les populations MMR- (Figure 1 de l'article) avec un nombre de transitions 28 fois supérieur en moyenne à celui des transversions. Dans les populations *mutT*-, les transversions sont majoritaires à cause de l'incorporation du 8-oxo-dGTP favorisant les mutations AC→GT (voir partie précédente). Ainsi, le taux transitions/transversions est proche de 0,015.

Pour étudier la distribution des mutations dans le génome, nous les avons rassemblées dans 100 fenêtres d'environ 48 000 nucléotides et nous avons tracé l'écart cumulé à la moyenne (Fi-

gure 2 de l'article). Cette méthode permet de visualiser facilement les variations d'accumulation de mutations à travers le génome. Ainsi dans la population Ara+3, un profil net est visible, où le génome est divisé en 4 régions dans lesquelles les tendances s'inversent successivement, avec moins de mutations autour du terminus et de l'origine de réplication. De telles tendances sont visibles dans les autres populations MMR-, mais dans une moindre mesure. Ces régions ne correspondent pas tout à fait aux macrodomaines définis dans le génome d'*E. coli* K-12 (Valens et al. 2004). Aucune tendance n'est visible dans les populations *mutT*-, présentant des profils opposés. Cependant, un autre biais topologique est visible, cette fois au niveau des brins. En effet, le chromosome est constitué de deux réplichores, séparés par l'origine et la région terminus de réplication, dans lesquels les brins directs sont opposés. Lorsque les mutations sont dénombrées en prenant uniquement en compte les brins directs, le nombre de T est de G mutés est nettement supérieur que pour leurs nucléotides complémentaires A et C dans les populations MMR-, alors qu'aucune différence n'est visible dans les populations *mutT*- (Table 2 de l'article). Ainsi, cette différence est propre aux transitions, et résulte d'un biais de réplication.

Outre la position d'un nucléotide, son entourage génomique immédiat peut également avoir une influence significative sur sa probabilité de mutation. En effet, lorsque les mutations ont été analysées en prenant en compte des triplets avec le nucléotide muté au centre et les nucléotides en 5' et 3' respectivement, nous avons observé que certains triplets étaient plus touchés que d'autres. Ainsi, le triplet GTA est jusqu'à 91 fois plus affecté par des mutations que le triplet CAA dans les populations MMR-, tandis que le triplet le plus affecté dans les populations *mutT*- est GAA (10 fois plus que GTG). Il a également été possible d'observer des tendances générales, comme une forte mutabilité des triplets XAC et GTX dans les populations MMR-, et XAA et TTX dans les populations *mutT*-. Ainsi, l'entourage nucléotidique influence la mutabilité d'un nucléotide.

L'ensemble des biais observés dans les populations MMR- recoupe ceux observés dans l'étude de Lee et al. (2012) au cours d'une expérience d'accumulation de mutations. En l'absence de système MMR, ces biais résultent donc de la réplication, soit à cause de mésappariements, soit suite à des dommages lorsque l'ADN est simple brin, comme la désamination. Au vu du nombre extrêmement élevé de mutations dans les lignées MMR-, Lee et al. (2012) ont suggéré que les désaminations ne pouvaient pas être assez fréquentes pour être la cause majeure de ces

biais, et ont ainsi proposé que les mésappariements en seraient responsables. Deux hypothèses sont alors possibles : soit les mésappariements sont plus fréquents sur un brin qu'un autre, soit l'activité de relecture est asymétrique. Bien qu'étudiée *in vivo* (Fijalkowska et al. 1998), cette question reste irrésolue.

Dans les populations *mutT*-, les biais peuvent, en plus de la réplication de l'ADN, résulter de biais de réparation par les enzymes MutY et MutM.

Cette étude est la première à s'intéresser aux biais mutationnels d'une population *mutT*- chez *E. coli*. De plus, elle se distingue par la présence de la sélection. Cependant, celle-ci s'applique surtout sur les mutations de fort effet (Lässig 2011, Maharjan et al. 2012), et les biais sont similaires lorsque seules les mutations synonymes sont prises en compte.

2 Article 2

Article

Discovery

Mutational patterns in *Escherichia coli* hypermutator populations during a long-term evolution experiment

Larissa Caudwell^{1,2}, Dominique Schneider^{*,1,2} and Olivier Tenaillon^{3,4}

¹Univ. Grenoble Alpes, Laboratoire Adaptation et Pathogénie des Microorganismes (LAPM),
F-38000 Grenoble, France

²Centre National de Recherche Scientifique (CNRS), LAPM, F-38000 Grenoble, France

³IAME Unité Mixte de Recherche (UMR) 1137, INSERM, F-75018 Paris, France

⁴IAME, UMR 1137, Université Paris Diderot, Sorbonne Paris Cité, F-75018 Paris, France

***Corresponding author:** E-mail: dominique.schneider@ujf-grenoble.fr

2.1 Abstract (up to 250 words)

Although mutations are random events, their patterns have been often reported to be biased, with for instance a higher frequency of transitions than transversions and an uneven distribution throughout the microbial genome. Here we report biases in the mutational pattern in six hypermutator populations from a long-term evolution experiment in *Escherichia coli*. During that experiment, an *E. coli* ancestral strain has evolved twelve times independently for more than 50,000 generations. Our study included a total of ~50,000 mutations, including non-synonymous and synonymous Single-Nucleotide Polymorphisms. The hypermutator phenotype of the six populations has evolved owing to two different types of deficiencies, one in the mismatch repair system (MMR-) and one in the 8-oxo-dGTP clearance system (*mutT*-). We showed that both hypermutator types had biases in nucleotide changes and mutation distribution, with particular influence of the neighboring nucleotides.

2.2 Introduction

Mutations are the primary source of variation for evolutionary processes. Mutations are stochastic events and may have different effects (deleterious, neutral, beneficial) on organismal fitness (Couce et al. 2013). Their occurrence, selection and fixation shape the genome architecture. Even if mutations occur randomly, their frequency has been shown to rely on both the nature of the base substitution and the local genomic context (Benzer 1961, Coulondre et al. 1978; Rogozin and Pavlov 2003; Lynch 2008; Lee et al. 2012). Indeed, important biases influence the ultimate emergence of a particular mutation, including transitions versus transversions, synonymous versus non-synonymous, and the CpG mutations.

In 1961, Benzer raised the question whether mutations were evenly distributed throughout the genome (Benzer 1961). Indeed, the microbial chromosome is a well-structured molecule organized in two replichores, that are distinct by their direction of replication and separated by the replication origin and terminus, and 6 different macrodomains that are differentially structured (Valens et al. 2004). Different genomic regions are differentially expressed, resulting in variable DNA-protein interactions which may expose DNA to damages (Friedberg et al. 1995). This complex organization may affect the distribution of mutations, whether they arise through mispairing during the replication or by DNA damage. In addition, biophysical contexts may favor nucleotide modifications, including methylated cytosine residues, which deaminate into thymine residues. Studies using the *rII* locus from phage T4 (Benzer 1961) and the *lacI* gene from *E. coli* (Coulondre et al. 1978) demonstrated the existence of mutational hotspots. However, whether these observations can be generalized is still an open question. The recent development of high-throughput genome sequencing technologies provides with the access to all genomic mutations, thereby allowing the study of mutational biases at the genomic scale. Recent studies have identified three main types of mutational biases (Lee et al. 2012; Foster et al. 2013; Long et al. 2015; Sung et al. 2015): the type of nucleotide change, the topologic distribution and the nucleotide context.

Six nucleotide changes are possible, including 2 transitions (Ts) and 4 transversions (Tv). However, the ratio Ts/Tv observed in most species is not equal to 0.5, but indicates a predominance of transitions (Lynch et al. 2007, Lynch 2010, Hershberg and Petrov 2010). In addition,

GC→AT mutations tend to be predominant while GC→CG changes are rare. The type of mutations also varies in hypermutator genotypes in which DNA repair systems are disrupted. For example, MMR- strains are characterized by a high number of GC→AT and AT→GC mutations (Lee et al. 2012) while *mutT*- hypermutators show a strong bias in favor of AT→CG changes (Maki and Sekiguchi 1992). The different mutational spectra and their respective importance shape the genome and its nucleotide composition (Rocha et al. 2006; Sharp et al. 2010).

Another bias includes the distribution of mutations throughout the genome. The first difference was observed by comparing the two strands of each chromosomal replicore in *Escherichia coli* (Lobry 1996). Replichores are characterized by their direction of replication and are separated by the DNA replication origin (Ori) and terminus (Ter) sites. In each replicore, one strand is replicated directly while the other is replicated using Okazaki fragments (Okazaki et al. 1968). These two modes of replication affect the mutational pattern, with a predominance of mutated T or G on the matrix strand compared to A and C. This mutational pattern results in strong biases in the base composition of the DNA leading and lagging strands known as the GC skew (Lobry 1996). It is strong enough in bacterial species to allow the localization of the origin and terminus sites of DNA replication. In addition to this bias due to replication, another topologic bias has recently been observed in *E. coli* (Foster et al. 2013). Using MMR- *E. coli* populations that were propagated through mutation accumulation experiments, they reported that mutations, when pooled in ~100-kbp bins, were distributed according to a wavelet model throughout the genome. Mutations are reported to be particularly dense close to the Ter sites (Foster et al. 2013). This wavelet model is not correlated to the macrodomain structure of the chromosome (Valens et al. 2004). The same pattern was observed in *Pseudomonas fluorescens* (Long et al. 2015) using 50-kbp bins. The wavelet model has been shown to be associated with global gene expression patterns, in particular with the density of HU-activated genes, FIS-repressed genes, and down-regulated genes in *gyrA* mutants, suggesting that mutability might be linked to DNA structure.

Contextual biases result from the influence of the neighboring bases on the probability of mutation of a particular nucleotide. While this bias was investigated by Benzer (1961) using the reporter *rII* locus, genome sequencing gives nowadays much more power. In the yeast *Saccharomyces cerevisiae*, cytosine residues are highly mutable in the CCG and TCG triplets

(with the underlined center nucleotide mutated, Zhu et al. 2014), while the triplets GTG are 67-fold more mutated than the AAA triplets in *P. fluorescens* (Long et al. 2015). This context dependence results in mutational hotspots along the genome. Among these hotspots, the most widely observed are CpG biases which result in 1/3 of the genetic diseases in humans (Cooper and Kravzack 1990; Duret 2009), methylated cytosine residues which spontaneously deaminate into thymine residues (Coulondre et al. 1978; Lee et al. 2012), and stretches of identical nucleotides which favour DNA polymerase slippage (Rogozin and Pavlov 2003).

The recent combination of experimental evolution and whole-genome sequencing now allows the investigation of the dynamics of mutational biases over entire genomes, which may help to understand how mutations shape microbial genomes. Such investigations focused on mutational biases that occurred during evolution by using different microbial models propagated under mutation accumulation (MA) conditions (*E. coli* in Lee et al. 2012 and Foster et al. 2013; *S. cerevisiae* in Zhu et al. 2014; *Bacillus subtilis* in Sung et al. 2015; *P. fluorescens* in Long et al. 2015). Evolution under MA uses single-cell bottlenecks in order to limit the action of natural selection and consequently allows to detect mutational biases free from selection. Since mutational events remain rare, these experiments also use mismatch-repair deficient strains (MMR-) to increase the number of mutations that can be analyzed. It is however not sure whether these mutational biases are strong enough to shape genome evolution in the presence of selection or whether selection will overwhelm the signal.

Here, we used Richard Lenski's Long-Term Evolution Experiment (LTEE) as a model study for the investigation of mutational biases. This experiment was initiated in 1988 from a single ancestor *E. coli* B that was used to propagate 12 populations in a minimal glucose medium by daily serial transfers. Samples were collected at 500-to 1000-generation intervals from each population and frozen at -80°C constituting a living fossil record. Individual evolved clones were isolated from each population at each of the time points. Genome sequences of the evolved clones sampled after 50,000 generations revealed more than 50,000 mutations, including ~15,000 single events. Therefore, this experiment constitutes a huge database for studying mutational biases. This high number is due to the fact that 6 of the 12 populations have evolved a hypermutator phenotype. Four, called Ara-2, Ara-3, Ara-4 and Ara+3, are altered in mismatch repair genes (MMR-): *mutL* and *uvrD* in Ara-2 and Ara-4, *mutS* in Ara+3 (Sniegowski et al. 1997), and *mutL*

in Ara-3 (Blount et al. 2012). These mutation rate changes occurred early in the experiment (between 2000 and 4000 generations) for Ara-2 and Ara+3, between 8000 and 9000 generations for Ara-4 and much later (after 30,000 generations) for Ara-3. The 2 other hypermutator populations Ara-1 and Ara+6 are altered in the oxidized dGTP, 8-oxo-dGTP, detoxification system and have a mutation in the *mutT* gene (Barrick et al. 2009, Wielgoss et al. 2013, Caudwell et al. to be published) which occurred early in Ara+6 (4000 generations) and much later in Ara-1 (26,000 generations). The 8-oxo-dGTP modified nucleotide pairs equally well with dATP or dCTP and thereby produces AT→CG mutations. Both Ara-1 and Ara+6 populations also have *mutY* mutations that have been shown to compensate the *mutT* hypermutator phenotype by increasing CG→AT mutations. Yet, these mutations appear late during evolution (~30,000 and ~40,000 generations for Ara-1 and Ara+6, respectively) and therefore have minor impact on the mutational patterns. These two types of hypermutator phenotypes, MMR- and *mutT*-, result from mutations in two different mechanisms of DNA repair, and we investigated how their disruption altered the mutational spectrum. This dataset therefore offers a unique opportunity to study mutational biases in populations undergoing natural selection and having two very distinct types of DNA repair alteration.

2.3 Results

Six hypermutator populations and shifted mutational spectra Two individual clones were sampled from each of the 12 populations of the long-term evolution experiment at 11 different time points, with a total of 264 clones (Table S1). Their genomes were sequenced and a total of 46,876 mutations were identified, including 13,961 unique events (Table 1).

In the non-mutator populations, the majority of the mutations (43%) were CG→TA transitions (Figure 1). Consequently, the mutational spectrum is biased towards transitions, with a mean ratio $Ts/Tv = 0.76$, much higher than 0.5 which is the theoretical value. Such bias towards transitions has already been reported (Lee et al. 2012). In the four MMR- populations, such a trend was also observed (Figure 1), with both GC→AT and AT→GC mutations (mean $Ts/Tv \sim 28$). In striking contrast, the two *mutT*- populations revealed ~90% of AT→CG transversions (Figure 1), with $Ts/Tv = 0.15$.

More detailed analyses required a higher number of mutations to have statistical power. Therefore, we focused all our analyses on the six hypermutator populations, which revealed large differences between the two types of populations with alterations in the MMR and 8-oxo-dGTP systems.

Mutation distribution We analyzed the distribution of the mutations throughout the genome by using a cumulative sum method with ~46,300-nucleotide bins (Figure 2). This is a scale-free method in which the cumulative deviation from the mean of a statistic is plotted along the chromosome. Increased and decreased levels revealed chromosomal regions with higher and lower than average value, respectively. The strongest pattern was found in Ara+3, one of the MMR populations that spent the longest period of time as a mutator (Figure 2). In that population, the genome can be divided in 4 regions with marked biases in mutation distribution. From bins 95 to 24 and 48 to 71 (positions 4398.5 to 111,132 kbp and 2222.4 to 3287.3 kbp, respectively), mutations were more frequent than average (1.28 and 1.14 fold, respectively) and were homogeneously distributed. In contrast, from bins 25 to 47 and 72 to 94, mutations were less frequent compared to the mean (by 0.82 and 0.62 fold, respectively, Table of *p*-values in Table S2). Interestingly, the genomic region from bins 24 to 47 corresponds to the Ter macrodomain.

Despite discrepancies in the mutational patterns of the four MMR- populations, they also

showed some similarities (Figure 2). First, the mutations tended to be denser in the bins 13 to 26 and 47 to 62 that correspond to the Right and Left macrodomains (Valens et al. 2004), respectively. The three populations Ara-2, Ara-3, Ara+3 harbored a consistent increase in the distribution of mutations in a region overlapping the Left macrodomain and the three populations Ara-3, Ara-4 and Ara+3 showed a consistent increase over the Right macrodomain. The amplitude of these fluctuations ranges from 0.65 to 1.28-fold the mean, some large regions being therefore about twice more mutagenic than others. Second, the mutations were less frequent around the origin of replication (Ori) corresponding to the bins 81 to 100. Third, two breaking points, where the density of mutations changed, were detected in all four populations albeit at slightly different positions: one between the Left and Ori macrodomains (although very blurry in Ara-4), and one in the Ori region, after *oriC*. Therefore, the mismatch mutations were not randomly distributed, with a higher frequency in some regions.

In the two *mutT*⁻ populations, the situation was much less clear, with different patterns of mutation distribution, suggesting that the chromosome structure had no major impact on the incorporation of 8-oxo-dGTP.

Strand bias We analyzed the mutations by taking into account only the leading strands, *i.e.* they were considered directly on the right replichore (from Ori to Ter) and on the complementary sequence in the left replichore (from Ter to Ori). We observed a strong mutational bias in all four MMR⁻ populations, which was absent when mutations were considered linearly throughout the genome (Table 2). Indeed, AT→GC transitions occurred predominantly at the T sites of the leading strand, while the GC→AT transitions were more frequent at the G sites. Interestingly, these biases were consistent with the GC skew found in most bacterial genomes, including *E. coli*, in which the leading strand is more G- and A-rich than the lagging strand.

We detected also a small bias in the two *mutT*⁻ populations for AT→CG mutations. The effect is however mild despite thousands of mutations, suggesting that, in contrast to mispairing, the integration of 8-oxo-dGTP in the DNA was mostly independent from the replication mode. Therefore, we performed all further analyses by taking into account the mutations on the leading strands, the ones on the left replichore being reverse-complemented.

Nucleotide context We observed the effect of neighboring nucleotides on the probability of mutation of each nucleotide (Figure 3). We considered a triplet with each mutated nucleotide at the center and its 5' and 3' neighbors. We then tested if the number of each mutated triplet deviated from its respective genomic abundance.

In contrast to strand biases, the context dependent-mutability was detected in both the MMR- and *mutT*- populations, with similar patterns in each mutator group (*mutT*- and MMR- populations, respectively). The context appeared to be mostly symmetric, *i.e.*, if X is complementary to Y, complementary XXX and YYY triplets presented the same over- or under-representation (Figure 3). In the MMR- hypermutator populations, the main biases involved the triplets GTN (with N being any nucleotide, and the underlined base being the mutated nucleotide), and GCG which are up to 3.7 and 2.6 fold more mutated than expected by chance, respectively ($p < 10^{-9}$, Fisher exact test). We found a similar over-representation for their symmetric triplets NAC and CGC. The TTN and AGN triplets were less mutated by 0.12- to 0.4-fold ($p < 10^{-9}$, Fisher exact test), which was also the case for the reverse-complement triplet NAA, but not for the NCT triplet, the number of mutations being really too low to allow statistically relevant analyses. These biases led to a 31-fold differences in rates among the most extreme contexts.

In the two *mutT*- populations, mutated triplets TTX (X = A, C or T) and symmetric XAA (X = A, G or T) were 1.7- to 2.5-fold more frequent than expected by chance, while GTN / NAC were under-represented by 0.25- to 0.55-fold, as well as the triplet ATM (with M = G or T) and its complement MTA. Only the CTC triplet presented an asymmetric pattern with no significant bias in GAG, although it was detected only in Ara+6 and not in Ara-1.

These biases may be important, with 91-fold more changes in GTA than CAA.

Traces of selection We used three different approaches to detect if selection may shape the observed signals. First, we analyzed synonymous mutations and found similar patterns as for the total mutations. As synonymous mutations are subject overall to much lower selective pressure, finding similar pattern between total and synonymous mutations suggested that the effect of selection is minor (Figure S1).

Second, the biases, at least in MMR- populations, were consistent with those found in mutation accumulation experiments (Lee et al. 2012, Foster et al. 2013), further suggesting that the

observed biases resulted from mere replication errors rather than from selection.

Third, we computed, over the entire genome, the ratio K_a/K_s (<http://libseq.sourceforge.net/>) of non-synonymous over synonymous mutation substitutions. This ratio is indicative of the fraction of non-synonymous mutations that have been filtered by the purifying action of natural selection. In the two *mutT*- populations, K_a/K_s was ~ 0.8 which suggested that $\sim 20\%$ of non-synonymous mutations have been removed by selection. In the four MMR- populations, K_a/K_s was higher than 0.9, suggesting that less than 10% of non-synonymous mutations were filtered out. The difference between the two types of hypermutator populations was likely to be driven by the type of mutations produced. Indeed, the MutT enzyme favors a transversion mutational event that generates less synonymous mutations and more drastic amino-acid changes. In contrast, the MMR system favors transitions that have a milder effect.

2.4 Discussion

We used the six hypermutator populations of the long-term evolution experiment (Wiser et al. 2013) to analyze mutational biases during long-term selection in a constant environment. Four of these six populations have alterations in the MMR system, and the other two in the 8-oxo-dGTP detoxification system. We detected, despite the presence of selection, three different types of mutational biases: the type of nucleotide change or mutational spectrum, the topologic distribution throughout the genome, and the nucleotide context of the surrounding bases within triplets.

In the four MMR- populations, our results were consistent with previous studies using mutation accumulation populations (Lee et al, 2012, Foster et al, 2013). We therefore showed that such mutational biases are strong enough to persist even in the presence of selection that gave rise to more than 50% fitness increase (Wiser et al. 2013). The mutational spectrum is naturally biased toward transversions (Fijalkowska et al. 1998, Lynch 2010), becoming even larger in the four MMR- populations. In the MMR- populations we followed, we found that both transitions were equally likely to occur. This contrasts with previous results on mutation accumulation that showed that AT→GC were much more likely to occur than GC→AT (Lee et al. 2012).

Transitions arise after errors in DNA replication, when A or T residues are mispaired with C or G residues, respectively. In the LTEE MMR- populations, the DNA polymerase seems equally likely to produce either one or the other transition since their respective frequency was ~50%. However, this was not the case in the non-mutator populations of the LTEE, implying that, for a given mispairing, the MMR system is more efficient in repairing one configuration than the other. For example, the A:C mismatch can give rise to either AT→GC or GC→AT transitions whether A or the C was the original base. Even if these two transitions are equally well produced by the DNA polymerase, the A:C mismatch is better repaired by the MMR system when A is on the leading strand (noted A:C) than in its opposite configuration (C:A).

The discrepancy in the ratio of AT→GC or GC→AT transitions in MMR- populations between the mutation accumulation experiment and the LTEE might be related to three non-exclusive hypotheses. First, the presence of selection in the latter: why selection should preferentially decrease the number of AT→GC transitions compared to a mutation accumulation

experiment remains unclear but it is important to remind that many selective hypotheses have been proposed to explain the differences in the GC percentage among bacteria (Hershberg and Petrov 2010). Second, the different media used for growth: the LTEE was performed in minimal medium in which replication time is lower than in the rich media used in the mutation accumulation experiment (Lenski et al. 1991, Lee et al. 2012, Foster et al. 2013). Differences in oxidation levels may affect the way DNA is altered, thereby affecting mutational biases. Third, the use of two different *E. coli* strains: they may have slightly different mutation rates and biases. Mutation rate is a complex quantitative trait involving tens of enzymes whose patterns of expression may be variable from strain to strain. Indeed, there is a large variability of mutation rates among natural isolates of *E. coli* (Bjedov et al. 2003). If this last hypothesis were true, our experiments showed that transitions are the primary type of errors the polymerase is producing, and that the MMR system repairs them quite efficiently being better on AT→GC changes.

In the MMR- populations, mutability was affected by the DNA strand. Indeed, the *E. coli* genome is divided in two replichores in which the leading and lagging strand are inverted. Using the sequence of the leading strands only, we demonstrated that T and G residues were preferentially mutated than A and C residues, respectively. Such a bias was also observed in mutation accumulation experiments using *E. coli* (Lee et al. 2012). Since strand bias is observed in MMR- populations, it therefore results from the replication process. Two hypotheses can be suggested: mutations may arise from either mispairing by DNA polymerase or cytosine deamination when DNA is single-stranded which occurs more often on the leading strand while it is used as a template for the lagging strand (Frank and Lobry 1999). Lee et al. (2012) suggested that the observed frequency of mutations was too high to result from cytosine deamination only. Therefore, the strand bias might essentially result from mispairing during DNA replication. Whether this bias results from a preferential mispairing on the leading or lagging strand has been discussed previously (Lee et al. 2012) and even studied *in vitro* (Fijalkowska et al. 1998), and remains unsolved. However, this bias may constitute a driving force for the strand composition (Rocha et al. 2006).

At a local scale, we found that the nucleotide context strongly influences mutation probability. In the MMR- populations, the mutation probability for a single nucleotide can vary up to 31-fold depending on its neighboring context, the maximum difference being observed between the

GTA and TTG triplets. In the absence of MMR, this bias results from replication only. One of the biases was the over-representation of AT→GC mutations in NAC or GTN triplets. The neighboring G:C pairing is strong and its presence may stabilize the A:C or T:G mispairing. However, such a bias was not observed for the NAG or CTN triplets, in which the mutations should be stabilized in the same way. The stabilizing theory therefore cannot explain how the neighboring nucleotides influence the local mutability.

Finally, at a global scale, mutations are unevenly distributed throughout the genome, with fewer mutations around Ori and Ter. The Ara+3 population, that remained mutator for the longer time, showed clear patterns, in contrast to the other MMR- populations. In a previous study, the distribution of the mutations was found to be correlated with FIS-activated genes and anti-correlated with HU- and gyrase-activated genes, suggesting a link with DNA topology (Foster et al. 2013). This however does not seem to be the case in our data set as the relevant regions do not match with the distribution of nucleotide-associated proteins (HU, H-NS, FIS) and RNA polymerase throughout the genome (data not shown, Grainger et al. 2006). Interestingly, we found that the regions with high and low mutation numbers matched broadly with the four macrodomains defined in *E. coli* (Figure 2, Valens et al. 2004). The physical organization of these macrodomains may alter the accessibility of DNA to either chemicals or repair enzymes of NER or BER, making them less easy to repair (Maher et al. 2013). Highly structured DNA might also slow down the replication fork, increasing the mutation rate at these locations (Foster et al. 2013).

To the best of our knowledge, this study is the first to focus on mutation biases in *mutT*-populations at the genome level. In the two *mutT*- populations, mutations are biased towards AT→CG transversions, as previously described (Schaaper et al. 1997). Indeed, the MutT protein is involved in the detoxification of the oxidized form of dGTP, called 8-oxo-dGTP, which can pair equally with dCTP or dATP, leading to AT→CG transversions. Mutations in *mutT*-populations were much less biased by DNA strands, A and T being almost equally mutated even when the replicore organization was taken into account. However, a context effect was detected. Mutations were more frequent in TTN or NAA triplets, and less in GTN or NAC, with a maximum difference of 91-fold between the TTC and GTG triplets. This bias may result from both replication, meaning that the incorporation of 8-oxo-dGTP would be context dependent,

and DNA-repair by MutM and MutY, with differential repair efficiency depending on context. However, a study of mutation biases in *Salmonella typhimurium* did not detect any context bias in *mutY*-*mutM*- lineages (Lind et al. 2008). If repair system biases are conserved between the two species, this would imply a differential incorporation of 8-oxo-dGTP by the DNA polymerase. Finally, we could not find a consistent pattern of the distribution of mutations throughout the genome in the *mutT*- populations.

Our study differs from previous ones owing to the presence of selection. However, selection applies to a limited number of loci where the mutations have strong effects, with the other loci evolving almost neutrally (Lässig 2011, Maharjan et al. 2012). Moreover, mutational patterns and strand biases are conserved when we considered only neutral mutations (Figure S1). Triplet also showed similar biases, albeit with a lower statistical signal due to the lower number of synonymous mutations. Therefore, we can observe biases in synonymous codon changes that are similar to those observed in the triplets taking into account all mutations when the mutated nucleotide is located at the third position (data not shown).

We also showed that purifying selection was low, especially in the MMR- populations. This observation would imply that most mutations produced by hypermutator genotypes are not efficiently counter-selected, elevating their mutational load and thereby reducing their fitness. Hypermutator genotypes should then be doomed to disappear, or alternatively to reduce their mutation rate through the selection of anti-mutator mechanisms (Wielgoss et al. 2013, McDonald et al. 2013), or through the reacquisition of functional repair mechanisms through horizontal gene transfer (Denamur et al. 2000).

In conclusion, we observe that mutations are not all equivalent, and their probability depends on their localization on the genome, the DNA strand and the neighboring nucleotides. These biases result mostly from DNA replication and not selection and may result in a slow decay of the fitness of mutator lineages in the long term.

2.5 Materials & Methods

Evolution experiment and genome sequencing Twelve populations were initiated by a common ancestor *E. coli* B REL606 and propagated by serial daily transfers (1/100 dilutions) in a minimal medium containing 25 $\mu\text{g/mL}$ of glucose (DM25, Lenski et al. 1991). Individual clones were isolated from samples collected at regular time points in the populations (Table S1). The genomes of the sampled clones were sequenced on the Illumina Genome Analyzer platform at the Centre National de Séquenage, Genoscope, with one lane of single-ended 75-bp reads per genome. Reads were compared to the ancestral genome to identify mutations, using the BRESEQ pipeline (Deatherage and Barrick, 2014). For our analysis, we considered single events, meaning that conserved events, which were found in several clones of a population, were only counted once.

Mutational spectrum analysis The number of each type of synonymous mutations (AT→CG, AT→GC, AT→TA, CG→GC, CG→TA, CG→AT) in 4-fold degenerated codons was counted in each population, and divided by the total number of mutations in 4-fold degenerated codons in the respective population to obtain frequencies.

Strand bias Mutations were annotated using the sequence of the leading strands only. In the right replicore, from the replication origin (position 3,923,883 in the *E. coli* B genome sequence, Jeong et al. 2009) to the middle of the Ter region (position 1,473,484), mutations were reverse-complemented (read on the complementary strand). On the left replicore, mutations were read as such. The total number of each nucleotide was determined according to the same method. The number of mutated nucleotides was compared to their relative abundance in the genome using a Fisher exact test (Fisher 1922).

Distribution analysis using cumulative sums The distribution of the mutations throughout the genome was analyzed using a cumulative sum method. The GC→AT and AT→GC mutations in Ara-2, Ara-3, Ara-4 and Ara+3, and AT→CG mutations in Ara-1 and Ara+6 were split into 100 bins of 463,000 nucleotides each and numerated. The difference between the number of mutations in the bin and their mean was summed cumulatively throughout the genome

and plotted.

Context analysis Mutations were pooled by triplets XYZ, Y being the mutated nucleotide at the center, and X and Z its 5' and 3' neighboring nucleotide, respectively. All possible triplets were also numbered in the genome using a 3-nucleotide sliding window. The relative frequency of each mutated triplet was compared to its relative abundance to other triplets NYN in the genome using a Fisher exact test (Fisher 1922), N being any nucleotide. For example, when studying the number of mutations in the triplet ATA, only other triplets with a T at the central position were considered, allowing us to avoid spectrum biases. Mutations were considered on the leading strands only.

Statistics and plotting Mutational data were annotated and organized using Python 2.7 in Spyder. Statistics were done using Fisher exact test in R 3.2.1 (R Studio 0.97.551 software). The *P*-value threshold is low to account for Bonferroni correction when necessary. Graphs were also plotted in R, using the pheatmap and grid packages for heatmaps.

Bibliography

- Barrick J, Lenski R. 2009. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harb Symp Quant Biol.*, 74: 119–129.
- Benzer S 1961. On the topography of the genetic fine structure. *Proc Natl Acad Sci U S A.*, 47: 403–415.
- Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature*, 489: 513–518.
- Cooper DN, Krawczak M. 1990. The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum Genet.*, 85: 55–74.
- Couce A, Guelfo JR, Blazquez J. 2013. Mutational spectrum drives the rise of mutator bacteria. *PLoS Genet.*, 9: e1003167.

- Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, 274: 775–780.
- Deatherage DE, Barrick JE. 2014. *Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq*, volume 1151 of *Methods in Molecular Biology*, pages 165–188. Springer New York.
- Duret L 2009. Mutation patterns in the human genome: More variable than expected. *PLoS Biol.*, 7.
- Fijalkowska IJ, Schaaper RM, Jonczyk P. 2012. DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. *FEMS Microbiol Rev.*, 36: 1105–1121.
- Fisher RA 1922. On the interpretation of χ^2 from contingency tables, and the calculation of p. *J R Stat Soc.*, 85: 87–94.
- Foster PL, Hanson AJ, Lee H, Popodi EM, Tang H. 2013. On the mutational topology of the bacterial genome. *G3 (Bethesda)*, 3: 399–407.
- Frank AC, Lobry JR. 1999. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene*, 238: 65 – 77.
- Friedberg E, Walker G, Siede W. 1995. *DNA Repair and Mutagenesis*. ASM Press, Washington, DC.
- Grainger DC, Hurd D, Goldberg MD, Busby SJW. 2006. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.*, 34: 4642–4652.
- Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.*, 6: e1001115.
- Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A.*, 109: E2774–E2783.

- Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat.*, 138: 1315–1341.
- Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci U S A.*, 105: 17878–17883.
- Lobry JR 1996. Asymmetric substitution patterns in the two dna strands of bacteria. *Mol Biol Evol.*, 13: 660–665.
- Long H, Sung W, Miller SF, Ackerman MS, Doak TG, Lynch M. 2015. Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol.*, 7: 262–271.
- Lynch M 2010. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A.*, 107: 961–968.
- Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A.*, 105: 9272–9277.
- Maharjan RP, Liu B, Li Y, Reeves PR, Wang L, Ferenci T. 2012. Mutation accumulation and fitness in mutator subpopulations of *Escherichia coli*. *Biol Lett.*, 9.
- Maki H, Sekiguchi M. 1992. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature*, 355: 273–275.
- Niki H, Yamaichi Y, Hiraga S. 2000. Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.*, 14: 212–223.
- Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A. 1968. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A.*, 59: 598–605.
- Rocha EP, Touchon M, Feil EJ. 2006. Similar compositional biases are caused by very different mutational effects. *Genome Res.*, 16: 1537–1547.

- Rogozin IB, Pavlov YI. 2003. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res Rev Mutat Res.*, 544: 65 – 85.
- Schiffels S, Szöllosi GJ, Mustonen V, Lässig M. 2011. Emergent neutrality in adaptive asexual evolution. *Genetics*, 189: 1361–1375.
- Sharp PM, Emery LR, Zeng K. 2010. Forces that influence the evolution of codon bias. *Philos Trans R Soc Lond B Biol Sci.*, 365: 1203–1212.
- Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*, 387: 703–705.
- Sung W, Ackerman MS, Gout J, Miller SF, Williams E, Foster PL, Lynch M. 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. *Mol Biol Evol.*, 32: 1672–1683.
- Valens M, Penaud S, Rossignol M, Cornet F, Boccard F. 2004. Macrodomain organization of the *Escherichia coli* chromosome. *The EMBO Journal*, 23: 4330–4341.
- Wiser M, Ribeck N, Lenski R. 2013. Long-term dynamics of adaptation in asexual populations. *Science*, 342: 1364–1367.
- Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci U S A.*, 111: E2310–E2318.

2.6 Figure Legends

Fig. 1. Relative frequency of each type of mutational change in the 12 populations of the long-term evolution experiment. Synonymous mutations in 4-fold degenerated codons were numbered and pooled by mutational spectrum, then divided by their total number. The total number of mutations and respective proportions of each spectrum are given in Table 1. Non-mutator populations were pooled together.

Fig. 2. Distribution of mutations throughout the genome. The difference between the number of mutations in each bin and their average was summed cumulatively. An increase or decrease of the cumulative sum indicates a region where mutations tend to be more or less frequent, respectively, than average. MMR- populations are plotted in purple and *mutT*- populations in green. Macrodomains are indicated by the background coloring: Right, Ter, Left and Ori regions from left to right, respectively. White represents the two unstructured macrodomains (Valens et al. 2004).

Fig. 3. Nucleotide context bias. The proportion of each mutated triplets was compared to their respective proportion in the genome using a Fisher exact test. Positive biases are indicated in red/orange, and negative ones in blue. From the darkest to the lightest, $p < 5 \times 10^{-9}$, $p < 5 \times 10^{-6}$ and $p < 5 \times 10^{-4}$. The mutated base is indicated as the x axis, while the 5' and 3' neighboring bases are shown on the left and right of each heat map, respectively.

2.7 Tables

Table 1. Mutation summary for each population. Hypermutator populations are written in bold. Mutational spectra are listed in percentage.

Population*	Non Synonymous	Synonymous	Intergenic	Total	AT->CG	AT->GC	AT->TA	CG->GC	CG->TA	CG->AT	Ts/Tv
Ara-1	1355	266	265	1886	0.885	0.004	0.004	-	0.012	0.095	0.02
Ara-2	1295	736	235	2266	0.004	0.449	0.009	0.013	0.524	0.002	32.85
Ara-3	611	323	127	1061	0.005	0.299	0.005	0.019	0.645	0.028	11.79
Ara-4	914	533	161	1608	-	0.448	-	0.017	0.523	0.012	26.83
Ara-5	78	17	21	116	-	-	0.083	-	0.667	0.250	0.86
Ara-6	61	9	16	86	0.111	0.111	-	-	0.444	0.333	0.89
Ara+1	69	14	22	105	0.333	0.167	0.083	-	0.250	0.167	0.63
Ara+2	40	7	18	65	-	0.167	-	0.167	0.333	0.333	0.96
Ara+3	1633	889	304	2826	-	0.457	0.004	0.006	0.528	0.006	44.04
Ara+4	42	9	21	72	0.333	-	-	-	0.556	0.111	0.82
Ara+5	55	10	12	77	0.333	-	0.167	-	0.333	0.167	1.17
Ara+6	3558	578	657	4793	0.955	0.002	-	-	0.009	0.034	0.01

Mutational spectra are given as percentages.

*Hypermutator populations are indicated in bold.

Table 2. Strand mutational bias. Excess of one nucleotide compared to its complementary was tested using Fisher exact test. *p-values* are listed in the table, without taking account of the replicore on the two left columns, and taking account of the replicore on the two right columns. Threshold for *p-values* was set at 10^{-4} to account for Bonferroni correction for multiple tests. Significant *p-values* are listed in bold.

Population	No replicore		With replicore	
-	A/T bias	C/G bias	A/T bias	C/G bias
Ara+1	3.30E-02	8.88E-01	5.05E-01	1
Ara+2	5.01E-02	1	2.94E-01	1
Ara+3	3.50E-01	3.98E-01	7.39E-64	2.22E-90
Ara+4	5.33E-01	1	6.43E-01	2.08E-01
Ara+5	2.96E-01	7.66E-01	7.27E-01	1.66E-02
Ara+6	2.55E-01	6.90E-02	3.69E-04	4.53E-01
Ara-1	1	4.30E-02	8.64E-03	1.33E-02
Ara-2	2.86E-01	1.50E-01	1.06E-45	1.53E-49
Ara-3	2.17E-01	2.78E-01	7.34E-11	6.96E-33
Ara-4	4.70E-01	4.91E-01	1.18E-24	2.65E-36
Ara-5	2.62E-01	5.38E-01	8.89E-01	2.68E-01
Ara-6	8.71E-01	6.71E-01	5.12E-01	2.54E-01

NOTE. - Excess of one nucleotide compared to its complementary was tested using Fisher exact test. *P-values* are listed in the table, either without or with taking into account the replicores. Threshold for *p-values* was set at 10^{-4} to account for Bonferroni correction for multiple tests.

2.8 Figures

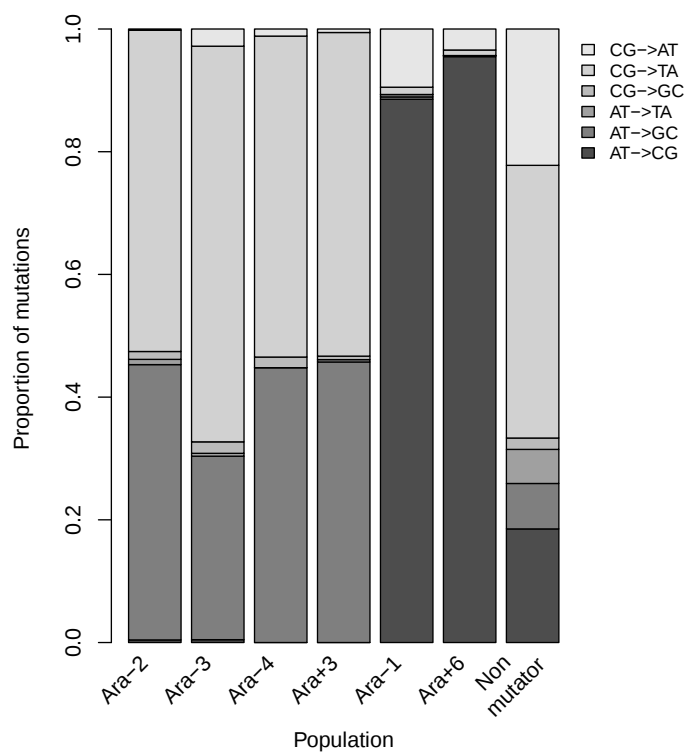


Fig. 1.

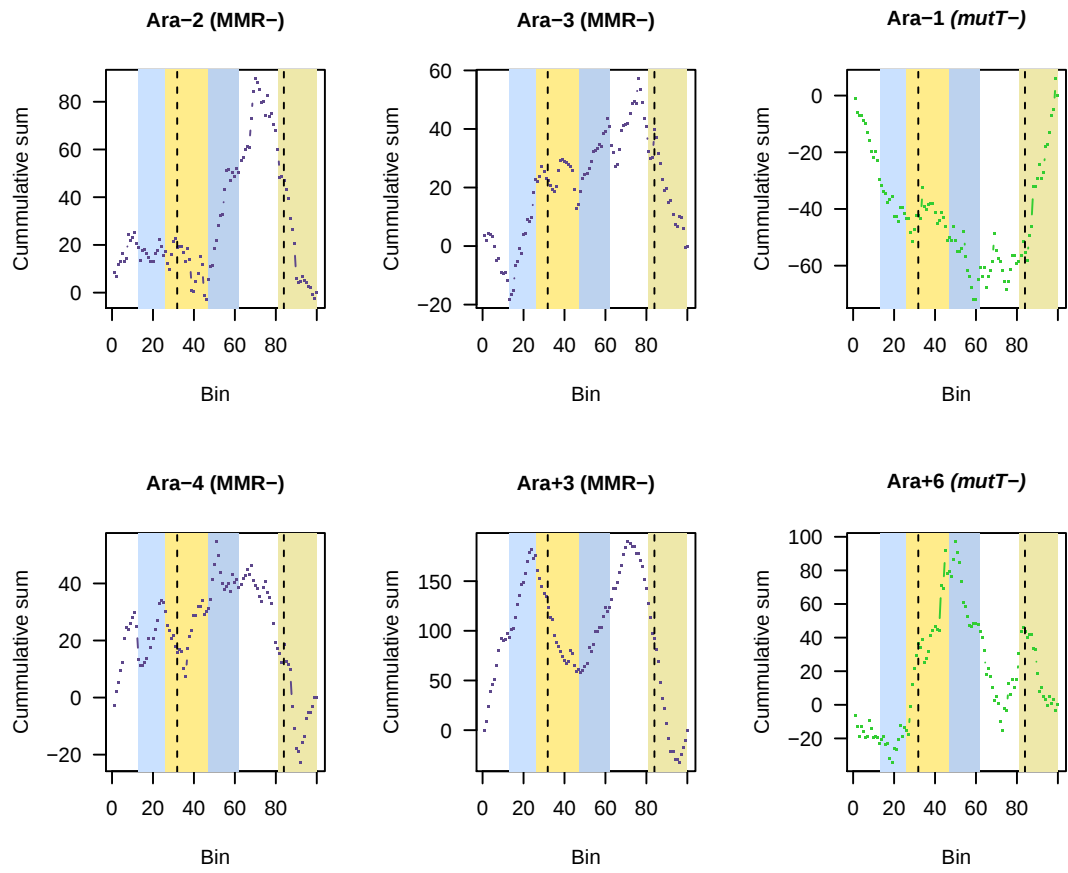


Fig. 2.

BIAIS DE MUTATIONS DANS LES POPULATIONS HYPERMUTATRICES DE
L'EXPÉRIENCE D'ÉVOLUTION À LONG TERME

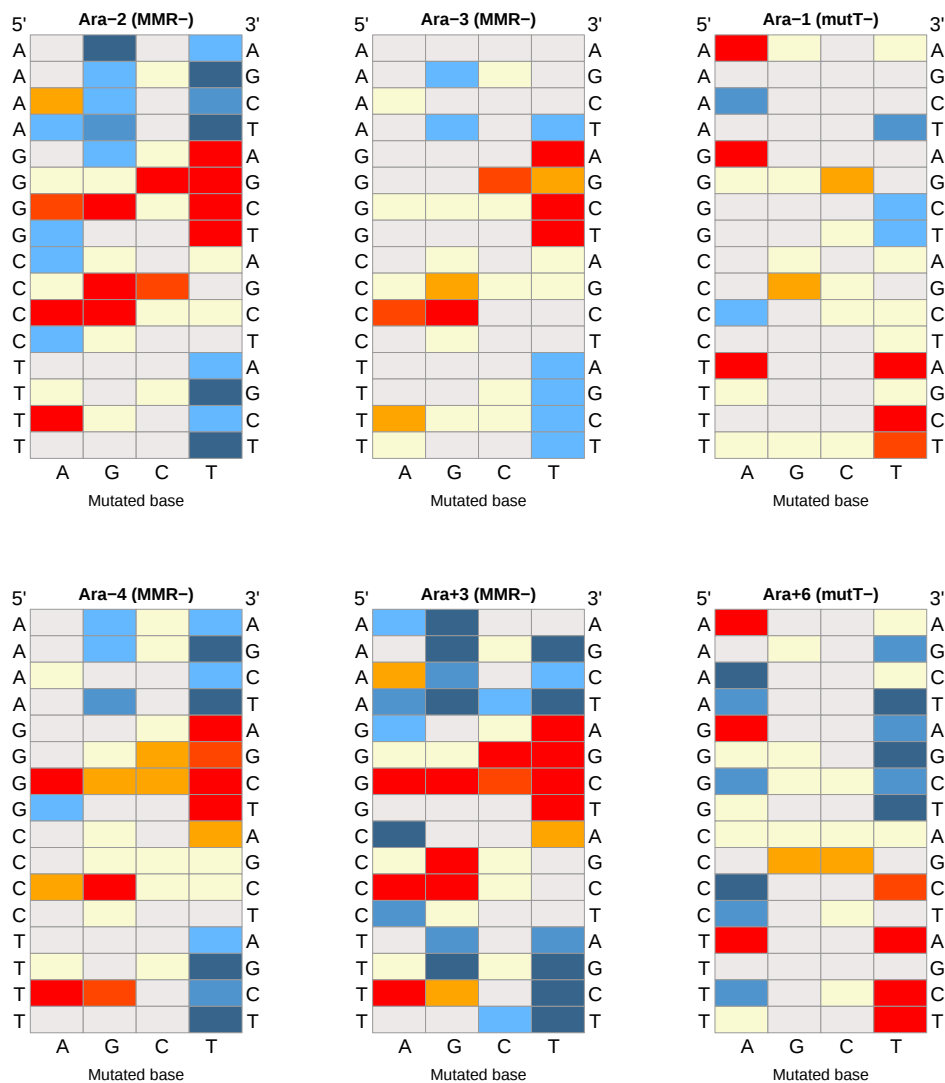


Fig. 3.

2.9 Supplementary Information

Mutational patterns in *Escherichia coli* hypermutator populations during a long-term evolution experiment

Larissa Caudwell, Dominique Schneider and Olivier Tenaillon

Table S1. List of strains used in this study

Table S2. Comparison, using *t*-tests, of the number of mutations in four chromosomal regions (see text)

Figure S1. Nucleotide context bias for synonymous mutations in the Ara+3 and Ara+6 populations.

Table S1. List of strains used in this study

Population*	Time point	Strain name
Ara+1	500	RELara+1_500gen_768A
Ara+1	500	RELara+1_500gen_768B
Ara+1	1000	RELara+1_1000gen_REL958A
Ara+1	1000	RELara+1_1000gen_958B
Ara+1	1500	RELara+1_1500gen_1062A
Ara+1	1500	RELara+1_1500gen_1062B
Ara+1	2000	RELara+1_2000gen_1158A
Ara+1	2000	RELara+1_2000gen_1158C
Ara+1	5000	RELara+1_5000gen_2173A
Ara+1	5000	RELara+1_5000gen_2173B
Ara+1	10,000	RELara+1_10000gen_4530A
Ara+1	10,000	RELara+1_10000gen_4530B
Ara+1	15,000	RELara+1_15000gen_7183A
Ara+1	15,000	RELara+1_15000gen_7183B
Ara+1	20,000	RELara+1_20000gen_9282A
Ara+1	20,000	RELara+1_20000gen_9282B
Ara+1	30,000	RELara+1_30000gen_10450
Ara+1	30,000	RELara+1_30000gen_10451
Ara+1	40,000	RELara+1_40000gen_11008
Ara+1	40,000	RELara+1_40000gen_11009
Ara+1	50,000	RELara+1_50000gen_11392
Ara+1	50,000	RELara+1_50000gen_11393
Ara+2	500	RELara+2_500gen_769A
Ara+2	500	RELara+2_500gen_769B
Ara+2	1000	RELara+2_1000gen_959A
Ara+2	1000	RELara+2_1000gen_959B
Ara+2	1500	RELara+2_1500gen_1063A

Ara+2	1500	RELARA+2_1500gen_1063B
Ara+2	2000	RELARA+2_2000gen_1159A
Ara+2	2000	RELARA+2_2000gen_1159B
Ara+2	5000	RELARA+2_5000gen_2174A
Ara+2	5000	RELARA+2_5000gen_2174B
Ara+2	10,000	RELARA+2_10000gen_4531A
Ara+2	10,000	RELARA+2_10000gen_4531B
Ara+2	15,000	RELARA+2_15000gen_7184A
Ara+2	15,000	RELARA+2_15000gen_7184B
Ara+2	20,000	RELARA+2_20000gen_8601A
Ara+2	20,000	RELARA+2_20000gen_8601B
Ara+2	30,000	RELARA+2_30000gen_10407
Ara+2	30,000	RELARA+2_30000gen_10408
Ara+2	40,000	RELARA+2_40000gen_10950
Ara+2	40,000	RELARA+2_40000gen_10951
Ara+2	50,000	RELARA+2_50000gen_11342
Ara+2	50,000	RELARA+2_50000gen_11343
Ara+3	500	RELARA+3_500gen_770A
Ara+3	500	RELARA+3_500gen_770B
Ara+3	1000	RELARA+3_1000gen_960A
Ara+3	1000	RELARA+3_1000gen_960B
Ara+3	1500	RELARA+3_1500gen_1064A
Ara+3	1500	RELARA+3_1500gen_1064B
Ara+3	2000	RELARA+3_2000gen_1160A
Ara+3	2000	RELARA+3_2000gen_1160B
Ara+3	5000	RELARA+3_5000gen_2175A
Ara+3	5000	RELARA+3_5000gen_2175B
Ara+3	10,000	RELARA+3_10000gen_4532A
Ara+3	10,000	RELARA+3_10000gen_4532B

BIAIS DE MUTATIONS DANS LES POPULATIONS HYPERMUTATRICES DE
L'EXPÉRIENCE D'ÉVOLUTION À LONG TERME

Ara+3	15,000	RELara+3_15000gen_7185A
Ara+3	15,000	RELara+3_15000gen_7185B
Ara+3	20,000	RELara+3_20000gen_8602A
Ara+3	20,000	RELara+3_20000gen_8602B
Ara+3	30,000	RELara+3_30000gen_10410
Ara+3	30,000	RELara+3_30000gen_10411
Ara+3	40,000	RELara+3_40000gen_10953
Ara+3	40,000	RELara+3_40000gen_10954
Ara+3	50,000	RELara+3_50000gen_11345
Ara+3	50,000	RELara+3_50000gen_11346
Ara+4	500	RELara+4_500gen_771A
Ara+4	500	RELara+4_500gen_771B
Ara+4	1000	RELara+4_1000gen_961A
Ara+4	1000	RELara+4_1000gen_961B
Ara+4	1500	RELara+4_1500gen_1065A
Ara+4	1500	RELara+4_1500gen_1065B
Ara+4	2000	RELara+4_2000gen_1161A
Ara+4	2000	RELara+4_2000gen_1161B
Ara+4	5000	RELara+4_5000gen_2176A
Ara+4	5000	RELara+4_5000gen_2176B
Ara+4	10,000	RELara+4_10000gen_4533A
Ara+4	10,000	RELara+4_10000gen_4533B
Ara+4	15,000	RELara+4_15000gen_7186A
Ara+4	15,000	RELara+4_15000gen_7186B
Ara+4	20,000	RELara+4_20000gen_8603A
Ara+4	20,000	RELara+4_20000gen_8603B
Ara+4	30,000	RELara+4_30000gen_10414
Ara+4	30,000	RELara+4_30000gen_10415
Ara+4	40,000	RELara+4_40000gen_10956

Ara+4	40,000	RELara+4_40000gen_10957
Ara+4	50,000	RELara+4_50000gen_11348
Ara+4	50,000	RELara+4_50000gen_11349
Ara+5	500	RELara+5_500gen_772A
Ara+5	500	RELara+5_500gen_772B
Ara+5	1000	RELara+5_1000gen_962A
Ara+5	1000	RELara+5_1000gen_962B
Ara+5	1500	RELara+5_1500gen_1066A
Ara+5	1500	RELara+5_1500gen_1066B
Ara+5	2000	RELara+5_2000gen_1162A
Ara+5	2000	RELara+5_2000gen_1162B
Ara+5	5000	RELara+5_5000gen_2177A
Ara+5	5000	RELara+5_5000gen_2177B
Ara+5	10,000	RELara+5_10000gen_4534A
Ara+5	10,000	RELara+5_10000gen_4534B
Ara+5	15,000	RELara+5_15000gen_7187A
Ara+5	15,000	RELara+5_15000gen_7187B
Ara+5	20,000	RELara+5_20000gen_8604A
Ara+5	20,000	RELara+5_20000gen_8604B
Ara+5	30,000	RELara+5_30000gen_10432
Ara+5	30,000	RELara+5_30000gen_10433
Ara+5	40,000	RELara+5_40000gen_10982
Ara+5	40,000	RELara+5_40000gen_10983
Ara+5	50,000	RELara+5_50000gen_11367
Ara+5	50,000	RELara+5_50000gen_11368
Ara+6	500	RELara+6_500gen_773A
Ara+6	500	RELara+6_500gen_773B
Ara+6	1000	RELara+6_1000gen_963A
Ara+6	1000	RELara+6_1000gen_963B

BIAIS DE MUTATIONS DANS LES POPULATIONS HYPERMUTATRICES DE
L'EXPÉRIENCE D'ÉVOLUTION À LONG TERME

Ara+6	1500	RELara+6_1500gen_1067A
Ara+6	1500	RELara+6_1500gen_1067B
Ara+6	2000	RELara+6_2000gen_1163A
Ara+6	2000	RELara+6_2000gen_1163B
Ara+6	5000	RELara+6_5000gen_2178A
Ara+6	5000	RELara+6_5000gen_2178B
Ara+6	10,000	RELara+6_10000gen_4535A
Ara+6	10,000	RELara+6_10000gen_4535B
Ara+6	15,000	RELara+6_15000gen_7188A
Ara+6	15,000	RELara+6_15000gen_7188B
Ara+6	20,000	RELara+6_20000gen_8605A
Ara+6	20,000	RELara+6_20000gen_8605B
Ara+6	30,000	RELara+6_30000gen_10417
Ara+6	30,000	RELara+6_30000gen_10418
Ara+6	40,000	RELara+6_40000gen_10985
Ara+6	40,000	RELara+6_40000gen_10986
Ara+6	50,000	RELara+6_50000gen_11370
Ara+6	50,000	RELara+6_50000gen_11371
Ara-1	500	RELara-1_500gen_762B
Ara-1	500	RELara-1_500gen_762C
Ara-1	1000	RELara-1_1000gen_964B
Ara-1	1000	RELara-1_1000gen_964C
Ara-1	1500	RELara-1_1500gen_1068B
Ara-1	1500	RELara-1_1500gen_1068C
Ara-1	2000	RELara-1_2000gen_1164A
Ara-1	2000	RELara-1_2000gen_1164B
Ara-1	5000	RELara-1_5000gen_2179A
Ara-1	5000	RELara-1_5000gen_2179B
Ara-1	10,000	RELara-1_10000gen_4536A

Ara-1	10,000	RELARA-1_10000gen_4536B
Ara-1	15,000	RELARA-1_15000gen_7177A
Ara-1	15,000	RELARA-1_15000gen_7177B
Ara-1	20,000	RELARA-1_20000gen_8593A
Ara-1	20,000	RELARA-1_20000gen_8593B
Ara-1	30,000	RELARA-1_30000gen_10391
Ara-1	30,000	RELARA-1_30000gen_10392
Ara-1	40,000	RELARA-1_40000gen_10938
Ara-1	40,000	RELARA-1_40000gen_10939
Ara-1	50,000	RELARA-1_50000gen_11330
Ara-1	50,000	RELARA-1_50000gen_11331
Ara-2	500	RELARA-2_500gen_763A
Ara-2	500	RELARA-2_500gen_763B
Ara-2	1000	RELARA-2_1000gen_965A
Ara-2	1000	RELARA-2_1000gen_965B
Ara-2	1500	RELARA-2_1500gen_1069A
Ara-2	1500	RELARA-2_1500gen_1069B
Ara-2	2000	RELARA-2_2000gen_1165A
Ara-2	2000	RELARA-2_2000gen_2000.4
Ara-2	5000	RELARA-2_5000gen_2180A
Ara-2	5000	RELARA-2_5000gen_2180B
Ara-2	10,000	RELARA-2_10000gen_4537A
Ara-2	10,000	RELARA-2_10000gen_4537B
Ara-2	15,000	RELARA-2_15000gen_7178A
Ara-2	15,000	RELARA-2_15000gen_7178B
Ara-2	20,000	RELARA-2_20000gen_20K-S1
Ara-2	20,000	RELARA-2_20000gen_20K-LA
Ara-2	30,000	RELARA-2_30000gen_30K-S1
Ara-2	30,000	RELARA-2_30000gen_30K-L1

BIAIS DE MUTATIONS DANS LES POPULATIONS HYPERMUTATRICES DE
L'EXPÉRIENCE D'ÉVOLUTION À LONG TERME

Ara-2	40,000	RELara-2_40000gen_11035
Ara-2	40,000	RELara-2_40000gen_11036
Ara-2	50,000	RELara-2_50000gen_11333
Ara-2	50,000	RELara-2_50000gen_11335
Ara-3	500	RELara-3_500gen_764A
Ara-3	500	RELara-3_500gen_764B
Ara-3	1000	RELara-3_1000gen_966A
Ara-3	1000	RELara-3_1000gen_966B
Ara-3	1500	RELara-3_1500gen_1070A
Ara-3	1500	RELara-3_1500gen_1070B
Ara-3	2000	RELara-3_2000gen_1166A
Ara-3	2000	RELara-3_2000gen_1166B
Ara-3	5000	RELara-3_5000gen_2181A
Ara-3	5000	RELara-3_5000gen_ZDB409
Ara-3	10,000	RELara-3_10000gen_4538A
Ara-3	10,000	RELara-3_10000gen_ZDB429
Ara-3	15,000	RELara-3_15000gen_ZDB446
Ara-3	15,000	RELara-3_15000gen_7179B
Ara-3	20,000	RELara-3_20000gen_ZDB458
Ara-3	20,000	RELara-3_20000gen_ZDB464
Ara-3	30,000	RELara-3_30000gen_ZDB16
Ara-3	30,000	RELara-3_30000gen_ZDB357
Ara-3	40,000	RELara-3_40000gen_10988
Ara-3	40,000	RELara-3_40000gen_10979
Ara-3	50,000	RELara-3_50000gen_11364
Ara-3	50,000	RELara-3_50000gen_11365
Ara-4	500	RELara-4_500gen_765A
Ara-4	500	RELara-4_500gen_765B
Ara-4	1000	RELara-4_1000gen_967A

Ara-4	1000	RELARA-4_1000gen_967B
Ara-4	1500	RELARA-4_1500gen_1071A
Ara-4	1500	RELARA-4_1500gen_1071B
Ara-4	2000	RELARA-4_2000gen_1167A
Ara-4	2000	RELARA-4_2000gen_1167B
Ara-4	5000	RELARA-4_5000gen_2182A
Ara-4	5000	RELARA-4_5000gen_2182B
Ara-4	10,000	RELARA-4_10000gen_4539A
Ara-4	10,000	RELARA-4_10000gen_4539B
Ara-4	15,000	RELARA-4_15000gen_7180A
Ara-4	15,000	RELARA-4_15000gen_7180B
Ara-4	20,000	RELARA-4_20000gen_8596A
Ara-4	20,000	RELARA-4_20000gen_8596B
Ara-4	30,000	RELARA-4_30000gen_10401
Ara-4	30,000	RELARA-4_30000gen_10402
Ara-4	40,000	RELARA-4_40000gen_10944
Ara-4	40,000	RELARA-4_40000gen_10945
Ara-4	50,000	RELARA-4_50000gen_11336
Ara-4	50,000	RELARA-4_50000gen_11337
Ara-5	500	RELARA-5_500gen_766A
Ara-5	500	RELARA-5_500gen_766B
Ara-5	1000	RELARA-5_1000gen_968A
Ara-5	1000	RELARA-5_1000gen_968B
Ara-5	1500	RELARA-5_1500gen_1072A
Ara-5	1500	RELARA-5_1500gen_1072B
Ara-5	2000	RELARA-5_2000gen_1168A
Ara-5	2000	RELARA-5_2000gen_1168B
Ara-5	5000	RELARA-5_5000gen_2183A
Ara-5	5000	RELARA-5_5000gen_2183B

BIAIS DE MUTATIONS DANS LES POPULATIONS HYPERMUTATRICES DE
L'EXPÉRIENCE D'ÉVOLUTION À LONG TERME

Ara-5	10,000	RELara-5_10000gen_4540A
Ara-5	10,000	RELara-5_10000gen_4540B
Ara-5	15,000	RELara-5_15000gen_7181A
Ara-5	15,000	RELara-5_15000gen_7181B
Ara-5	20,000	RELara-5_20000gen_8597A
Ara-5	20,000	RELara-5_20000gen_8597B
Ara-5	30,000	RELara-5_30000gen_10404
Ara-5	30,000	RELara-5_30000gen_10405
Ara-5	40,000	RELara-5_40000gen_10947
Ara-5	40,000	RELara-5_40000gen_10948
Ara-5	50,000	RELara-5_50000gen_11339
Ara-5	50,000	RELara-5_50000gen_11340
Ara-6	500	RELara-6_500gen_767A
Ara-6	500	RELara-6_500gen_767B
Ara-6	1000	RELara-6_1000gen_969A
Ara-6	1000	RELara-6_1000gen_969B
Ara-6	1500	RELara-6_1500gen_1073A
Ara-6	1500	RELara-6_1500gen_1073B
Ara-6	2000	RELara-6_2000gen_1169A
Ara-6	2000	RELara-6_2000gen_1169B
Ara-6	5000	RELara-6_5000gen_2184A
Ara-6	5000	RELara-6_5000gen_2184B
Ara-6	10,000	RELara-6_10000gen_4541A
Ara-6	10,000	RELara-6_10000gen_4541B
Ara-6	15,000	RELara-6_15000gen_7182A
Ara-6	15,000	RELara-6_15000gen_7182B
Ara-6	20,000	RELara-6_20000gen_8598A
Ara-6	20,000	RELara-6_20000gen_8598B
Ara-6	30,000	RELara-6_30000gen_10447

Ara-6	30,000	RELAra-6_30000gen_10448
Ara-6	40,000	RELAra-6_40000gen_11005
Ara-6	40,000	RELAra-6_40000gen_11006
Ara-6	50,000	RELAra-6_50000gen_11389
Ara-6	50,000	RELAra-6_50000gen_11390

The accession numbers for the genome sequences of the strains are given in XXXX.

*Hypermutator populations are highlighted in bold.

Table S2. T-test Comparison, using *t*-tests, of the number of mutations between the defined chromosomal 4 regions of the chromosome (see text). *p*-values are listed in the table.

Ara-2	95 → 24	25 → 47	48 → 71	72 → 94	
95 → 24		1.00	0.88	0.19	0.08
25 → 47		0.88	1.00	0.38	0.12
48 → 71		0.19	0.38	1.00	5.63E-03
72 → 94		0.08	0.12	5.63E-03	1.00
Ara-3	95 → 24	25 → 47	48 → 71	72 → 94	
95 → 24		1.00	0.47	0.19	0.39
25 → 47		0.47	1.00	0.79	0.18
48 → 71		0.19	0.79	1.00	0.04
72 → 94		0.39	0.18	0.04	1.00
Ara-4	95 → 24	25 → 47	48 → 71	72 → 94	
95 → 24		1.00	0.68	0.89	0.07
25 → 47		0.68	1.00	0.80	0.17
48 → 71		0.89	0.80	1.00	0.10
72 → 94		0.07	0.17	0.10	1.00
Ara+3	95 → 24	25 → 47	48 → 71	72 → 94	
95 → 24		1.00	9.97E-05	0.16	5.46E-07
25 → 47		9.97E-05	1.00	1.14E-03	0.14
48 → 71		0.16	1.14E-03	1.00	5.11E-06
72 → 94		5.46E-07	0.14	5.11E-06	1.00
Ara-1	95 → 24	25 → 47	48 → 71	72 → 94	
95 → 24		1.00	0.86	0.15	0.90
25 → 47		0.86	1.00	0.18	0.76
48 → 71		0.15	0.18	1.00	0.13
72 → 94		0.90	0.76	0.13	1.00
Ara+6	95 → 24	25 → 47	48 → 71	72 → 94	
95 → 24		1.00	0.81	0.04	0.15
25 → 47		0.81	1.00	0.02	0.10
48 → 71		0.04	0.02	1.00	0.84
72 → 94		0.15	0.10	0.84	1.00

p-values are listed in the table. The bin numbers are given (see text).

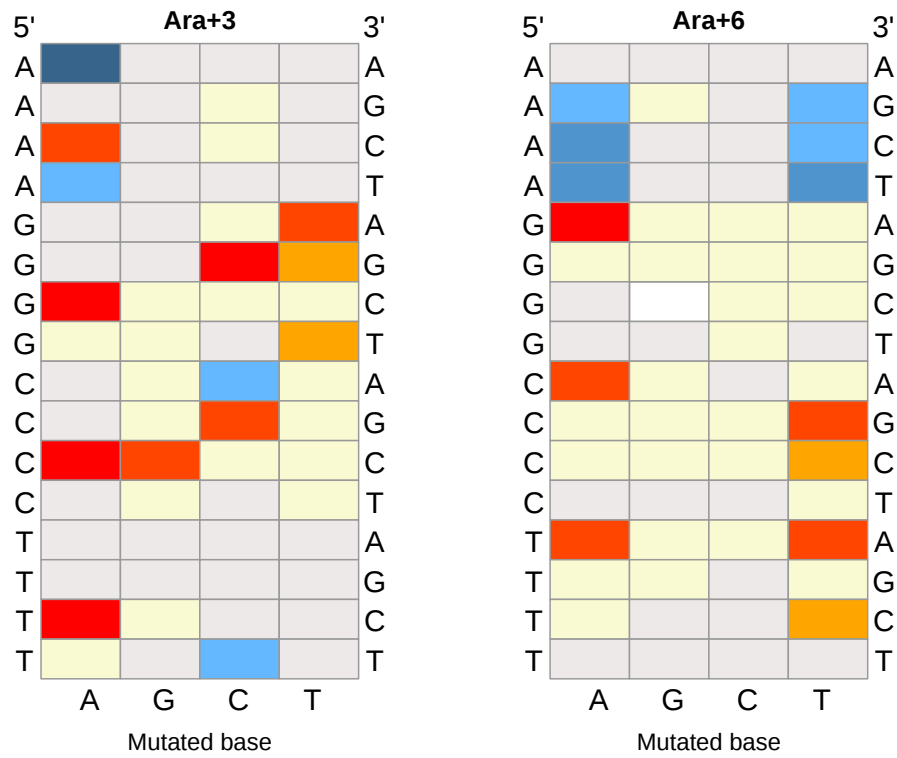


Figure S1. Nucleotide context bias for synonymous mutations in the Ara+3 and Ara+6 populations. Positive biases are indicated in red/orange, and negative ones in blue. From the darkest to the lightest, $p < 5 \times 10^{-9}$, $p < 5 \times 10^{-6}$ and $p < 5 \times 10^{-4}$. The mutated base is indicated as the x axis, while the 5' and 3' neighboring bases are shown on the left and right of each heat map, respectively.

Chapitre IV

Discussion

Depuis 1988, l'expérience d'évolution à long terme de Richard Lenski avec la bactérie *E. coli* a permis la mise en évidence et une compréhension approfondie de nombreux processus évolutifs, comme l'interférence clonale (Gerrish et Lenski 1998), la réorganisation de réseaux de régulation (Hindré et al. 2012, Philippe et al. 2007), l'émergence de polymorphismes stables à long terme (Blount et al. 2008, 2012, Rozen et Lenski 2000, Plucain et al. 2014), et le parallélisme phénotypique et génétique (Elena et Lenski 2003). En s'appuyant sur cette expérience, cette thèse apporte un nouveau regard sur l'évolution des taux de mutation et la distribution des mutations dans le génome.

Bien que ces deux questions aient été soulevées depuis près d'un siècle (dès 1930 pour les taux de mutation et 1960 pour la distribution des mutations), aucune n'avait encore été étudiée sur une expérience d'évolution sur un temps aussi long, la présente expérience étant la plus longue jamais initiée (depuis près de trente ans maintenant). De plus, le développement des techniques de séquençage de génome à haut débit sur ces dernières années a permis d'accéder aux mutations dans toutes les populations à différents temps d'évolution. Ainsi, j'ai pu mettre en évidence que le taux de mutation est un caractère variable et soumis à sélection naturelle, et que la distribution des mutations varie au sein d'un même génome. Dans une première partie, j'ai suivi l'évolution des taux de mutation au cours du temps dans une population qui a développé un caractère hypermutateur. Après une première forte augmentation des taux de mutation, ceux-ci tendent à diminuer au cours du temps au travers d'un processus d'essais/erreurs. Dans une deuxième partie, j'ai étudié la distribution des mutations dans les six populations connues pour avoir développé un phénotype hypermutateur, où j'ai pu mettre en évidence des biais mutationnels quelle que soit l'origine génétique du phénotype hypermutateur considéré. Les résultats obtenus dans cette thèse apportent une meilleure compréhension des mécanismes évolutifs liés aux mutations, ayant de fortes implications dans l'émergence de résistances aux antibiotiques ou l'établissement d'infections chroniques. La compréhension des processus mutationnels est essentielle pour comprendre les mécanismes évolutifs puisqu'il s'agit d'un des substrats de la sélection naturelle.

1 Considérations évolutives : évolution de l'évolution

Les mécanismes permettant l'évolution, tels que les mutations, la recombinaison ou les réarrangements peuvent eux aussi varier, être régulés et sélectionnés. Ils sont donc eux-mêmes soumis à l'évolution : on parle alors d'évolution de l'évolution. L'étude de la dynamique des taux de mutation s'inscrit donc dans ce cadre.

1.1 Dynamique au cours du temps

Le taux de mutation est susceptible de varier au cours du temps et cela a été observé chez des microorganismes aussi bien dans des conditions de laboratoire (Wielgoss et al. 2013, McDonald et al. 2012, ...) que dans des conditions de l'environnement (Oliver et al. 2000, Marvig et al. 2015, ...). Les taux de mutation résultent d'un équilibre entre stabilité et adaptabilité, deux forces pouvant jouer sur des échelles de temps différentes : à court terme pour l'adaptabilité ce qui se traduit par l'émergence rapide d'hypermutateurs par entraînement génétique (auto-stop) avec des mutations bénéfiques (souvent produites d'ailleurs par le génotype hypermutateur), et à plus long terme pour la stabilité reposant alors sur la contre-sélection de clones ayant accumulé une charge mutationnelle trop lourde.

Ainsi, j'ai observé une augmentation rapide des taux de mutation dans la population Ara+6, en quelques milliers de générations (moins de deux ans de propagation en laboratoire), tandis que leur diminution est plus lente, s'étendant sur plusieurs dizaines de milliers de générations. De plus, et contrairement à ce qui avait été décrit dans les études précédentes, comme dans la population Ara-1 (Wielgoss et al. 2013) et diverses simulations (Leigh 1970, Johnson 1999), cette diminution des taux de mutation n'est pas linéaire. Elle se produit en effet par un processus qui pourrait être qualifié d' « essais et erreurs », au cours duquel certains mécanismes de compensation se produisent mais ne sont pas fixés au sein des populations. Au cours de cette dynamique, le taux de mutation génomique, qui représente un taux moyen de mutation relatif au temps écoulé depuis la modification du taux de mutation, diminue par paliers, tandis que le taux de mutation mesuré phénotypiquement, qui lui donne une mesure du taux de mutation instantané au sein d'un clone, varie considérablement que ce soit dans les clones isolés ou dans les mélanges de population.

Ces fortes variations peuvent résulter du phénomène d'interférence clonale entre des clones à

taux de mutation élevés, produisant plus de mutations bénéfiques, et des clones ayant des mécanismes de compensation et donc une charge mutationnelle moindre. Ces derniers sont d'autant plus avantagés quand les populations sont déjà adaptées à leur environnement, c'est-à-dire à des temps plus tardifs d'évolution, lorsque les mutations bénéfiques accessibles diminuent et la charge mutationnelle augmente. Ce phénomène d'interférence clonale est d'autant plus fort avec le phénotype mutateur qui produit une grande diversité de clones qui vont alors interférer les uns avec les autres. Ainsi, le taux de mutation est un paramètre finement modulable au cours du temps et soumis lui-même à pression de sélection. Cependant, il peut aussi révéler des variations dans l'espace, c'est-à-dire le long du génome.

1.2 Dynamique spatiale, influence du contexte

Les analyses informatiques des mutations trouvées dans les 6 lignées hypermutatrices de l'expérience d'évolution au long terme ont démontré que la distribution des mutations était biaisée à plusieurs niveaux : le spectre de mutations largement dépendant du mécanisme génétique à l'origine du phénotype hypermutateur (c'est-à-dire du type de gène de réparation de l'ADN qui est altéré), la distribution topologique en fonction du brin et de la localisation, et enfin l'entourage nucléotidique. Dans les populations déficientes pour le système de réparation des mésappariements (MMR-), ces biais résultent des mécanismes de réplication, tandis que dans les populations *mutT*-, ils peuvent résulter également des mécanismes de réparations par les protéines MutM et MutY. Des biais de réplication entre les brins d'ADN ont été observés *in vitro* (Fijalkowska et al. 1998) et confirment la différence existant entre la mutabilité des brins direct et indirect. Au niveau de l'influence des nucléotides avoisinant, il a fréquemment été suggéré que les mutations peuvent être stabilisées par un appariement G:C en 3', qui est plus stable qu'un appariement A:T (Rogozin et Pavlov 2003, Lee et al. 2012). Cependant, il semblerait que les nucléotides C et G ne soient pas interchangeable et n'aient pas la même influence sur la mutabilité d'un locus. Ainsi, il est probable que d'autres mécanismes entrent en jeu. Une précédente étude des biais de mutations chez *E. coli* précisait le rôle des résidus cytosine méthylés comme points chauds de mutation, due à une désamination spontanée (Lee et al. 2012). Ici, la souche d'*E. coli* B utilisée ne présente pas de méthylation des séquences CCAGG (Coulondre et al. 1978), et donc aucun

biais n'est visible à ce niveau.

Les études mettant en évidence les biais mutationnels au niveau génomique se sont multipliées ces dernières années, grâce au développement du séquençage de génomes à haut débit (Lee et al. 2012, Foster et al. 2013, Sung et al. 2015, Long et al. 2015...). Hormis des différences au niveau de la déméthylation des résidus cytosine et du contexte des transitions CG→TA, j'ai également retrouvé des similitudes avec la précédente étude ayant été réalisée chez *E. coli* au travers d'une expérience d'accumulation de mutations (Lee et al. 2012), avec un biais important pour les triplets XAC et GTX. De façon intéressante, des biais sont retrouvés chez *P. fluorescens*, mais touchent les triplets CAX et XTG (Long et al. 2015). Les biais sont différents et moins marqués chez *B. subtilis* (Sung et al. 2015). Ainsi, les biais mutationnels dépendent non seulement du type de mutateur, mais aussi du type de bactéries analysées. Ces différences pourraient aussi refléter la sélection, mais les mêmes biais sont retrouvés dans une moindre mesure au niveau des mutations synonymes et des mutations non conservées.

Une autre différence majeure qui caractérise cette étude par rapport aux données existant dans la littérature est l'analyse de populations *mutT*-. Des analyses de biais de mutations dans l'incorporation de 8-oxo-dGTP ont précédemment été faites *in vivo* sur quelques triplets (Watanabe et al. 2001). D'autres ont analysé ces biais dans des pseudogènes humains, mettant en évidence un fort biais au niveau des dinucléotides AA et TT (Pozdniakov et al. 1997, Rogozin et Pavlov 2003). Dans cette étude, je trouve des biais similaires, mais au niveau des dinucléotides AA et TT. Ceci peut résulter de la réplication comme de la réparation par MutY et MutM. Une récente étude a démontré que la base voisine en 5' avait une influence sur l'efficacité de la réparation de la lésion par MutM (Sung et al. 2012). Aucun biais de brin n'a été observé dans ces populations contrairement aux populations MMR-.

L'ensemble de ces biais créent des zones génomiques à faible ou forte mutabilité et peuvent impacter sur l'évolution des génomes, en privilégiant la variabilité de certaines régions plutôt que d'autres. Combinées à la variation du taux de mutation global, les mutations peuvent donc être modulées finement au cours de l'évolution. La variation et la sélection de ces paramètres évolutifs contribuent donc à l'évolution de l'évolution.

1.3 Modulation des taux de mutation et évolution de l'évolution

Les taux de mutations au cours du temps ou à travers le génome sont non seulement variables, mais aussi sélectionnables (sélection de second ordre). Ainsi, un génotype hypermutateur peut émerger dans la population grâce à l'entraînement génétique, mais il peut aussi en être de même au niveau des biais. Par exemple, la modification de facteurs de fidélité de la réplication ou de la réparation de l'ADN, ou l'existence de séquences hypermutables dans certains gènes peuvent être sélectionnés. C'est particulièrement le cas pour des gènes dont la variation représente un enjeu important, comme certains épitopes ou même les gènes de réparation de l'ADN eux-mêmes. En effet, dans les deux populations *mutT*-, les mutations à l'origine du phénotype hypermutateur sont identiques : il s'agit de la délétion d'un C dans une séquence de nucléotides répétés de 6 C. Les séquences répétées de nucléotides sont connues pour être des points chauds de mutations par suite de la possibilité d'un glissement de l'ADN polymérase (Rogozin et Pavlov 2003). L'existence d'un tel point chaud de mutation dans le gène *mutT* a pu être sélectionnée et conservée car il pourrait permettre d'accroître la variabilité allélique de ce gène, et par conséquent permettre de générer une fluctuation des taux de mutation. Ainsi, lors de périodes d'adaptation, des génotypes hypermutateurs pourraient être facilement générés, augmentant les taux de mutation et accélérant l'adaptation. Le retour des taux de mutation à une valeur ancestrale est quant à lui assez lent, mais cela peut être facilité par d'autres mécanismes évolutifs comme la recombinaison (Denamur et al. 2000, Syvanen 2012). Ainsi, le taux de mutation peut être facilement et finement modulé, permettant de maintenir l'équilibre entre stabilité et variabilité.

Du type de mutateurs dépend un type de mutations et un biais correspondant. Ainsi, chaque mutateur n'a pas accès aux mêmes trajectoires évolutives. Chez les hypermutateurs MMR-, un biais mutationnel existe pour les transitions, mais aussi pour les petites insertions ou délétions ayant un effet plus grand que les mutations ponctuelles. Bien que les génotypes hypermutateurs *mutT*- produisent un grand nombre de mutations, celles-ci ont un spectre extrêmement réduit. Ces biais peuvent également avoir d'autres conséquences, comme au niveau de la composition du génome et de chaque brin d'ADN, même si les mutations ne représentent qu'un facteur parmi d'autres (comme la désamination des cytosines ou la recombinaison) dans ces évolutions (Rocha et al. 2006).

Enfin, l'évolution de l'évolution peut être considérée comme un concept encore plus large car les autres mécanismes même de l'évolution, comme la recombinaison, les réarrangements génomiques, ou la réorganisation des réseaux de régulation, sont également soumis à la sélection naturelle (Raeside et al. 2014, Hindré et al. 2012). A cela s'ajoute une variabilité au niveau de l'expression des gènes qui permet alors une variation phénotypique. Cette thèse apporte un éclairage supplémentaire dans la compréhension de la régulation des taux de mutation comme mécanisme évolutif, en œuvre dans l'environnement comme au laboratoire. Ces résultats peuvent également trouver écho dans des situations médicales comme l'évolution de pathogènes lors d'infections chroniques.

2 Implications cliniques

Des génotypes hypermutateurs ont été isolés à des fréquences élevées dans divers échantillons cliniques : jusqu'à 12% pour *E. coli* lors d'infections urinaires (Denamur et al. 2002), environ 20% chez *Streptococcus pneumoniae* (Morosini et al. 2003)... La fréquence des génotypes hypermutateurs est particulièrement élevée lors d'infections chroniques liées à *P. aeruginosa* chez des patients atteints de mucoviscidose, où elle peut dépasser les 50% (Ciofu et al. 2005, Oliver et Mena 2010), alors qu'elle est inférieure à 1% dans les cas d'infections aiguës. Une étude génomique a mis en évidence que cette fréquence pouvait même atteindre 90% au cours d'une infection chronique (Feliziani et al. 2015). Cette corrélation entre chronicité et hypermutabilité a été particulièrement étudiée.

L'invasion des génotypes hypermutateurs repose sur des entraînements génétiques successifs avec des mutations bénéfiques. Lors d'une infection chronique, les bactéries sont exposées pendant une longue période à des stress réguliers tels que des attaques du système immunitaire, des antibiotiques, ou la concurrence avec d'autres pathogènes. Tous ces facteurs créent une forte pression de sélection et produisent également des environnements fluctuants connus pour favoriser la présence de génotypes hypermutateurs (Folkesson et al. 2012). Le nombre de mutations bénéfiques potentielles, avec lesquelles les allèles hypermutateurs vont pouvoir être entraînés et augmenter en fréquence, est donc élevé. L'état d'infection chronique favorise donc l'émergence de génotypes hypermutateurs (Oliver et Mena 2010).

De même, la présence de génotypes hypermutateurs favorise l'infection chronique, car ces souches produisent plus fréquemment des mutations de résistance aux antibiotiques, les rendant difficilement traitables (Ciofu et al. 2005, Ferroni et al 2009, Oliver et Mena 2010). Différentes études ont en effet démontré que l'aspect hypermutateur est un gros atout dans l'acquisition des résistances aux antibiotiques tels que la rifampicine, la streptomycine ou la ciprofloxacine. Outre les mutations provoquant ces résistances, le phénotype hypermutateur favorise aussi d'autres mutations qui réduisent les coûts parfois élevés associés à ces résistances (Perron et al. 2010). En effet, l'adaptation des souches à leur environnement est également associée à d'autres caractères comme l'adhésion ou la formation de biofilms, pour lesquels les génotypes hypermutateurs présentent une meilleure variabilité (Mena et al. 2008). Cette adaptation au milieu pulmonaire se fait aux dépens de la virulence, les souches adaptés devenant moins pathogènes et transmissibles (Mena et al. 2007).

Les génotypes hypermutateurs observés lors de ces infections chroniques résultent généralement de mutations dans les gènes du système MMR (Oliver et Mena 2010). Les hypermutateurs MMR- se caractérisent par un spectre de mutation plus large que les mutateurs *mutT*-, ce qui leur permet donc de produire une plus grande variété de mutations. Chaque type de mutateur possède son propre paysage évolutif, et ceci est particulièrement visible sur l'adaptation aux antibiotiques : des souches *mutT*- produisent des clones résistants à la streptomycine ayant une meilleure valeur sélective que ceux résultant d'un phénotype *mutY*- (Couce et al. 2013). Ainsi, le type de mutateur conditionne le profil d'adaptation des souches. J'ai pu aussi observer que certaines régions, étant plus mutables que d'autres, peuvent conduire à des zones de mutations privilégiées. Ceci pourrait permettre une adaptation ciblée et rapide des bactéries, plus simples que d'autres systèmes tels que la variation de phase ou d'antigènes (Wisniewski-Dyé et Vial 2008).

Ainsi, l'étude des taux de mutations et l'invasion de clones hypermutateurs a une véritable importance au niveau clinique. En effet, l'utilisation d'antibiotiques favorise l'émergence de clones hypermutateurs, qui développent alors rapidement d'autres résistances, diminuant l'efficacité des traitements. Or une fois les hypermutateurs sélectionnés, j'ai pu mettre en évidence que, en l'absence de recombinaison, le retour à un taux de mutation inférieur peut être très lent et se joue à une échelle de temps bien plus longue que leur sélection. Ainsi, une fois sélectionnés,

les souches pathogènes hypermutatrices sont susceptibles de se maintenir sur le long terme, les rendant difficilement curables, même dans un environnement stable.

L'ensemble des résultats obtenus au cours de ma thèse apporte un nouveau regard sur la dynamique des taux de mutation au cours de l'évolution à une échelle de temps assez longue. Cependant, plusieurs questions subsistent : d'une part au niveau mécanistique sur l'établissement de ces biais mutationnels et la compensation des taux de mutation, et d'autre part au niveau de leurs implications évolutives.

3 Perspectives

3.1 Mécanistiques

L'étude de la population Ara+6 a permis de mettre en évidence l'existence d'autres mécanismes de compensation que ceux déjà décrits, c'est-à-dire après mutations des gènes *mutM*, *mutY* (Maki et Sekiguchi 1992) ou de gènes codant des pompes à efflux (Guelfo et al. 2010). Une particularité de ces mécanismes est qu'ils ne modifient pas le spectre de mutations, contrairement aux compensations associées au gène *mutY*. Ainsi, la sélection de ces clones à taux de mutation compensés a été faite uniquement sur la diminution des taux. Parmi les pistes possibles, des mutations ont été retrouvées dans les gènes *dnaQ*, codant la sous-unité de l'ADN polymérase III responsable de l'activité de relecture, ainsi que *mutS* et *mutL*, impliqués dans le système de réparation MMR. Ces trois mutations sont retrouvées dans tous les clones séquencés à partir de 30 000 générations et pourraient être à l'origine de la diminution des taux de mutation mis en évidence par la simulation de Monte-Carlo vers 25 000 générations. Ceci pourrait être confirmé par échange allélique entre des clones évolués à 30 000 générations et le clone mutateur de référence 10K-B. Deux autres événements de compensation restent cependant sans candidat dans les clones 20K-B et 40K-B. L'identification des gènes responsables, sans signature mutationnelle particulière, risque d'être laborieuse dans les 272 et 162 mutations propres à chaque clone respectivement. La possibilité que la compensation de ces deux clones soit réalisée par la modification du même gène reste non négligeable et pourrait faciliter son identification.

Un autre apport de l'étude de la population Ara+6 se situe au niveau des biais de mutations.

En effet, les biais précédemment décrits pour des hypermutateurs MMR- résultent essentiellement du système de réplication de l'ADN. Dans une population *mutT*-, à la réplication de l'ADN se rajoute le système de réparation associé aux protéines MutM et MutY. Ces biais ne dépendent pas du brin d'ADN, mais il est possible qu'ils soient liés à des biais de réparations. Des différences ont déjà été mises en évidence pour les nucléotides situés en 5' sur l'efficacité de réparation de MutM grâce à l'utilisation d'un rapporteur génique (Watanabe et al. 2001). Une meilleure étude de ces biais pourrait être réalisée par la synthèse d'oligonucléotides avec toutes les séquences possibles et de tester les différences d'activité de MutM et MutY sur la réparation de l'ADN. Il serait aussi possible de mesurer l'activité de la protéine MutM modifiée suite à une mutation de *mutM* dans le clone 50K-A, pour vérifier si cette protéine mutante modifie le taux de mutation, comme le suggèrent les tests de fluctuation. Enfin, cette thèse dessine aussi d'autres perspectives au niveau évolutif.

3.2 Evolutives

Tout d'abord, il serait intéressant de mesurer la fréquence des mutations *mutY* dans la population Ara+6 à différents temps, afin de suivre leur apparition et leur élimination de la population. Cette étude ne comportant que deux clones isolés à chaque temps, elle ne laisse entrevoir qu'une infime partie de l'évolution de la population. Aussi, cela permettrait de voir si la dynamique des taux de mutations a pu être plus élevée que celle mesurée ou modélisée. Pour cela, il serait possible de réaliser un métagénome de la population à différents temps, ou plus simplement de réaliser un séquençage en profondeur des mélanges de populations. Ceci permettrait aussi de voir les différentes interactions clonales au cours de cette évolution.

L'expérience d'évolution au long terme se caractérise par un milieu extrêmement stable au cours du temps, provoquant une dynamique d'adaptation particulière, au cours de laquelle pour la première fois, des taux de mutations ont été suivis sur une période aussi longue. Il serait intéressant d'observer de telles variations sur un long terme avec un milieu fluctuant, qui favoriserait alors en théorie un maintien de taux de mutation à des valeurs élevées, peut-être aux dépens de la charge mutationnelle.

Enfin, grâce aux mesures de taux de mutations et de leur biais, il est possible de modéliser

plus finement les processus évolutifs. Il devient possible de réaliser des expériences d'évolution *in silico* avec les paramètres obtenus au cours de cette expérience et d'observer si les scénarii évolutifs convergent ou non, avec de moindres contraintes que dans une expérience d'évolution en laboratoire. Cela peut être possible notamment en utilisant le logiciel *Aevol* (Batut et al. 2013).

Bibliographie

- André J, Godelle B. 2006. The evolution of mutation rate in finite asexual populations. *Genetics*, 172 : 611–626.
- Avery OT, MacLeod CM, McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med.*, 79 : 137–158.
- Baele G, de Peer YV, Vansteelandt S. 2010. Modelling the ancestral sequence distribution and model frequencies in context-dependent models for primate non-coding sequences. *BMC Evol Biol.*, 10 : 244.
- Barrick J, Lenski R. 2009. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harb Symp Quant Biol.*, 74 : 119–129.
- Barrick J, Yu D, Yoon S, Jeong H, Oh T, Schneider D, Lenski R, Kim J. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461 : 1243–1247.
- Batut B, Parsons D, Fischer S, Beslon G, Knibbe C. 2013. In silico experimental evolution : a tool to test evolutionary scenarios. *BMC Bioinformatics*, 14 : S11.
- Benzer S 1961. On the topography of the genetic fine structure. *Proc Natl Acad Sci U S A.*, 47 : 403–415.
- Bhatnagar SK, Bullions LC, Lew G, Bessman MJ. 1990. Characterization of the defect in the *Escherichia coli* mutT1 mutator gene. *J Bact.*, 172 : 2802–2803.
- Bjedov I, Tenaillon O, Gérard B, Souza V, Denamur E, Radman M, Taddei F, Matic I. 2003. Stress-induced mutagenesis in bacteria. *Science*, 300 : 1404–1409.
- Blomfield I, Vaughn V, Rest R, Eisenstein B. 1991. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis* *sacB* gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol*, 5 : 1447–1457.

- Blount ZD, Borland CZ, Lenski RE. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A.*, 105 : 7899–7906.
- Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature*, 489 : 513–518.
- Carlton BC, Brown BJ. 1981. *Manual of Methods for General Bacteriology*, chapter Gene mutation., pages 222–242. American Society for Microbiology, Washington, D.C.
- Chao L, Cox E. 1983. Competition between high and low mutating strains of *Escherichia coli*. *Evolution*, 37 : 125–134.
- Ciofu O, Riis B, Pressler T, Poulsen HE, Hoiby N. 2005. Occurrence of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. *Antimicrob Agents Chemother.*, 49 : 2276–2282.
- Clune J, Misevic D, Ofria C, Lenski R, Elena S, R S. 2008. Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. *PLoS Comput Biol*, 4 : e1000187.
- Cooper DN, Krawczak M. 1990. The mutational spectrum of single base-pair substitutions causing human genetic disease : patterns and predictions. *Hum Genet.*, 85 : 55–74.
- Cooper VS, Bennett AF, Lenski RE. 2001. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. *Evolution*, 55 : 889–896.
- Couce A, Guelfo JR, Blazquez J. 2013. Mutational spectrum drives the rise of mutator bacteria. *PLoS Genet.*, 9 : e1003167.
- Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, 274 : 775–780.
- Cox EC, Gibson TC. 1974. Selection for high mutation rates in chemostats. *Genetics*, 77 : 169–184.

- Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics*, 169 : 523–532.
- Crozat E, Winkworth C, Gaffé J, Hallin PF, Riley MA, Lenski RE, Schneider D. 2010. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Mol Biol Evol.*, 27 : 2113–2128.
- Darmon E, Leach DR. 2014. Bacterial genome instability. *Microbiol Mol Biol Rev.*, 78 : 1–39.
- Dawson KJ 1998. Evolutionarily stable mutation rates. *J Theo Biol.*, 194 : 143 – 157.
- Deatherage DE, Barrick JE. 2014. *Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq*, volume 1151 of *Methods in Molecular Biology*, pages 165–188. Springer New York.
- Denamur E, Lecointre G, Darlu P, Tenaillon O, Acquaviva C, Sayada C, Sunjevaric I, Rothstein R, Elion J, Taddei F, Radman M, Matic I. 2000a. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell*, 103 : 711–721.
- Denamur E, Lecointre G, Darlu P, Tenaillon O, Acquaviva C, Sayada C, Sunjevaric I, Rothstein R, Elion J, Taddei F, Radman M, Matic I. 2000b. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell*, 103 : 711–721.
- Denamur E, Bonacorsi S, Giraud A, Duriez P, Hilali F, Amorin C, Bingen E, Andremont A, Picard B, Taddei F, Matic I. 2002. High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *J Bacteriol*, 184 : 605–609.
- Desai M, Fisher D. 2011. The balance between mutators and nonmutators in asexual populations. *Genetics*, 188 : 977–1014.
- Dorman CJ 2009. Nucleoid-associated proteins and bacterial physiology. *Adv Appl Microbiol.*, 67 : 47–64.
- Drake J 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A*, 88 : 7160–7164.

- Duret L 2009. Mutation patterns in the human genome : More variable than expected. *PLoS Biol.*, 7.
- Eyre-Walker A, Keightley P. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet*, 8 : 610–6018.
- Feher T, Bogos B, Mehi O, Fekete G, Csorgo B, Kovacs K, Posfai G, Papp B, Hurst LD, Pal C. 2012. Competition between transposable elements and mutator genes in bacteria. *Mol Biol Evol.*, 29 : 3153–3159.
- Feldman MW, Liberman U. 1986. An evolutionary reduction principle for genetic modifiers. *Proc Natl Acad Sci U S A.*, 83 : 4824–4827.
- Ferroni A, Guillemot D, Moumile K, Bernede C, Le Bourgeois M, Waernessyckle S, Descamps P, Sermet-Gaudelus I, Lenoir G, Berche P, Taddei F. 2009. Effect of mutator *P. aeruginosa* on antibiotic resistance acquisition and respiratory function in cystic fibrosis. *Pediatr Pulmonol.*, 44 : 820–825.
- Fijalkowska IJ, Schaaper RM, Jonczyk P. 2012a. DNA replication fidelity in *Escherichia coli* : a multi-DNA polymerase affair. *FEMS Microbiol Rev.*, 36 : 1105–1121.
- Fijalkowska IJ, Schaaper RM, Jonczyk P. 2012b. DNA replication fidelity in *Escherichia coli* : a multi-DNA polymerase affair. *FEMS Microbiol Rev.*, 36 : 1105–1121.
- Fisher R 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- Fisher RA 1922. On the interpretation of χ^2 from contingency tables, and the calculation of p. *J R Stat Soc.*, 85 : 87–94.
- Foster PL, Hanson AJ, Lee H, Popodi EM, Tang H. 2013. On the mutational topology of the bacterial genome. *G3 (Bethesda)*, 3 : 399–407.
- Fowler R, White S, Koyama C, Moore S, Dunn R, Schaaper R. 2003. Interactions among the *Escherichia coli* *mutT*, *mutM*, and *mutY* damage prevention pathways. *DNA Repair*, 2 : 159–173.

- Fowler RG, Schaaper RM. 1997. The role of the *mutT* gene of *Escherichia coli* in maintaining replication fidelity. *FEMS Microbiol Rev*, 21 : 43–54.
- Frank A, Lobry J. 1999. Asymmetric substitution patterns : a review of possible underlying mutational or selective mechanisms. *Gene*, 238 : 65 – 77.
- Friedberg E, Walker G, Siede W. 1995. *DNA Repair and Mutagenesis*. ASM Press, Washington, DC.
- Gentile CF, Yu SC, Serrano SA, Gerrish PJ, Sniegowski PD. 2011. Competition between high- and higher-mutating strains of *Escherichia coli*. *Biol Lett.*, 7 : 422–424.
- Gerrish P, Colato A, Perelson A, Sniegowski P. 2007. Complete genetic linkage can subvert natural selection. *Proc Natl Acad Sci U S A*, 104 : 6266–6271.
- Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102-103 : 127–144.
- Gibson TC, Scheppe ML, Cox EC. 1970. Fitness of an *Escherichia coli* mutator gene. *Science*, 169 : 686–688.
- Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, Taddei F. 2001. Costs and benefits of high mutation rates : adaptive evolution of bacteria in the mouse gut. *Science*, 291 : 2606–2608.
- Grainger DC, Hurd D, Goldberg MD, Busby SJW. 2006. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.*, 34 : 4642–4652.
- Gross MD, Siegel EC. 1981. Incidence of mutator strains in *Escherichia coli* and coliforms in nature. *Mutat Res.*, 91 : 107–110.
- Guelfo J, Rodríguez-Rojas A, Matic I, Blázquez J. 2010. A MATE-family efflux pump rescues the *Escherichia coli* 8-oxoguanine-repair-deficient mutator phenotype and protects against H₂O₂ killing. *PLoS Genet*, 6 : e1000931.

- Haldane JBS 1927. A mathematical theory of natural and artificial selection, part v : Selection and mutation. *Mat Proc Cambridge*, 23 : 838–844.
- Hall B, Ma C, Liang P, Singh K. 2009. Fluctuation AnaLysis CalculatOR (FALCOR) : a web tool for the determination of mutation rate using Luria-Delbrück fluctuation analysis. *Bioinformatics*, 25 : 1564–1565.
- Harfe BD, Jinks-Robertson S. 2000. DNA mismatch repair and genetic instability. *Annu Rev Genet.*, 34 : 359–399.
- Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.*, 6 : e1001115.
- Hess ST, Blake JD, Blake R. 1994. Wide variations in neighbor-dependent substitution rates. *J Mol Biol.*, 236 : 1022 – 1033.
- Hindre T, Knibbe C, Beslon G, Schneider D. 2012. New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nat Rev Microbiol.*, 10 : 352–365.
- Hwang DG, Green P. 2004. Bayesian markov chain monte carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci U S A.*, 101 : 13994–14001.
- Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Segurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF. 2009. Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). *J Mol Biol.*, 394 : 644–652.
- Johnson T 1999. The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *Proc Biol Sci*, 266 : 2389–2397.
- Joseph N, Duppatla V, Rao DN. 2006. Prokaryotic DNA mismatch repair. 81 : 1–49.
- Kimura M 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genet Res.*, 9 : 23–34.

- Kisker C, Kuper J, Van Houten B. 2013. Prokaryotic nucleotide excision repair. *Cold Spring Harb Perspect Biol.*, 5 : a012591.
- Koch RE 1971. The influence of neighboring base pairs upon Base-Pair substitution mutation rates. *Proc Natl Acad Sci U S A.*, 68 : 773–776.
- Kondrashov AS 1995. Dynamics of unconditionally deleterious mutations : Gaussian approximation and soft selection. *Genet Res.*, 65 : 113–121.
- Krokan HE, Bjoras M. 2013. Base excision repair. *Cold Spring Harb Perspect Biol.*, 5 : a012583.
- Le Gac M, Plucain J, Hindré T, Lenski RE, Schneider D. 2012. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci U S A.*, 109 : 9487–9492.
- LeClerc J, Li B, Payne W, Cebula T. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science*, 274 : 1208–1211.
- Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A.*, 109 : E2774–E2783.
- Leigh E 1970. Natural selection and mutability. *Am Nat*, 104 : 301–305.
- Lenski RE, Travisiano M. 1994. Dynamics of adaptation and diversification : a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci U S A.*, 20 : 259–270.
- Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat*, 138 : 1315–1341.
- Li GM 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res.*, 18 : 85–98.
- Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci U S A.*, 105 : 17878–17883.

- Link A, Phillips D, Church G. 1997. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli* : Application to open reading frame characterization. *J Bacteriol*, 179 : 6228–6237.
- Lobry JR 1996. Asymmetric substitution patterns in the two dna strands of bacteria. *Mol Biol Evol.*, 13 : 660–665.
- Long H, Sung W, Miller SF, Ackerman MS, Doak TG, Lynch M. 2015. Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol.*, 7 : 262–271.
- Lynch M 2010. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A.*, 107 : 961–968.
- Lynch M 2011. The lower bound to the evolution of mutation rates. *Gen Biol Evol.*, 3 : 1107–1118.
- Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A.*, 105 : 9272–9277.
- Ma W, Sandri G.V. H, Sarkar S. 1992. Analysis of the Luria-Delbrück distribution using discrete convolution powers. *J Appl Probab*, 29 : 255–267.
- Maharjan RP, Liu B, Li Y, Reeves PR, Wang L, Ferenci T. 2012. Mutation accumulation and fitness in mutator subpopulations of *Escherichia coli*. *Biol Lett.*, 9.
- Maki H, Sekiguchi M. 1992. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature*, 355 : 273–275.
- Marvig RL, Johansen HK, Molin Sr, Jelsbak L. 2013. Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet.*, 9 : e1003741.
- Marvig RL, Sommer LM, Jelsbak L, Molin S, Johansen HK. 2015. Evolutionary insight from whole-genome sequencing of *Pseudomonas aeruginosa* from cystic fibrosis patients. *Future Microbiol.*, 10 : 599–611.

- Matic I, Radman M, Taddei F, Picard B, Doit C, Bingen E, Denamur E, Elion J. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science*, 277 : 1833–1834.
- Maynard-Smith J 1982. *Evolution and the Theory of Games*. Cambridge University Press.
- McDonald M, Hsieh Y, Yu Y, Chang S, Leu J. 2012. The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. *Curr Biol*, 22 : 1235–1240.
- Mena A, Macia MD, Borrell N, Moya B, de Francisco T, Perez JL, Oliver A. 2007. Inactivation of the mismatch repair system in *Pseudomonas aeruginosa* attenuates virulence but favors persistence of oropharyngeal colonization in cystic fibrosis mice. *J Bacteriol.*, 189 : 3665–3668.
- Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM, Perez JL, Oliver A. 2008. Genetic adaptation of *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients is catalyzed by hypermutation. *J Bacteriol.*, 190 : 7910–7917.
- Mendel J 1865. Versuche über pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 4 : 3–47.
- Meyer RR, Laine PS. 1990. The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev.*, 54 : 342–380.
- Michaels M, Tchou J, Grollman A, Miller J. 1992. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry*, 31 : 10964–10968.
- Miller H, Perrino F. 1996. Kinetic mechanism of the 3'→5' proofreading exonuclease of dna polymerase III. analysis by steady state and pre-steady state methods. *Biochemistry*, 35 : 12919–12925.
- Miller JH 1998. Mutators in escherichia coli. *Mut Res./DNA Repair*, 409 : 99 – 106.
- Morgan TH 1911. Chromosomes and associative inheritance. *Science*, 34 : 636–638.
- Morosini MI, Baquero MR, Sanchez-Romero JM, Negri MC, Galan JC, del Campo R, Perez-Diaz JC, Baquero F. 2003. Frequency of mutation to rifampin resistance in *Streptococcus pneumoniae* clinical strains : hexA and hexB polymorphisms do not account for hypermutation. *Antimicrob Agents Chemother.*, 47 : 1464–1467.

- Muller HJ 1964. The relation of recombination to mutational advance. *Mutat Res.*, 106 : 2–9.
- Newton C, Graham A, Heptinstall L, Powell S, Summers C, Kalsheker N, Smith J, Markham A. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*, 17 : 2503–2516.
- Niki H, Yamaichi Y, Hiraga S. 2000. Dynamic organization of chromosomal dna in escherichia coli. *Genes & Dev.*, 14 : 212–223.
- Notley-McRobb L, Seeto S, Ferenci T. 2002. Enrichment and elimination of *mutY* mutators in *Escherichia coli* populations. *Genetics*, 162 : 1055–1062.
- Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A. 1968. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A.*, 59 : 598–605.
- Oliver A, Mena A. 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. *Clin Microbiol Infection*, 16 : 798 – 808.
- Oliver A, Cantón R, Campo P, Baquero F, Blázquez J. 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science*, 288 : 1251–1254.
- Orr H 2000. The rate of adaptation in asexuals. *Genetics*, 155 : 961–968.
- Pal C, Macia MD, Oliver A, Schachar I, Buckling A. 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, 450 : 1079–1081.
- Parker BO, Marinus MG. 1992. Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc Natl Acad Sci U S A.*, 89 : 1730–1734.
- Pelosi L, Kühn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics*, 173 : 1851–1869.
- Perron GG, Hall AR, Buckling A. 2010. Hypermutability and compensatory adaptation in antibiotic-resistant bacteria. *Am Nat.*, 176 : pp. 303–311.

- Philippe N, Crozat E, Lenski RE, Schneider D. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *Bioessays*, 29 : 846–860.
- Philippe N, Pelosi L, Lenski RE, Schneider D. 2009. Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with *Escherichia coli*. *J Bact.*, 191 : 909–921.
- Raessens C, Gaffe J, Deatherage DE, Tenaillon O, Briska AM, Ptashkin RN, Cruveiller S, Medigue C, Lenski RE, Barrick JE, Schneider D. 2014. Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *MBio*, 5 : e01377–01314.
- Raynes Y, Sniegowski P. 2014. Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity*, 113 : 375–380.
- Raynes Y, Gazzara MR, Sniegowski PD. 2012. Contrasting dynamics of a mutator allele in asexual populations of differing size. *Evolution*, 66 : 2329–2334.
- Rocha EP, Touchon M, Feil EJ. 2006. Similar compositional biases are caused by very different mutational effects. *Genome Res.*, 16 : 1537–1547.
- Rogozin IB, Pavlov YI. 2003. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res Rev Mutat Res.*, 544 : 65 – 85.
- Rozen DE, Lenski RE. 2000. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am Nat.*, 155 : pp. 24–35.
- Rozen DE, Schneider D, Lenski RE. 2005. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *J Mol Evol.*, 61 : 171–180.
- Schaaper R 1996. Suppressors of *Escherichia coli mutT* : antimutators for dna replication errors. *Mutat Res*, 350 : 17–23.
- Schiffels S, Szölloši GJ, Mustonen V, Lässig M. 2011. Emergent neutrality in adaptive asexual evolution. *Genetics*, 189 : 1361–1375.
- Sharp PM, Emery LR, Zeng K. 2010. Forces that influence the evolution of codon bias. *Philos Trans R Soc Lond B Biol Sci.*, 365 : 1203–1212.

- Shaver AC, Dombrowski PG, Sweeney JY, Treis T, Zappala RM, Sniegowski PD. 2002. Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics*, 162 : 557–566.
- Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res.*, 23 : 23–35.
- Sniegowski P, Gerrish P. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philos Trans R Soc Lond B Biol Sci*, 365 : 1255–1263.
- Sniegowski P, Gerrish P, Lenski R. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*, 387 : 703–705.
- Söderberg R, Berg O. 2011. Kick-starting the ratchet : the fate of mutators in an asexual population. *Genetics*, 187 : 1129–1137.
- Sturtevant A 1937. Essays on evolution. I. On the effects of selection on mutation rate. *Q Rev Biol*, 12 : 464–467.
- Su SS, Grilley M, Thresher R, Griffith J, Modrich P. 1989. Gap formation is associated with methyl-directed mismatch correction under conditions of restricted DNA synthesis. *Genome*, 31 : 104–111.
- Sung W, Ackerman MS, Gout J, Miller SF, Williams E, Foster PL, Lynch M. 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. *Mol Biol Evol.*, 32 : 1672–1683.
- Syvanen M 2012. Evolutionary implications of horizontal gene transfer. *Annu Rev Genet.*, 46 : 341–358.
- Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutator alleles in adaptive evolution. *Nature*, 387 : 700–702.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6 : Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*, 30 : 2725–2729.

- Tenaillon O, Toupance B, Le Nagard H, Taddei F, Godelle B. 1999. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics*, 152 : 485–493.
- Tenaillon O, Le Nagard H, Godelle B, Taddei F. 2000. Mutators and sex in bacteria : conflict between adaptive strategies. *Proc Natl Acad Sci U S A*, 97 : 10465–10470.
- Tenaillon O, Taddei F, Radmian M, Matic I. 2001. Second-order selection in bacterial evolution : selection acting on mutation and recombination rates in the course of adaptation. *Res Microbiol.*, 152 : 11–16.
- Travis J, Travis E. 2002. Mutator dynamics in fluctuating environments. *Proc Biol Sci.*, 269 : 591–597.
- Treffers HP, Spinelli V, Belser NO. 1954. A Factor (or Mutator Gene) Influencing Mutation Rates in *Escherichia Coli*. *Proc Natl Acad Sci U S A.*, 40 : 1064–1071.
- Tröbner W, Piechocki R. 1984. Selection against hypermutability in *Escherichia coli* during long term evolution. *Mol Gen Genet*, 198 : 177–178.
- Valens M, Penaud S, Rossignol M, Cornet F, Boccard F. 2004. Macrodomain organization of the *Escherichia coli* chromosome. *The EMBO Journal*, 23 : 4330–4341.
- Watanabe T, van Geldorp G, Najrana T, Yamamura E, Nunoshiba T, Yamamoto K. 2001. Miscoding and misincorporation of 8-oxo-guanine during leading and lagging strand synthesis in *Escherichia coli*. *Mol Gen Genet.*, 264 : 836–841.
- Watson JD, Crick FH. 1953. The structure of DNA. *Cold Spring Harb Symp Quant Biol.*, 18 : 123–131.
- Wielgoss S, Barrick J, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Médigue C, Lenski R, Schneider D. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3 (Bethesda)*, 1 : 183–186.
- Wielgoss S, Barrick J, Tenaillon O, Wiser M, Dittmar W, Cruveiller S, Chane-Woon Ming B, Médigue C, Lenski R, Schneider D. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A*, 110 : 222–227.

- Wiser M, Ribeck N, Lenski R. 2013. Long-term dynamics of adaptation in asexual populations. *Science*, 342 : 1364–1367.
- Wisniewski-Dyé F, Vial L. 2008. Phase and antigenic variation mediated by genome modifications. *Antonie Van Leeuwenhoek*, 94 : 493–515.
- Yamada M, Shimizu M, Katafuchi A, Grúz P, Fujii S, Usui Y, Fuchs R, Nohmi T. 2012. *Escherichia coli* DNA polymerase III is responsible for the high level of spontaneous mutations in *mutT* strains. *Mol Microbiol*, 86 : 1364–1375.
- Yang W 2011. Surviving the sun : repair and bypass of DNA UV lesions. *Protein Sci.*, 20 : 1781–1789.
- Yanisch-Perron C, Vieira J, Messing J. 1985. Improved M13 phage cloning vectors and host strains : nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33 : 103–119.
- Yanofsky C, Cox EC, Horn V. 1966. The unusual mutagenic specificity of an *E. coli* mutator gene. *Proc Natl Acad Sci U S A.*, 55 : 274–281.
- Zeth K 2012. Dps biomineralizing proteins : multifunctional architects of nature. *Biochem J.*, 445 : 297–311.
- Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci U S A.*, 111 : E2310–E2318.

Résumé

Les mutations constituent une des principales sources de variation sur lesquelles agit la sélection naturelle, permettant ainsi l'évolution des organismes vivants. Comprendre la dynamique d'accumulation des mutations, ainsi que les biais pouvant influencer leur apparition, est donc indispensable pour mieux appréhender les processus évolutifs. Dans cette thèse, j'ai exploré ces deux aspects dans un contexte évolutif.

Dans une première partie, je me suis intéressée à la dynamique des taux de mutation au cours du temps évolutif. En effet, les mutations pouvant être bénéfiques, neutres ou délétères, la dynamique des taux de mutation est régie par deux forces opposées que sont l'adaptabilité (la capacité à évoluer) et la stabilité du génome. Cette dynamique a été très étudiée de façon théorique, mais les études expérimentales sont plus limitées, et surtout à des périodes de temps courtes.

Dans une seconde partie, je me suis intéressée aux biais mutationnels. En effet, de précédentes études ont montré que les taux de mutation pouvaient varier au sein d'un même génome. Ainsi, certaines mutations peuvent se produire de façon plus fréquente que d'autres, le taux de mutation d'un nucléotide pouvant par exemple être influencé par les nucléotides avoisinants. Ces analyses ont été réalisées dans le contexte de l'expérience d'évolution à long terme initiée en 1988 par Richard Lenski (Michigan State University, USA). Douze populations ont été initiées à partir d'un ancêtre commun *Escherichia coli* et sont propagées depuis plus de 25 ans par repiquages quotidiens dans un milieu frais. Des échantillons ont été prélevés et le génome de clones évolués séquencé à différents temps, permettant une étude phénotypique et génomique des taux de mutations sur plus de 50 000 générations.

J'ai ainsi pu mettre en évidence une dynamique importante des taux de mutation, avec l'émergence de génotypes hypermutateurs suivie de phénomènes de compensation multiples. D'autre part, j'ai pu observer des biais mutationnels importants dont l'impact des nucléotides avoisinant les mutations fixées ou non dans les populations.

Abstract

Mutations are the ultimate source of variation which allows living organisms to adapt through natural selection. Understanding the dynamics of mutation accumulation and how they are biased stands as a keystone for the comprehension of the evolution process. In this work, I explore these two aspects of mutation accumulation using an evolution experiment.

First, I study the dynamics of mutation rate along time. Mutations can be either beneficial, neutral or deleterious, thus their accumulations depends on two opposite forces: adaptability or the ability to evolve on one hand, and genome stability on the other hand. The resulting dynamics has been widely studied theoretically but experimental studies are mostly limited to short periods of time. Second, I focus on mutational bias. Previous studies showed that the mutation rate could vary within a genome, under the influence of their position or the neighboring nucleotides for example.

These studies were done using the long term evolution experiment which was started in 1988 by Richard Lenski (Michigan State University, USA). Twelve populations were initiated for a single ancestor *Escherichia coli* and have been propagated for more than 25 years by daily transfer in fresh medium. Samples were taken and some clones were sequenced at regular time points, allowing both the phenotypic and genomic study of the mutation rate for more than 50,000 generations. In this study, I show that mutation rate can be highly dynamic: the emergence of hypermutator genotypes is followed by multiple compensation events. I also observe large mutational bias, with the particular impact of the neighboring nucleotides on resulting aminoacid changes.