



HAL
open science

Vers la compréhension du rôle de NucS dans l'évolution du génome de Streptomyces

Oyut Dagva

► **To cite this version:**

Oyut Dagva. Vers la compréhension du rôle de NucS dans l'évolution du génome de Streptomyces. Microbiologie et Parasitologie. Université de Lorraine, 2024. Français. NNT : 2024LORR0031 . tel-04638247

HAL Id: tel-04638247

<https://theses.hal.science/tel-04638247>

Submitted on 8 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**UNIVERSITÉ
DE LORRAINE**

**BIBLIOTHÈQUES
UNIVERSITAIRES**

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE DE LORRAINE

Mention : Ecotoxicologie, Biodiversité, Ecosystèmes

par **Oyut DAGVA**

Sous la direction conjointe de **Claire BERTRAND, Annabelle THIBESSARD**
et **Pierre LEBLOND**

Vers la compréhension du rôle de NucS dans l'évolution du génomme de *Streptomyces*

Soutenance publique le 14 mars 2024

Composition du Jury :

Rapporteurs :

Mme. Stéphanie BURY-MONE, Professeure, Université Paris-Saclay

M. Fabrice CONFALONIERI, Professeur, Université Paris-Saclay – Président du Jury

Examineurs :

M. François LECOINTE, Chargé de recherche, INRAE, Jouy-en-Josas

Mme. Claire BERTRAND, Maître de Conférences, Université de Lorraine (Co-directrice)

M. Pierre LEBLOND, Professeur, Université de Lorraine (Directeur de thèse)

Invités :

M. Hannu MYLLYKALLIO, Directeur de recherche, CNRS, École Polytechnique, Paris

Mme. Annabelle THIBESSARD, Maître de Conférences, Université de Lorraine (Co-encadrante)

Laboratoire Dynamique des Génomes et Adaptation Microbienne – UMR 1128 INRAE
Université de Lorraine, Faculté des Sciences et Technologies
Boulevard des Aiguillettes, BP 70239, 54506 Vandœuvre-lès-Nancy

Le plus important, c'est d'être ensemble

Remerciements

Je voudrais tout d'abord remercier les membres du jury, Pr. Stéphanie BURY-MONE, Pr. Fabrice CONFALONIERI et Dr. François LECOINTE, d'avoir accepté d'examiner mon travail de thèse.

Aux Pr. Pierre Leblond, Dr. Claire Bertrand et Dr. Annabelle Thibessard, un immense merci pour votre accompagnement tout au long de cette route parsemée d'embûches et de rebondissements qu'est la thèse. Je comprends mieux, aujourd'hui, pourquoi la résilience est une qualité essentielle dans la vie d'un chercheur (et dans la vie tout court?). C'est toujours au bout du chemin que l'on réalise de la hauteur du mur escaladé. Je vous remercie de votre écoute, merci de m'avoir laissé la liberté de me faire mon propre jugement. Merci pour toutes les discussions scientifiques et parfois moments de rire dans la bibliothèque du laboratoire. Surtout, je vous serai toujours reconnaissante pour votre soutien infaillible durant ce dernier mois. Quand il fût sombre, seul votre clairvoyance m'a permis d'avancer. Là, j'ai été témoin d'une expérience de vie à laquelle je ne m'attendais pas. Quel que soit le chemin emprunté, j'essaierai, à l'avenir, de m'attacher aux mots que vous m'avez insufflés. Encore merci à vous trois ; d'avoir cru en moi, d'avoir partagé un bout de vous. L'expérience humaine que représente la thèse transcende parfois nos propres limites et en ça, il s'agit d'une belle leçon.

Je tiens également à remercier tous les membres du laboratoire DynAMic. Merci au corps technique de faire tourner le labo ; surtout merci à Emilie pour ses conseils techniques concernant *Streptomyces* mais aussi d'avoir partagé ces instants envolés au bureau.

Une pensée spéciale pour Manon, Caroline et Robine, merci à vous pour votre écoute, votre disponibilité. Merci de m'avoir tendu la main et de votre bienveillance. J'espère qu'à l'avenir, nos routes se recroiseront !

Merci aussi aux enseignants-chercheurs du laboratoire pour nos échanges, vos encouragements.

Je veux écrire ces quelques lignes pour remercier Baptiste. Les mots me semblent faibles pour décrire comme tu as été important dans cette traversée. Tu as connu mes peines, mes joies et mes désillusions. Merci de m'avoir épaulé sans jamais me juger. Merci de m'avoir donné quelque peu de courage quand je n'en avais plus. Merci d'avoir été si compréhensif dans mes choix. J'aurais vécu cette thèse bien différemment sans toi à mes côtés. Peu importe ce que nous réserve l'avenir, « *l'important c'est d'être ensemble* ».

Pour finir et le plus important sûrement, je veux dédicacer cette thèse à ma mère pour laquelle je me bats tous les jours. Il n'y a qu'une seule chose dont je suis sûre : si j'ai la force de continuer à affronter les vents et marées de la vie, c'est bien pour toi. Je n'oublie pas d'où je viens et les sacrifices consentis. Merci pour ta combativité, ton courage. Merci pour l'éducation et les valeurs humaines que tu m'as inculquées. Merci de m'avoir toujours laissé une grande liberté de faire mes choix de vie et de m'avoir toujours soutenu quoi qu'il en coûte. Je t'admire et je continuerai à t'admirer, maman.

LISTE DES ABREVIATIONS

ADN : Acide DésoxyRibonucléique

ADNc : Acide DésoxyRibonucléique complémentaire

AICE : Actinomycete Integrative and Conjugative Element

ARN : Acide RiboNucléique

ARNg : Acide RiboNucléique guide

ATP : Adenosine Triphosphate

BER : Base Excision Repair

BIR : Break Induced Replication

Chi : Crossover Hot-spot Instigator

CRISPR-BEST : Clustered Regularly Interspaced Short Palindromic Repeats Base Edition SysTem

DSB : Double-Strand Break

DSBR : Double-Strand Break Repair

EndoMS : Mismatch Specific Endonuclease

GOC : Gene Order Conservations

HGT : Horizontal Gene Transfer

HT : Hicket Tresner

ICE : Integrative and Conjugative Element

kb : kilobase

Lig-Dom : domaine ligase

Mb : mégabase

MM : mismatch

MMR : MisMatch repair

MRN : Mre11/Rad50/Nbs1 complexe

NER : Nucleotide Excision Repair

NHEJ : Non-Homologous End-Joining

Nuc-Dom : domaine nucléase

NucS : Nuclease for single-stranded DNA

oriC : origine de répliation

pb : paire de bases

PCNA : Proliferating Cell Nuclear Antigen

PCR : Polymerase Chain Reaction

PFGE : Pulsed Field Gel Electrophoresis

Pol-Dom : domaine polymérase

SDSA : Synthesis Dependant Strand Annealing

SFM : Soy Mannitol Flour

SSA : Single-Strand Annealing

HGT : Horizontal Gene Transfer

HT : Hicket Tresner

qPCR : PCR quantitative

TIR : Terminal Inverted Repeats

UV : UltraViolet

VSP : Very Short Patch

SOMMAIRE

| | |
|--|--------------|
| PREAMBULE | |
| INTRODUCTION | 1 |
| I. <i>Streptomyces</i>: modèle d'étude de la plasticité génomique | 2 |
| 1. Le chromosome de <i>Streptomyces</i> est hautement compartimenté | 3 |
| 2. L'instabilité remarquable du chromosome de <i>Streptomyces</i> | 5 |
| II. Les forces évolutives du génome bactérien | 6 |
| 1. Les transferts horizontaux, voie d'entrée de l'ADN exogène..... | 6 |
| 1.1. La conjugaison bactérienne, principal mécanisme de HGT chez <i>Streptomyces</i> | 7 |
| 1.2. Le maintien de l'ADN exogène dans le génome | 7 |
| 2. La recombinaison dans les systèmes de réparation de l'ADN..... | 8 |
| 2.1. Les sources de DSB | 8 |
| 2.2. La recombinaison homologue..... | 9 |
| 2.3. La recombinaison illégitime..... | 13 |
| 3. Les mécanismes de recombinaison façonnent le chromosome des <i>Streptomyces</i> | 15 |
| III. Une nouvelle voie de MMR : origine de DSB endogènes ?..... | 15 |
| 1. Les erreurs de réplication..... | 16 |
| 2. Le MMR canonique..... | 16 |
| 2.1. Le MMR canonique méthyl-dépendant..... | 17 |
| 2.2. Le MMR indépendant de la méthylation..... | 18 |
| 2.3. Les protéines du MMR canonique en lien avec des voies de réparation de l'ADN | 20 |
| 3. Le rôle biologique du MMR canonique..... | 21 |
| 4. L'existence d'une voie de MMR non canonique | 22 |
| 4.1. La découverte de la protéine NucS..... | 22 |
| 4.2. NucS, acteur clé du MMR non-canonique et générateur de DSB..... | 23 |
| 4.3. L'intervention de NucS dans d'autres systèmes de réparation..... | 28 |
| OBJECTIFS DE LA THESE..... | 30 |

| | |
|---|------------|
| RESULTATS | 32 |
| I. Chapitre 1 : Caractérisation de NucS chez <i>S. ambofaciens</i> | 34 |
| 1. Synthèse du manuscrit..... | 34 |
| 2. Résultats complémentaires | 78 |
| II. Chapitre 2 : Lien entre l'activité de NucS et la plasticité du génome de <i>S. ambofaciens</i> | 82 |
| 1. Création d'un contexte génétique augmentant la fréquence de mésappariements... | 82 |
| 1.1. Ciblage de l'activité de relecture de l'ADN polymérase | 83 |
| 1.2. La mutation <i>dnaE_{H44Y}</i> confère un phénotype mutateur..... | 85 |
| 1.3. Instabilité génomique de l'hypermutateur <i>dnaE_{H44Y}</i> | 86 |
| 2. Implication de NucS dans l'augmentation de la fréquence des réarrangements génomiques | 90 |
| 2.1. Approche 1 : abolition de l'activité de NucS dans une souche hypermutatrice <i>dnaE_{H44Y}</i> | 91 |
| 2.1.1. Mutants <i>nucS_{W49*}</i> dans le contexte sauvage | 91 |
| 2.1.2. Mutants <i>nucS_{W49*}</i> dans le contexte mutateur <i>dnaE_{H44Y}</i> | 92 |
| 2.2. Approche 2 : surexpression de <i>nucS</i> dans un contexte hypermutateur..... | 94 |
| 2.2.1. Stratégie employée..... | 94 |
| 2.2.2. Les mutants <i>dnaE_{H44Y}</i> surexpriment <i>nucS</i> | 96 |
| 3. Perspectives à court terme | 97 |
| 4. Matériel et Méthodes | 97 |
| 4.1. Mutagénèse dirigée avec CRISPR-cBEST..... | 97 |
| 4.2. Conjugaison intergénérique..... | 99 |
| 4.3. Préparation des lamelles, digestion enzymatique et PFGE | 100 |
| 4.4. PCR quantitative..... | 101 |
| 4.4.1. Extraction des ARN | 101 |
| 4.4.2. Rétro-transcription et PCR quantitative | 101 |
| DISCUSSION ET PERSPECTIVES | 104 |
| I. Existence d'un MMR non-canonique chez <i>Streptomyces</i>..... | 106 |
| 1. Phénotype mutateur lié à la délétion du gène <i>nucS</i> | 106 |

| | |
|---|------------|
| 2. NucS est une endonucléase spécifique des mésappariements | 107 |
| 3. Spectre mutationnel lié à la délétion de <i>nucS</i> | 108 |
| II. NucS et la réparation des DSB | 108 |
| 1. Voies de réparation des cassures potentiellement médiées par NucS..... | 108 |
| 1.1. La recombinaison homologue..... | 109 |
| 1.2. Autres mécanismes candidats | 110 |
| 2. Contrôle de la recombinaison par NucS..... | 110 |
| 3. NucS contribue-t-elle à la plasticité du chromosome des <i>Streptomyces</i> ?..... | 111 |
| III. Evolutivité et adaptation | 113 |
| REFERENCES | 116 |

PREAMBULE

Depuis l'avènement des technologies de séquençage haut-débit, l'accès au contenu génomique des nombreuses espèces bactériennes a révélé une diversité génétique à tous les niveaux taxonomiques et ce jusqu'à l'échelle de la population, celle du microhabitat dans la niche écologique. Les sources de la diversité génétique chez les bactéries sont maintenant identifiées : l'acquisition de gènes par transfert horizontal, les réarrangements chromosomiques auxquels s'ajoutent l'accumulation de mutations ponctuelles façonnent les génomes bactériens sous la pression sélective de l'écosystème. Les transferts horizontaux et les réarrangements d'ADN sont les événements les plus féconds en termes d'innovation et de vitesse évolutive. Ainsi l'acquisition de gènes « prêts à l'emploi » et les grands remaniements du génome changent profondément les capacités d'adaptation. Par contraste, l'accumulation de mutations apparaît comme un phénomène lent et incertain. Les individus d'un groupe taxonomique partagent les gènes primordiaux à leur développement mais diffèrent par de nombreuses fonctions dites accessoires et spécifiques. L'évolution ainsi accélérée constitue un mécanisme d'adaptation aussi efficace que la régulation de l'expression génique, en temps réel et non à l'échelle des temps géologiques. La diversité générée permet de faire émerger des individus et des populations au potentiel plus avantageux. « Car de nombreux chemins mènent à presque n'importe quelle Rome » (Stephen J. Gould, *Le Pouce du Panda*, 1980).

Une autre dimension de l'apport de la diversification génétique a été mise à jour récemment chez les bactéries, c'est le « partage du travail » au sein d'une population. Partager sans posséder ... le patrimoine génétique d'une population dépasse de loin celle d'un individu en nature et diversité. Depuis peu, des phénomènes mimant les comportements sociaux des animaux ont été identifiés chez les bactéries. Dans le modèle que nous étudions, les *Streptomyces*, bactéries de la rhizosphère, la diversité génétique au sein d'une population permet la production de « biens communs », c'est-à-dire la fourniture, pour l'ensemble de la population par quelques individus, de molécules importantes pour l'accès aux nutriments ou encore la lutte contre les compétiteurs (ex. activités antibiotiques). Ces phénomènes constituent un front de science et soulève des questions clé en évolution microbienne : Quelle est l'origine de la diversité ? Ces mécanismes ont-ils été sélectionnés pour produire une diversité utile à la population bactérienne ?

Dans cette thèse, nous focalisons nos recherches sur l'origine des cassures d'ADN double brin du génome. En effet, c'est la réparation de ces dommages majeurs et létaux pour la cellule qui induit la recombinaison. Cette dernière favorise l'intégration d'ADN étranger (transfert horizontal) et les remaniements du génome. Une activité enzymatique portée par la protéine NucS a été identifiée chez

les archées et les actinomycètes (classe à laquelle appartiennent les *Streptomyces*) ; cette enzyme pourrait être la source endogène de cassures de l'ADN chez les *Streptomyces* et constituer le moteur de l'évolution rapide du génome.

INTRODUCTION

I. *Streptomyces* : modèle d'étude de la plasticité génomique

Les bactéries du genre *Streptomyces* sont des bactéries filamenteuses à coloration Gram positive appartenant au phylum des actinobactéries. Elles sont naturellement présentes dans les couches superficielles des sols où elles jouent un rôle écologique important. Ce genre bactérien produit en effet une grande diversité de métabolites secondaires avec des activités antibiotiques, antitumorales, antifongiques et insecticides, souvent exploitées par l'Homme à des fins biotechnologiques en médecine, en agriculture ou dans l'industrie agroalimentaire. (Viaene *et al.*, 2016; Khushboo *et al.*, 2022). Au niveau écologique, les *Streptomyces* et leur arsenal enzymatique contribuent à limiter la croissance de compétiteurs d'origine fongique ou bactérienne dans les sols (Chater, 2006; Palaniyandi *et al.*, 2013). Elles sont également capables de promouvoir directement ou indirectement la croissance des plantes (Vurukonda *et al.*, 2018) par exemple *via* la production d'analogues de phytohormones ou *via* la solubilisation de phosphate, ou l'acquisition de fer (Marschner *et al.*, 2011). Ainsi, les *Streptomyces* sont des acteurs essentiels du microbiote racinaire dans lequel elles co-habitent avec de nombreux autres microorganismes. Ces communautés microbiennes sont soumises à divers stress biotiques et abiotiques (pH, azote, composition de sol) et leur capacité d'adaptation rapide aux modifications de leur environnement constitue en atout majeur pour leur persistance dans les sols.

Les *Streptomyces* se distinguent par un cycle de vie généralement qualifié de « complexe » (**Figure 1**). Sur milieu solide, une spore contenant une seule copie du chromosome germe en formant le mycélium végétatif pluri-génomique. Ce mycélium végétatif est capable de se ramifier et sa croissance est radiale. Lorsque le milieu s'appauvrit en nutriments, le mycélium végétatif se différencie en mycélium aérien dont les hyphes subissent de multiples septations donnant lieu à des chaînes de spores matures qui se dissémineront et initieront en conditions favorables un nouveau cycle (Hopwood, 2006).

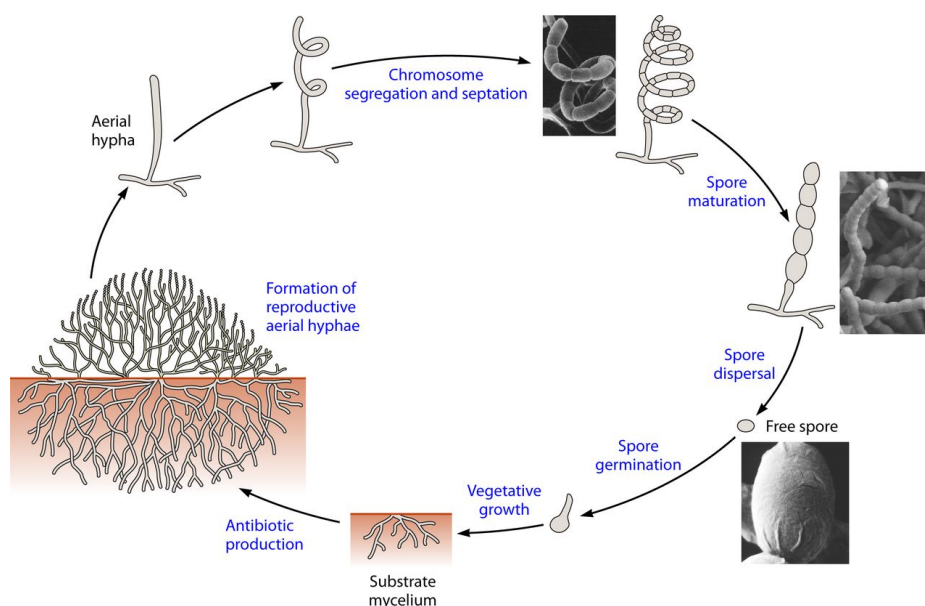


Figure 1. Cycle de développement de *Streptomyces* sur milieu solide (D'après Barka *et al.*, 2016).

1. Le chromosome de *Streptomyces* est hautement compartimenté

D'un point de vue génétique, les *Streptomyces* sont dotées d'un chromosome original. Contrairement à la majorité des procaryotes, elles possèdent un chromosome linéaire enrichi en GC (72%) et de grande taille (Leblond *et al.*, 1993; Lin *et al.*, 1993) pouvant aller de 6 Mb à 12 Mb selon les espèces (Tidjani *et al.*, 2019a). La caractéristique d'un chromosome linéaire n'est pas exclusive aux *Streptomyces*. En effet, parmi les actinobactéries, certaines espèces présentent un chromosome linéaire telles que *Nocardia asteroides* ou *Micromonospora chalcone*, tandis que d'autres comme *Mycobacterium tuberculosis* et *Corynebacterium glutamicum* possèdent un chromosome circulaire (Paradkar *et al.*, 2003).

Le génome de *Streptomyces* est hautement compartimenté : la partie centrale globalement conservée au sein du genre, se distinguant des extrémités chromosomiques, dont le contenu est plus variable. La recherche de gènes dits « core » c'est-à-dire conservés dans l'ensemble des génomes du genre montrent que ceux-ci se concentrent dans la région centrale, de fait nommée région « core » (**Figure 2A**) (Lorenzi *et al.*, 2021). Il est à noter que ces gènes « core » contiennent notamment les gènes impliqués dans le métabolisme primaire de la cellule. Les deux régions flanquant la région core sont appelées « bras chromosomiques » ; ils sont, par définition, dépourvus de gènes core et sont composés de séquences variables entre les espèces et se terminent par des séquences répétées inversées nommées TIR (Terminal Inverted Repeat) (Hopwood, 2006). Parmi ces séquences variables on remarque la présence de nombreux clusters impliqués dans la biosynthèse de métabolites spécialisés (Bury-Moné *et al.*, 2023). Des études de génomique comparative sur une collection de 125 génomes, représentative du genre

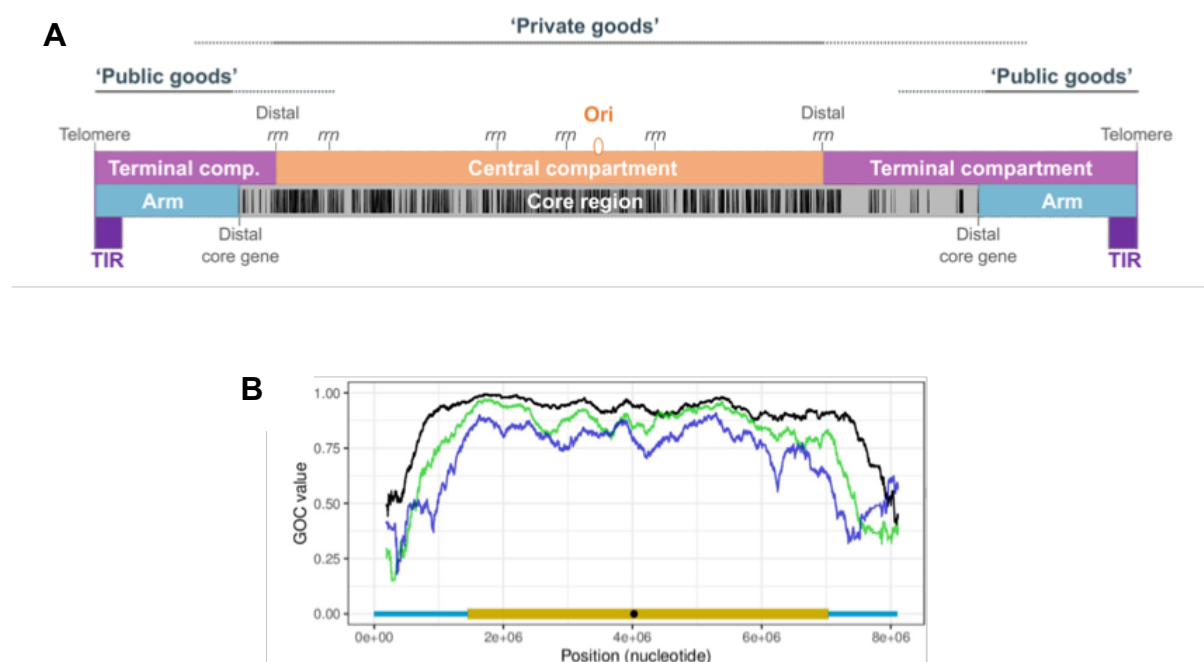


Figure 2. La compartimentation du chromosome de *Streptomyces*. (A) L'origine de réplication (*ori*) est située approximativement au milieu du chromosome linéaire de *Streptomyces*. Il est composé d'un compartiment central comportant des gènes hautement conservés (désignés par les barres verticales noires), considérés comme « biens privés » car nécessaires pour la survie cellulaire. Les régions terminales qui le bordent sont enrichies en « biens publics » c'est-à-dire de gènes conférant un avantage pour la colonie. Les gènes « core » sont partagés par tous les *Streptomyces* et sont retrouvés dans la région « core » ; cette dernière est séparée des bras chromosomiques par la localisation des gènes « core » distaux. Récemment, une nouvelle délimitation de la partie centrale a été proposée sur la base des opérons distaux codant pour des ARN ribosomiaux (*rrn*) (Lorenzi *et al.*, 2022). TIR : terminal inverted repeat. (D'après Bury-Moné *et al.*, 2023). (B) Profil de la conservation de l'ordre des gènes (GOC pour *Gene Order Conservation*) le long du chromosome de *S. ambofaciens* ATCC 23877. Les valeurs de GOC ont été obtenues en comparant *S. ambofaciens* ATCC 23877 avec 3 espèces ayant une distance phylogénétique proche (*S. coelicolor* A3(2), courbe noire), moyenne (*S. reticuli* TUE45, courbe verte), distante (*S. albus* BK3 25, courbe bleue). Les valeurs de GOC ont été calculées en utilisant une fenêtre glissante de taille égale à 5% du contenu en gènes chromosomiques de *S. ambofaciens* ATCC 23877. (D'après Lorenzi *et al.*, 2021).

Streptomyces, ont démontré que la région « core » et les subtélomères n'évoluent pas avec la même dynamique : la région « core » constitue un bloc ancestral moins sujet aux remaniements que les bras chromosomiques (Figure 2B) (Lorenzi *et al.*, 2021). La même observation fut décrite par Choulet *et al.* (2006) lors d'une analyse deux à deux des subtélomères de *Streptomyces ambofaciens* et de *Streptomyces coelicolor* (Choulet *et al.*, 2006b). Il a été mis en évidence l'existence d'un gradient croissant de la distribution des insertions/délétions (indels) le long des bras en direction des extrémités du chromosome. Ce phénomène a pour conséquence la diminution de la synténie (conservation de l'ordre des gènes) dans les régions terminales (Choulet *et al.*, 2006a, 2006b; Lorenzi *et al.*, 2021; Tidjani *et al.*, 2020). Récemment, un nouveau regard sur la compartimentation du génome de *Streptomyces* a été jeté. Il repose sur la localisation des opérons codant les ARN ribosomiaux (*rrn*). Ces derniers (au nombre de 5 à 8 selon les espèces) sont localisés au sein de la région « core » du chromosome, y définissant une

région plus restreinte dénommée le compartiment central et qui coïnciderait avec une région chromosomique dont le niveau d'expression la distinguerait des compartiments dits "terminaux", plus faiblement transcrits (Lorenzi *et al.*, 2022; Liroy *et al.*, 2021).

2. L'instabilité remarquable du chromosome de *Streptomyces*

L'instabilité génétique est un trait connu dans le monde bactérien, mais les *Streptomyces* se démarquent par une plasticité remarquablement plus élevée (Birch *et al.*, 1989). Dans les conditions de laboratoire, il a été montré que dans la descendance de la souche sauvage de *S. ambofaciens*, des variants phénotypiques spontanés affectés dans la pigmentation apparaissent à haute fréquence allant jusqu'à 1% (Martin *et al.*, 1999 ; Leblond *et al.*, 1989). Cette instabilité est liée à la formation de réarrangements d'ADN localisés au niveau des régions subterminales comme des événements de délétions, translocations, circularisations ou encore d'amplifications (Volf and Altenbuchner, 1998). Il a été montré que la circularisation s'accompagne de la perte des deux subtélomères alors que les mutants ayant perdu un seul bras présentent un chromosome linéaire. Dans deux cas de translocation terminale engendrant l'accroissement de la taille des TIR de 210 kb chez la souche sauvage à 480 kb et 850 kb chez deux mutants spontanés de *S. ambofaciens* (**Figure 3**), l'implication de la recombinaison homologue a été avérée. En effet, deux gènes, *hasL* et *hasR*, partageant 99% d'identité et étant situés chacun dans un bras sont le siège des événements de recombinaison (Fischer *et al.*, 1998). Ces remaniements résultent de la délétion d'un des deux bras chromosomiques et son remplacement par le second bras (par un mécanisme de *Break Induced Replication*, voir paragraphe sur la recombinaison homologue).

Cet événement de remplacement de bras a été décrit dans d'autres espèces de *Streptomyces* comme *S. griseus* et *S. coelicolor* A3(2) chez lesquels la recombinaison entre deux gènes de lipoprotéines et deux transposons a été impliquée, respectivement (Uchida *et al.*, 2003 ; Widenbrant *et al.*, 2007). La recombinaison illégitime a également été évoquée dans l'analyse d'un mutant dont le chromosome serait le résultat de la fusion en orientation inversée de deux chromosomes délétés partiellement du bras droit. Cet événement entraîne la formation d'un génome dont la taille est presque deux fois plus grande que la souche sauvage (8 Mb *versus* 14-15 Mb) (Wenner *et al.*, 2003). La formation de cette structure dupliquée du chromosome initie un cycle de transmission de l'instabilité génétique, sans doute consécutive à la présence de deux origines de réplication, *oriC*, et/ou de deux sites de partitionnement chromosomique. En effet, l'héritage d'un tel chromosome confère un état mutateur à la descendance avec l'apparition de près de 30% de variants phénotypiques.

En somme, la plasticité du génome des *Streptomyces* trouverait en partie son origine dans les mécanismes de réparation de l'ADN. Ces systèmes sont impliqués dans la réparation des cassures double-brin (DSB).

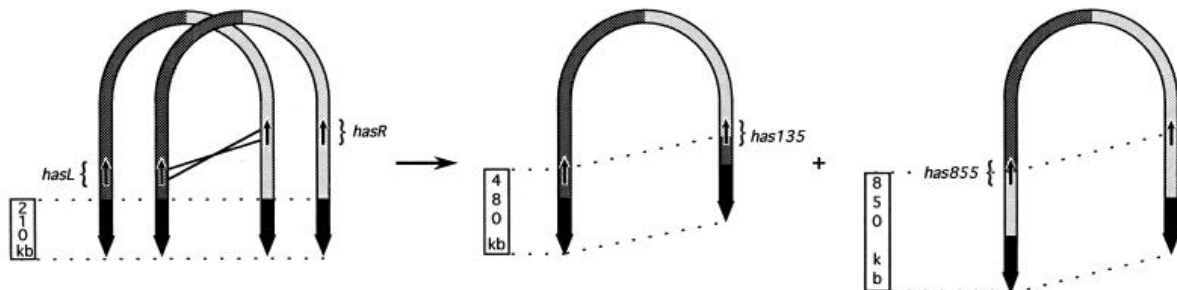


Figure 3. Echange chromosomique entre chromosomes sœurs. Schéma explicatif de la recombinaison entre les loci *hasL* et *hasR* localisés sur deux chromosomes sauvages de *S. ambofaciens*. Il en résulte un chromosome avec des TIR de 480 kb et un autre qui présente des TIR de 850 kb (Fischer *et al.*, 1998).

II. Les forces évolutives du génome bactérien

1. Les transferts horizontaux, voie d'entrée de l'ADN exogène

Les bactéries sont des organismes unicellulaires se multipliant par scissiparité : une cellule mère se divise en deux cellules filles. On parle de transfert vertical de l'information génétique. Le transfert horizontal de gènes (HGT pour *Horizontal Gene Transfer*), quant à lui, peut être défini comme l'acquisition de matériel génétique entre individus non-ascendants. Ce mode de transfert d'information génétique est un moyen rapide et efficace de dissémination de gènes conférant des avantages adaptatifs et donc d'augmenter le *fitness* des individus, c'est-à-dire leur valeur sélective dans leur environnement. Si les mécanismes évolutifs basés sur l'accumulation de mutations ponctuelles conduisent rarement et lentement à l'acquisition de nouvelles fonctions, le HGT permet d'acquérir rapidement des gènes « prêts à l'emploi » en un seul événement. Il revêt, de ce fait, un rôle crucial dans la diversification génétique et représente une force d'évolution rapide des génomes.

Trois principaux mécanismes de transfert horizontal de gènes ont été décrits : la transduction qui consiste au transfert de matériel génétique entre bactéries par l'intermédiaire d'un phage; la transformation naturelle définie comme l'absorption d'ADN libre dans l'environnement, et la conjugaison qui se traduit par le transfert de l'ADN par contact direct entre deux bactéries suite à la formation d'un pore de conjugaison entre les deux cellules (Lederberg and Tatum, 1946). L'ADN exogène

entrant dans la cellule receveuse est maintenu soit en se répliquant de manière autonome, si celui-ci en a la capacité, soit en s'intégrant dans le génome hôte par recombinaison. En somme, l'entrée de l'ADN exogène ainsi que son maintien dans la cellule sont deux facteurs clé de la dynamique des génomes.

1.1. La conjugaison bactérienne, principal mécanisme de HGT chez *Streptomyces*

La conjugaison bactérienne serait le principal mécanisme de transfert horizontal chez les *Streptomyces*. En effet, peu d'éléments sont connus sur la transformation naturelle des bactéries du genre *Streptomyces* (Roelants *et al.*, 1976) et seulement quelques phages transducteurs ont été décrits dans le genre tel que ϕ SV1 chez *Streptomyces venezuelae* (Stuttard, 1979; Burke *et al.*, 2001). Par conséquent, il est admis, chez *Streptomyces*, que le HGT est majoritairement médié par la conjugaison (Tidjani *et al.*, 2019b; Choufa *et al.*, 2022). Par ailleurs, de nombreux éléments intégratifs et conjugatifs ont été identifiés au sein du genre (Choufa *et al.*, 2022 ; Tidjani *et al.*, 2019b) confirmant que la conjugaison est le mécanisme majeur d'acquisition de séquences exogènes chez *Streptomyces*.

Chez les *Streptomyces*, la conjugaison est médiée par une protéine nécessaire et suffisante, TraB (Vogelmann *et al.*, 2011). Cette dernière est homologue de FtsK, une enzyme impliquée dans la ségrégation des chromosomes lors de la division cellulaire (Vogelmann *et al.*, 2011). TraB est une ATPase qui en s'oligomérisant forme le pore conjugatif par lequel l'ADN est transloqué. Le gène *traB* est porté par des plasmides ou par des ICE spécifiques aux actinomycètes appelées AICE pour Actinomycete ICE. Après excision, l'AICE est reconnu par TraB qui ne génère aucune coupure et il est transloqué sous forme circulaire et double-brin vers l'hyphe récepteur (Possoz *et al.*, 2001). La reconnaissance de l'élément conjugatif par TraB serait médiée par un court motif nucléotidique (8 nt) appelé *clt* (*cis-acting locus of transfer*) (Thoma and Muth, 2016). Après le transfert de l'élément mobile de la cellule donneuse à la cellule receveuse, ce dernier est propagé au sein du mycélium récepteur. Cette propagation dépend d'un complexe multiprotéique composé de TraB et de protéines de diffusion appelées Spd (*spread proteins*). La diffusion intra-mycélienne de l'élément conjugatif induit un retard de croissance local du mycélium, et elle se traduit par l'apparition de halos circulaires appelées *pocks*. (Thoma *et al.*, 2015)

1.2. Le maintien de l'ADN exogène dans le génome

Une fois transféré de la cellule donneuse à la receveuse, le matériel génétique se réplique de manière autonome s'il abrite sa propre origine de répllication, sinon il doit être intégré dans le génome hôte afin d'être maintenu. Ce processus est médié par les systèmes de recombinaison comme la recombinaison homologue, la recombinaison illégitime ou la recombinaison site-spécifique. Si la recombinaison

homologue requiert des séquences homologues au génome hôte pour l'intégration de l'ADN exogène, la recombinaison illégitime, si elle est opérationnelle dans l'organisme receveur, lève cette contrainte de la présence d'une matrice homologue, en procédant à la ligature d'extrémités d'ADN dissemblables. Ce mécanisme est mutagène ; il génère des mutations 'ponctuelles' au site de la recombinaison, et engendre des réarrangements chromosomiques de grande ampleur (Thibessard *et al.*, 2021 ; Bertrand *et al.*, 2019).

Si l'ADN exogène est fixé dans le génome, il peut conférer un avantage à l'hôte. S'il s'agit d'une phase codante, elle devra s'insérer dans le réseau de contrôle de l'expression pour éviter les effets délétères (Thomas and Nielsen, 2005). Des mécanismes de contrôle de l'expression des gènes exogènes ont ainsi été mis en évidence chez plusieurs espèces bactériennes dont les *Streptomyces*. En outre, au cours de l'évolution, les bactéries ont mis en place des barrières à l'intégration de l'ADN exogène, comme les systèmes de restriction-modification, les systèmes CRISPR (Getino and de la Cruz, 2018) ou encore la modulation de la recombinaison homéologue (Spies and Fishel, 2015; Rayssiguier *et al.*, 1989).

2. La recombinaison dans les systèmes de réparation de l'ADN

Les remaniements du chromosome sont médiés par les systèmes de recombinaison qui interviennent dans la réparation des cassures double-brin. Il existe deux grands mécanismes de réparation des DSB : la recombinaison homologue et la recombinaison illégitime.

Dans ce paragraphe, nous aborderons l'origine des DSB et nous décrirons les aspects mécanistiques de la recombinaison homologue et de la recombinaison illégitime. Enfin, les résultats mettant en évidence la réparation des DSB comme moteur de l'évolution chez *Streptomyces* seront présentés.

2.1. Les sources de DSB

La cassure double-brin est la lésion de l'ADN la plus délétère pour une cellule vivante, menant à la mort cellulaire si elle n'est pas réparée. Les DSB peuvent avoir des origines exogènes ou endogènes. Des facteurs de nature soit physique comme les radiations ionisantes et les rayons ultraviolets (UV), soit chimique tels que les agents alkylants (ex. la mitomycine C) ou les agents radiomimétiques (ex. la bléomycine) (Bonura and Smith, 1975; Pfeiffer *et al.*, 2000) induisent des DSB entre autres dommages.

Les DSB peuvent aussi trouver leur origine dans le métabolisme cellulaire, notamment par le biais d'espèces réactives de l'oxygène telles que les radicaux libres. Ces composés oxydent les composants cellulaires et modifient notamment la structure de certaines bases (ex. la guanine peut être convertie en 8-oxo-guanine). Des cassures simple-brin peuvent alors résulter de leur réparation par le mécanisme de réparation par excision de base (BER pour *Base Excision Repair*). L'accumulation de ces cassures simple-brin ou le passage de la fourche de réplication peut convertir ces dommages en DSB. De plus, le

processus de réplication de l'ADN est également une source de DSB. En effet, un arrêt prolongé de la machinerie de réplication peut entraîner une cassure simple-brin ; si cette dernière n'est pas réparée, elle peut conduire à la formation d'une DSB lors du redémarrage de la machinerie de réplication (Pfeiffer *et al.*, 2000).

Par ailleurs, certaines DSB peuvent être programmée par la cellule elle-même, comme par exemple chez les vertébrés, lors de la création de diversité au sein des récepteurs antigéniques des lymphocytes B et T. Dans ce processus contrôlé, la réparation des DSB générées favorise la formation de récepteurs antigéniques diversifiés, conférant au système immunitaire sa flexibilité et son efficacité face à la multitude des antigènes auxquels les organismes sont exposés (Jackson and Jeggo, 1995).

2.2. La recombinaison homologue

La recombinaison homologue est un mécanisme bien caractérisé chez les modèles bactériens *E. coli* et *Bacillus subtilis* et la levure *Saccharomyces cerevisiae*, modèle eucaryote. Quel que soit le modèle d'étude, la protéine RecA (Rad51 est l'homologue eucaryote) constitue un acteur central dans la recombinaison homologue. En ce qui concerne les *Streptomyces*, peu d'éléments sont connus à ce jour autant sur l'aspect mécanistique que fonctionnel. Plusieurs modèles de prise en charge des DSB (**Figure 4**) ont été proposés pour répondre de la réparation des DSB dans différentes situations chromosomiques et cellulaires (tissus somatiques, méiose,...). Ainsi, le SDSA, Synthesis-Dependent Strand Annealing (Nassif *et al.*, 1994), est conservateur et prédomine pour la réparation des cassures de l'ADN dans les cellules somatiques. Le BIR (*Break-Induced Replication*) (Formosa and Alberts, 1986), intervient sur une cassure d'ADN en l'absence d'une seconde extrémité disponible comme celles générées par l'effondrement d'une fourche de réplication. Ce mécanisme est notamment responsable dans la formation de duplications segmentales chez la levure (Liti and Louis, 2005), les eucaryotes (ref) et aussi chez les bactéries notamment les *Streptomyces* (Hoff *et al.*, 2018). Le modèle *Double-strand Break Repair* (DSBR) proposé par Jack W. Szostak en 1983 répond le mieux aux observations et conséquences de la recombinaison chez les procaryotes lors des phases réplcatives que chez les eucaryotes en méiose. Si les acteurs de la recombinaison homologue diffèrent selon les organismes, les étapes et intermédiaires de recombinaison du modèle DSBR, présentés ci-dessous, sont communes. Trois étapes majeures sont distinguées :

- **La phase pré-synaptique**

Lors de cette phase, la résection de la DSB conduit à la formation d'extrémités 3' sortantes. Chez *E. coli* et *B. subtilis*, cette maturation des extrémités est assurée par le complexe hélicase-nucléase RecBCD et AddAB, respectivement. Le complexe dégrade l'ADN double-brin jusqu'à la rencontre d'un motif

nucléotidique court, appelé Chi (pour *Crossover Hot-spot Instigator*), à partir de laquelle seule l'extrémité 5' est dégradée, conduisant ainsi à la formation d'extrémités 3' sortantes. Bien que des homologues de RecBCD existent chez *Mycobacterium*, le mutant de délétion $\Delta recBCD$ ne montre aucune sensibilité aux UV suggérant, de ce fait, l'implication d'un autre mécanisme de résection de l'ADN. Chez les mycobactéries, cette activité est assurée par un nouveau complexe hélicase-nucléase, nommé AdnAB (Sinha *et al.*, 2009). Le mutant *adnAB* est viable et il présente une sensibilité accrue aux radiations ionisantes. Cette observation suggère que AdnAB joue un rôle dans la réponse au stress génotoxique et la réparation de l'ADN.

Chez les bactéries du genre *Streptomyces*, aucun homologue de RecBC, ni de AddAB n'a été identifié (Rocha *et al.*, 2005). En revanche, les gènes *adnAB* existent chez tous les *Streptomyces* analysés. AdnAB présente une forte conservation des domaines hélicase et nucléase (Zhang *et al.*, 2014). Les analyses fonctionnelles ont montré que la délétion de *adnAB* est létale soulignant son rôle essentiel dans la survie de la cellule. Les tests de complémentation ont démontré que l'ajout de *adnAB* dans une souche $\Delta recB$ de *E. coli* augmente le taux de survie d'un facteur 10. De plus, ce mutant de complémentation présente une résistance accrue aux rayons UV avec un taux de survie 60 fois supérieur à celui de la souche $\Delta recB$. Ainsi, *adnAB* de *Streptomyces* peut restaurer la capacité de la souche *E. coli* $\Delta recB$ à réparer les lésions induites par le rayonnement UV. Ces résultats montrent que AdnAB constitue un homologue fonctionnel de RecBCD de *E. coli* et est impliqué dans la réparation de l'ADN.

- **La phase synaptique**

L'échange du brin endommagé avec le brin intact est une étape centrale dans la réparation des DSB par recombinaison homologue. En ce sens, la recombinase RecA (Rad51 chez les eucaryotes et RadA chez les archées) joue un rôle crucial dans le processus (Cox, 1999). La liaison des protéines RecA à l'ADN simple-brin 3' sortant formé lors de l'étape pré-synaptique constitue le nucléofilament ou filament RecA. Celui-ci assure la recherche d'homologie et l'envahissement d'un duplex d'ADN homologue conduisant ainsi à une structure intermédiaire appelée « D-loop » (ou D-boucle).

L'activité synaptique de la protéine bactérienne RecA est retrouvée chez tous ses homologues fonctionnels (Rad51, RadA). RecA représente donc un acteur majeur de la recombinaison homologue. La délétion de *recA* est létale ou tolérée bien que délétère selon les organismes (Huang and Chen, 2006).

- **La phase post-synaptique**

L'échange des brins catalysé par RecA conduit à la formation d'une structure à 4 brins connue sous le nom de « jonctions de Holliday ». Le déplacement de ces intermédiaires, appelé migration de la branche, est assurée par le complexe RuvAB ou par l'hélicase RecG. Cette étape génère au sein des duplexes, la formation de zones d'ADN hétéroduplexe, c'est-à-dire, où les deux brins de la double hélice ne sont pas complémentaires à toutes les positions (positions correspondant aux variations alléliques entre les molécules parentales). La correction de ces mésappariements relève du système *MisMatch Repair* (MMR) et participe de la conversion génique accompagnant tous les événements de recombinaison homologue. Enfin, les jonctions de Holliday sont résolues par l'endonucléase RuvC, et les molécules recombinées, c'est-à-dire ayant subi un crossing-over, ou non sont ségréguées.

Chez les *Streptomyces*, les gènes *ruvABC* et *recG* sont fortement conservés (Rocha *et al.*, 2005). Au sein du laboratoire, les analyses génétiques chez *S. ambofaciens* ont mis en lumière le rôle de RuvABC et RecG dans la recombinaison post-conjugative (Hoff *et al.*, 2017). En effet, la double mutation $\Delta ruvABC \Delta recG$ confère une efficacité de recombinaison jusqu'à 15 fois inférieure à la souche sauvage. Notons que le phénotype du double mutant n'est pas aussi marqué que celui du mutant $\Delta recA$ chez lequel aucun transconjugant n'a été observé. Par conséquent, il a été suggéré l'existence de protéines alternatives capables de prendre en charge la migration de la branche et la résolution des jonctions de Holliday chez *Streptomyces* (Hoff *et al.*, 2017).

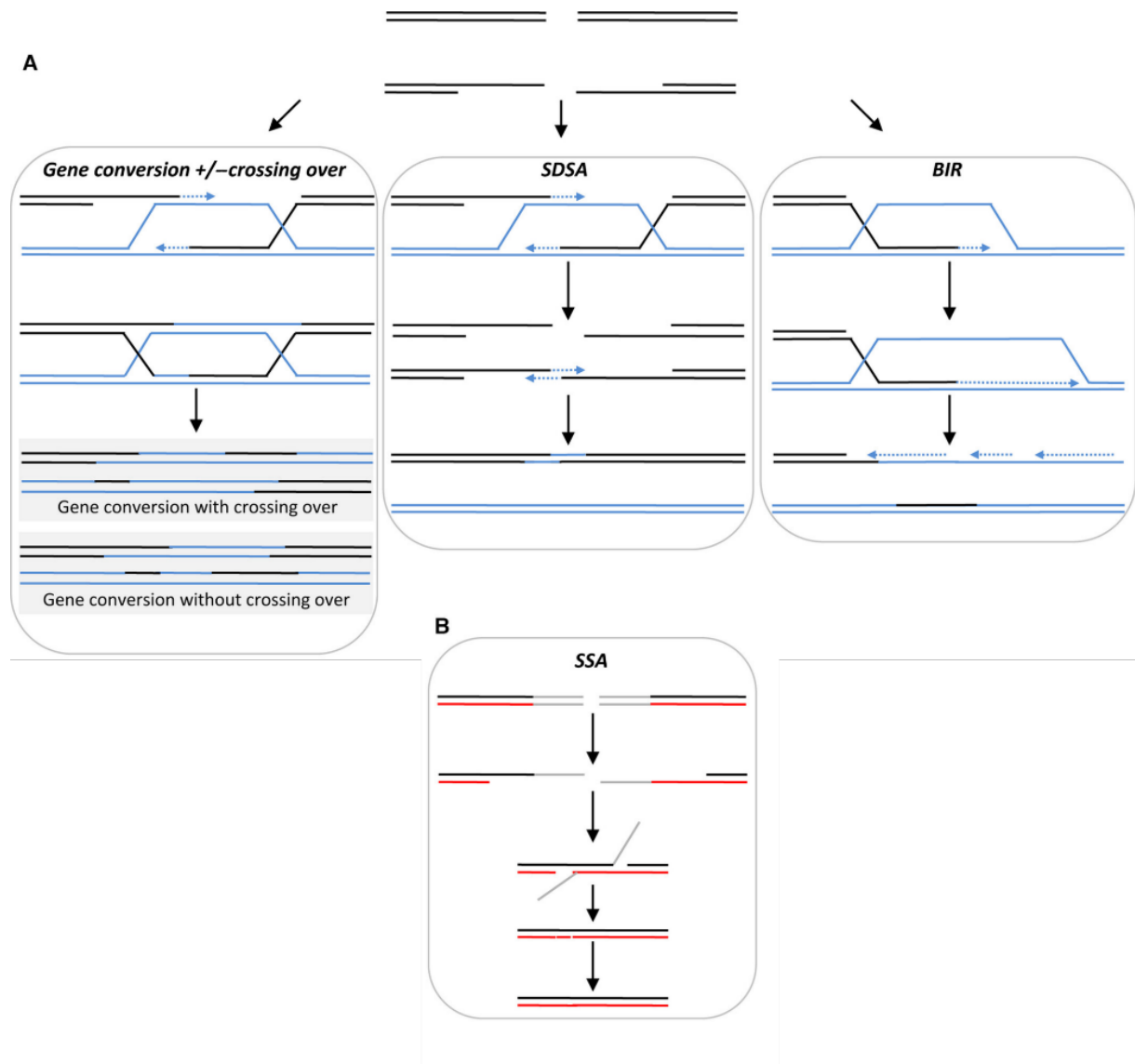


Figure 4. Différentes voies de réparation d'une DSB par recombinaison homologue. Lorsqu'une DSB survient, la première étape de résection générant des extrémités 3' simple-brin est commune aux 4 mécanismes décrits. **(A)** Sont schématisés la conversion génique, le *Synthesis-Dependant Strand Annealing* (SDSA) et le *Break-Induced Replication* (BIR). Pour la **conversion génique** (panel gauche), un brin simple-brin envahit un ADN double-brin homologue intacte par le biais de RecA. L'invasion du brin génère une structure en forme de boucle appelée « D-loop » (boucle de déplacement). Cette structure comporte des molécules d'ADN hétéroduplex et des intermédiaires cruciformes connus sous le nom de « jonction de Holliday ». Le brin 3' envahissant amorce la synthèse de l'ADN en utilisant l'ADN intact comme matrice ce qui entraîne la migration de la jonction de Holliday. La résolution des jonctions de Holliday peut se faire avec ou sans cross-over. Dans le **SDSA** (panel du milieu), le brin d'ADN néoformé est déplacé de la matrice, s'associe au brin complémentaire et est relié avec l'autre extrémité de la molécule d'ADN cassée. Le processus de **BIR** (panel droite) commence par une DSB à une extrémité du chromosome. Une D-loop est formée suite à l'envahissement du brin simple-brin 3'. Ce dernier est utilisé pour initier la réplication. La fourche de réplication migre le long du chromosome jusqu'à atteindre l'extrémité du chromosome. **(B)** Si une DSB survient entre deux séquences répétées, elle peut être réparée par la voie **SSA** : les extrémités 3' de la cassure sont maturées et la séquence répétée du brin sortant située en amont de la cassure s'hybride avec sa séquence complémentaire située sur le brin sortant en aval de la cassure. La résolution des intermédiaires et le comblement des lacunes entraîne la perte de la séquence localisée entre les deux répétitions. (D'après So *et al.*, 2017)

2.3. La recombinaison illégitime

Chez les bactéries, le principal mécanisme de réparation des DSB est la recombinaison homologue. Cependant, la séquence d'ADN endommagée ne peut être réparée par recombinaison homologue qu'en présence d'une séquence homologue, par exemple provenant d'un autre chromosome (au sein des hyphes polygénomiques des *Streptomyces*) ou de la chromatide sœur lors des phases répliquatives (phase végétative). Or, lors de la phase stationnaire de croissance ou dans les spores, une seule copie du génome existe (Bury-Moné *et al.*, 2023). En outre ces phases sont propices aux agressions du génome (ex. accumulation des dommages oxydants, DSB liées à la dessiccation) (Dupuy *et al.*, 2019). Dans ce cas, les DSB peuvent être prises en charge par le *Non-Homologous End-Joining* (NHEJ), un mécanisme connu de longue date, prégnant et bien documenté chez les eucaryotes (De Bragança *et al.*, 2023) mais encore mal compris chez les bactéries (Thibessard *et al.*, 2021; Bertrand *et al.*, 2019). Le NHEJ implique la jonction d'extrémités d'ADN provenant de séquences partageant peu ou pas d'homologies. Le NHEJ est réputé mutagène avec à la jonction la possibilité de mutations ponctuelles, d'insertions et délétions, et générateur de grands réarrangements chromosomiques.

- **Le NHEJ eucaryote**

Chez les eucaryotes, la réparation des DSB par NHEJ fait intervenir de nombreux facteurs, avec plus de huit acteurs participant à ce processus (Pannunzio *et al.*, 2018). Le complexe hétérodimérique Ku70–Ku80, responsable de la reconnaissance et de la stabilisation des extrémités d'ADN, se lie aux extrémités de la cassure. Le complexe recrute une ADN-kinase, la DNA-PK, qui recrute à son tour le complexe ligase IV–XRCC4–XLF pour assurer le pontage et la ligature des extrémités.

D'autres facteurs accessoires sont également recrutés au site de cassure pour la maturation des extrémités comme Artemis (une 5'-3' exonucléase) ou encore les polymérases Pol μ , Pol λ pour le comblement des lacunes.

- **Le NHEJ bactérien**

Le NHEJ n'est pas un mécanisme conservé chez les bactéries ; seulement 20% des génomes bactériens séquencés présentent des homologues de Ku (McGovern *et al.*, 2016; Sharda *et al.*, 2020). Des gènes homologues de *ku* ont été identifiés dans différentes espèces bactériennes comme les mycobactéries, *Bacillus subtilis* ou encore le genre *Streptomyces* (Weller *et al.*, 2002; Aravind and Koonin, 2001). Le premier NHEJ bactérien décrit est celui de *Mycobacterium tuberculosis* qui fait intervenir deux protéines, Ku et LigD ((Stephanou *et al.*, 2007) ; modèle expliqué en **Figure 5**). Ce système est considéré minimal car il ne fait intervenir que 2 acteurs protéiques : Ku et la protéine LigD. Cette dernière, chez *M. tuberculosis* est une protéine multifonctionnelle composée de 3 domaines : (i) un domaine Pol-Dom qui

assure une activité ADN et ARN polymérase, ainsi qu'une activité primase, c'est-à-dire l'ajout de nucléotides aux extrémités de l'ADN en absence de matrice ; (ii) un domaine ligase (Lig-Dom) et (iii) un domaine 3'-5'nucléase (Nuc-Dom) (Pitcher *et al.*, 2007; Bertrand *et al.*, 2019). Cependant, chez d'autres espèces comme *Sinorhizobium meliloti*, le NHEJ apparaît plus complexe avec plusieurs mécanismes co-existants et intervenant en réponse à des stimuli distincts (Dupuy *et al.*, 2019).

De même chez les *Streptomyces*, le NHEJ est complexe. Chez *S. ambofaciens*, alors que 3 homologues de Ku (KuA, KuB, et KuC) ont été identifiés, aucun homologue de LigD présentant les 3 domaines fonctionnels n'a été trouvé (Hoff *et al.*, 2016). En revanche, les différents domaines fonctionnels de LigD sont trouvés sous forme de polypeptides codés de façon indépendante, et ce, en plusieurs exemplaires dans le génome (Hoff *et al.*, 2016; Bertrand *et al.*, 2019; Thibessard *et al.*, 2021). Les analyses génétiques et fonctionnelles ont montré que les 3 homologues de Ku, 2 homologues de Lig-Dom et 2 homologues de Pol-Dom de LigD sont impliqués dans la réponse aux dommages à l'ADN (Hoff *et al.*, 2016). Par ailleurs, il a été montré que les protéines Ku s'accumulent dans les spores suggérant que le NHEJ pourrait jouer un rôle dans la résistance au cours de la phase de sporulation, la dissémination et/ou la germination, c'est-à-dire lorsqu'une seule copie du chromosome existe. De plus, les tests *in vitro* ont montré que les 3 protéines Ku de *S. ambofaciens* sont capables de protéger les extrémités d'ADN cassées et de stimuler leur ligation. L'ensemble de ces résultats tend à montrer qu'il existe un NHEJ fonctionnel chez les *Streptomyces*. En revanche, le rôle des acteurs 'redondants' du NHEJ des *Streptomyces* dans la réparation des DSB ou d'autres dommages de l'ADN (ex. le BER) demeurent investiguées.

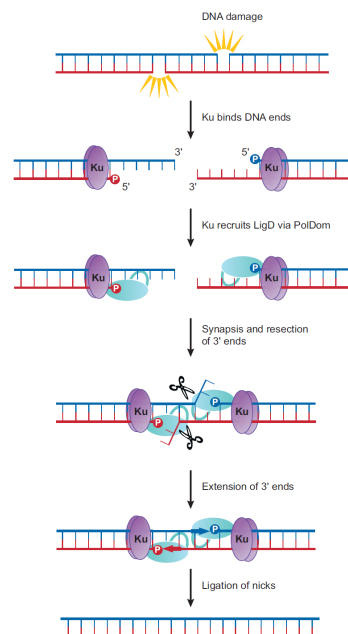


Figure 5. Modèle de réparation d'une DSB par NHEJ chez *Mycobacterium*. Les extrémités cassées de la DSB sont reconnues par l'homodimère Ku, qui recrute LigD *via* Pol-Dom. La maturation des extrémités est assurée par les activités polymérase, nucléase et ligase de LigD. La religature des extrémités est assurée par Lig-Dom. (Pitcher *et al.*, 2007).

3. Les mécanismes de recombinaison façonnent le chromosome des *Streptomyces*

Comme présenté dans la section I., les bactéries du genre *Streptomyces* possèdent une plasticité génétique élevée corrélée à la formation de grands réarrangements terminaux tels que des délétions de grandes tailles, des amplifications ou encore des translocations (Thibessard and Leblond, 2014). Il a également été montré que le chromosome des *Streptomyces* présentait un gradient d'insertions et délétions croissant dans les régions terminales en direction des extrémités chromosomiques (Lorenzi *et al.*, 2021). Ainsi, les régions subtélomériques correspondent à des points chauds pour l'intégration de matériel exogène issu du transfert horizontal (principalement la conjugaison chez *Streptomyces*) et le remaniement de séquences. Notre hypothèse de travail est l'existence d'un lien entre la plasticité du chromosome et la réparation des cassures double brin. Dans des travaux antérieurs, des DSB ont été introduites artificiellement dans la région centrale et les bras chromosomiques de *S. ambofaciens* (grâce à l'expression hétérologue de la méganucléase I-SceI et l'introduction d'un site unique de reconnaissance de l'endonucléase), permettant d'observer la structure chromosomique des survivants (Hoff *et al.*, 2018). L'analyse du génome des survivants montre que dans la région centrale, les DSB sont réparées de façon mutagène, et que ces mutations ne sont pas produites dans un contexte déficient pour KuA ou LigC et LigD démontrant l'existence d'un mécanisme NHEJ fonctionnel chez *S. ambofaciens* (Hoff *et al.*, 2018). Par ailleurs, toutes les structures chromosomiques observées chez les mutants spontanés de *S. ambofaciens* sont retrouvées chez les survivants. Ainsi, la réparation des DSB engendre tous les remaniements observés dans les génomes des variants phénotypiques spontanés suggérant que la plasticité est directement liée à la réparation des DSB. Enfin, des événements d'insertion de séquences d'ADN (de 30 bp à 18,8 kb) ont été identifiés au site de réparation de la DSB. La capacité du NHEJ à favoriser l'insertion de matériel génétique avait déjà été montrée chez *M. smegmatis* (Aniukwu *et al.*, 2008), chez *Sinorhizobium* (Dupuy *et al.*, 2019) et chez la levure *S. cerevisiae* (Ricchetti *et al.*, 1999). Le NHEJ pourrait donc favoriser l'intégration de matériel génétique issu du transfert horizontal (Hoff *et al.*, 2018).

III. Une nouvelle voie de MMR : origine de DSB endogènes ?

L'hypothèse d'un lien entre réparation des DSB et plasticité génomique chez *Streptomyces* soulève la question de l'origine des DSB, et si ce lien établit des causes de leur distribution le long du génome. Dans la section II.2.1, les facteurs exogènes et endogènes pouvant générer des DSB ont été décrits. Dans le cadre de nos travaux, nous nous sommes intéressés à une autre source potentielle de DSB, découverte récemment chez les actinobactéries et les archées : la voie non-canonique de réparation des

mésappariements. Cette dernière fait intervenir une endonucléase appelée NucS/EndoMS. Dans cette partie, nous exposerons brièvement la genèse des mésappariements post-réplicatifs, le MMR canonique, puis la découverte et le mécanisme de la voie non-canonique. Nous focaliserons notamment sur les différences identifiées jusqu'à présent entre les deux mécanismes et leurs conséquences.

1. Les erreurs de réplication

La réplication est un mécanisme semi-conservatif, il repose sur la néo-synthèse d'un brin d'ADN par copiage du brin parental (Lewis *et al.*, 2016). Ce processus, hautement fidèle, est régi par trois mécanismes majeurs : la spécificité d'incorporation des nucléotides par l'ADN polymérase, l'élimination des nucléotides mal appariés par l'exonucléase 3' – 5' de correction associée à la polymérase (appelée activité *proofreading*) et la réparation post-réplivative des mésappariements. Chez *E. coli*, la réplication est assurée par l'ADN polymérase III (Robinson and van Oijen, 2013). Cette dernière fait partie des polymérases dites « *error-free* » car elle commet rarement des erreurs d'incorporation de bases et elle possède une fonction de relecture qui augmente la fidélité de réplication. La polymérase III est composée de différents partenaires protéiques dont la sous-unité α (également appelée DnaE) qui assure l'activité polymérase 5' – 3' et la sous-unité ϵ qui porte l'activité exonucléase 3' – 5' nécessaire pour l'excision des nucléotides mal appariés. La sous-unité α interagit avec le facteur de processivité β , ce qui augmente sa processivité. Cette sous-unité α comprend quatre domaines dont notamment le domaine PHP (Polymerase and Histidinol Phosphate) qui assure la stabilité de la polymérase. Il a été montré chez *Mycobacterium* que la sous-unité ϵ , codée par le gène *dnaQ*, ne possède pas une activité exonucléase 3' – 5' de relecture. Dans ce cas de figure, c'est le domaine PHP de la sous-unité α (DnaE) qui assure cette activité *proofreading* (Rock *et al.*, 2015), rendue possible grâce à un centre zinc tri-nucléaire coordonné par neuf résidus d'acides aminés conservés. Bien que l'activité de relecture augmente la fidélité d'un facteur 1000, des mésappariements peuvent s'échapper. S'ils ne sont pas corrigés, ils engendrent des mutations ponctuelles au cours du cycle de réplication suivant. Ces mésappariements sont corrigés par le système MMR, un mécanisme important de surveillance de l'ADN au cours de la division cellulaire.

2. Le MMR canonique

Le MMR est un des mécanismes les plus importants assurant l'intégrité du génome. Il a pour rôle de prendre en charge les petites insertions-délétions (1-4 bases) et les erreurs d'incorporation de bases qui ont échappées à l'activité exonucléase 3'-5' de la polymérase lors de la réplication. En corrigeant ces mésappariements résiduels, le MMR contribue à accroître la fidélité de la réplication d'environ 3 ordres de grandeur (On and Welch, 2021).

Le MMR canonique aussi appelé « MutSL » peut être subdivisé en deux mécanismes distincts mettant en jeu les deux protéines majeures que sont MutS et MutL : le MMR dépendant de la méthylation et le MMR indépendant de la méthylation.

2.1. Le MMR canonique méthyl-dépendant

Le MMR méthyl-dépendant est retrouvé dans les gamma-protéobactéries, y compris les *Enterobacteriales* et les *Vibrionales* ; son mode d'action a été largement caractérisé chez la bactérie modèle *E. coli* au cours des dernières décennies (Modrich, 1987). Contrairement aux cellules eucaryotes, le système repose sur l'hémi-méthylation de l'ADN lors de la réplication (Putnam, 2021). Chez *E. coli*, la méthyltransférase Dam assure l'ajout d'un groupement méthyl à l'adénosine des séquences palindromiques d(GATC) retrouvées dans le génome. Après la réplication de l'ADN, les sites Dam sont transitoirement hémi-méthylés avec le brin parental méthylé et le brin néo-synthétisé non méthylé. Cette hémi-méthylation temporaire est le signal de distinction entre le brin parental et le brin nouvellement synthétisé dans le MMR canonique méthyl-dépendant. Cette distinction est cruciale puisqu'un phénotype mutateur est observé en cas de perte ou de surexpression de la méthyltransférase Dam (Herman and Modrich, 1981).

Dans le MMR canonique, l'homodimère MutS est hautement conservé et il interagit avec le facteur de processivité du réplisome, le β -clamp. La reconnaissance et la liaison de MutS à un mésappariement provoquent un changement conformationnel qui permet le recrutement de MutL. Le complexe MutS/MutL active l'endonucléase MutH capable de cliver le brin nouvellement synthétisé non-méthylé au site d(GATC) le plus proche (Li et al., 2008). Une fois l'incision réalisée, l'hélicase UvrD et les exonucléases interviennent pour dérouler l'ADN et dégrader le brin erroné en direction du mésappariement. Le brin parental monocaténaire est protégé par la protéine Ssb tandis que l'ADN polymérase III synthétise le brin complémentaire au niveau de la brèche ; enfin la ligase A procède à la ligature (**Figure 6**).

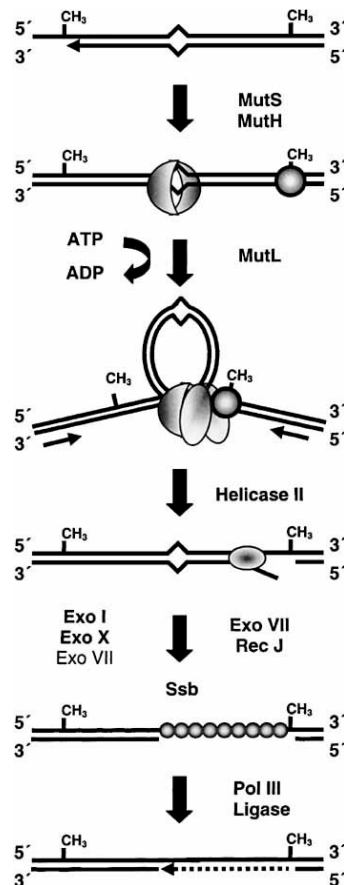


Figure 6. Schéma du mécanisme MMR méthyl-dépendant chez *E. coli*. Les mésappariements qui ne sont pas corrigés par l'activité de relecture de la polymérase sont reconnus par la protéine dimérique MutS. L'enzyme MutH est fixée aux sites hémiméthylés. La liaison d'une molécule d'ATP induit la formation du complexe MutS-MutL. L'hydrolyse de l'ATP active l'endonucléase MutH qui génère une incision sur le brin néosynthétisé au niveau de sites hémiméthylés. L'hélicase II (UvrD) déroule l'ADN permettant ainsi la dégradation du brin fille par des exonucléases. Le brin parental est protégé par les Ssb (*single strand binding proteins*) puis la synthèse du brin naissant est rendu possible par l'intervention de l'ADN polymérase III et les extrémités sont ligaturées par la ligase A (Marti *et al.*, 2002).

2.2. Le MMR indépendant de la méthylation

Chez les organismes utilisant le MMR canonique méthyl-dépendant, l'état de méthylation du génome est un signal indispensable pour guider le clivage simple-brin de MutH et déclencher la suite de la cascade. Or, les cellules eucaryotes ainsi que la majorité des bactéries sont dépourvues d'homologue de MutH, suggérant l'existence d'un autre mécanisme de discrimination des brins d'ADN (Fukui *et al.*, 2010). Les études du MMR humain ont révélé que l'homologue de MutL présente une activité endonucléasique simple-brin intrinsèque due à un domaine de liaison au métal situé en C-terminale de la protéine (Guarné, 2012) (**Figure 7**). Ainsi, les incisions provoquées par MutL pourraient servir de point de départ de la correction des mésappariements. L'activité endonucléase de MutL doit être régulée pour éviter les clivages incontrôlés pouvant entraîner des DSB qui mettraient en péril la stabilité du génome. Des travaux chez *B. subtilis* montrent que l'interaction de MutL avec le β -clamp augmente l'activité

endonucléasique de MutL (Pillon *et al.*, 2011). Bien que le mode de régulation de MutL reste à élucider, il semblerait que le facteur de processivité β -clamp joue un rôle dans la régulation de l'activité de clivage de MutL.

Pour les organismes dépourvus de MutH, le processus de discrimination des brins reste inconnu. Des analyses cristallographiques suggèrent que l'interaction de MutL avec le β -clamp induit le bon positionnement du domaine endonucléase de MutL sur le brin néo-synthétisé pour le clivage (Putnam, 2021). Une autre hypothèse chez *B. subtilis* est basée sur la présence de cassures (*nick*) sur le brin contenant le mésappariement qui pourrait servir de signal pour le distinguer du brin parental. Ces cassures peuvent provenir de l'élimination des ribonucléotides incorporés pendant la réplication de l'ADN. Lors de la réplication du génome, les rNTP (ribonucléoside triphosphates) sont occasionnellement utilisés et incorporés à la place des dNTP. Leur élimination par le RER (Ribonucleotide Excision Repair) génère une cassure simple-brin qui pourrait servir de signal de distinction entre brin néosynthétisé et le brin parental (Yao *et al.*, 2013; Lenhart *et al.*, 2016).

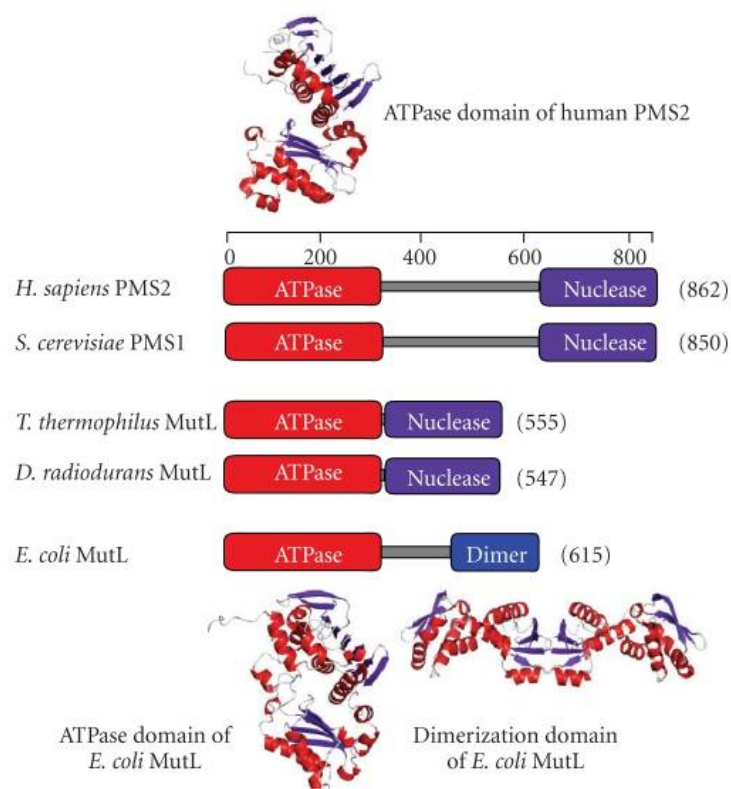


Figure 7. Représentation schématique des domaines des homologues MutL. PMS2 et PMS1 sont des homologues de MutL chez *Homo sapiens* et *Saccharomyces cerevisiae* respectivement. Chez ces derniers ainsi que chez *Thermococcus thermophilus* et *Deinococcus radiodurans*, la protéine MutH n'est pas retrouvée. En revanche, la protéine MutL de ces organismes comporte un domaine endonucléase qui est manquant chez *E. coli*. Les organismes dépourvus de MutH font intervenir l'activité endonucléase de MutL pour la correction de mésappariements (Fukui, 2010).

2.3. Les protéines du MMR canonique en lien avec des voies de réparation de l'ADN

Un certain nombre de preuves mettent en évidence l'interconnexion entre le MMR et des mécanismes de réparation de l'ADN comme les systèmes d'excision de bases (BER) et de nucléotides (NER).

- **Le BER**

Le stress oxydant engendre des espèces réactives de l'oxygène qui attaquent toutes les biomolécules de la cellule (Cadet et al., 2003). L'ADN est une molécule hautement sensible aux attaques par les radicaux de l'oxygène. Ces agressions peuvent, entre autres, entraîner des bases oxydées comme la 8-oxoGuanine (8-oxoG). Ce type de lésions engendre lors du cycle de réplication suivant des mutations de type transversions C:G→A:T et A:T→C:G. Elles sont le résultat de la capacité de la 8-oxoG à s'apparier efficacement non seulement avec la cytosine, mais aussi avec l'adénine (David *et al.*, 2007). La réparation de ces lésions mutationnelles nécessite le recrutement d'ADN glycosylases qui initient la voie de réparation des bases par excision (BER : base excision repair). Chez *E. coli* les ADN glycosylases, MutY et MutM (les homologues eucaryotes sont MYH et OGG1, respectivement), sont capables d'éliminer respectivement la 8-oxoG et l'adénine des paires 8-oxoG:C et 8-oxoG:A. Des études chez l'Homme et *E. coli* ont révélé que MutS peut reconnaître la paire de bases 8-oxoG:A et augmenter l'activité glycosylase de MYH/MutY, suggérant une coopération entre le MMR et le BER (Gu *et al.*, 2002; Bai and Lu, 2007). Comme MutY et MutS interagissent avec le β -clamp, ce dernier pourrait coordonner ces processus pour la réparation des bases endommagées.

- **Le NER**

Contrairement au BER, le système NER est un mécanisme réparant par excision d'un segment d'ADN ; il est retrouvé dans tous les domaines du vivant (Nudler, 2023). Cette voie répare une pléthore de lésions dont les adduits d'ADN, les dimères de pyrimidine résultant de l'exposition aux UV et les pontages inter-brins. UvrA, UvrB et UvrC sont les trois protéines bactériennes clés du système NER. UvrA reconnaît un dommage à l'ADN et recrute UvrB qui interagit avec l'endonucléase UvrC. Cette dernière clive le brin endommagé, de part et d'autre de la lésion. Le segment d'ADN contenant le dommage est libéré par l'hélicase UvrD et le brin complémentaire est synthétisé par l'ADN polymérase I. Chez *E. coli*, il a été montré *in vivo* que UvrB interagit avec MutS et MutL en présence de la 5-azacytidine, un analogue de base provoquant des transversions (Pitsikas et al., 2009). L'étude révèle que le taux de mutations est augmenté dans les simples mutants *mutS*, *mutL* ou *uvrB*. Les auteurs proposent qu'en présence de la 5-azacytidine, les protéines Mut recruteraient les protéines du NER pour la réparation des

mésappariements conduisant aux transversions, qui sont peu pris en charge par le MMR canonique. (Polosina and Cupples, 2010, 2010).

3. Le rôle biologique du MMR canonique

Les mésappariements n'apparaissent pas seulement lors de la réplication ; d'autres processus peuvent en être l'origine comme la recombinaison homéologue (recombinaison entre séquences d'ADN divergentes). L'implication des protéines du MMR dans le contrôle de la recombinaison homéologue, ou l'anti-recombinaison, a été mise en évidence pour la première fois par des tests de croisements entre *Salmonella typhimurium* et *E. coli* dont les génomes présentent une divergence de séquence d'environ 20% (Rayssiguier *et al.*, 1989; Štambuk and Radman, 1998; Matic *et al.*, 1995). L'inactivation du MMR provoque une augmentation d'un facteur 1000 du taux de recombinaison homéologue entre les deux espèces (Rayssiguier *et al.*, 1989). Il a été montré plus tard que même une faible divergence (<1%) peut significativement inhiber la recombinaison autant chez les procaryotes (Zahrt *et al.*, 1994; Majewski and Cohan, 1998; Vulić *et al.*, 1997) que chez les levures (Datta *et al.*, 1996, 1997). Le mécanisme par lequel la recombinaison homéologue est rejetée par les protéines du MMR demeure obscur. Les investigations visant à reconstituer *in vitro* l'anti-recombinaison chez les eucaryotes suggèrent que MutS, en se fixant sur les mésappariements des intermédiaires de recombinaison catalysés par RecA, empêche l'extension de l'hétéroduplex. D'autres travaux *in vitro* suggèrent que c'est le complexe MutS-MutL qui inhibe la migration de branches d'ADN (Tham *et al.*, 2013; Honda *et al.*, 2014).

En raison de son activité anti-recombinaison, le MMR apparait comme un mécanisme de contrôle du transfert horizontal de gènes et par conséquent, il ralentirait l'évolution des génomes.

Des expériences d'évolution sur le long terme ont permis de dresser le spectre mutationnel de différentes espèces bactériennes comme *E. coli*, *Vibrio*, *Pseudomonas fluorescens* ou encore *Bacillus subtilis* (Lee *et al.*, 2012; Long *et al.*, 2014; Dillon *et al.*, 2017; Sung *et al.*, 2015; Long *et al.*, 2014). Les résultats révèlent l'existence d'un biais vers les mutations de type transitions dans les souches déficientes pour le MMR. En effet, les transitions A:T>G:C sont davantage accumulées comparé aux autres mutations. Ce résultat met en lumière un biais favorisant l'augmentation du contenu en GC dans les génomes en absence du MMR. Par conséquent, le MMR permettrait non seulement le maintien d'un taux de mutation bas, mais serait également impliqué dans l'évolution des chromosomes en évitant leur dérive vers des taux en GC excessivement élevés.

La voie MMR est un mécanisme étroitement lié à diverses voies de métabolisme de l'ADN telles que la réplication ou la recombinaison. Son rôle biologique est multiple ; il ne se limite pas à la correction des erreurs issues de la réplication, il est également impliqué dans le contrôle de la recombinaison et dans

la prévention de la dérive du taux en GC. Les souches mutatrices, déficientes pour le MMR, tendent à accumuler un plus grand nombre de mutations, parmi lesquelles certaines peuvent être létales. Néanmoins, cette hypermutabilité et cette capacité hyper-recombinogène leur confèrent une opportunité accrue d'acquérir des mutations avantageuses et d'intégrer de l'ADN exogène, favorisant ainsi une meilleure adaptation à leur environnement.

4. L'existence d'une voie de MMR non canonique

Il est couramment admis que le MMR canonique est un mécanisme ubiquitaire dans les 3 domaines du vivant. Néanmoins, la recherche des gènes *mutS* et *mutL* à travers les différentes espèces révèle leur absence chez une partie des archées et la plupart de actinobactéries (Sachadyn, 2010). Ces organismes ne présentent néanmoins pas un taux de mutation particulièrement plus élevé. Cette observation a suscité l'hypothèse de l'existence d'une voie de MMR non-canonique. Les circonstances qui ont conduit à la découverte du MMR non-canonique, les acteurs de ce système alternatif ainsi que ses singularités sont détaillés dans cette partie.

4.1. La découverte de la protéine NucS

En recherchant des endonucléases résolvant les intermédiaires de recombinaison chez l'archée *Pyrococcus abyssi*, Ren et ses collègues (2009) ont identifié une nouvelle protéine appartenant à la famille des nucléases RecB-like, dénommée NucS (nuclease for single-stranded DNA) (Ren *et al.*, 2009). Cette découverte fût possible grâce à l'identification d'un motif « PIP » dans sa séquence (pour PCNA-interacting protein), la classant ainsi comme une protéine interagissant avec l'homologue eucaryote du β -clamp, appelé PCNA (Proliferating Cell Nuclear Antigen) (Meslet-Cladière *et al.*, 2007). Les analyses structurales dévoilent que NucS de *P. abyssi* forme un dimère dont la région N-terminale abrite un domaine de liaison à l'ADN et la partie C-terminale renferme un site catalytique « RecB-like nuclease domain » (Ren *et al.*, 2009). Les tests d'activité ont montré que NucS clive l'ADN simple-brin des structures d'ADN branchées (telles que « flapped DNA » ou « splayed-arm DNA ») (Ren *et al.*, 2009; Creze *et al.*, 2012). Elle est capable de couper en 5' et en 3' à la jonction ADN simple-brin/ADN double-brin, faisant d'elle une « flap endonucléase » ou SSE (pour *Structure Specific Endonuclease*) (Dehé and Gaillard, 2017). Les analyses phylogénétiques révèlent que le gène *nucS* est retrouvé non seulement chez les archées mais aussi chez les bactéries à haut taux de GC comme les actinobactéries (Castañeda-García *et al.*, 2017).

4.2. NucS, acteur clé du MMR non-canonique et générateur de DSB

La recherche d'enzymes possédant une activité endonucléasique spécifique des mésappariements chez l'archée *Pyrococcus furiosus* a conduit Ishino *et al.* à la découverte de EndoMS (Mismatch Specific Endonuclease) (Ishino *et al.*, 2016). Cette protéine, hautement conservée chez les archées et certaines bactéries, se révèle être l'homologue de NucS précédemment identifiée chez *P. abyssi*. Les outils de génétique et de biologie moléculaire étant peu développés chez *P. abyssi*, l'équipe a décidé de poursuivre les investigations sur l'homologue de EndoMS chez l'archée *Thermococcus kodakarensis*.

Dans un souci de clarté, EndoMS sera appelé « NucS » dans la suite du manuscrit.

- **Chez les archées**

La caractérisation de l'activité de NucS de *T. kodakarensis* a dévoilé qu'elle reconnaît et clive les deux brins d'ADN au niveau d'un mésappariement en produisant une extrémité 5' sortante de 5 nucléotides (Ishino *et al.*, 2016). Par conséquent, elle induit la formation d'une DSB au site du mésappariement. Les essais montrent qu'elle coupe efficacement les mésappariements de type G/T, T/T, G/G tandis que les substrats T/C et A/G sont moins bien pris en charge (**Figure 8**).

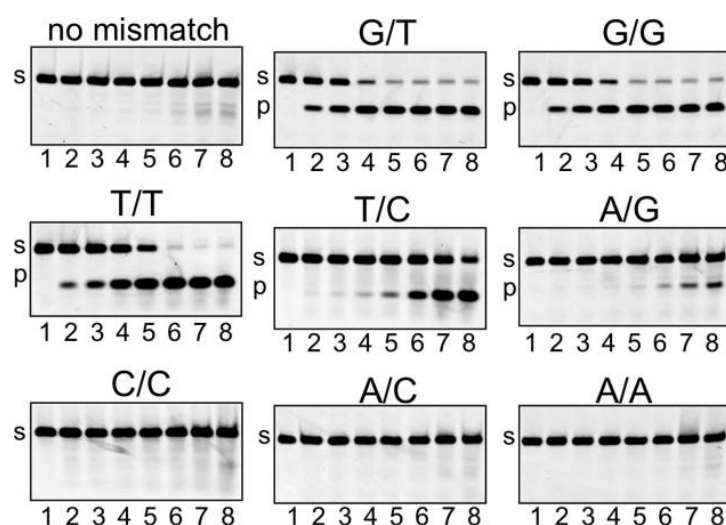


Figure 8. Activité endonucléasique de EndoMS/NucS de *T. kodakarensis* sur les mésappariements de paires de bases. Cinq nanomoles d'ADNdb marqué au Cy5 (45 pb), contenant des mésappariements d'une seule paire de bases (G/T, G/G, T/T, T/C, A/G, C/C, A/C, A/A), ont été incubées avec différentes concentrations de EndoMS/NucS de *T. kodakarensis* (voies 1, pas de protéine ; 2, 0.5 nM ; 3, 1 nM ; 4, 2,5 nM ; 5, 5 nM ; 6, 10 nM ; 7, 25 nM ; 8, 50 nM, sous forme de dimère) à 55°C pendant 5 min. Les paires de bases mésappariées sont indiquées en haut de chaque panneau. Les produits ont été séparés par PAGE à 10 %. L'affectation des bandes est indiquée sur le côté des panneaux, s, substrats ; p, produits clivés. (Ishino *et al.*, 2016)

Tout comme NucS de *P. abyssi*, l'homologue de *T. kodakarensis* clive l'ADN simple brin des structures d'ADN « flap » et « splayed » ; et interagit avec PCNA ce qui augmente significativement son activité

endonucléasique (Ishino *et al.*, 2016). Ces données laissent à penser que les homologues de NucS sont en réalité des endonucléases capables de cliver à la fois l'ADN simple-brin des structures branchées et l'ADN double-brin au site d'un mésappariement. Les tests d'activité et de spécificité de NucS des archées de l'ordre des *Sulfolobales* tels que *Sulfolobus islandicus* REY15A et *Sulfolobus acidocaldarius* ont révélé une activité endonucléasique similaire à celle de NucS de *T. kodakarensis* (Ahmad *et al.*, 2020). A la lumière de ces résultats, la protéine NucS est considérée comme un acteur clé du MMR non-canonique. Cette voie alternative apparaît singulière, car elle se distingue du système MutSL en induisant une DSB, une altération extrêmement dommageable pour la cellule. D'un point de vue évolutif, il est intrigant qu'un système de MMR induise une lésion aussi grave pour corriger des erreurs d'incorporation de bases. Au regard des résultats *in vitro* de l'activité de NucS, le mécanisme de correction des mésappariements casserait le chromosome pour le réparer. A ce jour, la voie prenant en charge les DSB produites par NucS n'est pas connue. Les chercheurs supposent que ces DSB pourraient être réparées par la recombinaison homologue. Cette hypothèse s'appuie sur l'organisation en opéron de *nucS* avec *radA* (l'homologue eucaryote de RecA). Cependant, la proximité des deux gènes sur le chromosome n'est vraie que chez les *Thermococcales* (Ishino *et al.*, 2016). On peut aussi imaginer que l'interaction entre NucS et PCNA peut favoriser le recrutement des systèmes de recombinaison pour corriger le mésappariement.

- **Chez les actinobactéries**

Étant donné l'absence de gènes *mutSL* au sein du phylum des actinobactéries, les chercheurs se sont intéressés à la protéine NucS dans ces organismes. Leurs travaux révèlent des résultats similaires à ceux trouvés chez les archées. A ce jour, NucS a été décrite chez *Corynebacterium glutamicum* et différentes espèces de *Mycobacterium* (Ishino *et al.*, 2016; Takemoto *et al.*, 2018; Castañeda-García *et al.*, 2017, 2020; Fressatti Cardoso *et al.*, 2022). Castañeda et ses collègues ont été les premiers à rapporter le rôle de NucS dans le MMR non-canonique *in vivo* chez *Mycobacterium smegmatis* (Castañeda-García *et al.*, 2017). Cette même équipe a réalisé une analyse taxonomique de la distribution de NucS en déterminant, dans un total de 3 942 protéomes de référence lesquels présentaient ou non NucS. Les bactéries étaient représentées par 2 709 protéomes (68 %), les archées par 132 (4 %), les 1 101 (28%) restants appartenant aux eucaryotes et aux virus. NucS s'est avéré présente dans 370 organismes : 60 espèces d'archées et 310 espèces de bactéries, parmi lesquelles 300 actinobactéries (sur les 302 analysées) et absente des eucaryotes et des virus. La représentation phylogénétique de cette distribution est illustrée en **Figure 9**.

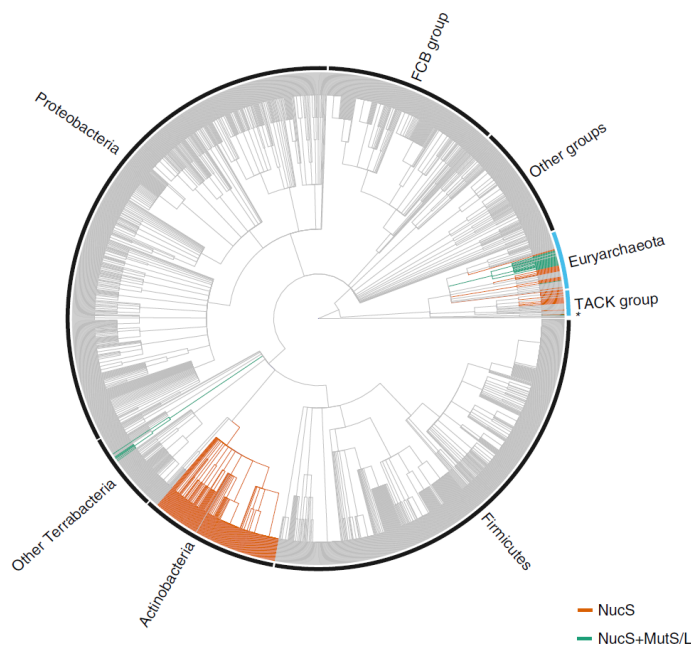


Figure 9. Arbre phylogénétique de NucS. Arbre construit à partir de 2186 espèces de bactéries (contour noir) et d'archées (contour bleu). Les organismes présentant uniquement NucS sont représentés en orange alors que ceux pourvus de NucS et MutS-MutL sont affichés en vert. Bactéries : Actinobactéries, Firmicutes, Proteobactéries, FCB (Fibrobactères, Chlorobi and Bacteroidetes), Terrabactéries (Armatimonadetes, Chloroflexi, Cyanobactéries, Deinococcus-Thermus, Tenericutes et Terrabactéries non classées) et autres groupes (Acidobactéries, Aquificae, Caldiseptica, Chrysiogenetes, Deferribactères, Dictyoglomi, Elusimicrobes, Fusobactéries, Nitrospirae, PVC groupe, Spirochaetes, Synergistetes, Thermodesulfobactéria, Thermotogae and unclassified bacteria). Archées : Euryarchaeota, TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota) et des espèces d'archées non classées (*). Comme NucS est absent chez les eucaryotes et les virus, ils ont été supprimés de l'arbre. (Castañeda-García *et al.*, 2017)

De plus, ces auteurs ont démontré le phénotype hypermutateur conféré par l'absence de NucS. Les souches $\Delta nucS$ de *M. smegmatis* et *S. coelicolor* A3(2) affichent un taux de mutation 150 fois et 108 fois plus élevé que la souche sauvage, respectivement. Ils ont aussi montré l'activité anti-recombinaison de NucS de *M. smegmatis* avec une augmentation d'un log du taux de recombinaison dans la souche $\Delta nucS$ comparativement à la souche sauvage (Castañeda-García *et al.*, 2017). Bien que l'équipe n'ait pas réussi à caractériser biochimiquement l'activité endonucléasique de NucS, l'ensemble de ces données suffisent pour projeter NucS comme acteur clé du MMR non-canonique chez les mycobactéries.

Plus tard, l'équipe a déterminé le profil mutationnel associé à la perte de NucS de *M. smegmatis* par des expériences d'accumulation de mutations (Castañeda-García *et al.*, 2020). Après avoir laissé évoluer 11 lignées sauvages et 11 lignées $\Delta nucS$ pendant une dizaine de milliers de générations, le spectre de mutations dévoile que la délétion de *nucS* augmente globalement le taux de mutation (par nucléotide et par génération) de 31 fois chez *M. smegmatis*. Cette augmentation concerne principalement les transitions (x65) ; les taux de transversions et d'indels n'ont été que légèrement accrus (**Figure 10**). Cette observation est confortée par l'analyse du profil mutationnel des souches délétées de *nucS* chez

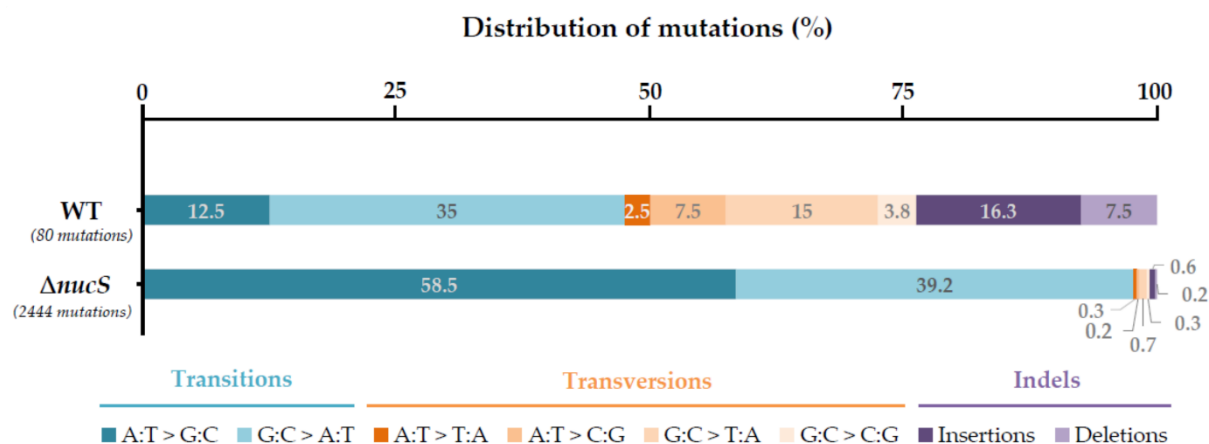


Figure 10. Distribution des mutations de l'expérience d'accumulation de mutations sur *M. smegmatis* sauvage et $\Delta nucS$. Les pourcentages sont calculés en tenant compte du nombre total de mutations dans la souche sauvage (80 mutations) et dans la souche *nucS*-déficiente (2444 mutations) (données de Castañeda-García *et al.*, 2017 ; figure tirée de Cebrián-Sastre *et al.*, 2021).

C. glutamicum (Takemoto *et al.*, 2018). Ces dernières présentent aussi un fort taux de transitions (plus de 50 fois supérieur à celui de la souche sauvage) et les effets de la mutation *nucS* sur les transversions et indels sont minimes. En revanche, les mutants du gène *dnaE* accumulent aussi bien transitions et transversions (x50). En somme, le MMR non-canonique impliquant NucS apparaît plus enclin à prévenir les transitions que les transversions ou les indels.

En 2018, deux équipes ont décrits pour la première fois l'activité enzymatique *in vitro* de NucS chez l'actinobactérie *C. glutamicum* (Ishino *et al.*, 2018; Takemoto *et al.*, 2018). Comme son homologue archée, NucS de *C. glutamicum* possède une activité endonucléase conférée par le domaine « RecB-like nuclease » situé en C-terminale. Elle est retrouvée en dimère en solution et dispose d'un domaine de liaison à l'ADN et de dimérisation dans la région N-terminale. Elle est capable d'interagir avec le β -clamp *via* le motif de liaison au clamp (CBM pour clamp-binding motif) retrouvé à l'extrémité C-terminale de sa séquence protéique. Son activité endonucléasique, stimulée par le clamp, est spécifique des mésappariements G/T, T/T et G/G seulement alors que NucS de *T. kodakarensis* est capable de cibler en plus les substrats G/A et T/C. NucS de *C. glutamicum* coupe les deux brins d'ADN en 5' du mésappariement. Le site de clivage reste encore à confirmer car les deux équipes rapportent des résultats divergents sur le sujet. En effet, les travaux de Ishino *et al.* (2018) révèlent une cassure dont les extrémités sont cohésives 5' sortantes de 5 nucléotides. Cela suggère que NucS de *C. glutamicum* clive 2 nucléotides en 5' de part et d'autre du mésappariement (Ishino *et al.*, 2018). Quant à Takemoto et ses collaborateurs (2018), ils mettent en évidence des extrémités cohésives 5' sortantes de 3 nucléotides ; la protéine couperait, dans ce cas, 1 nucléotide en 5' de part et d'autre du mésappariement (Takemoto *et al.*, 2018). Notons que l'interaction entre le β -clamp et NucS de *C. glutamicum* augmente significativement son activité enzymatique. Takemoto *et al.* (2018) affirment que NucS de *C. glutamicum*

est impliquée dans la correction des erreurs issues de la réplication en se basant sur l'effet synergique observé dans le double mutant $\Delta nucS dnaE$. Dans ce dernier, en plus de la perte de *nucS*, l'activité *proofreading* de l'ADN polymérase DnaE est altérée. Si NucS de *C. glutamicum* n'est pas en lien avec la réplication, le taux de mutation du double mutant devrait correspondre à l'addition des taux de mutation des souches $\Delta nucS$ d'une part et *dnaE* d'autre part. Ceci n'étant pas le cas, les auteurs concluent que NucS de *C. glutamicum*, acteur du MMR non-canonique, est associé à la réplication (Takemoto *et al.*, 2018).

Dans une revue portant sur l'ensemble des données concernant NucS, et notamment sur son rôle dans la stabilité et l'évolution des génomes d'actinobactéries, Cebrián-Sastre *et al.* (2021) ont illustré son intervention sur les mésappariements post-réplicatifs dans la figure ci-dessous (**Figure 11**).

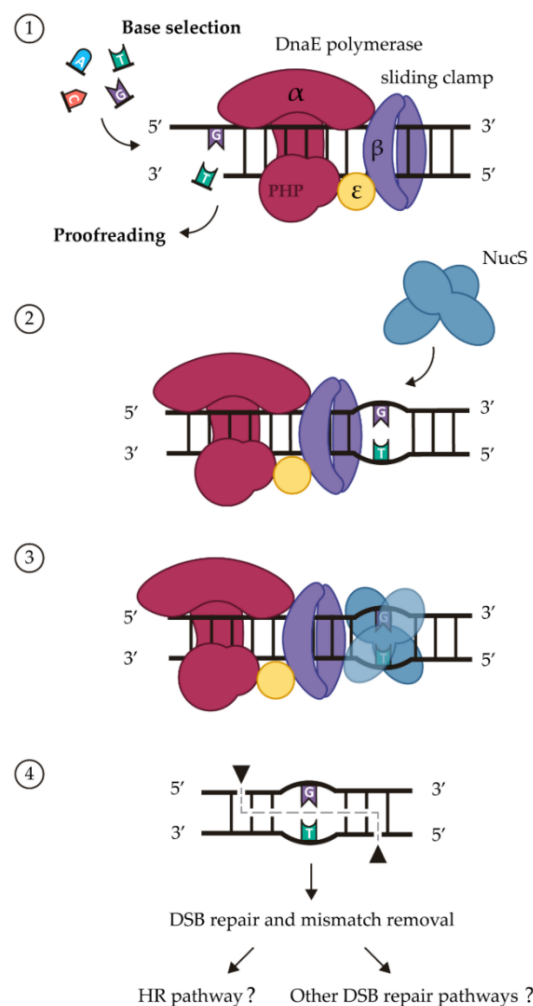


Figure 11. Modèle d'action de la voie MMR non-canonique chez les actinobactéries. Core Polymérase DnaE (sous-unité α , rouge ; sous-unité ϵ , jaune), pince coulissante β -clamp (sous-unités β , violet) et dimère NucS (bleu). **(1)** Au cours de la réplication, la polymérase DnaE incorpore les bases et, par l'intermédiaire de son domaine PHP, effectue une activité de relecture (3'-5' exonucléase). Chez les mycobactéries, la sous-unité ϵ n'a pas d'activité de relecture. **(2)** Les mésappariements qui échappent à ce processus de correction sont les substrats de NucS. **(3)** NucS se lie à l'ADNdb contenant un mésappariement et son activité est stimulée par l'interaction avec β -clamp. **(4)** NucS coupe les deux brins autour du mésappariement, ce qui crée une DSB. Enfin, la DSB et le mésappariement peuvent être réparés par la voie HR ou par d'autres mécanismes de réparation des DSB. (Cebrián-Sastre *et al.*, 2021)

4.3. L'intervention de NucS dans d'autres systèmes de réparation

Outre son activité spécifique des mésappariements post-réplicatifs, il a été montré que NucS de *T. kodakarensis* clive les substrats d'ADN contenant des bases désaminées comme l'hypoxanthine (Ishino *et al.*, 2016), prises en charge par le BER chez *E. coli* (Krokan and Bjørås, 2013). Ce résultat a été corroboré par des données plus approfondies sur l'activité de NucS de *Thermococcus gammatolerans* (Zhang *et al.*, 2020a). Les tests d'activité ont mis en lumière la capacité de NucS de *T. gammatolerans* à cliver l'ADN double-brin contenant de l'hypoxanthine (H) et de l'uracile (U), notamment les substrats H/T, H/G, U/C et U/G. Ainsi comme pour l'interaction entre le MMR canonique et le BER, NucS semble intervenir dans le processus de réparation des bases endommagées.

Récemment, une équipe a investigué le rôle *in vivo* de NucS de l'archée *Sulfolobus acidocaldarius* dans la réparation des dommages à l'ADN (Suzuki and Kurosawa, 2019). Le mutant déficient pour *nucS* montre une sensibilité accrue à divers agents génotoxiques comme la cisplatine et le métronidazole. Ces drogues sont connues pour induire des lésions de type pontage inter-brins et formation d'adduits (Fichtinger-Schepman *et al.*, 1985; Malinge *et al.*, 1999). Ces dommages arrêtent la progression de la fourche de réplication. Une flap endonucléase est recrutée au site du dommage et procède au clivage de la fourche ce qui induit la formation d'une DSB. La réparation de celle-ci serait ensuite assurée par recombinaison homologe. La 3' flap endonucléase XPF interviendrait dans les premières étapes de ce processus (Grogan, 2015). Les analyses génétiques de Suzuki & Kurosawa (2019) suggèrent que NucS de *S. acidocaldarius* pourrait aussi intervenir *via* son activité flap endonucléase dans la réparation des fourches bloquées. Ces données apportent des éléments en faveur d'une interaction entre l'action de NucS et la réparation des fourches de réplication par recombinaison homologe.

OBJECTIFS de la THESE

Le chromosome de *Streptomyces* est hautement plastique, il est sujet à nombreux remaniements d'ampleur notamment au niveau des régions terminales du chromosome. En effet, que ce soit au niveau du genre que de la population, nous avons observé une fréquence croissante des évènements de recombinaison dans les régions terminales (Lorenzi *et al.*, 2021; Tidjani *et al.*, 2019b; Choulet *et al.*, 2006a). L'équipe StrAda a montré que la recombinaison homologue et le NHEJ co-existent et sont fonctionnels chez *S. ambofaciens* (Zhang *et al.*, 2014; Hoff *et al.*, 2017, 2016, 2018). Ces derniers intervenant dans la réparation des DSB, nous avons cherché à comprendre s'il existe chez *Streptomyces* une source de DSB endogène dont la réparation pourrait expliquer la dynamique remarquable du chromosome de la bactérie.

Comme la majorité des actinobactéries, aucun homologue des gènes *mutS* et *mutL* n'a été retrouvé dans le génome des *Streptomyces*. La récente mise en évidence de l'existence du MMR non-canonique ainsi que l'activité de clivage en double-brin d'ADN de NucS chez l'actinobactérie *C. glutamicum* (Castañeda-García *et al.*, 2017; Ishino *et al.*, 2018; Takemoto *et al.*, 2018) ont encouragé la recherche d'un homologue de la protéine chez notre bactérie modèle, *Streptomyces ambofaciens*.

L'objectif de mon travail de thèse a été d'explorer le rôle de NucS dans la plasticité du chromosome de *Streptomyces*. Pour ce faire, deux objectifs ont été fixés :

- 1) **La caractérisation et la mise en évidence de l'activité endonucléase de NucS chez *Streptomyces ambofaciens* (NucS_{Sam}).** Par des approches de biochimie (*in vitro*), de génétique fonctionnelle (création de contextes mutants) et de génomique (analyse de l'accumulation de mutations), nous déterminerons le phénotype de souches déficientes pour NucS_{Sam}, et questionnerons sa participation au fonctionnement du MMR non-canonique.
- 2) **L'établissement d'un lien entre l'activité de NucS et la plasticité du chromosome de *Streptomyces*.** Afin de corrélérer l'activité de NucS avec l'éventuelle formation de cassures double brin dans le chromosome, nous construirons un contexte 'hypermutateur', dans lequel la fréquence d'apparition mésappariements est augmentée (activité *proofreading*) ; ces mésappariements pouvant être les cibles de NucS_{Sam}, la stabilité du chromosome, lorsque NucS est présente, surexprimée, ou absente sera observée.

RESULTATS

I. Chapitre 1 : Caractérisation de NucS chez *S. ambofaciens*

Les résultats obtenus sur la caractérisation et le rôle de la protéine NucS chez *S. ambofaciens* ont été rédigés sous la forme d'un article qui à cette date est en révision dans le journal « *Nucleic Acids Research* ».

1. Synthèse du manuscrit

La découverte de la protéine NucS a été un jalon majeur dans la compréhension du MMR non-canonique. A ce jour, NucS a été caractérisée chez quelques archées et chez les genres actinobactérien *Mycobacterium* et *Corynebacterium*. Les recherches ont révélé que NucS est une endonucléase spécifique des mésappariements capable de cliver *in vitro* les deux brins d'ADN au site du mésappariement avec une préférence marquée pour les mésappariements G/T, G/G et T/T (Ishino et al., 2018 ; Takemoto et al., 2018). Son activité endonucléasique est stimulée par la présence du β -clamp, un facteur crucial pour la réplication de l'ADN. La perte de *nucS* est associée à une augmentation significative du taux de transitions aussi bien chez les archées que chez *Mycobacterium* et *Corynebacterium*. Le MMR non-canonique impliquant NucS partage des caractéristiques avec le système MMR canonique dépendant de MutSL, mais mène à la formation d'une DSB au niveau des mésappariements substrats contrairement au système canonique.

Chez *Streptomyces*, peu d'informations sont disponibles sur le rôle de NucS. Castañeda et ses collègues ont publié que NucS est largement retrouvé chez les actinobactéries et qu'elle est impliquée dans l'évitement de mutation chez *S. coelicolor* (Castaneda et al., 2017). Nous avons entrepris dans cette étude de caractériser le rôle de NucS chez notre bactérie modèle *S. ambofaciens*. La recherche d'orthologues de *nucS* dans le génome d'une collection de 125 espèces représentatives de la diversité du genre *Streptomyces* (Lorenzi et al., 2021) révèle qu'une seule copie de *nucS* est retrouvée dans la partie centrale du génome de toutes les espèces interrogées. La séquence protéique est fortement conservée avec des identités allant de 82% à 97%.

Pour mieux comprendre le rôle de NucS chez les *S. ambofaciens*, quatre lignées indépendantes délétées de *nucS* ont été construites. Pour chaque lignée, une souche de complémentation correspondante a été également créée : le gène *nucS* et son promoteur natif ont été introduits en *trans* à l'aide du plasmide intégratif pSET152.

L'analyse du phénotype des mutants $\Delta nucS$ a révélé une caractéristique commune à toutes les lignées : la formation à haute fréquence de colonies à secteurs. La proportion de ces colonies est significativement plus grande dans les souches $\Delta nucS$ comparé à la souche sauvage (en moyenne respectivement

82% versus 7%). Le pourcentage de colonies à secteurs dans les mutants de complémentation est similaire à celui observé chez la souche sauvage. Le phénotype particulier des souches $\Delta nucS$ est le résultat de l'accumulation de mutations qui affectent l'aspect visuel de la colonie et suggère donc un phénotype hypermutateur du mutant.

Pour évaluer le phénotype mutateur conféré par la délétion de *nucS*, le taux de mutation des souches a été estimé par des tests de fluctuations basées sur l'émergence d'individus spontanément résistants à la rifampicine (méthode Luria et Delbrück). Étant donné que les 4 lignées $\Delta nucS$ présentent le même phénotype colonial, nous avons déterminé le taux de mutation d'une seule lignée *nucS*-déficiente ($\Delta nucS$ -2) et de la souche de complémentation correspondante. Une augmentation d'un facteur de 130 du taux de mutation a été observée dans le mutant $\Delta nucS$ -2 comparé à la souche sauvage alors que le mutant de complémentation présente un taux de mutation similaire à celui de sauvage. L'ensemble de ces résultats suggère que la protéine NucS de *S. ambifaciens* ($NucS_{sam}$) joue un rôle crucial dans l'évitement de mutation et la stabilité du génome.

Des essais biochimiques visant à déterminer l'activité de $NucS_{sam}$ ont été entrepris. Pour cela, la protéine ainsi que le facteur de processivité de la réplication β -clamp ont été purifiés à l'aide d'une étiquette histidine en N-terminale. Les tests de clivage montrent que, comme son homologue archéen ou corynebactérien, $NucS_{sam}$ coupe les deux brins d'ADN au niveau des mésappariements G/T, G/G et T/T. Son activité endonucléasique est stimulée en présence du β -clamp. Outre les mésappariements issus de la réplication, $NucS_{sam}$ prend aussi en charge les bases désaminées telles que G/U et T/H. L'ensemble de ces données est en accord avec la littérature et suggère l'intervention de $NucS_{sam}$ dans le MMR non-canonique et aussi dans le BER (Zhang et al., 2020).

Afin d'appréhender *in vivo* l'implication de NucS dans la maintenance du génome, nous avons conduit une expérience d'accumulation de mutations (MA) sur 60 cycles de sporulation. Un total de 12 lignées de la souche sauvage et 12 lignées des mutants $\Delta nucS$ (3 lignées issues de chacun des 4 mutants *nucS*-déficients) a été généré. Chaque semaine, un sous-clonage des 24 lignées a été effectué : une colonie sporulante bien isolée est striée pour la réalisation de l'isolement bactérien suivant et une semaine d'incubation correspondant à un cycle de sporulation. Au bout de 60 cycles, le génome des lignées évoluées et des souches « parentales » a été séquencé. L'analyse révèle que la pression de sélection au cours de l'expérience MA était négligeable. Si les résultats montrent que les lignées sauvages présentent un taux de mutation de 2.8×10^{-8} par nucléotide et par cycle, celui des mutants $\Delta nucS$ est en moyenne 32 fois plus grand avec un taux de mutation moyen de 9.2×10^{-7} par nucléotide et par cycle. Par ailleurs, l'analyse du profil mutationnel des lignées $\Delta nucS$ a mis en lumière l'existence d'un biais vers les transitions. En effet, le taux de transitions des mutants $\Delta nucS$ (97.2% des substitutions identifiées) est 67 fois plus grand que celui des souches sauvages. Les mutations de type transition apparaissent lorsque

les mésappariements G/T ou A/C ne sont pas corrigés. Or notre étude a montré que seul le mésappariement G/T est reconnu et clivé *in vitro* par NucS_{Sam}. Cette préférence a également été mise en évidence pour d'autres bactéries et archées pourvues d'un MMR non-canonique. Ainsi, dans les mutants $\Delta nucS$, la hausse des transitions serait issue d'une réduction de la prise en charge des G/T, qui représenteraient alors la majorité des mésappariements apparaissant au cours de la réplication, et le MMR non-canonique serait spécialisé dans leur élimination.

Nous avons ensuite déterminé la distribution des mutations de type substitution (BPS pour base pair substitutions) le long du chromosome de *S. ambofaciens*. Pour ce faire, à partir de l'origine de réplication du chromosome, située approximativement à son centre, jusqu'à ses extrémités, nous avons compté le nombre de BPS dans une fenêtre non-chevauchante de 100 kb. Les résultats dans les lignées $\Delta nucS$ montrent une augmentation progressive du nombre de BPS vers les extrémités et l'existence d'une corrélation positive entre le nombre de BPS et la position génomique. Par ailleurs, le même résultat est obtenu quand on considère uniquement les transitions. Ceci signifierait que non seulement NucS est plus prompt à corriger les mésappariements donnant lieu aux transitions mais elle serait aussi plus active au niveau des subtélomères, régions où les transitions semblent davantage s'accumuler. La même analyse a été effectuée pour les lignées sauvages. Cependant aucune corrélation n'a pu être établie, probablement à cause du faible nombre de BPS dans ces lignées. Face à ce constat, nous avons estimé l'accumulation des transitions en utilisant des souches environnementales issues du sol et dont le génome a été entièrement séquencé. Ces souches partagent un ancêtre commun et sont phylogénétiquement très proches, mais ont divergé pendant un temps bien supérieur à 60 cycles de sporulation. Nous avons choisi d'étudier les souches RLB1-8 et RLB17-3 (Tidjani et al., 2019) qui partagent un ANI (average nucleotide identity) de 99.13% qui est suffisant pour dresser la distribution des BPS entre les deux souches. La comparaison du chromosome de ces deux souches indique, comme dans les lignées $\Delta nucS$, une augmentation croissante des transitions vers les régions terminales. En somme, ces résultats révèlent que même en présence du MMR non-canonique et à une courte échelle de temps, une accumulation des transitions dans les subtélomères peut être observée. Cette donnée semble importante du point de vue de l'évolution et de la compartimentation du chromosome de *Streptomyces*. En effet, l'augmentation graduelle des transitions vers les extrémités du chromosome laisse penser que NucS pourrait être davantage sollicitée dans ces régions et induirait, de par son activité, des DSB. Ainsi, les points chauds de recombinaison des régions terminales pourraient trouver source dans l'activité de la protéine NucS. Il reste néanmoins à déterminer si cette dernière induit effectivement des DSB *in vivo* et si celles-ci sont recombinogènes.

MANUSCRIPT TITLE

Correction of non-random mutational biases along a linear bacterial chromosome by the mismatch repair endonuclease NucS

AUTHORS

Oyut Dagva¹, Annabelle Thibessard¹, Jean-Noël Lorenzi², Victor Labat¹, Emilie Piotrowski¹, Nicolas Rouhier³, Hannu Myllykallio⁴, Pierre Leblond^{1*}, Claire Bertrand^{1*}

¹ Université de Lorraine, INRAE, UMR 1128 DynAMic, 54000 Nancy, France

² CNRS, Institut Jacques Monod, 75013 Paris, France

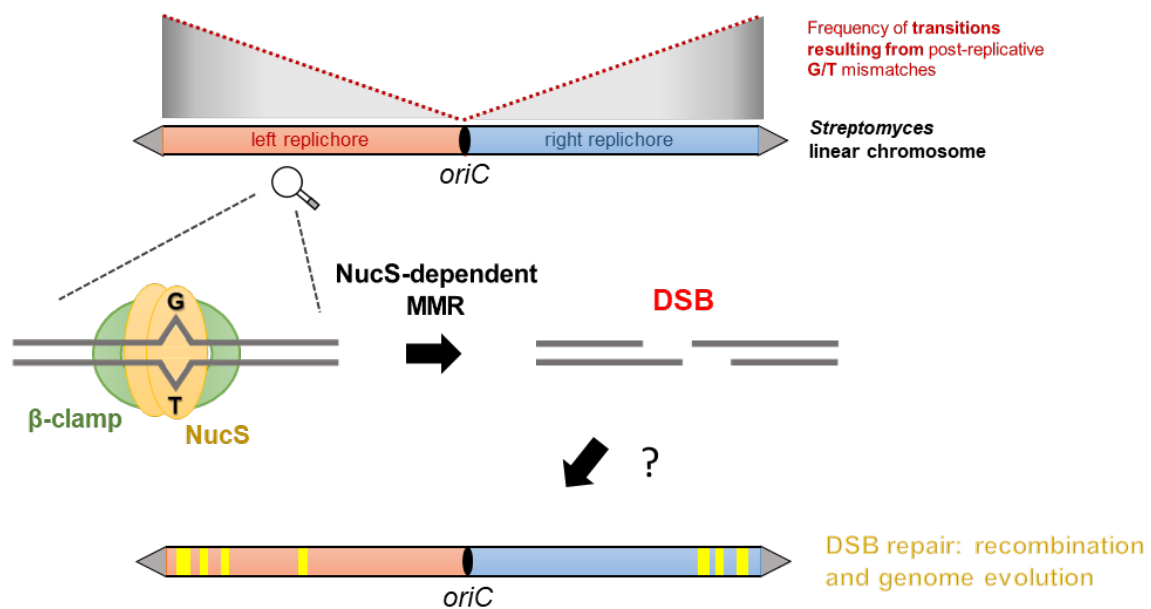
³ Université de Lorraine, INRAE, UMR 1136 IAM, 54000 Nancy, France

⁴ Ecole Polytechnique, INSERM U696-CNRS UMR 7645 LOB, 91128 Palaiseau, France

* To whom correspondence should be addressed. Tel: +33 3 72 74 51 44; Email: pierre.leblond@univ-lorraine.fr

* correspondence may also be addressed to. Tel: +33 3 72 74 51 89; Email: claire.bertrand@univ-lorraine.fr

GRAPHICAL ABSTRACT



ABSTRACT

The linear chromosome of *Streptomyces* exhibits a highly compartmentalized structure with a conserved central region flanked by variable arms. As double strand break (DSB) repair mechanisms play a crucial role in shaping the genome plasticity of *Streptomyces*, we investigated the role of EndoMS/NucS, a recently characterized endonuclease involved in a non-canonical Mismatch Repair (MMR) mechanism in archaea and actinobacteria, that singularly corrects mismatches by creating a DSB. We showed that *Streptomyces* mutants lacking NucS display a marked colonial phenotype and a drastic increase in spontaneous mutation rate. *In vitro* biochemical assays revealed that NucS cooperates with the replication clamp to efficiently cleave G/T, G/G and T/T mismatched DNA by producing DSBs. These findings are consistent with the transition-shifted mutational spectrum observed in the mutant strains and reveal that NucS-dependent MMR specific task is to eliminate G/T mismatches generated by the DNA polymerase during replication. Interestingly, our data unveil a crescent-shaped distribution of the transition frequency from the replication origin towards the chromosomal ends, shedding light on a possible link between NucS-mediated DSBs and *Streptomyces* genome evolution.

INTRODUCTION

In all organisms, the fidelity of DNA replication is crucial for the accurate transmission of genetic information. During replication, despite the accuracy of the DNA polymerase, base-pairing errors result in mismatches that are detected and corrected either by the proofreading 3'-5' exonuclease activity of the DNA polymerase or by the post-replicative Mismatch Repair (MMR) (1). The canonical MMR system has been extensively studied in *Escherichia coli* and is well characterized both biochemically and genetically (2). Methyl-directed MMR (MutSL dependent) in *E. coli* uses transient DNA hemi-methylation to identify the daughter strand and excise the mismatched nucleotide(s). The DNA repair process is subsequently completed by a template-dependent re-synthesis in order to restore the original genetic information (3). MMR deficient *E. coli* mutants exhibit a 100- to 1,000-fold increase in mutation rate resulting in a hypermutable phenotype (4). Although the canonical MMR system is widespread, bioinformatic studies surprisingly failed to identify homologs of the *mutSL* genes in the genomes of many archaeal species and of almost all members of the actinobacteria phylum (5). However, these organisms exhibit a mutation rate comparable to that of canonical MMR-bearing bacterial species (6, 7), suggesting the existence of an alternative MMR pathway.

NucS, discovered in the archaeon *Pyrococcus abyssi*, was reported as a novel endonuclease that degrades single-stranded regions on branched DNA (8). Structural analysis of *P. abyssi* NucS revealed a two-domain enzyme that adopts a dimeric conformation. The N-terminal domain is involved in dimerization and DNA binding, while the C-terminal domain carries a minimal RecB-like domain and contains the catalytic site (8). A recent study identified the orthologs of NucS, named EndoMS (mismatch-specific endonuclease) from the hyperthermophilic euryarchaea *Pyrococcus furiosus* and *Thermococcus kodakarensis*, which specifically recognizes double-stranded DNA (dsDNA) containing a mismatch (9). Strikingly, biochemical characterization revealed that EndoMS from *T. kodakarensis* introduces double-strand breaks (DSBs) by cleaving both strands of mismatched substrates. While dsDNA containing G/T, G/G, T/T, T/C and A/G mismatches is cleaved with variable efficiency, no activity on A/C, C/C, and A/A mispairs is detected *in vitro*. This endonuclease activity is considerably enhanced by the association with the sliding clamp PCNA (proliferating cell nuclear antigen). Sliding clamps are found in all organisms and are considered to be a universal platform that recruits numerous DNA-processing enzymes during replication and repair (10, 11). In *E. coli*, the replication clamp interacts with MutS and MutL, as well as with the DNA polymerase III, suggesting a coupling between DNA replication and MMR (12, 13). NucS was in fact historically discovered as an interacting partner of PCNA in *P. abyssi* (14) and this interaction modulates NucS structure and activity (8, 15). *P. abyssi* NucS and *T. kodakarensis* EndoMS exhibit a PCNA-interacting protein (PIP) box in their C-terminal domain (14). A similar box is also known as the clamp-binding motif CBM in bacteria and has a general [EK]-[LY]-[TR]-L-F consensus sequence that

tolerates greater variability than the PIP-box (16). The crystal structure elucidation of EndoMS dimer with dsDNA confirmed a preferred binding to specific mismatches and a preliminary complex model with PCNA was proposed (17). This interaction of a restriction enzyme with the sliding clamp is a remarkable specificity that has never been reported before.

In actinobacteria, the biological role and the enzymatic activity of EndoMS/NucS (referred to as NucS hereafter) have been mainly studied in *Mycobacterium* species and *Corynebacterium glutamicum*. Deletion of *nucS* in *Mycobacterium smegmatis* and in *C. glutamicum* leads to the increase of spontaneous mutation rates and causes a transition-biased mutation spectrum, highlighting its role in mutation avoidance (18–22). Cleavage assays have shown that NucS of *C. glutamicum*, like archaeal enzymes, generates a DSB by preferentially cleaving G/T mismatched DNA substrate. The activity of the enzyme is dependent on the presence of divalent cation cofactors, Mn^{2+} or Mg^{2+} , and its efficiency is greatly enhanced by the addition of β -clamp, the bacterial homolog of PCNA. However, such cleavage has not yet been demonstrated in *Mycobacterium* species. Taken together, these results suggest that NucS corrects mismatches that have escaped the proofreading activity of the polymerase and represents an important component of the non-canonical MMR system. Most importantly, unlike the canonical MMR, NucS creates a DSB that, if left unprocessed, should be dangerous to the cell. DSBs are indeed the most detrimental type of DNA damage in all living cells (23). They can be repaired by homologous recombination when a DNA template is available or by illegitimate recombination for which a homologous template is not necessary (24–26). How the DSB induced by NucS during MMR is repaired and how the correct base pair is restored has yet to be elucidated.

Streptomyces are soil bacteria belonging to the phylum of actinobacteria and are known as the most prolific producers of specialized metabolites used in medicine and other industries (27). These bacteria have original genomic features including a high GC content (~72%) and a large linear chromosome (6–12 Mb) ended by terminal inverted repeats (TIR). The genome of *Streptomyces* is highly compartmentalized with a conserved central region flanked by variable chromosomal arms (28, 29). Preliminary genomic comparison studies revealed that the loss of synteny at the terminal arms results from DNA insertions and deletions (30). Recently, a more comprehensive analysis performed on 125 *Streptomyces* genomes highlighted that the great flexibility of the chromosomal arms is due to gain and loss of genes resulting from horizontal gene transfer or genome shuffling (29). This phenomenon is directly linked to large genomic rearrangements spontaneously occurring in the arms, including large deletions (up to 2 Mb), chromosome circularization and arm replacement events (31–34). Recent work has shown that repair of DSBs in the chromosomal arms resulted in all the chromosomal rearrangements observed in the spontaneous variants (35). DSBs in *Streptomyces* can be repaired by homologous recombination but also by Non-Homologous End-Joining (NHEJ), a widespread illegitimate

recombination mechanism among eukaryotes that is not ubiquitous in bacteria (35–38). The repair of DSBs in the central part of the chromosome can be mutagenic when mediated by NHEJ, with deletions of up to 20 kb, whereas repair at the terminal regions can result in huge genomic alterations such as deletions of up to 2.1 Mb, or arm replacement (mediated by a single crossing-over between two homologous sequences present in both arms, or by break-induced replication events). Some scar analyses revealed insertion events of ectopic sequences ranging from tens to thousands of nucleotides, highlighting possibilities of potential integration of foreign DNA in a natural environment. Therefore, DSB repair has been proposed as a driving force for the huge plasticity and evolution of *Streptomyces* genome. Various exogenous sources such as UV light, ionizing radiation or mutagenic chemicals can induce DSBs, but vast amount of these are thought to arise from internal cellular processes, for instance during DNA synthesis when the replication fork collapses (39–41). The recent discovery of the non-canonical MMR unveils a new putative internal origin of DSBs that could have a role in *Streptomyces* genome plasticity. In this study, we used mutation accumulation experiment to demonstrate the role of NucS in mutation avoidance in *Streptomyces ambofaciens* and we showed the enzyme cleavage activity on diverse base pair mismatches, particularly on G/Ts, generating a DSB. Our results notably highlight a transition bias that increases from the replication origin towards the extremities of the chromosome of strains lacking NucS, suggesting that the non-canonical MMR system is more active in the terminal regions of *Streptomyces* linear chromosome.

MATERIAL AND METHODS

Strains, plasmids, and growth conditions

DH5 α *E. coli* strain was used as the cloning host for plasmid construction. Plasmid pGEM[®]-T Easy (Promega) was used as a primary cloning vector. *E. coli* BL21 (DE3) strain was used for *nucS* and *dnaN* heterologous expression. Plasmid pET15B is the expression vector used in this study, and pSBET is a helper plasmid for heterologous expression. *E. coli* ET12567 pUZ8002 was used as a donor for intergeneric conjugation to introduce DNA constructs into *S. ambofaciens*. Plasmid pUZ8002 carries the conjugative functions. *E. coli* strains were grown in Luria-Bertani (LB) medium at 37°C, except for the BW25113 pKD20 thermosensitive strain used for PCR targeting which was grown at 30°C. Plasmid pKD20 expresses the phage λ Red recombination system. The pOSV508 plasmid allowing the expression of *int* and *xis* from pSAM2 (42) is used in *Streptomyces* to excise genetic cassettes by site-specific recombination. The integrative plasmid pSET152 is used for complementation in *Streptomyces*. The *Streptomyces* mutants mentioned in this study were derived from our reference strain *S. ambofaciens* ATCC23877. *S. ambofaciens* strains were grown on Soya Flour Mannitol agar (SFM) for spore harvest. Single colony isolation for phenotypic observation and growth for DNA extraction were performed on Hickey-Tresner (HT) agar and liquid medium, respectively.

Deletion of *nucS* in *S. ambofaciens*

The *nucS*-deficient mutants were generated by PCR targeting (43). Briefly, the deletion was carried out in a recombinant bacterial artificial chromosome (BAC) containing a 35 kb region containing *nucS* gene, which was replaced by an *oriT-aac(3)IV* apramycin resistance cassette using the lambda-red system in the highly recombinogenic *E. coli* BW25113/pKD20 strain. This recombinant BAC was then conjugated into *S. ambofaciens* using, as a donor strain, the non-methylating *E. coli* ET12567 which harbors the mobilizing pUZ8002 plasmid. Double-crossover events leading to the replacement of genomic *nucS* by the cassette were monitored by screening for both apramycin resistance and kanamycin sensitivity (selection of the loss of the BAC vector). As the disruption cassette is flanked by pSAM2 *attL* and *attR* sites, its excision by site-specific recombination can be induced thanks to the introduction of pOSV508 plasmid allowing the expression of *int* and *xis* from pSAM2 (42). The excision event was verified by PCR in the exconjugants. The Δ *nucS*-1 and Δ *nucS*-2 mutants originate from two well-isolated colonies of the same conjugation experiment, whereas Δ *nucS*-3 and Δ *nucS*-4 mutants originate from a second conjugation. These two conjugations were performed with two modified BACs. The resulting four *nucS*-deficient strains were therefore considered as independent.

Complementation of $\Delta nucS$ mutants

S. ambofaciens nucS locus (SAM23877_5135) including the promoter and the coding region was amplified by PCR using primers *nucS_compl_F/R* containing an EcoRI restriction site (Supplementary Table S1). The PCR fragment was ligated into the pGEM-T Easy vector, sequenced and then released by EcoRI digestion in order to be inserted into the conjugative and integrative plasmid pSET152, resulting in the pSET152-*nucS* complementation vector. The latter was introduced into the four $\Delta nucS$ mutants by intergeneric conjugation as described above. As pSET152 plasmid is integrative at the $\phi C31$ chromosomal *attB* site, there is only one complementation copy of *nucS* gene per chromosome. The integration of the pSET152-*nucS* plasmid in the chromosome was verified by PCR.

Estimation of mutation rate by fluctuation tests

Spontaneous mutation rates were estimated using fluctuation assays (44). Isolated colonies of wild-type (WT) and $\Delta nucS-2$ strains were excised after 7 and 5 days of growth on SFM at 30°C, respectively, and vortexed vigorously in water to homogenize spores in solution. Three volumes of 100 μ L of each suspension were plated on three SFM plates supplemented with rifampicin (50 μ g/mL) and appropriate dilutions were plated on SFM in order to determine the total viable cell number. Rifampicin-resistant colonies were counted after 4 days of incubation at 30°C. The mutation rate was determined using the Ma-Sandri-Sarkar maximum likelihood method (45) and confidence intervals were calculated according to the literature (46). These calculations were performed using the FALCOR web tool available at <https://lianglab.brocku.ca/FALCOR/> (47).

Heterologous expression and purification of His-tagged NucS_{Sam} and β -clamp_{Sam} proteins

The coding sequences of *nucS* (SAM23877_5135) and of *dnaN* (SAM23877_3780), which code for NucS_{Sam} and β -clamp_{Sam}, respectively, were amplified by PCR from *S. ambofaciens* chromosome. The amplification of *nucS* and *dnaN* was carried out with the primer couples *nucS_F/R* and *dnaN_F/R*, respectively (Supplementary Table S1). The PCR fragments were ligated into the pGEM®-T Easy vector and were verified by sequencing. The digestion by NdeI and BamHI released *nucS* and *dnaN* sequences, which were further cloned into the pET15b expression vector. Using these constructs, a His-tag is added to the N-terminal end of NucS_{Sam} and β -clamp_{Sam}. The *E. coli* BL21 (DE3) strain, containing the helper plasmid pSBET (48), was used for protein expression. Expression and purification of NucS_{Sam} and β -clamp_{Sam} were performed using the same procedure. Bacteria were grown at 37°C to exponential phase in LB medium supplemented with 50 μ g/mL kanamycin and 100 μ g/mL ampicillin. The expression of chaperones was boosted by adding ethanol to the bacterial culture to a final concentration of 0.5% and

by cooling it at 4°C for 3 h. IPTG was then added to a final concentration of 100 µM and cells were further grown for 20 h at 20°C. Cells were collected, diluted in buffer A (30 mM pH 8.0 Tris-HCl, 200 mM NaCl, 10 mM imidazole, 0.05% polyethyleneimine) and disrupted by sonication. The soluble extract was applied to an IMAC-Select affinity gel column (Sigma Aldrich, St. Louis, MO, USA) equilibrated in buffer A. After extensive washing, the protein was eluted using the same buffer A with an imidazole concentration of 100 mM. Fractions containing the protein of interest were pooled, concentrated using an Amicon centrifugal filter (Millipore) with a 10 kDa cut-off and loaded onto a Superdex 200 pg HiLoad 16/600 (GE Healthcare, 1 mL/min). Sample purity was assessed by 12% SDS-PAGE analysis and sample concentration was determined by spectrophotometry using a theoretical molar extinction coefficient at 280 nm of 20,190 M⁻¹.cm⁻¹ and 19,940 M⁻¹.cm⁻¹ for NucS_{Sam} and β-clamp_{Sam}, respectively.

Cleavage assay

Heteroduplex DNA was prepared by annealing oligonucleotides at 95°C for 10 min. The sequence and the different combinations of oligonucleotides are shown in Supplementary Table S2. The cleavage reaction of the dsDNA substrate was generally performed at 30°C for 60 min in a buffer containing 2.4 µM NucS_{Sam} and/or 1.2 µM β-clamp_{Sam}, 25 mM Bis-Tris pH 6.4, 2.5 mM MnCl₂, 0.1% Triton X-100. For cleavage reaction performed with Mg²⁺, either 2.5 mM MnCl₂ was replaced by 2.5 mM MgCl₂ in the above conditions, or the reaction conditions described by Takemoto *et al.* (21) were used as follows, 20 mM Tris-HCl, pH 8.0, 2.5 or 10 mM MgCl₂, 4 mM dithiothreitol, 0.071 mg/ml BSA, and 30 or 100 mM K[OAc]. Monomeric NucS_{Sam} fraction was used in these assays. The products were analyzed by 10% native PAGE. At least three biological replicates were done for each mismatch configuration.

Mutation accumulation experiments

The WT strain and the four independent *nucS*-deficient mutants were used for mutation accumulation (MA) experiments. Three lineages per $\Delta nucS$ mutants and twelve WT lines were generated. Each line was generated by streaking a single colony. A sporulation cycle was initiated by streaking a single well-isolated colony from each lineage onto a new SFM plate and incubating at 30°C for 7 days. This procedure was repeated every week for 60 cycles. To keep track of evolved lineages, colonies streaked at each new cycle were picked from the agar plate, resuspended in 20% glycerol and stored at -20°C.

Genomic DNA preparation and whole genome sequencing analysis

Parental and evolved strains at the end of the 60th sporulation cycle were grown in liquid HT medium to extract genomic DNA (49). Whole genome sequencing was performed using an Illumina Genome Analyzer to obtain 300-bp paired-end reads (Mutualized Platform for Microbiology (P2M), Pasteur Institute, Paris). Data analysis was performed using Geneious Prime 2023.2.1. Reads from each lineage

were trimmed based on their quality (Minimum quality = 20; primers, indexes and reads less than 20 bp in length were removed). Trimmed reads from parental strains only were mapped to the *S. ambofaciens* ATCC23877 reference genome (GenBank: CP012382.1) using Geneious mapper with the following parameters: sensitivity = "medium-low", fine tuning = "iterate up to 5 times", minimum mapping quality = "30", only map paired reads which = "map closely nearby". The mean depth coverage of each assembly ranged from 113 to 481x. Consensus sequence was generated from alignments using the highest quality parameter as a threshold and "N" was assigned to sites with coverage < 20. Annotation was performed with Prokka (50) using a custom database containing all the genes from *S. ambofaciens* ATCC 23877 chromosome. These annotated genomes were then used as references for mapping trimmed reads from evolved lines. The same Geneious alignment settings as above were applied. The mean depth coverage of each mapping ranged from 45 to 565x. Variant calling was performed using the in-built "Find Variations/SNPs" feature. A single-nucleotide polymorphism (SNP) was called if the position was covered by at least 20 reads, was found at a frequency of at least 0.9, and was not found in a niche (>3 SNPs in a sliding 100-bp window with a 1-bp step size). For a duplicated sequence, the frequency of at least 0.9 would not be observed if only one of the two copies carries the SNP, and if the two copies carry it, the calling could not be assigned to any of them. For this reason, all the genomic analyses were done on the whole genome of *S. ambofaciens* ATCC23877 with the exception of repeated sequences, including the 200 kb long TIR of *S. ambofaciens*. Note that 89 kb circular plasmid pSAM1 was also excluded from the analyses. Therefore, in the manuscript, when we refer to the chromosome, it is without the above sequences.

Genomic comparison of *Streptomyces* close environmental strains

Genomes of RLB1-8 and RLB3-17 (accession numbers CP041650 and CP041610, respectively) environmental strains (51) were used in this study. ANIb (Average Nucleotide Identity using BLAST) was calculated between both genomes: we averaged the two reciprocally obtained ANIb scores (52). For transition analysis in whole genome, both sequences were aligned using the progressive Mauve algorithm (53). The SNP call table was exported with MAUVE tool 'export SNPs'. For transition analyses in coding sequences, both genome sequences were annotated using RAST (54), with the RAST pipeline (<https://rast.nmpdr.org/>) using RASTtk (v1.3.0) with all parameters by default. After extraction, RAST annotations (with GenBank files) were formatted to a usable format for the subsequent pipeline. BLASTp was performed between each gene coding for a protein of one genome against that of the other genome and reciprocally. Orthologs were defined by identifying BLASTp reciprocal best hits (BBH) with at least 40% identity, 70% coverage, and an E-value lower than 1e-10. For each orthologous gene pair, the nucleotide sequences of both genes were retrieved and aligned with MAFFT (55). In these alignments (one per orthologous gene pair), identified non-gap differences corresponded to SNPs. RLB1-8 genome

was used as reference. The sequence of the 5 identified integrative and conjugative elements present in RLB1-8 genome (56) were removed from the transition analysis to avoid bias caused by their possible redundant nucleotide sequences. The 357 kb TIR of RLB1-8 genome was also excluded, as in the MA study.

Statistical analyses of mutation distribution along the chromosome

For the analyses of mutation distribution along the chromosome, the number of mutations was counted within a sliding non-overlapping 100 kb window, starting from the replication origin *oriC*, extending towards the left or the right end of the chromosome. The number in windows of the right replichore was added to the number in windows at the same distance of *oriC* in the left replichore. The sum was represented as a function of the distance from *oriC*. As the left replichore is slightly longer than the right one in both *S. ambofaciens* ATCC23877 and RLB1-8 genomes, values corresponding to the last 100 kb windows at the extremity of the left replichore were excluded from the analyses. Correlation analyses between the number of mutations and the distance from the origin of replication in the genomes of evolved lines or between the genomes of environmental strains were done with R software (<https://www.R-project.org/>). Pearson correlation test was used when the distribution of the residues was compatible with the normal law and Kendall'T coefficient was calculated when the distribution of the residues was not compatible with the normal law.

Reagents

All cloning reagents, including restriction enzymes (FastDigest), T4 DNA ligase, and DNA polymerases (DreamTaq, Phusion) were provided by Thermo Fisher Scientific. All molecular biology kits (PCR purification, gel extraction and plasmid miniprep) were also provided by Thermo Fisher Scientific.

RESULTS

NucS is highly conserved in *Streptomyces*

The persistence of *nucS* across *Streptomyces* genus was determined by searching for *nucS* orthologs in the genomes of a collection of 125 species representative of the genus diversity (29). A single copy of *nucS* was present in all species and, as every core gene, is localized in the central region of the chromosome. NucS protein sequence consists of 223 amino acids and is highly conserved with identities ranging from 86% to 97% when using NucS from *S. ambofaciens* (NucS_{Sam}) as a reference. The ubiquity of *nucS* through *Streptomyces* genus suggests its important role in cell development and survival in *Streptomyces*. When compared to NucS from *M. smegmatis* and *Mycobacterium tuberculosis* (18), or from *C. glutamicum* (20) the identity drops to 75%, 78%, and 72% respectively, highlighting the significant conservation of the protein among actinobacteria. NucS_{Sam} shares structural homologies with NucS from *P. abyssi* (8), bearing an N-terminal DNA binding domain connected by a linker to the C-terminal RecB-like nuclease domain. The well-known RecB motifs for nuclease activity are present in NucS_{Sam} and most importantly all the catalytic residues are conserved across actinobacteria (Supplementary Figure S1). The consensus clamp-binding KLRLF motif is also present in the C-terminal part of all *Streptomyces* NucS proteins. These data strongly suggest that NucS_{Sam} may have a similar activity to NucS_{Pab}.

NucS_{Sam} is involved in mutation avoidance

To assess the effect of *nucS* deletion, four independent mutants deleted for *nucS* ($\Delta nucS$) were generated by PCR targeting method in *S. ambofaciens* ATCC 23877 strain. For each independent $\Delta nucS$ mutant, a complemented strain, expressing *nucS* WT copy under the control of its own promoter, was obtained using the integrative plasmid pSET152. The $\Delta nucS$ mutants were grown for 8 days at 30°C on HT solid medium for phenotype analysis. Compared to the WT, all *nucS*-deleted strains shared the same remarkable phenotype consisting of a high frequency of sectorized colonies in their progeny (Figure 1A and B). This phenotype is heritable. The proportion of sector-harboring colonies was quantified in the 4 independent $\Delta nucS$ mutants, in the WT and in the complemented strains. The occurrence of sectorized colonies was significantly greater in $\Delta nucS$ strains compared to the WT strain (overall 82% versus 7% respectively; Z-test for proportions with Bonferroni correction, $P < 10^{-3}$; Figure 1C). In the complemented strains, this frequency was drastically reduced to levels comparable to those of the WT (on average 9% vs 7% respectively). This indicates that expressing *nucS* behind its original promoter restored the WT phenotype. In the WT background, it has been shown that the spontaneous variability of the colonial phenotype originates from mutations, affecting phenotypical traits in *Streptomyces* such

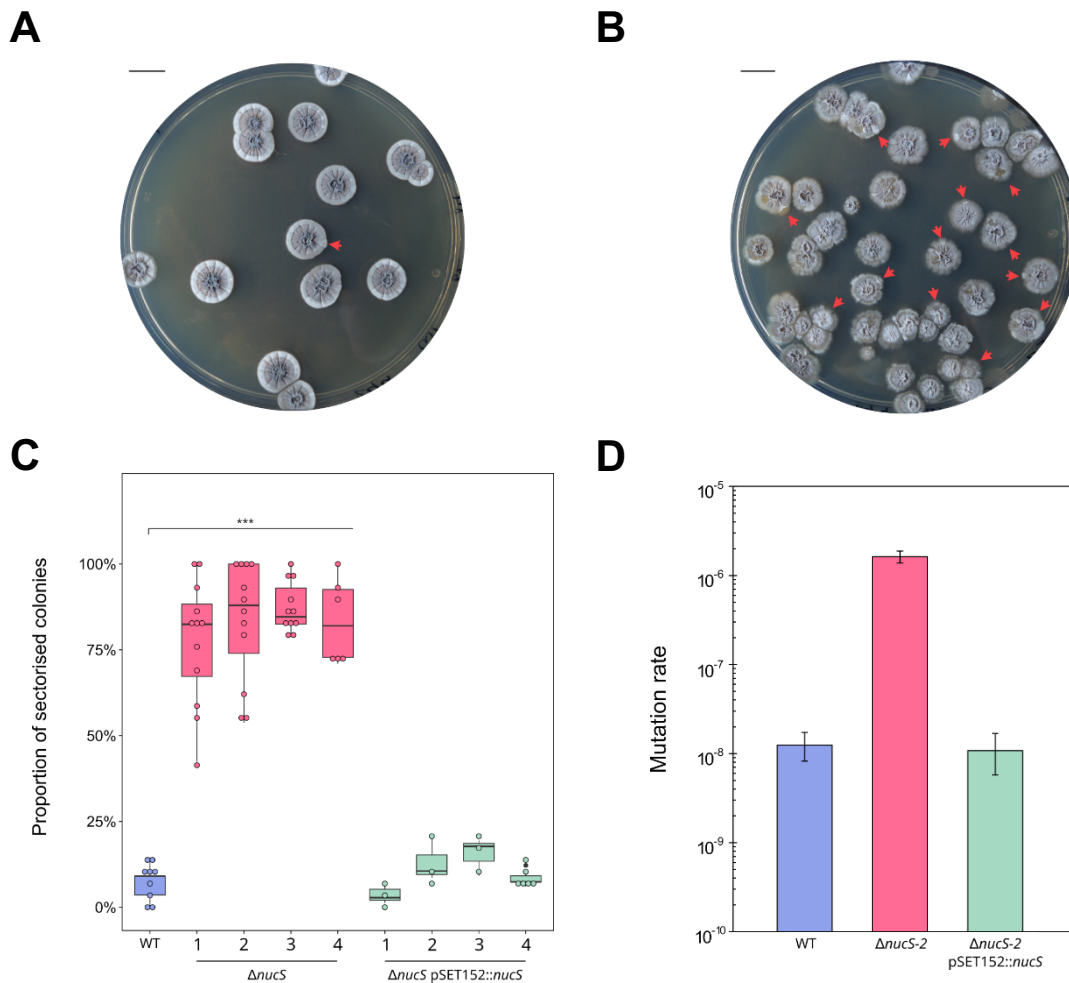


Figure 1. Colonial phenotype of *nucS*-deleted mutants. (A) *S. ambofaciens* WT strain and (B) $\Delta nucS$ mutant were grown on HT agar for 8 days at 30°C. A sectorized colonial phenotype was observed for *nucS*-deleted strain (red arrows). Scale bar represents 1 cm. (C) Proportion of sector-harboring colonies counted in the WT, the 4 independent $\Delta nucS$ mutants and in their complemented strains denoted as $\Delta nucS$ pSET152::*nucS* (z-test for proportions with Bonferroni correction, *** $P < 0.001$). (D) The rate of spontaneous mutations conferring rifampicin resistance was determined by fluctuation assay in WT, $\Delta nucS-2$ and its complemented strain $\Delta nucS-2$ pSET152::*nucS*. Error bars represent 95% confidence intervals.

as pigmentation or sporulation (31, 57). We therefore assumed that the specific phenotype observed for *nucS*-defective strains was the result of mutation accumulation.

To confirm that *nucS* deletion results in a hypermutator phenotype, the spontaneous emergence of rifampicin-resistant colonies was monitored by fluctuation assays. Originally described by Luria and Delbruck (44), fluctuation tests are widely used for estimating mutation rates. As all 4 $\Delta nucS$ strains exhibited the same colonial phenotype, $\Delta nucS-2$ was selected for further assays. The same procedure was applied to the corresponding complemented strain. Deletion of *nucS* resulted in an increase in mutation rate by a 130-factor compared to the WT strain (1.63×10^{-6} versus 1.24×10^{-8} mutations per

cell, respectively, Figure 1D). Of note, a mutation rate comparable to that of the WT strain was recovered in the complemented strain (1.24×10^{-8} versus 1.08×10^{-8} mutations per cell, respectively) confirming that the absence of NucS was responsible for the observed increase in the mutation rate. Similar findings were reported in *Streptomyces coelicolor*, in which inactivation of *nucS* resulted in a 2-log increase in mutation rate, although no colonial phenotype was reported in this study (18). Altogether, these results showed that *nucS* inactivation caused a hypermutator phenotype and highlighted the key role of NucS in mutation avoidance.

Purification of His-tagged NucS_{Sam} and His-tagged β -clamp_{Sam} proteins

NucS from the actinobacterium *C. glutamicum* has been recently demonstrated as capable of cleaving mismatched dsDNA (20, 21). Yet, Castañeda-García and colleagues (18) have shown that NucS from *M. smegmatis* has ssDNA-binding activity but no cleavage activity was detected neither on ssDNA nor on mismatched dsDNA. For this reason, the activity of NucS_{Sam} *in vitro* was investigated. To achieve this, we purified N-terminal His-tagged NucS_{Sam} by cloning the *nucS* coding sequence (SAM23877_5135) in the pET15b expression vector for heterologous expression in *E. coli* BL21. The fusion protein was purified using affinity chromatography and gel filtration. Based on the amino acid sequence, the calculated molecular weight of the protein is 24,762.44 Da. The gel filtration chromatogram showed a main peak at an elution volume corresponding to an estimated molecular weight of 41 kDa (Supplementary Figure S2). Additional peaks were observed with estimated molecular weights of 61 and 115 kDa, indicating the presence of multimeric forms. Because it was difficult to firmly conclude about the oligomerization state of the major form of purified NucS_{Sam}, the fractions corresponding to the main peak were collected and subjected to size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). The analysis revealed two main peaks with molecular weights of 30,026 Da and 61,151 Da, which would correspond to monomeric and dimeric forms of NucS_{Sam}, respectively. Because the homogeneity of these peaks was moderate, molecular weights were determined considering the right side of the peaks. The gel filtration profile and SEC-MALS analysis indicate that the protein may exist in an equilibrium between a monomeric and dimeric state, with the monomeric form being more prevalent under the purification conditions employed in this work. *C. glutamicum* NucS is also found as a monomer and as a dimer in solution, with a majority of dimeric forms (20). One hypothesis is that the assembly of two monomers into a dimer could be stimulated by the presence of a DNA substrate. However, *in vitro* studies have never shown the presence of NucS in monomeric form for archaeal enzymes, suggesting that DNA substrate is not necessary for dimerization (8, 9, 17).

The processivity-promoting factor β -clamp from *S. ambofaciens* (β -clamp_{Sam}), which is encoded by the *dnaN* gene (SAM23877_3780), was also purified following the same protocol as the one used for

NucS_{Sam}. Its calculated molecular weight is 39,882.35 Da. The gel filtration chromatogram revealed a single peak with an estimated molecular weight of approximately 97 kDa while SEC-MALS indicated a molecular weight of approximately 84 kDa. This highlighted the homodimeric nature of β-clamp_{Sam} in solution and confirmed previous results obtained for β-clamp from *C. glutamicum* and *T. kodakarensis* (9, 20).

NucS_{Sam} mismatch-specific endonuclease activity results in DSB formation *in vitro*

Cleavage assays were performed *in vitro* to test whether purified NucS_{Sam} possesses an endonuclease activity on mismatched dsDNA. The DNA substrate used in this work was a 43 bp-long dsDNA containing different combinations of mismatched base pairs at the 21st nucleotide pair. One strand of the substrate was fluorescently labeled at the 5' end with 6FAM dye. As a control, a dsDNA substrate with no mismatch was also subjected to NucS_{Sam} activity. The cleavage products generated by NucS_{Sam} were analyzed qualitatively using a 10% native PAGE. No cleavage product was observed when using the control substrate. NucS_{Sam} displayed cleavage activity on substrates containing G/T, T/T, and G/G mismatches. The cleaved product co-migrated with a control dsDNA containing a 5' protrusion of a five nucleotide-single-stranded DNA and expected to correspond to the product after a cleavage within the mismatch (an example is shown on Supplementary Figure S3) (20). No activity was detected with substrates containing T/C, A/A, A/C, G/A, or C/C mismatches (Figure 2A shows the results of one of the three biological replicates). Notably, the addition of β-clamp_{Sam} to the reaction greatly enhanced the cleavage

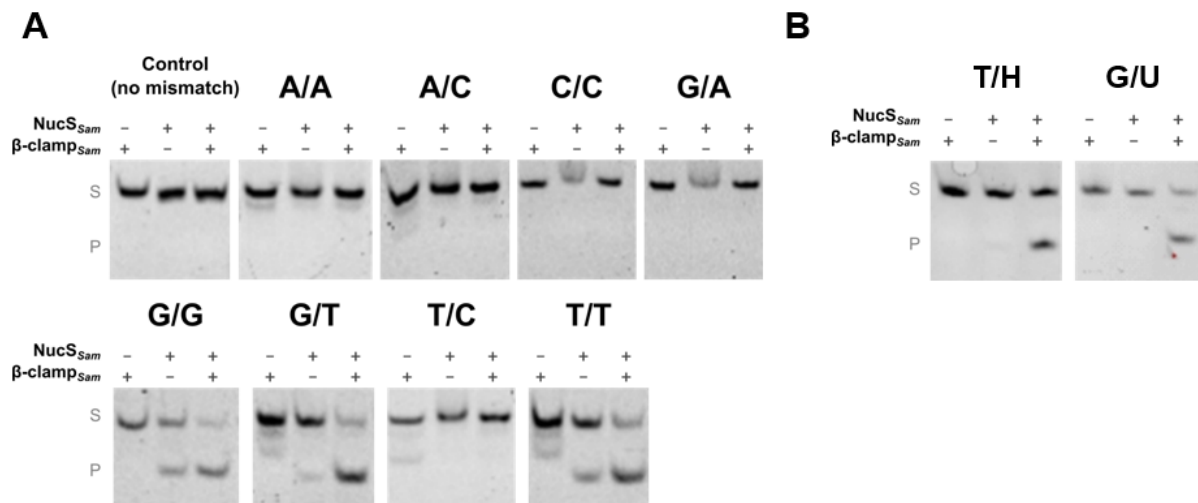


Figure 2. NucS_{Sam} mismatch specific cleavage activity. (A) *In vitro* cleavage assays were performed with 6-FAM-labeled dsDNA substrates (43-bp, 50 nM) containing no mismatch (control) or single base-pair mismatches (A/A, A/C, C/C, G/A, G/G, G/T, T/C or T/T) or (B) single deaminated bases (H instead of A or U instead of C). Each substrate was incubated for 60 min at 30°C, either with 2.4 μM of NucS_{Sam} or with 1.2 μM of β-clamp_{Sam} or with both proteins at the above concentrations. Products were separated by 10% native PAGE. Substrates and cleavage products are indicated by "S" and "P" letters on the left side of the panels. The figure is representative of three experiments.

efficiency of NucS_{Sam}, as evidenced by the presence of a more intense band corresponding to the cleaved substrate. The presence of β -clamp_{Sam} did not alter the substrate specificity of NucS_{Sam}. The presence of the metallic enzymatic cofactor Mn²⁺ was always required for NucS_{Sam} cleavage activity but no activity was observed with Mg²⁺ cofactor (data not shown).

Recently, an interesting study (58) showed that NucS from the euryarchaeon *Thermococcus gammatolerans* is able to cleave dsDNA containing deaminated bases such as uracil and hypoxanthine. Deamination process consists in the removal of an amino group from a molecule. In cells, deamination of DNA bases can occur spontaneously and produce mutagenic DNA lesions. Specifically, the removal of the amino groups from cytosine and adenine results in the formation of uracil (U) and hypoxanthine (H), respectively. If not repaired, these lesions can cause transition mutations since uracil pairs with adenine and hypoxanthine pairs with cytosine. We assessed whether NucS_{Sam} had an endonuclease activity on dsDNA containing deaminated bases. *In vitro* cleavage tests showed that NucS_{Sam} alone had no activity against G/U and T/H substrates, in contrast to the results reported by Zhang and collaborators (58) (Figure 2B). Interestingly, the activation of such mismatch-specific endonuclease activity was observed when β -clamp_{Sam} was added to the reaction. These results are original as they suggest that NucS_{Sam} participates in a process of DNA repair other than mismatch repair such as the base excision repair pathway.

Mutation rate is affected in *nucS*-deleted strains

To investigate the *in vivo* activity and specificity of NucS_{Sam}, a mutation accumulation (MA) assay was conducted. Twelve independent WT lines and three independent lines for each Δ *nucS* mutant were tracked for 60 sporulation cycles. Weekly, a single colony from each lineage was streaked onto fresh growth medium. By performing whole-genome sequencing on both parental and evolved strains, we identified the mutations that occurred throughout the MA experiment. To ensure accurate read mapping, a rigorous approach was implemented by excluding repetitive sequences from our analysis (as the 200 kb long TIR of *S. ambofaciens*). As a result, the sequence coverage for each line accounted for more than 89% of the complete *S. ambofaciens* ATCC 23877 genome. A total of 156 mutations were found in the WT lines, corresponding to an overall mutation rate of 2.84×10^{-8} mutation per nucleotide per sporulation cycle (Table 1). The number of generations occurring during a sporulation cycle of 7 days in *Streptomyces* was estimated to be approximately 24. Thus, the WT mutation rate would be around 1.18×10^{-9} per nucleotide per generation, which is comparable to that observed in *M. smegmatis* or *C. glutamicum* (19, 21). The ratio of non-synonymous to synonymous mutations (dN/dS) determined by the MA assay in WT lines was 2.96 (71/24), which was not significantly different from that expected (2.54),

Table 1. Total number of mutations and mutation rates determined by MA assays.

| Strain | Number of lineages | Average sequence coverage† (%) | Total number of mutations | Mutation rate per nucleotide per cycle ($\times 10^{-8}$) |
|---------------------------------|--------------------|--------------------------------|---------------------------|---|
| WT | 12 | 90.23 \pm 2.49 | 156 | 2.84 |
| $\Delta nucS$ | 12 | 91.34 \pm 1.78 | 5067 | 92.23 |
| $\Delta nucS-1$ | 3 | 91.34 \pm 0.28 | 1372 | 100.64 |
| $\Delta nucS-2$ | 3 | 92.20 \pm 0.51 | 1226 | 88.29 |
| $\Delta nucS-3$ | 3 | 92.70 \pm 0.41 | 1319 | 94.52 |
| $\Delta nucS-4$ | 3 | 89.14 \pm 2.30 | 1150 | 85.39 |

† The sequence coverage corresponds to the percentage of the reference genome of *S. ambofaciens* (ATCC 23877: 8,303,940 bp) analyzed for mutation call between evolved lines and their parental strains

based on the codon usage in *S. ambofaciens* ATCC 23877 strain (chi-squared test $\chi^2 = 0.4182$, $P = 0.51$), suggesting that the selective pressure in our MA protocol was minimal.

Overall, the 12 $\Delta nucS$ lines exhibited a total of 5,067 mutations (Table 1). The 3 evolved lines derived from each $\Delta nucS$ independent mutant exhibited an accumulation of mutations ranging from 1,150 to 1,372 (Table 1). The deletion of *nucS* resulted in a significant 32-fold average increase in the mutation rate per nucleotide per cycle, with rates of 2.8×10^{-8} for the WT and an average of 9.2×10^{-7} for the four $\Delta nucS$ mutants. The ratio of dN/dS mutations in $\Delta nucS$ lines was 3.13 (3280/1048), which was significantly different from that expected of 2.54 (chi-squared test $P < 10^{-5}$) but not significantly different from the dN/dS obtained in WT lines (chi-squared test $P = 0.11$). Thus, the results obtained with the mutants appear to be close to, but slightly above, the expected result for an unbiased distribution of non-synonymous and synonymous mutations.

Mutational spectrum is shifted upon NucS_{Sam} loss

The mutational profiles and rates of base pair substitutions (BPSs) in *S. ambofaciens* WT and $\Delta nucS$ lines are shown in Table 2. In the WT MA lines, a total of 147 BPSs was identified, corresponding to a rate of 2.67×10^{-8} BPS per nucleotide per cycle. This rate is consistent with the previously reported BPS rate in *S. coelicolor*, which was about 1.50×10^{-8} BPS per nucleotide per cycle (59). Furthermore, in the 12 $\Delta nucS$ lines, a total of 5,048 BPSs was detected, resulting in a BPS mutation rate of 9.19×10^{-7} per nucleotide per cycle. This substantial increase represents a remarkable 34-fold difference compared to the WT MA lines. Additionally, our analysis unveiled the presence of 9 indels in the WT lines, ranging from 1 to 4 bp in length, while we detected 19 indels (of 1 to 2 bp) in the absence of NucS_{Sam}. Therefore, the loss of NucS_{Sam} did not considerably affect the rate of indels (5×10^{-10} vs 1.6×10^{-9} for WT and $\Delta nucS$, respectively).

Table 2. Spectrum of Base Pair Substitutions (BPSs) and indels

| | WT | | | $\Delta nucS$ | | |
|---------------|--------|----------------|---|---------------|----------------|---|
| | Number | Percentage (%) | Mutation rate per nucleotide per cycle ($\times 10^{-8}$) | Number | Percentage (%) | Mutation rate per nucleotide per cycle ($\times 10^{-8}$) |
| BPSs | 147 | 100 | 2.67 | 5048 | 100 | 91.88 |
| Transitions | 73 | 49.66 | 1.33 | 4908 | 97.23 | 89.33 |
| A:T>G:C | 32 | 21.77 | 0.58 | 2986 | 59.15 | 54.35 |
| G:C>A:T | 41 | 27.89 | 0.75 | 1922 | 38.07 | 34.98 |
| Transversions | 74 | 50.34 | 1.35 | 140 | 2.77 | 2.55 |
| A:T>T:A | 5 | 3.55 | 0.09 | 12 | 0.24 | 0.22 |
| A:T>C:G | 6 | 3.55 | 0.11 | 20 | 0.40 | 0.36 |
| G:C>T:A | 18 | 11.35 | 0.33 | 34 | 0.67 | 0.62 |
| G:C>C:G | 45 | 31.21 | 0.82 | 74 | 1.47 | 1.35 |
| Indels | 9 | 100 | 0.16 | 19 | 100 | 0.35 |
| Insertions | 3 | 33.33 | 0.05 | 9 | 47.37 | 0.16 |
| Deletions | 6 | 66.67 | 0.11 | 10 | 52.63 | 0.18 |

Regarding the BPSs in the WT, the transition/transversion ratio was 1.01 (49.7% vs. 50.3% respectively). The most prevalent mutation type was G:C>C:G, accounting for 31.2% of total BPSs. Of note, G:C>A:T and A:T>G:C transitions were the second and third most frequent mutation types (27.9% and 21.8% of the BPSs, respectively). The inactivation of *NucS_{Sam}* caused a significant bias in the spectrum towards transitions, which accounted for 97.2% of the substitutions whereas only 2.8% of the BPSs were transversions. Consequently, this resulted in a high transition/transversion ratio of 36.1, which was 36-fold higher than that of the WT lines. In addition, $\Delta nucS$ lines exhibited a remarkable 67-fold increase in the transition mutation rate, while the transversion mutation rate underwent a modest 2-fold increase. Overall, the rates of mutation for all substitutions were higher in *nucS*-deleted lines compared to WT (Figure 3), but it is noteworthy that the mutation rate of the A:T>G:C and G:C>A:T substitutions increased strikingly in the mutant with 94-fold and 47-fold increases, respectively. Altogether, the deletion of *nucS* resulted in a significant shift in the BPS bias, particularly towards A:T>G:C transitions, which accounted to 59.2% of all observed BPSs.

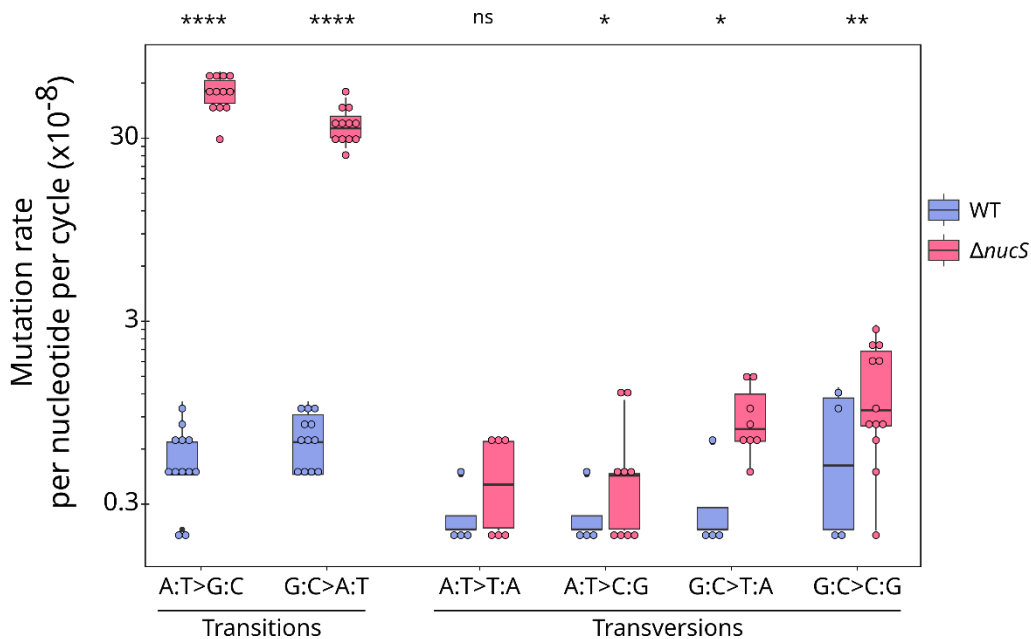


Figure 3. Transition and transversion rates in WT and $\Delta nucS$ lines. The top of each boxplot indicates 75% percentile, the bottom 25% percentile, and the thick bar inside the box is the median. The minimum and maximum values are shown by the whiskers. Individual data values for each line are represented by circles. Statistical significance levels are denoted as follows: * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$ and ns stands for non-significant (Wilcoxon rank sum test).

Transition frequency increases towards the extremities of the chromosome

The *S. ambifaciens* chromosome replicates from a centrally located origin of replication towards its ends, defining two replichores of equivalent size. To assess the distribution of transitions along the *S. ambifaciens* chromosome in the twelve evolved lines, the number of transitions was counted within a sliding non-overlapping 100 kb window, starting from *dnaA* gene (at position 4,021,374, approximately corresponding to the replication origin *oriC*), extending towards the left or the right end of the chromosome and plotted as a function of the distance from *dnaA* (Figure 4A). A significant positive correlation was observed between the number of transitions and the genomic position for $\Delta nucS$ lines (Pearson's correlation coefficient $r=0.457$, $P=0.004$). This suggests that in the absence of NucS, transitions have a crescent-shaped distribution with a minimum at the replication origin of the chromosome. There was no significant correlation for the transition distribution in the chromosomes of the twelve WT lines, probably because the sample size is too small (73 transitions in total). Since 97.3% of all the BPSs are transitions in $\Delta nucS$ lines, an increase in the BPS number towards the extremities was also observed (Supplementary Figure S4A). When only considering the transitions in coding regions (Figure 4B), the same increasing trend is again noticed (Pearson's correlation coefficient $r=0.468$,

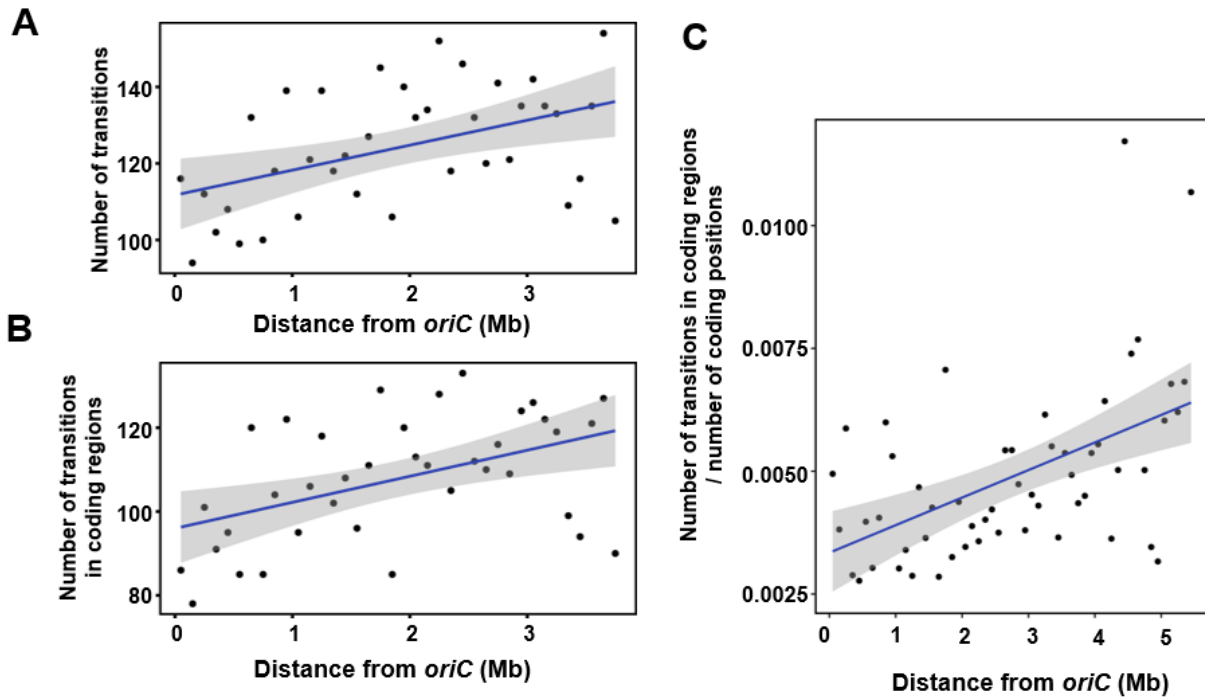


Figure 4. Transition distribution along *Streptomyces* chromosome. (A and B) Transitions along the chromosome of the 12 *S. ambofaciens* evolved $\Delta nucS$ lines were counted within a non-overlapping 100 kb window, starting from *dnaA* gene (located in the middle of the chromosome at position 4,021,374 and approximately corresponding to the origin of replication, *oriC*) and sliding towards the chromosome extremities. The number of transitions in windows of the right replichore was added to the number of transitions of windows at the same distance of *dnaA* in the left replichore. The sum of total transitions **(A)** or the sum of transitions restricted to the coding sequences **(B)** was represented as a function of the distance from *dnaA*. Significant positive correlations were observed between the genomic position and the transition count (Pearson's correlation coefficient $r=0.457$, $P=0.004$) or the number of transitions within the coding sequences (Pearson's correlation coefficient $r=0.468$, $P=0.003$). **(C)** Transition distribution was estimated along the chromosome of environmental strain RLB1-8 compared to the environmental strain RLB3-17. As above, transitions within coding regions, on orthologous pairs of genes (BLASTp reciprocal best hits with at least 40% identity, 70% coverage, and an E-value lower than $1e-10$), were counted within a non-overlapping 100 kb window starting from *dnaA* gene (corresponding to *oriC*, approximately at position 6,204,383 in RLB1-8). The number of transitions in windows of the right replichore was added to the number of transitions of windows at the same distance of *dnaA* in the left replichore and the sum was normalized to the total number of coding positions of the window. A significant positive correlation was observed between the transition count and the genomic position (Kendall rank correlation test $r=0.38$ $P=0.00005$).

$P=0.003$). As a low occurrence of mutations in the central part of the chromosome could be due to negative selection of mutations in essential genes, non-synonymous and synonymous mutations within coding sequences were counted along the chromosome of $\Delta nucS$ lines and plotted as above. Non-synonymous mutation frequency (Supplementary Figure S4B) increased significantly towards the arm ends (Pearson's correlation coefficient $r=0.499$, $P=0.001$), but no trend was observed for synonymous

mutations (Supplementary Figure S4C). However, the positive correlation found between the non-synonymous mutation number and the genomic position was lost when the non-synonymous mutation number was plotted against the number of synonymous mutations for each position (Supplementary Figure S4D). Indeed, the ratio between non-synonymous and synonymous mutations is homogeneous along the genome, suggesting that there is no clear selection difference depending on the genomic position.

As BPS accumulation is low in the WT evolved lines, we decided to estimate the accumulation of transitions by comparing WT genomes of strains that have been diverging for much longer than 60 sporulation cycles. Therefore, the genomes of *Streptomyces* environmental strains RLB1-8 and RLB3-17 (51) sharing an average nucleotide identity (ANI) of 99.19% were compared in terms of SNP distribution. Orthologs were defined by pairwise comparison of both strain genomes and detection of transitions was first specifically done in orthologous pairs. As above, the number of transitions was counted within a sliding non-overlapping 100 kb window, starting from *oriC* (estimated at position 6,204,383 by the *dnaA* gene location in RLB1-8 genome, which was used as the reference), extending towards the left or right end of the chromosome. The sum of equidistant windows on the left and the right replichores was normalized to the total number of coding positions included in the windows (to consider the possible differential density of orthologous genes along the chromosome) and plotted as a function of the distance from *oriC*. As shown in Figure 4C, a clear increase of transition number in orthologous genes is observed when the distance from the origin of replication increases (Kendall rank correlation test $r=0,38$ $P=0,00005$). When considering the whole genome and not only the coding regions, the transitions are also clearly more frequent when moving away from *oriC* (Kendall rank correlation test $r=0,33$ $P=0,00045$) (Supplementary Figure S5). The comparison of genomes of other conspecific strains isolated from the same soil grain showed a similar increasing trend towards the chromosome termini (data not shown). Altogether, these results indicate that a short time evolutionary scale is sufficient to observe transition accumulation with a crescent-shaped distribution along the chromosome of WT strains possessing a functional non-canonical MMR.

Occurrence of BPSs has a strong DNA-strand bias

When considering the 5' to 3' nucleotide sequence ("top strand") of *S. ambofaciens* chromosome, the left replichore (starting at the left end to the *oriC* position) corresponds to the lagging strand while the right replichore (starting at *oriC* to the right end) corresponds to the leading strand. The A:T>G:C transitions occur when C is mispaired with template A or G is mispaired with template T. Considering one strand, A:T>G:C transitions are either A>G or T>C, and our results in $\Delta nucS$ evolved lines (Supplementary Table S3) showed that in the leading strand (right replichore), A>G substitutions were

three times more frequent than T>C substitutions (1150 versus 412, respectively). This bias was reversed in the lagging strand (left replichore) where mispaired Ts occurred more often than mispaired As (1031 Ts versus 393 As). This means that during replication, A:T>G:C transitions are more likely to originate from a C in the lagging strand mispaired to template A and a G in the leading strand mispaired to a template T than the contrary. Likewise, a similar strand bias was found for G:C>A:T transitions with more C>T than G>A substitutions in the right replichore (715 versus 274, respectively); the reversed trend was observed in the left replichore (234 versus 699, respectively). Thus G:C>A:T transitions are more likely to originate from a A in the lagging strand mispaired to template C and a T in the leading strand mispaired to a template G than the contrary. This transition strand bias is consistent with the observations made in MMR defective *E. coli* (4), *M. smegmatis* (19), and also *C. glutamicum* (21). Since this finding was made in the mutant lines lacking a functional MMR, and since NucS_{Sam} was able to target G/T but not A/C mismatches *in vitro*, it is possible that transitions arise more frequently from a G mispaired to a T on the lagging strand template than a T on the leading strand template, which would indicate a nucleotide misincorporation bias by the DNA polymerase, as suggested in *E. coli* (60). No strand bias could be observed in the WT evolved lines, because the mutation rate is too low. Therefore, further investigation is needed to clarify a possible strand-specific activity of the non-canonical MMR.

DISCUSSION

Streptomyces species, like the majority of actinobacteria, do not possess the signature proteins of the canonical MMR pathway, MutSL. The discovery of NucS in archaea as part of an alternative MMR pathway prompted further investigation. NucS is not only found in many archaea but also widely conserved among most actinobacteria (9, 18). In our analysis, we found that NucS_{Sam} is highly conserved within the *Streptomyces* genus, establishing *nucS* as a core gene in these bacteria. *S. ambofaciens* Δ *nucS* mutant displayed a hypermutable phenotype characterized by the appearance of sectorized colonies. A sector on a colony (with a radial growth) is a visible different hyphal phenotype resulting from a mutation that occurred during the colony growth. The frequency of sectors is therefore a visual indication of mutation rate. The increased emergence of rifampicin-resistant colonies in the progeny of Δ *nucS* mutant also indicated a significantly increased spontaneous mutation rate. These results are consistent with previous studies conducted in other actinobacterial species, and provide additional evidence for the role of NucS in genome maintenance (18, 20, 21). Our biochemical study demonstrated that NucS_{Sam} targets G/T, T/T, and G/G dsDNA substrates by cleaving both strands, resulting thus in a DSB. This cleavage mechanism and preference are in accordance with those described in archaeal species and in *C. glutamicum* (9, 20, 21). The Mn²⁺ enzymatic cofactor was required for NucS_{Sam} activity, in the presence or absence of β -clamp_{Sam}. However, no cleavage activity was observed with Mg²⁺ for NucS_{Sam}. This differs from NucS from *T. kodakarensis* and *C. glutamicum* which use both Mn²⁺ and Mg²⁺ (9, 20, 21). Such activity has not yet been shown in *Mycobacterium* species, and one of the suggested hypotheses was the need for other partners (18). Indeed, the endonuclease activity of NucS_{Sam} was significantly enhanced by the presence of the β -clamp, a crucial protein involved in DNA replication. This interaction with β -clamp is essential for efficient mutation avoidance and genome stability in *C. glutamicum* (20, 21). It has been hypothesized that the β -clamp could induce a conformational change that favors DNA binding or induce the catalytic step (17) but most importantly, it could tether the MMR enzyme to the replication zone where mismatches occur (21). The coupling of canonical MMR to the β -clamp has also been demonstrated as the interaction of MutS and MutL with β -clamp is necessary for a functional MMR in *E. coli* and in *B. subtilis* (12, 61, 62).

Our biochemical study also revealed an additional endonuclease activity of NucS_{Sam} on deaminated substrates such as G/U and T/H mismatches. Interestingly, this activity was only observed in the presence of β -clamp. This suggests that NucS_{Sam}, together with β -clamp, plays a role in the repair of DNA lesions resulting from deamination similar to the base excision repair (BER) pathway. Conversion of cytosine to uracil leads to G:C>A:T base pair mutation in the subsequent DNA replication. Uracil-DNA glycosylase (UDG) is a key enzyme that cleaves the N-glycosidic bond thereby initiating the BER pathway. It has been shown that UDG genes are highly conserved across *Streptomyces* genus and that the deletion

of these genes is not lethal in *Streptomyces lividans* (63), nor in *E. coli* (64). Nevertheless, it is conceivable that, in addition to UDGs, NucS_{Sam} may play a significant role in the BER-like pathway in *Streptomyces*. This hypothesis is supported by structural and biochemical investigations of NucS from the hyperthermophilic archaeon *T. gammatolerans*, which is thought to be involved in an alternative pathway for the repair of deaminated bases (58). The study has highlighted that the ends of cleaved dsDNA products could be utilized by homologous recombination proteins. However, the involvement of PCNA was not investigated in this study. Our results suggest that the activity of NucS_{Sam} on deaminated bases is coupled to replication, as it requires the presence of β -clamp. A possibility is that NucS_{Sam} plays a role in a mechanism similar to pre-replicative BER. Pre-replicative BER, mainly described in eukaryotes, is thought to target and repair base lesions on DNA template in front of the replication fork and this function has been linked to the DNA-glycosylase NEIL1 and to PCNA in human cells (65). Interestingly, *M. tuberculosis* Nei2, a BER glycosylase that removes oxidized base lesions from ssDNA and replication fork-mimicking substrates, has recently been shown to interact with β -clamp, which enhances its activity (66, 67). NucS_{Sam}, through its interaction with β -clamp_{Sam}, could have a similar role in targeting deaminated bases in front of the replication forks. However, it remains unclear whether and how NucS_{Sam} targets deaminated base sites that arise spontaneously independently of replication, for example during oxidative stress.

A significant transition-biased spectrum was observed in $\Delta nucS$ strains through MA assays. This is in line with previous works in actinobacteria and archaea (19, 21, 68). Such transition shift is a well-established hallmark of MMR deficiency as observed in organisms bearing a canonical MMR such as *E. coli* (4). It is indeed known that transition mispairs are less efficiently proofread than transversions, making them specific targets of MMR machineries (69). It is important to point out that $\Delta nucS$ lines accumulate more A:T>G:C than G:C>A:T transitions whereas the opposite is observed in WT lines. Thus, the non-canonical MMR could prevent the accumulation of GCs in *Streptomyces* genomes, which already have a high GC content, as suggested for GC-rich mycobacterial genomes (19), and also with opposite observations in *E. coli* genome (50% GC) (4). A very interesting result is the correlation that can be made with our *in vivo* and *in vitro* findings. Indeed, *in vitro* activities revealed that the G/T mismatch is clearly one of the preferred substrates of NucS_{Sam}, whereas no cleavage activity on A/C mispair could be detected. Such striking preference was also observed with NucS from *T. kodakarensis* (9) and *C. glutamicum* (20, 21). Furthermore, recent oligonucleotide recombination experiments in *M. smegmatis* yet to be published (70) revealed that *in vivo* NucS would only correct G/T but not A/C mismatches in this species. In this case, transition events in MMR-deficient $\Delta nucS$ lines are due to uncorrected G/T mismatches. For example, during replication, a G>A transition can arise from a mispaired G/T or a mispaired C/A (Figure 5, upper panel). Since NucS only targets G/T mismatches, the increase of G>A

transition mutation in the absence of NucS is due to accumulation of G/Ts (Figure 5, lower panel), and this is also true for C>T, A>G and T>C transitions. Therefore, the drastic transition accumulation in *Streptomyces ΔnucS* lines reflects the frequency of G/T mismatch occurrence while the negligible frequency of uncorrected A/C mispair events can be estimated by the residual transition accumulation in genomes of WT lines possessing a functional non-canonical MMR. These results enable to discriminate which transition mispairing is the most frequent during incorporation of nucleotides by DNA polymerase. The replication complex is thereby prone to generate far more G/T than C/A mismatches.

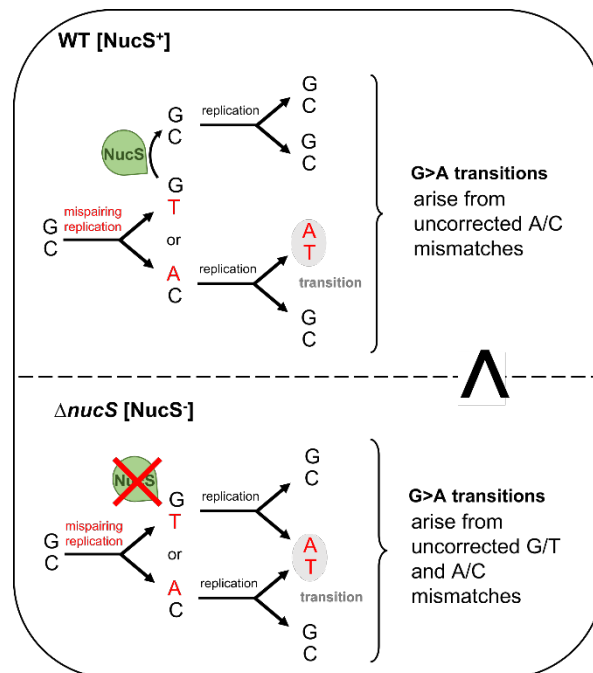


Figure 5. Implication of NucS mismatch preference in transition events during replication. The diagram illustrates the events leading to the base substitution of a G into an A, which is one of the four transitions that can occur on one strand during replication (the others being A>G, C>T and T>C). The sequence of events is described in WT bacteria possessing a functional non-canonical MMR (NucS endonuclease is active) and in a $\Delta nucS$ strain (lacking NucS-dependent MMR). Mutated $\Delta nucS$ lines accumulated far more G>A transitions than WT lines (shown with a superscript >).

Our results, supported by MA and biochemical studies on *C. glutamicum* (20, 21), highlight a remarkable *in vivo* property of the DNA polymerase. The hypothesis that G/T mismatches are much more likely to occur *in vivo* has previously been proposed for canonical MMR-bearing bacteria, since MutS protein preferentially binds G/T over A/C (71). MA assays in yeast showed that G pairing with T is the most common mispair generated by DNA polymerase (72). Many studies on DNA polymerase fidelity have been performed *in vitro*, and recently a high throughput analysis based on next-generation sequencing showed general DNA polymerase preference for dGTP and dTTP misincorporation at T and G bases (73). This preference can be explained by tautomer occurrence and stability, and spatial constraints in the

polymerase active site during catalysis of the phosphodiester bond (74–78). The DNA bases exist as tautomers (interchangeable pairs of isomers), and A/T and G/C base pairs contain each base in its predominant tautomeric form. Watson and Crick (79) proposed that if nucleotide bases adopted their energetically less favorable tautomeric forms, mismatches could arise in a Watson–Crick (WC)-like geometry and consequently lead to spontaneous mutations. However, it is only in the last decade that this 70-year-old model has been experimentally validated, and, in particular, G/T tautomers were demonstrated to exist during DNA polymerization and to have a role in base pair mis-incorporation and transition occurrence (80). Altogether, these data underline the propensity of DNA polymerases to produce G/T mispairs and our *in vivo* results together with observations in *C. glutamicum* (20, 21) demonstrate that these G/T mismatches are specifically eliminated by a NucS-dependent MMR mechanism, which has evolved to correct these frequent transition mismatches in cells.

Another remarkable observation is that accumulation of mutations in evolved $\Delta nucS$ lines correlates with the distance to the origin of replication. It is noteworthy that the frequency of transitions, which represent the vast majority of mutations in the mutant, follows the same trend. This crescent-shaped BPS distribution in the absence of a functional MMR reflects a variable mutation rate of DNA polymerase (due to variable accuracy or an imbalance in the nucleotide pools) and/or a differential selection along the chromosome. It is known that core genes are located in the central compartment of the chromosome in *Streptomyces* (29) and, although there is no data on the distribution of essential genes, it can be argued that most essential genes are among the core genes, so it is likely that some mutations in the central region of the chromosome are counter-selected. A commonly used way of assessing selection bias is the ratio of non-synonymous to synonymous mutations, assuming that synonymous mutations are relatively neutral. Our results showed that there was no correlation between the dN/dS ratio and the genomic position in the evolved $\Delta nucS$ lines. Therefore, the BPS distribution in the absence of a functional MMR reflects the distribution of mismatches that occur during replication and suggests an increasing mutation rate of the DNA polymerase towards the extremities (Figure 6). Given the propensity of DNA polymerase to generate G/T mismatches (as discussed above), this implies that G/T mispairs accumulate with the distance from the replication origin, a finding that, to our knowledge, has never been documented for bacterial linear chromosomes. Furthermore, our results of differential mutation rate along the chromosome suggest that different areas of the genome evolve at different rates in *Streptomyces*, implying that the genes located closer to the replication origin are less prone to evolution than those closer to the termini, which is consistent with the compartmentalization of *Streptomyces* chromosome (29). In WT evolved lines, the low total number of mutations prevented us from building a robust statistical model for transition distribution. However, the comparison of genomes of two close *Streptomyces* individuals isolated from a soil grain and deriving from a recent common ancestor revealed

that the transition rate increases significantly with the distance from the origin. This finding suggests that this short time evolutionary scale is long enough to observe that transitions in WT environmental strains tend to accumulate in the extremities of the chromosome, where transition mismatches are more likely to occur and/or escape proofreading and MMR activities. Pioneering studies on *E. coli* and *Salmonella typhimurium* genomes indicated that genes located further from the replication origin have higher mutation rates than those closer to the origin (81, 82), and some studies suggest that it is the proximity to the fragile replication termination sites that is responsible for this phenomenon (83, 84). In the last decade, MA studies on lines lacking MMR in *E. coli* or *Bacillus subtilis* (85, 86) and in *Vibrio* and *Pseudomonas* species (87). revealed a more precise wave-like pattern that is symmetrical about the replication origin. Such wave-like pattern of mutation distribution was not visible in the genome of evolved $\Delta nucS$ lines, either because the total number of mutations in this study is not high enough to build such a model, or because the *Streptomyces* linear chromosome has a different tridimensional structure (see below). In *E. coli*, it has been observed that intrinsic errors by the DNA polymerase occur more frequently near the replication origin, but also that error correction by the proofreading activity of the DNA polymerase or canonical MMR is more effective close to the *oriC* region. More importantly, it has been hypothesized that DNA replication becomes inaccurate in regions of high superhelical density and that nucleoid-associated protein (NAP) binding increases mutability by impeding DNA repair and/or interfering with DNA polymerase processivity during replication (84, 86, 88). *Streptomyces* NAPs are numerous and their binding along the chromosome is temporally and spatially variable throughout the bacterial life cycle (89–91) such that the association between NAPs and mutation density is complicated. An interesting example is the HU family-protein HupS which participates in chromosomal spatial rearrangement and compaction at the onset of sporulation in *Streptomyces venezuelae* (90). Increased binding of HupS to the subterminal regions organize them into a closed structure, that might be prone to a higher mutation rate.

Although several studies revealed a cleavage activity of NucS on mismatched DNA and the generation of a DSB *in vitro*, such activity has not been demonstrated *in vivo* and further investigations are needed to elucidate this particularity of the non-canonical MMR. The mechanism by which NucS-mediated DSBs would be repaired remains elusive. The involvement of homologous recombination to process the unwanted DSB is the most obvious hypothesis and would allow a mismatch repair without strand discrimination. The recruitment of homologous recombination to DNA damage could be achieved *via* the replication clamp which is a known platform that recruits numerous DNA-processing enzymes including DNA repair enzymes (92). Besides, unlike most bacteria, *Streptomyces* species also have the ability to repair DSBs by illegitimate recombination pathway. Our previous work has demonstrated the existence of a functional NHEJ pathway in *Streptomyces* (37). Hence, NucS-dependent DSBs may also be

repaired by NHEJ in *Streptomyces*. Very recently, a preprint manuscript (70) suggested that mycobacterial NucS participates in a MMR reaction involving short patch repair confined within 5-6 bp of the mispaired nucleotides, which appears to be mechanistically distinct from repair by homologous recombination or NHEJ DSB repair mechanisms. A similar study in *M. smegmatis*, not approved by peer review yet, has also shown that a NucS-promoted DSB is processed by a 5'-3' exonuclease, but is independent of both RecA, RadA, and NHEJ functions Ku and LigD (93).

The high level of genetic instability observed in the *Streptomyces* genus is one of its most intriguing features and DSB repair is thought to be the main driver of this instability (35). The increasing occurrence of mismatches towards the ends of the chromosome suggests that NucS activity is more required further from the origin and consequently that more DSBs occur in the subtelomeric regions (Figure 6). Several studies have pointed to the existence of a recombination gradient towards the chromosomal arms highlighting the compartmentalization of the *Streptomyces* genome (29, 51). Our results support an increase of DSBs towards the ends of the chromosome, contributing to increased recombination events in the extremities and thus to the unique organization of the *Streptomyces* chromosome. Yet, how the NucS-dependent DSBs are repaired remains to be elucidated. Additional aspects such as differential efficiency of DSB repair along the chromosome as well as the involvement of the spatial conformation of the chromosome (94) or the role of chromatin-organizing proteins (90) need to be considered to gain a comprehensive understanding of *Streptomyces* genome dynamics.

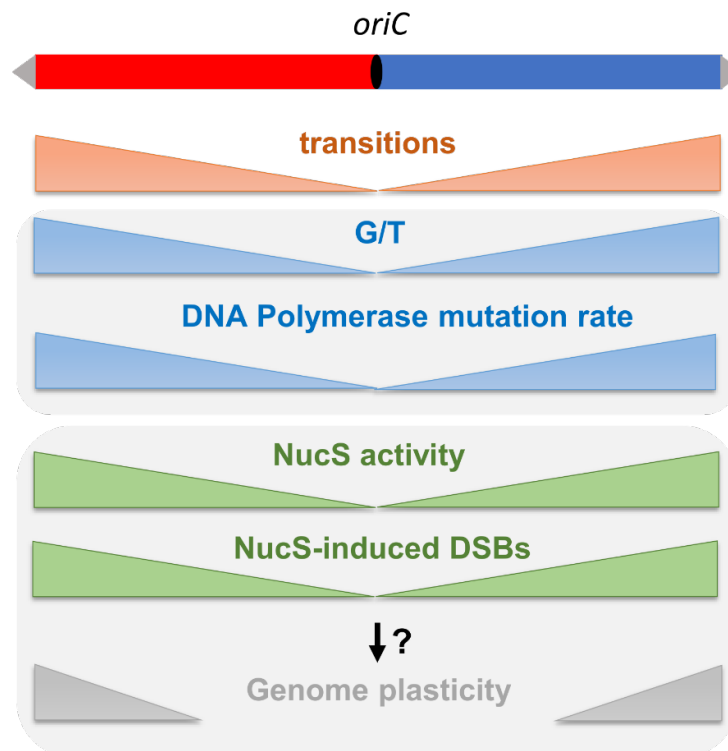


Figure 6. DNA polymerase mutation rate and non-canonical MMR activity along the chromosome during replication in *Streptomyces*. Red and blue boxes represent left and right replichores of the chromosome. Grey arrows represent the terminal inverted repeats at each chromosome extremity. In MMR defective ($\Delta nucS$) lines, the transition density increases towards the extremities of the chromosome. Because $NucS_{Sam}$ cleaves G/T but not A/C mismatches *in vitro*, this transition trend suggests an increase of G/T mispairing and an increase in DNA polymerase mutation rate towards the extremities (in blue). This observed mutation profile also suggests that in MMR proficient (WT) lines, NucS activity increases with the distance of the replication origin, and thus NucS-induced DSBs are more frequent towards the extremities of the chromosome (in green). Assuming that these DSBs are repaired by recombination, this profile is in accordance with the recombination gradient driving the genome plasticity described in *Streptomyces* (illustrated in grey).

DATA AVAILABILITY

The sequencing data underlying this article have been deposited with links to BioProject accession number PRJNA1047308 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>).

FUNDING

This work was supported by the French National Research Agency program MMRDNABREAK (ANR-22-CE12-0042-02).

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the ASIA platform and Tiphaine Dhalleine for the technical expertise with SEC-MALS analyses.

REFERENCES

1. Ganai,R.A. and Johansson,E. (2016) DNA Replication—A Matter of Fidelity. *Molecular Cell*, **62**, 745–755.
2. Li,G.-M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res*, **18**, 85–98.
3. Putnam,C.D. (2021) Strand discrimination in DNA mismatch repair. *DNA Repair*, **105**, 103161.
4. Lee,H., Popodi,E., Tang,H. and Foster,P.L. (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.
5. Sachadyn,P. (2010) Conservation and diversity of MutS proteins. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **694**, 20–30.
6. Busch,C.R. and DiRuggiero,J. (2010) MutS and MutL Are Dispensable for Maintenance of the Genomic Mutation Rate in the Halophilic Archaeon *Halobacterium salinarum* NRC-1. *PLoS ONE*, **5**, e9045.
7. Kucukyildirim,S., Long,H., Sung,W., Miller,S.F., Doak,T.G. and Lynch,M. (2016) The Rate and Spectrum of Spontaneous Mutations in *Mycobacterium smegmatis*, a Bacterium Naturally Devoid of the Postreplicative Mismatch Repair Pathway. *G3 (Bethesda)*, **6**, 2157–2163.
8. Ren,B., Kühn,J., Meslet-Cladiere,L., Briffotiaux,J., Norais,C., Lavigne,R., Flament,D., Ladenstein,R. and Myllykallio,H. (2009) Structure and function of a novel endonuclease acting on branched DNA substrates. *EMBO J*, **28**, 2479–2489.
9. Ishino,S., Nishi,Y., Oda,S., Uemori,T., Sagara,T., Takatsu,N., Yamagami,T., Shirai,T. and Ishino,Y. (2016) Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Res*, **44**, 2977–2986.
10. Maga,G. and Hubscher,U. (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci*, **116**, 3051–3060.
11. Moldovan,G.-L., Pfander,B. and Jentsch,S. (2007) PCNA, the maestro of the replication fork. *Cell*, **129**, 665–679.
12. Saro,F.J.L. de, Marinus,M.G., Modrich,P. and O'Donnell,M. (2006) The β Sliding Clamp Binds to Multiple Sites within MutL and MutS. *Journal of Biological Chemistry*, **281**, 14340–14349.
13. Pluciennik,A., Burdett,V., Lukianova,O., O'Donnell,M. and Modrich,P. (2009) Involvement of the β Clamp in Methyl-directed Mismatch Repair *in Vitro*. *Journal of Biological Chemistry*, **284**, 32782–32791.
14. Meslet-Cladière,L., Norais,C., Kuhn,J., Briffotiaux,J., Sloostra,J.W., Ferrari,E., Hübscher,U., Flament,D. and Myllykallio,H. (2007) A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen. *J Mol Biol*, **372**, 1137–1148.
15. Creze,C., Ligabue,A., Laurent,S., Lestini,R., Laptinok,S.P., Khun,J., Vos,M.H., Czjzek,M., Myllykallio,H. and Flament,D. (2012) Modulation of the *Pyrococcus abyssi* NucS Endonuclease Activity by Replication Clamp at Functional and Structural Levels *. *Journal of Biological Chemistry*, **287**, 15648–15660.

16. Dalrymple,B.P., Kongsuwan,K., Wijffels,G., Dixon,N.E. and Jennings,P.A. (2001) A universal protein–protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences*, **98**, 11627–11632.
17. Nakae,S., Hijikata,A., Tsuji,T., Yonezawa,K., Kouyama,K., Mayanagi,K., Ishino,S., Ishino,Y. and Shirai,T. (2016) Structure of the EndoMS-DNA Complex as Mismatch Restriction Endonuclease. *Structure*, **24**, 1960–1971.
18. Castañeda-García,A., Prieto,A.I., Rodríguez-Beltrán,J., Alonso,N., Cantillon,D., Costas,C., Pérez-Lago,L., Zegeye,E.D., Herranz,M., Płociński,P., *et al.* (2017) A non-canonical mismatch repair pathway in prokaryotes. *Nat Commun*, **8**.
19. Castañeda-García,A., Martín-Blecua,I., Cebrián-Sastre,E., Chiner-Oms,A., Torres-Puente,M., Comas,I. and Blázquez,J. (2020) Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies. *Sci. Adv.*, **6**, eaay4453.
20. Ishino,S., Skouloubris,S., Kudo,H., l’Hermitte-Stead,C., Es-Sadik,A., Lambry,J.-C., Ishino,Y. and Myllykallio,H. (2018) Activation of the mismatch-specific endonuclease EndoMS/NucS by the replication clamp is required for high fidelity DNA replication. *Nucleic Acids Research*, **46**, 6206–6217.
21. Takemoto,N., Numata,I., Su’etsugu,M. and Miyoshi-Akiyama,T. (2018) Bacterial EndoMS/NucS acts as a clamp-mediated mismatch endonuclease to prevent asymmetric accumulation of replication errors. *Nucleic Acids Research*, **46**, 6152–6165.
22. Fressatti Cardoso,R., Martín-Blecua,I., Pietrowski Baldin,V., Meneguello,J.E., Valverde,J.R., Blázquez,J. and Castañeda-García,A. (2022) Noncanonical Mismatch Repair Protein NucS Modulates the Emergence of Antibiotic Resistance in *Mycobacterium abscessus*. *Microbiol Spectr*, 10.1128/spectrum.02228-22.
23. Iliakis,G. (1991) The role of DNA double strand breaks in ionizing radiation-induced killing of eukaryotic cells. *Bioessays*, **13**, 641–648.
24. Kuzminov,A. (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol Mol Biol Rev*, **63**, 751–813, table of contents.
25. Glickman,M.S. (2014) Double-Strand DNA Break Repair in Mycobacteria. *Microbiol Spectr*, **2**.
26. Bertrand,C., Thibessard,A., Bruand,C., Lecointe,F. and Leblond,P. (2019) Bacterial NHEJ: a never ending story. *Mol. Microbiol.*, **111**, 1139–1151.
27. Barka,E.A., Vatsa,P., Sanchez,L., Gaveau-Vaillant,N., Jacquard,C., Meier-Kolthoff,J.P., Klenk,H.-P., Clément,C., Ouhdouch,Y. and van Wezel,G.P. (2016) Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev*, **80**, 1–43.
28. Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Ōmura,S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol*, **21**, 526–531.
29. Lorenzi,J.-N., Lespinet,O., Leblond,P. and Thibessard,A. (2021) Subtelomeres are fast-evolving regions of the *Streptomyces* linear chromosome. *Microb Genom*, **7**, 000525.
30. Choulet,F., Aigle,B., Gallois,A., Mangenot,S., Gerbaud,C., Truong,C., Francou,F.-X., Fourrier,C., Guérineau,M., Decaris,B., *et al.* (2006) Evolution of the Terminal Regions of the *Streptomyces* Linear Chromosome. *Molecular Biology and Evolution*, **23**, 2361–2369.
31. Leblond,P., Demuyter,P., Moutier,L., Laakel,M., Decaris,B. and Simonet,J.M. (1989) Hypervariability, a new phenomenon of genetic instability, related to DNA amplification in *Streptomyces ambofaciens*. *J Bacteriol*, **171**, 419–423.

32. Leblond,P., Demuyter,P., Simonet,J.M. and Decaris,B. (1990) Genetic instability and hypervariability in *Streptomyces ambofaciens*: towards an understanding of a mechanism of genome plasticity. *Molecular Microbiology*, **4**, 707–714.
33. Birch,A., Häusler,A. and Hütter,R. (1990) Genome rearrangement and genetic instability in *Streptomyces* spp. *J Bacteriol*, **172**, 4138–4142.
34. Fischer,G., Wenner,T., Decaris,B. and Leblond,P. (1998) Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proceedings of the National Academy of Sciences*, **95**, 14296–14301.
35. Hoff,G., Bertrand,C., Piotrowski,E., Thibessard,A. and Leblond,P. (2018) Genome plasticity is governed by double strand break DNA repair in *Streptomyces*. *Scientific Reports*, **8**, 5272.
36. Bury-Moné,S. Dynamics of the *Streptomyces* chromosome: chance and necessity. *Trends in Genetics*.
37. Hoff,G., Bertrand,C., Zhang,L., Piotrowski,E., Chipot,L., Bontemps,C., Confalonieri,F., McGovern,S., Lecoïnte,F., Thibessard,A., *et al.* (2016) Multiple and Variable NHEJ-Like Genes Are Involved in Resistance to DNA Damage in *Streptomyces ambofaciens*. *Front. Microbiol.*, **7**.
38. McGovern,S., Baconnais,S., Roblin,P., Nicolas,P., Drevet,P., Simonson,H., Piétrement,O., Charbonnier,J.-B., Le Cam,E., Noirot,P., *et al.* (2016) C-terminal region of bacterial Ku controls DNA bridging, DNA threading and recruitment of DNA ligase D for double strand breaks repair. *Nucleic Acids Res*, **44**, 4785–4806.
39. Bonura,T., Smith,K.C. and Kaplan,H.S. (1975) Enzymatic induction of DNA double-strand breaks in gamma-irradiated *Escherichia coli* K-12. *Proc Natl Acad Sci U S A*, **72**, 4265–4269.
40. Kuzminov,A. (1995) Instability of inhibited replication forks in *E. coli*. *Bioessays*, **17**, 733–741.
41. Khan,S.R. and Kuzminov,A. (2012) Replication forks stalled at ultraviolet lesions are rescued via RecA and RuvABC protein-catalyzed disintegration in *Escherichia coli*. *J Biol Chem*, **287**, 6250–6265.
42. Raynal,A., Karray,F., Tüphile,K., Darbon-Rongere,E. and Pernodet,J.-L. (2006) Excisable Cassettes: New Tools for Functional Analysis of *Streptomyces* Genomes. *Applied and Environmental Microbiology*, **72**, 4839–4844.
43. Gust,B., Challis,G.L., Fowler,K., Kieser,T. and Chater,K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A*, **100**, 1541–1546.
44. Luria,S.E. and Delbrück,M. (1943) Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*, **28**, 491–511.
45. Sarkar,S., Ma,W.T. and Sandri,G.H. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica*, **85**, 173–179.
46. Stewart,F.M. (1994) Fluctuation tests: how reliable are the estimates of mutation rates? *Genetics*, **137**, 1139–1146.
47. Hall,B.M., Ma,C.-X., Liang,P. and Singh,K.K. (2009) Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis. *Bioinformatics*, **25**, 1564–1565.
48. Schenk,P.M., Baumann,S., Mattes,R. and Steinbiss,H.H. (1995) Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA polymerase and rare Arg^tRNAs. *Biotechniques*, **19**, 196–198, 200.

49. Kieser, T. ed. (2000) Practical *Streptomyces* genetics Innes, Norwich.
50. Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068–2069.
51. Tidjani, A.-R., Lorenzi, J.-N., Toussaint, M., van Dijk, E., Naquin, D., Lespinet, O., Bontemps, C. and Leblond, P. (2019) Massive Gene Flux Drives Genome Diversity between Sympatric *Streptomyces* Conspecifics. *mBio*, **10**.
52. Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P. and Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*, **57**, 81–91.
53. Darling, A.C.E., Mau, B., Blattner, F.R. and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*, **14**, 1394–1403.
54. Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., et al. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*, **42**, D206-214.
55. Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**, 772–780.
56. Choufa, C., Tidjani, A.-R., Gauthier, A., Harb, M., Lao, J., Leblond-Bourget, N., Vos, M., Leblond, P. and Bontemps, C. (2022) Prevalence and mobility of integrative and conjugative elements within a *Streptomyces* natural population. *Frontiers in Microbiology*, **13**.
57. Martin, P., Dary, A., André, A., Fischer, G., Leblond, P. and Decaris, B. (1999) Intraclonal polymorphism in the bacterium *Streptomyces ambofaciens* ATCC23877: evidence for a high degree of heterogeneity of the wild type clones. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **430**, 75–85.
58. Zhang, L., Shi, H., Gan, Q., Wang, Y., Wu, M., Yang, Z., Oger, P. and Zheng, J. (2020) An alternative pathway for repair of deaminated bases in DNA triggered by archaeal NucS endonuclease. *DNA Repair*, **85**, 102734.
59. Zhang, Z., Shitut, S., Claushuis, B., Claessen, D. and Rozen, D.E. (2022) Mutational meltdown of putative microbial altruists in *Streptomyces coelicolor* colonies. *Nat Commun*, **13**, 2266.
60. Maslowska, K.H., Makiela-Dzbenka, K., Mo, J.-Y., Fijalkowska, I.J. and Schaaper, R.M. (2018) High-accuracy lagging-strand DNA replication mediated by DNA polymerase dissociation. *Proc Natl Acad Sci U S A*, **115**, 4212–4217.
61. Simmons, L.A., Davies, B.W., Grossman, A.D. and Walker, G.C. (2008) Beta clamp directs localization of mismatch repair in *Bacillus subtilis*. *Mol Cell*, **29**, 291–301.
62. Pillon, M.C., Miller, J.H. and Guarné, A. (2011) The endonuclease domain of MutL interacts with the β sliding clamp. *DNA Repair (Amst)*, **10**, 87–93.
63. Zhang, Y., Yun, K., Huang, H., Tu, R., Hua, E. and Wang, M. (2021) Antisense RNA Interference-Enhanced CRISPR/Cas9 Base Editing Method for Improving Base Editing Efficiency in *Streptomyces lividans* 66. *ACS Synth Biol*, **10**, 1053–1063.
64. Amado, L. and Kuzminov, A. (2013) Low-molecular-weight DNA replication intermediates in *Escherichia coli*: mechanism of formation and strand specificity. *J Mol Biol*, **425**, 4177–4191.
65. Rangaswamy, S., Pandey, A., Mitra, S. and Hegde, M.L. (2017) Pre-Replicative Repair of Oxidized Bases Maintains Fidelity in Mammalian Genomes: The Cowcatcher Role of NEIL1 DNA Glycosylase. *Genes (Basel)*, **8**, 175.

66. Lata,K., Afsar,M. and Ramachandran,R. (2017) Biochemical characterization and novel inhibitor identification of *Mycobacterium tuberculosis* Endonuclease VIII 2 (Rv3297). *Biochemistry and Biophysics Reports*, **12**, 20–28.
67. Lata,K., Vishwakarma,J., Kumar,S., Khanam,T. and Ramachandran,R. (2022) *Mycobacterium tuberculosis* Endonuclease VIII 2 (Nei2) forms a prereplicative BER complex with DnaN: Identification, characterization, and disruption of complex formation. *Molecular Microbiology*, **117**, 320–333.
68. Ahmad,S., Huang,Q., Ni,J., Xiao,Y., Yang,Y. and Shen,Y. (2020) Functional Analysis of the NucS/EndoMS of the Hyperthermophilic Archaeon *Sulfolobus islandicus* REY15A. *Front Microbiol*, **11**.
69. Schaaper,R.M. (1993) Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *J Biol Chem*, **268**, 23762–23765.
70. Tanjina Islam and Eric A. Josephs (2023) Genome Editing Outcomes Reveal Mycobacterial NucS Participates in a Short-Patch Repair of DNA Mismatches. *bioRxiv*, 10.1101/2023.10.23.563644.
71. Su,S.S. and Modrich,P. (1986) *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc Natl Acad Sci U S A*, **83**, 5057–5061.
72. Lujan,S.A., Clausen,A.R., Clark,A.B., MacAlpine,H.K., MacAlpine,D.M., Malc,E.P., Mieczkowski,P.A., Burkholder,A.B., Fargo,D.C., Gordenin,D.A., *et al.* (2014) Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res*, **24**, 1751–1764.
73. de Paz,A.M., Cybulski,T.R., Marblestone,A.H., Zamft,B.M., Church,G.M., Boyden,E.S., Kording,K.P. and Tyo,K.E.J. (2018) High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. *Nucleic Acids Research*, **46**.
74. Batra,V.K. Structures of DNA Polymerase Mispaiored DNA Termini Transitioning to Pre-catalytic Complexes Support an Induced-Fit Fidelity Mechanism.
75. Wang,W., Hellinga,H.W. and Beese,L.S. (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc Natl Acad Sci U S A*, **108**, 17644–17648.
76. Kimsey,I.J., Petzold,K., Sathyamoorthy,B., Stein,Z.W. and Al-Hashimi,H.M. (2015) Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature*, **519**, 315–320.
77. Kimsey,I.J., Szymanski,E.S., Zahurancik,W.J., Shakya,A., Xue,Y., Chu,C.-C., Sathyamoorthy,B., Suo,Z. and Al-Hashimi,H.M. (2018) Dynamic basis for dG•dT misincorporation via tautomerization and ionization. *Nature*, **554**, 195–201.
78. Slocombe,L., S. Al-Khalili,J. and Sacchi,M. (2021) Quantum and classical effects in DNA point mutations: Watson–Crick tautomerism in AT and GC base pairs. *Physical Chemistry Chemical Physics*, **23**, 4141–4150.
79. Watson,J.D. and Crick,F.H.C. (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, **171**, 737–738.
80. Goodman,M.F. (2018) Smoking gun for a rare mutation mechanism. *Nature*, **554**, 180–181.
81. Sharp,P.M., Shields,D.C., Wolfe,K.H. and Li,W.H. (1989) Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science*, **246**, 808–810.
82. Mira,A. and Ochman,H. (2002) Gene location and bacterial sequence divergence. *Mol Biol Evol*, **19**, 1350–1358.

83. Mei,Q., Fitzgerald,D.M., Liu,J., Xia,J., Pribis,J.P., Zhai,Y., Nehring,R.B., Paiano,J., Li,H., Nussenzweig,A., *et al.* (2021) Two mechanisms of chromosome fragility at replication-termination sites in bacteria. *Sci Adv*, **7**, eabe2846.
84. Badel,C., Samson,R.Y. and Bell,S.D. (2022) Chromosome organization affects genome evolution in *Sulfolobus archaea*. *Nat Microbiol*, **7**, 820–830.
85. Foster,P.L., Hanson,A.J., Lee,H., Popodi,E.M. and Tang,H. (2013) On the mutational topology of the bacterial genome. *G3 (Bethesda)*, **3**, 399–407.
86. Niccum,B.A., Lee,H., MohammedIsmail,W., Tang,H. and Foster,P.L. (2019) The Symmetrical Wave Pattern of Base-Pair Substitution Rates across the *Escherichia coli* Chromosome Has Multiple Causes. *mBio*, **10**, e01226-19.
87. Kivisaar,M. (2019) Mutation and Recombination Rates Vary Across Bacterial Chromosome. *Microorganisms*, **8**, 25.
88. Warnecke,T., Supek,F. and Lehner,B. (2012) Nucleoid-associated proteins affect mutation dynamics in *E. coli* in a growth phase-specific manner. *PLoS Comput Biol*, **8**, e1002846.
89. Szafran,M.J., Jakimowicz,D. and Elliot,M.A. (2020) Compaction and control-the role of chromosome-organizing proteins in *Streptomyces*. *FEMS Microbiol Rev*, **44**, 725–739.
90. Szafran,M.J., MałeckiT., Strzałka,A., Pawlikiewicz,K., Duława,J., Zarek,A., Kois-Ostrowska,A., Findlay,K.C., Le,T.B.K. and Jakimowicz,D. (2021) Spatial rearrangement of the *Streptomyces venezuelae* linear chromosome during sporogenic development. *Nat Commun*, **12**, 5222.
91. Strzałka,A., Kois-Ostrowska,A., Kędra,M., Łebkowski,T., Bieniarz,G., Szafran,M.J. and Jakimowicz,D. (2022) Enhanced binding of an HU homologue under increased DNA supercoiling preserves chromosome organisation and sustains *Streptomyces* hyphal growth. *Nucleic Acids Res*, **50**, 12202–12216.
92. Henry,C., Kaur,G., Cherry,M.E., Henrikus,S.S., Bonde,N.J., Sharma,N., Beyer,H.A., Wood,E.A., Chitteni-Pattu,S., van Oijen,A.M., *et al.* (2023) RecF protein targeting to post-replication (daughter strand) gaps II: RecF interaction with replisomes. *Nucleic Acids Res*, **51**, 5714–5742.
93. Iris V. Rivera-Flores, Emily X. Wang, and Kenan C. Murphy (2023) *Mycobacterium smegmatis* NucS-promoted DNA Mismatch Repair involves limited resection by a 5'-3' exonuclease and is independent of homologous recombination and NHEJ. *bioRxiv*, 10.1101/2023.11.26.568737.
94. Lioy,V.S., Lorenzi,J.-N., Najah,S., Poinignon,T., Leh,H., Saulnier,C., Aigle,B., Lautru,S., Thibessard,A., Lespinet,O., *et al.* (2021) Dynamics of the compartmentalized *Streptomyces* chromosome during metabolic differentiation. *Nat Commun*, **12**, 5221.

SUPPLEMENTARY DATA

Supplementary Table S1. Primers used in this study.

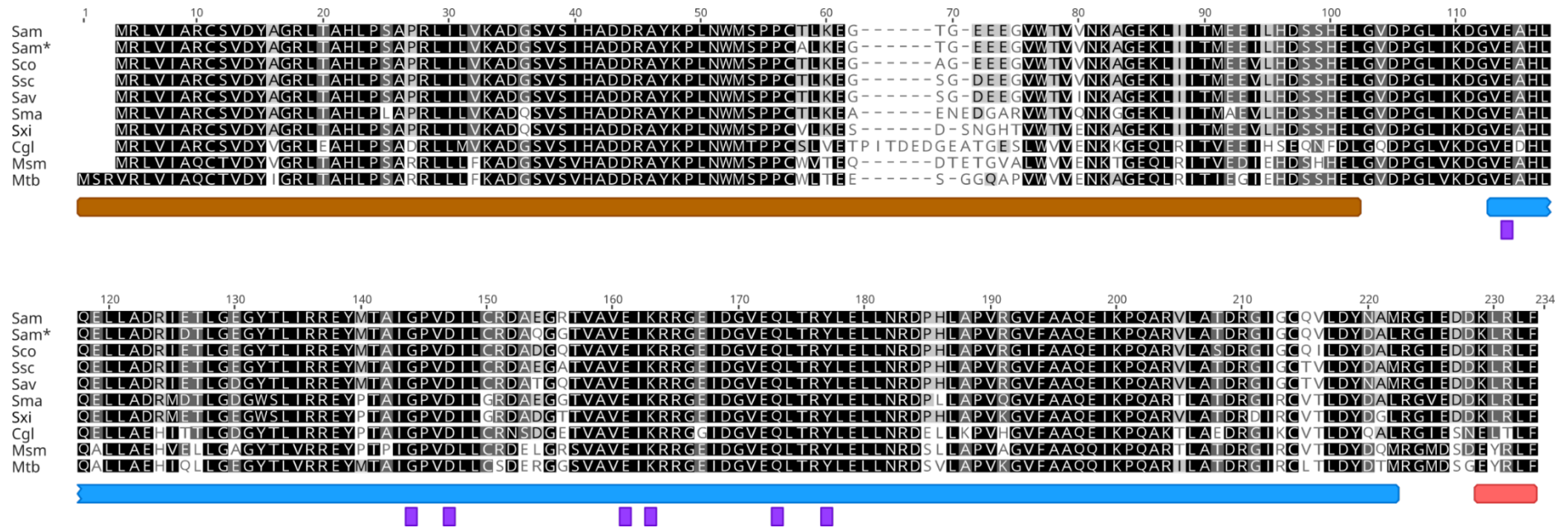
| Name | Sequence (5'-3') | Purpose |
|--------------|----------------------------------|---|
| nucS_compl_F | AGGGGAATTCAGACGTCGAGCGACATGGGT | Cloning <i>nucS</i> in pSET152 for the creation of complemented strains |
| nucS_compl_R | CGAGGAATTCGACGGAGCAGATCATCTGAGC | |
| nucS_F | GTTGCCGCATATGCGTCTCGTCATTGCCCGCT | Cloning <i>nucS</i> in pET15b for heterologous expression |
| nucS_R | TGTAGGATCCTCAGAACAGCCGACGCTTG | |
| dnaN_F | TAGCATATGAAGATCCGGGTGGAACGC | Cloning <i>dnaN</i> in pET15b for heterologous expression |
| dnaN_R | ACTGGATCCTCAGCCGCTCAGCCGCAC | |

Supplementary Table S2. Sequence of oligonucleotides used for cleavage assays (left) and combination of the oligonucleotides to form heteroduplex DNA with a mismatch (right). Asterisk means that the sequence is 6FAM labeled at 5'

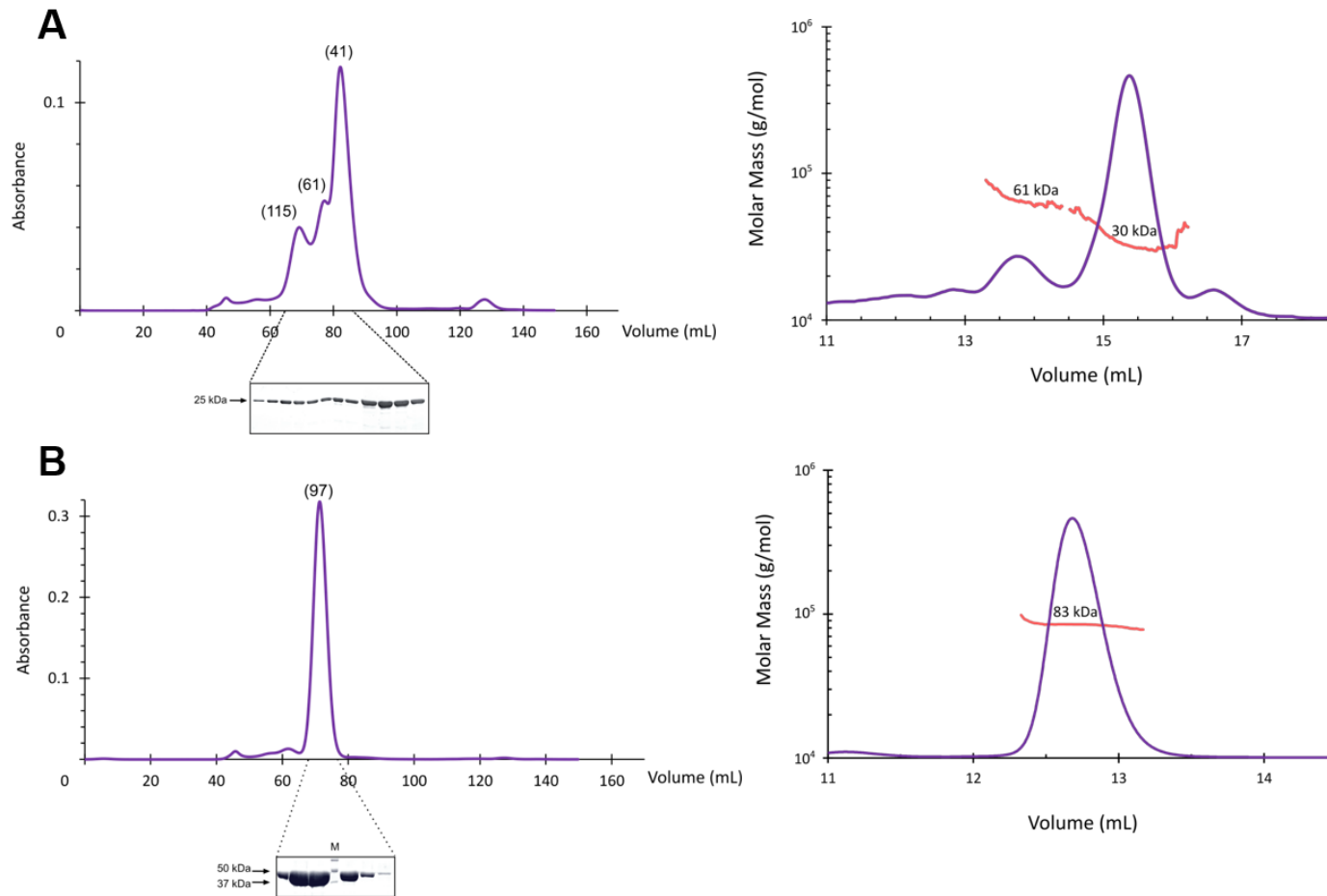
| Number | Sequence (5'-3') |
|--------|---|
| 1* | ACCCAGTACTCGCTGCTGAATGGAGCCGCGGGCTGAAGGACA |
| 2* | TGTCCTTCAGCCGCGGGCTCCGTTACAGCAGCAGTACTGGGT |
| 3* | ACCCAGTACTCGCTGCTG |
| 4* | TGTCCTTCAGCCGCGGGCT |
| 5 | TGTCCTTCAGCCGCGGGCTCCATTACAGCAGCAGTACTGGGT |
| 6 | TGTCCTTCAGCCGCGGGCTCCGTTACAGCAGCAGTACTGGGT |
| 7 | TGTCCTTCAGCCGCGGGCTCCTTTCAGCAGCAGTACTGGGT |
| 8 | TGTCCTTCAGCCGCGGGCTCCCTTCAGCAGCAGTACTGGGT |
| 9 | TGTCCTTCAGCCGCGGGCTCCAATCAGCAGCAGTACTGGGT |
| 10 | TGTCCTTCAGCCGCGGGCTCCAATCAGCAGCAGTACTGGGT |
| 11 | ACCCAGTACTCGCTGCTGAAAGGAGCCGCGGGCTGAAGGACA |
| 12 | ACCCAGTACTCGCTGCTGAAAGGAGCCGCGGGCTGAAGGACA |
| 13 | ACCCAGTACTCGCTGCTGAAACCGAGCCGCGGGCTGAAGGACA |
| 14 | ACCCAGTACTCGCTGCTGAAUGGAGCCGCGGGCTGAAGGACA |
| 15 | TGTCCTTCAGCCGCGGGCTCCHTTACAGCAGCAGTACTGGGT |
| 16 | CCGTTACAGCAGCAGTACTGGGT |
| 17 | AATGGAGCCGCGGGCTGAAGGACA |

| Mismatch or marker name | Combination |
|-------------------------|-------------|
| none | 1+5 |
| G/T | 1+6 |
| T/T | 1+7 |
| T/C | 1+8 |
| A/A | 1+9 |
| A/C | 1+10 |
| G/G | 2+11 |
| G/A | 2+12 |
| C/C | 2+13 |
| G/U | 2+14 |
| T/H | 1+15 |
| Marker 1 (M1) | 3+16 |
| Marker 2 (M2) | 4+17 |

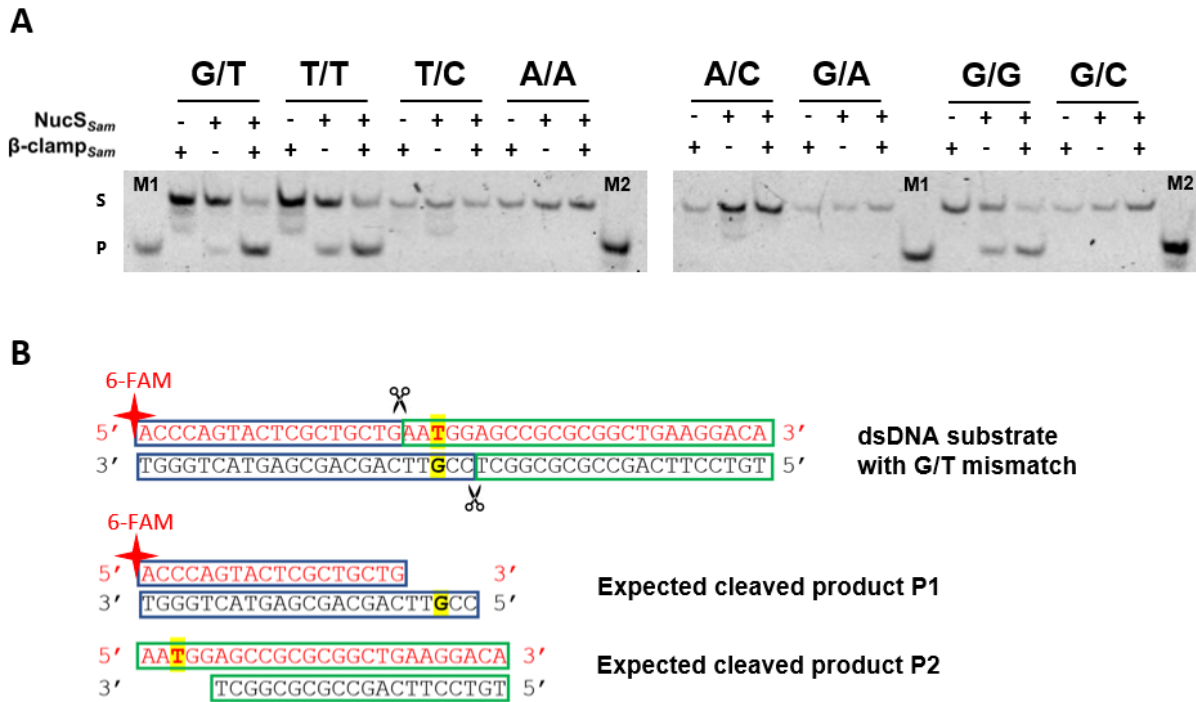
*6FAM labeled oligonucleotide at the 5'-end



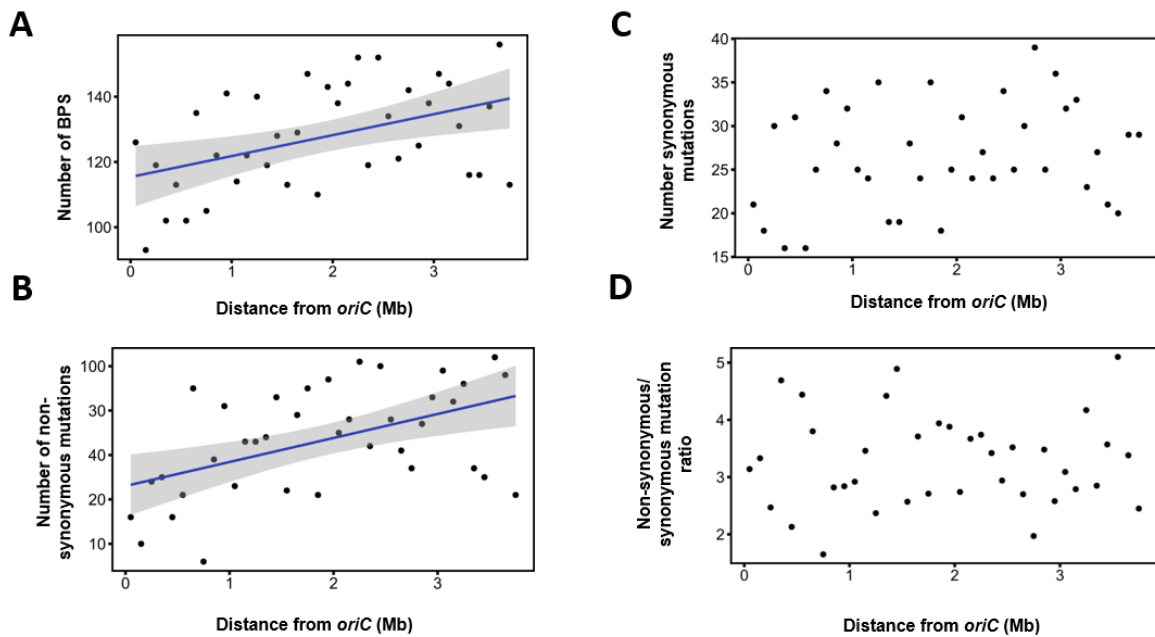
Supplementary Figure S1. Conservation of NucS protein through actinobacteria. Alignment of NucS protein sequences from different species of *Streptomyces* and *Mycobacterium*. *S. ambofaciens* ATCC 28777 (Sam), *S. ambofaciens* DSM 40697 (Sam*), *S. coelicolor* A3(2) (Sco), *S. scabiei* NCPPB 4086 (Ssc), *S. avermitilis* MA 4680 (Sav), *S. xiamenensis* 318 (Sxi), *S. marincola* strain SCSIO 03032 (Sma), *C. glutamicum* ATCC 13032 (Cgl), *M. smegmatis* mc² 155 (Msm), *M. tuberculosis* H38rv (Mtb). Black, dark grey, light grey and blank boxed residues correspond to 100%, 80-100%, 60-80% and less than 60% sequences presenting the same residue. Colored boxes under the alignment indicate protein domains according to *Pyrococcus abyssi* NucS structure (Ren et al., 2009). DNA-binding and catalytic domains are shown in brown and blue respectively. PIP-box motif is represented in red. Essential residues for nuclease activity are shown in purple.



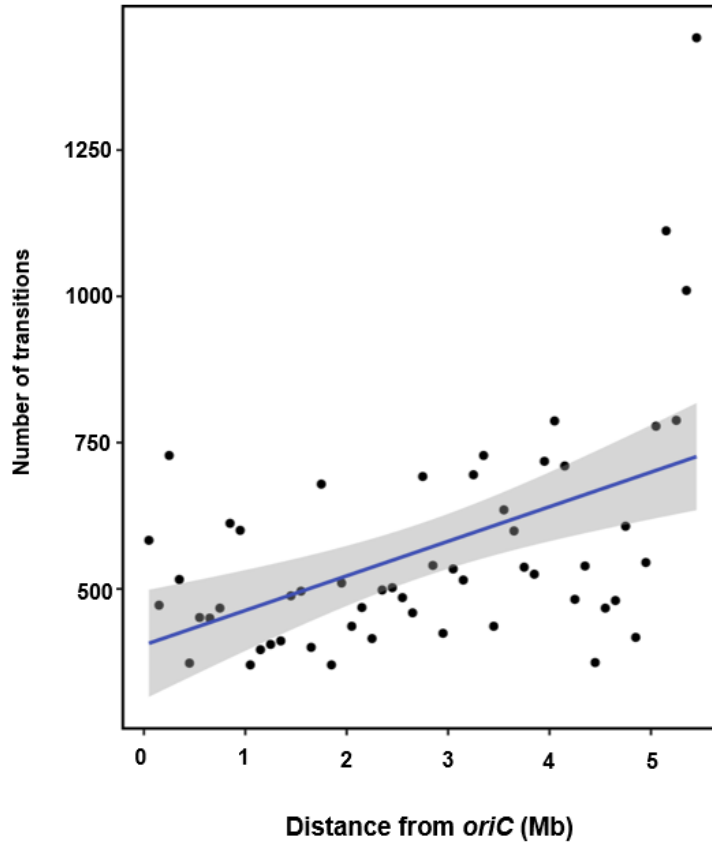
Supplementary Figure S2. Purification and oligomeric states of NucS_{Sam} and β -clamp_{Sam}. Gel filtration profile and SEC-MALS analysis (left and right panels, respectively) of (A) NucS_{Sam} and (B) β -clamp_{Sam}. For gel filtration, the samples were loaded onto the Superdex 200 16/600 column. Number in brackets correspond to estimated molecular weight (kDa) using calibration curve. Aliquots of the 12 elution fractions from 63 to 87 mL for NucS_{Sam} and the 6 elution fractions from 67 to 79 mL for β -clamp_{Sam} were subjected to 12% SDS-PAGE (lower part). M is the protein ladder "Precision Plus Protein Standard" (Biorad). The SEC-MALS chromatograms display the UV absorbance at 280 nm and red lines indicate the molar mass distribution. Because of the reasonable homogeneity of NucS_{Sam} peaks, the molecular weight was estimated at the right extremity of each peak.



Supplementary Figure S3. Estimation of the size of the cleaved products. (A) *In vitro* cleavage assays were performed with 6-FAM-labeled dsDNA substrates (43-bp, 50 nM) containing no mismatch (control) or single base-pair mismatches (A/A, A/C, C/C, G/A, G/G, G/T, T/C or T/T). Each substrate was incubated for 60 min at 30°C, either with 2.4 μM of NucS_{Sam} or with 1.2 μM of β-clamp_{Sam} or with both proteins at the above concentrations. Products were separated by 10% native PAGE. Substrates and cleavage products are indicated by “S” and “P” letters on the left side of the panel. M1 and M2 DNA markers (see also supplementary Table 2) were used to estimate the products size. (B) DNA products expected after cleavage of a dsDNA substrate containing a G/T mismatch. For G/T substrate, it is the upper strand that is 6-FAM-labeled, therefore only P1 expected product will be detected. Depending on the 6-FAM labeled oligonucleotide used to create the substrate, 2 different DNA markers were used : M1 marker was used to estimate the size of P1 product for G/T, T/T, T/C, A/A, A/C and T/H mismatches while M2 marker was used to estimate the size of P2 product for G/G, G/A, C/C, and G/U mismatches. M1 and M2 markers have been designed with the same composition as P1 and P2 products.



Supplementary Figure S4. Distribution of mutations along the chromosome in $\Delta nucS$ lines. Base pair substitutions (BPSs), non-synonymous and synonymous mutations within coding sequences, were counted within a non-overlapping 100 kb window, starting from *dnaA* gene (located in the middle of the chromosome at position 4,021,374 and approximately corresponding to the origin of replication, *oriC*) and sliding towards the chromosome extremities. The number of these mutations in windows of the right replichore were added to the number in windows at the same distance of *dnaA* in the left replichore. The sum of BPSs (A), non-synonymous mutations (B), or synonymous mutations (C) is represented as a function of the distance from *dnaA*. The ratio of non-synonymous to synonymous mutations (dN/dS) was calculated for each 100 kb window and plotted as a function of the genomic position (D). Significant positive correlations were observed between the genomic position and the BPS count (Pearson's correlation coefficient $r=0.451$, $P=0.004$), or between the genomic position and the non-synonymous mutations (Pearson's correlation coefficient $r=0.499$, $P=0.001$). No correlation was noted for synonymous mutations or for dN/dS ratio.



Supplementary Figure S5. Transition distribution along the chromosome of *Streptomyces* environmental strain RLB1-8 compared to *Streptomyces* environmental strain RLB3-17. Transitions were counted within a non-overlapping 100 kb window starting from *dnaA* gene (corresponding to *oriC*, approximately at position 6,204,383 in RLB1-8). The number of transitions in windows of the right replichore was added to the number of transitions of windows at the same distance of *dnaA* in the left replichore, and represented as a function of the distance from *dnaA* gene. A significant positive correlation was observed between the transition count and the genomic position (Kendall rank correlation test $r=0,33$ $P=0,00045$).

Supplementary Table S3. Distribution of BPSs in the replichores

| | WT | | <i>ΔnucS</i> | |
|----------------------|-----------------|------------------|-----------------|------------------|
| | Left Replichore | Right Replichore | Left Replichore | Right Replichore |
| Transitions | 30 | 43 | 2357 | 2551 |
| A:T>G:C | 11 | 21 | 1424 | 1562 |
| A -> G | 3 | 14 | 393 | 1150 |
| T -> C | 8 | 7 | 1031 | 412 |
| G:C>A:T | 19 | 22 | 933 | 989 |
| G -> A | 16 | 12 | 699 | 274 |
| C -> T | 3 | 10 | 234 | 715 |
| Transversions | 36 | 38 | 47 | 93 |
| A:T>T:A | 1 | 4 | 1 | 11 |
| A -> T | 1 | 2 | 0 | 7 |
| T -> A | 0 | 2 | 1 | 4 |
| A:T>C:G | 2 | 4 | 8 | 12 |
| A -> C | 1 | 1 | 7 | 2 |
| T -> G | 1 | 3 | 1 | 10 |
| G:C>T:A | 12 | 6 | 14 | 18 |
| G -> T | 11 | 2 | 10 | 5 |
| C -> A | 1 | 4 | 4 | 13 |
| G:C>C:G | 21 | 24 | 24 | 52 |
| G -> C | 11 | 2 | 9 | 29 |
| C -> G | 10 | 22 | 15 | 23 |
| Total | 66 | 81 | 2404 | 2644 |

2. Résultats complémentaires

A la suite de la démonstration *in vitro* du clivage double-brin de NucS_{Sam} sur les mésappariements, nous avons voulu aller plus loin en déterminant le site de clivage de la protéine. Il existe dans la littérature des divergences quant au site de clivage de NucS de *C. glutamicum*. En effet, les travaux de Ishino et al. (2018) ont montré que NucS de *C. glutamicum* clive 2 nucléotides en 5' de part et d'autre du mésappariement générant ainsi des extrémités cohésives 5' sortantes contenant 5 nucléotides libres alors que Takemoto et ses collaborateurs (2018) ont mis en évidence que la protéine produit des extrémités cohésives 5' sortantes à 3 nucléotides (Ishino et al., 2018; Takemoto et al., 2018). Face à ces résultats discordants, nous avons cherché à déterminer *in vitro* le site de clivage de NucS_{Sam} au niveau de ses substrats, en analysant la taille des produits de clivage par PAGE en conditions dénaturantes (Urée 8M-15% PAGE dans tampon TBE 89 mM Tris, 89 mM borate and EDTA 2,5 mM, pH 8,3). Cette technique permet d'estimer la taille des produits sous leur forme simple brin, à l'aide de marqueurs de taille appropriés. Pour ce faire, les substrats double-brin contenant les mésappariements G/G, G/T et T/T ont été utilisés car ce sont les seuls qui sont clivés par NucS_{Sam}. Seul un des deux oligonucléotides formant le substrat d'ADN double-brin est marqué en 5' (**Figure 12A**). Pour définir la taille des produits de clivage, cinq oligonucléotides (notés M17 à M21) marqués au 6-FAM en 5' dont la taille varie de 17 à 21 nucléotides ont été utilisés comme marqueurs (**Figure 12B**). Cet ensemble de marqueurs devait nous permettre d'estimer la taille des produits simple-brin attendus selon les résultats précédents sur *T. kodakarensis* et *C. glutamicum* (Ishino et al., 2016 ; Ishino et al., 2018) à -2 et +2 nucléotides du mésappariement. En effet, étant donné la position du mésappariement dans les substrats que nous avons utilisés, nous nous attendons à un produit de 18 b issu du brin supérieur et un produit de 20 b issu du brin inférieur (**Figure 12A**).

Malheureusement, il nous a été impossible d'utiliser ces 5 marqueurs pour tous les produits de clivage. En effet, la **Figure 13** montre que la distance de migration des marqueurs ne correspond pas à celle attendue, le marqueur M19 par exemple, migrant plus loin que le marqueur M17. Or il s'avère que M17 et M18 n'ont pas la même composition nucléotidique que M19, M20 et M21 (**Figure 12B**). Ainsi nous avons pu observer que la migration des oligonucléotides en conditions dénaturantes dépendait non seulement de leur taille mais également de leur composition en nucléotides, particularité qui a déjà été décrite (Tang et al., 2015). Dès lors, les marqueurs M17 et M18 n'ont été utilisés que pour estimer la taille des produits de clivage issus de substrats marqués sur le brin supérieur (G/T, T/T), et les marqueurs M19, M20 et M21 n'ont été utilisés que pour estimer la taille du produit de clivage issus de substrats marqués sur le brin inférieur (G/G).

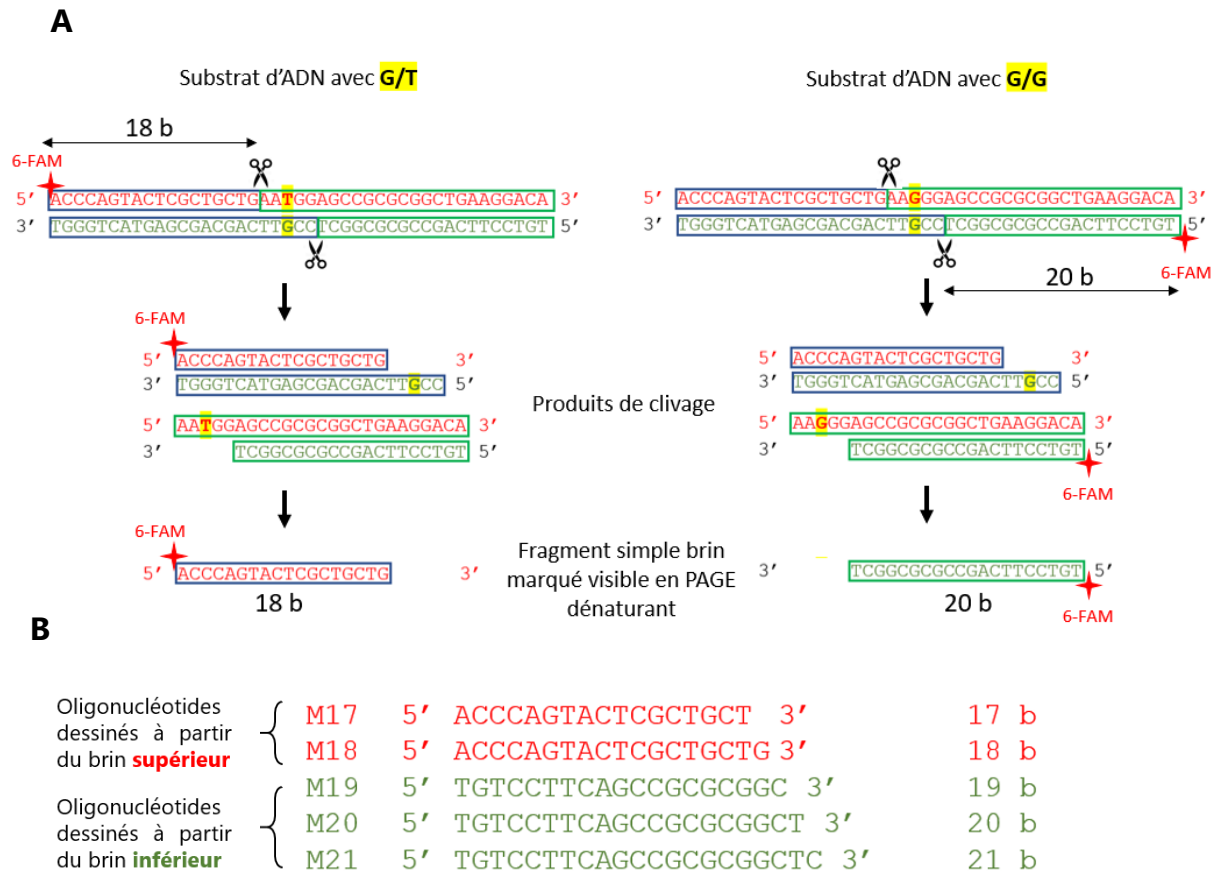


Figure 12. Estimation de la taille des produits de clivage au cours de la réaction de clivage de NucS sur les substrats d'ADN double-brin contenant un mésappariement. (A) Exemples de produits de clivage simple-brin marqués attendus après clivage de substrats d'ADN double-brin contenant un mésappariement G/T ou G/G. Le tableau des différentes combinaisons d'oligonucléotides utilisés pour la réaction de clivage pour tous les mésappariements possibles se trouve en section « supplementary data » de l'article (Table S3). Dans l'exemple ci-dessus, le brin marqué au 6-FAM est le brin supérieur (rouge) pour le substrat G/T et le brin inférieur (vert) pour le substrat G/G. Les produits de clivage double-brin obtenus représentés sont ceux attendus selon Ishino *et al.* (2016 et 2018). En PAGE dénaturant, seuls les brins marqués seront révélés. **(B)** Marqueurs de taille utilisés dans cette étude pour estimer la taille des produits de clivage simple-brin marqués. Cinq oligonucléotides de 17, 18, 19, 20 et 21 b ont été dessinés à partir de la composition du brin supérieur (rouge) ou du brin inférieur (vert). Ces cinq oligonucléotides ont été marqués au 6-FAM en 5' de la molécule.

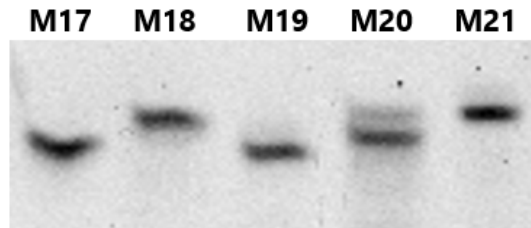


Figure 13. Comparaison de la distance de migration en PAGE dénaturant des marqueurs utilisés pour estimer la taille des produits de clivage de NucS. M17, M18, M19, M20 et M21 sont des oligonucléotides simple-brin composés de 17, 18, 19, 20 et 21 nucléotides, marqués par 6-FAM à leur extrémité 5'. Ils ont été séparés sur un PAGE dénaturant (Urée 8M, 15% polyacrylamide).

Après réaction de clivage en présence de $2,4 \mu\text{M}$ de NucS_{Sam} et $1,2 \mu\text{M}$ β -clamp_{Sam}, dans les conditions décrites dans l'article, puis séparation des produits de clivage en PAGE dénaturant (urée 8M), les résultats obtenus sont illustrés sur la **Figure 14**. Les produits simple-brin marqués issus du clivage des substrats portant les mésappariements G/T ou T/T semblent être de la même taille que le marqueur de 18 b. Ainsi, comme illustré en **Figure 12A**, le produit simple-brin issu de brin supérieur du substrat serait de la taille attendue si NucS_{Sam} coupe 2 nucléotides en 5' du mésappariement. De la même façon, le produit simple-brin marqué issu du clivage du substrat portant le mésappariement G/G semble migrer à la même distance que le marqueur de 20 b, ce qui correspondrait à un produit simple-brin issu du brin inférieur du substrat après coupure à 2 nucléotides en 5' du mésappariement sur ce brin. Ces résultats semblent donc correspondre à ceux des travaux d'Ishino *et al.* en 2016 et en 2018. Néanmoins les répliquats réalisés n'ont pas permis d'être plus résolutifs et cette expérience est à réitérer avec plus de marqueurs adaptés à chaque produit, sur un gel dénaturant plus grand.

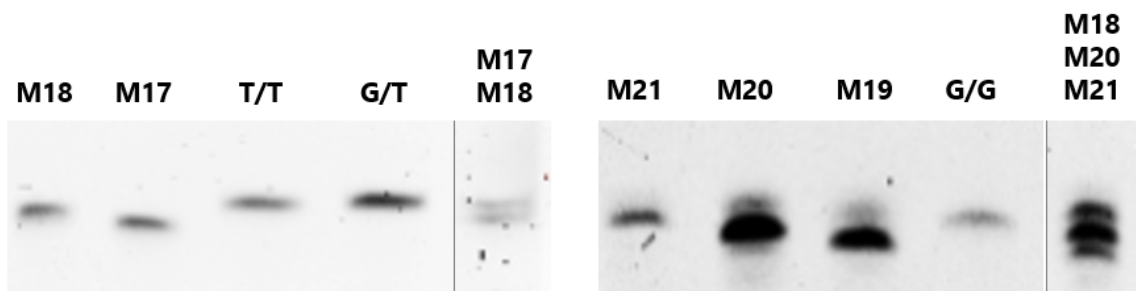


Figure 14. Estimation la taille des produits de clivage de NucS. Un substrat d'ADN double-brin (43 pb) marqué au 6-FAM et contenant un mésappariement de type G/T, T/T, ou G/G a été incubé avec $2,4 \mu\text{M}$ de NucS_{Sam} et $1,2 \mu\text{M}$ β -clamp_{Sam}. Les produits de clivage ont été séparés sur un PAGE 15% dénaturant (Urée 8M). Les marqueurs M17 et M18, de taille 17 b et 18 b respectivement sont utilisés pour estimer la taille des produits de clivage issus de substrats portant un mésappariement G/T ou T/T, tandis que les marqueurs M19, M20 et M21 de 19, 20 et 21 nucléotides chacun, sont utilisés pour estimer la taille du produit de clivage issu du substrat portant un mésappariement G/G.

II. Chapitre 2 : Lien entre l'activité de NucS et la plasticité du génome de *S. ambofaciens*

Dans le précédent chapitre, nous avons mis en évidence le rôle de NucS dans le MMR non-canonique chez *Streptomyces*. En effet, nos expériences ont révélé que NucS est capable, *in vitro*, d'induire une DSB en clivant des substrats d'ADN double-brin portant les mésappariements G/T, T/T et G/G. Ces résultats corroborent nos données *in vivo*. En l'absence de NucS, les mutations de type transitions s'accumulent davantage que les transversions. Par ailleurs, l'analyse de la distribution des SNP le long du chromosome a mis en évidence une augmentation progressive de ceux-ci vers les extrémités, suggérant une augmentation de la fréquence des DSB vers les extrémités. Le (ou les) mécanisme(s) impliqué(s) dans la réparation de ces DSB demeure(nt) encore indéterminé(s) : ces cassures seraient-elles prises charge par d'autres acteurs encore non identifiés de ce nouveau mécanisme de MMR (des partenaires spécifiques qui prendraient le relais après l'action de NucS), ou bien par les mécanismes de réparation des DSB que sont la recombinaison homologe et le NHEJ ? Si au moins une partie de la réparation des DSB est prise en charge par les mécanismes de recombinaison, on peut s'interroger sur le lien qu'il y aurait entre l'augmentation de la fréquence de mésappariements vers les extrémités du chromosome et l'augmentation des indels mise en évidence par différentes approches de génomique comparée dans les études antérieures (Choulet *et al.*, 2006b, Tidjani *et al.*, 2019 et Lorenzi *et al.*, 2021).

1. Création d'un contexte génétique augmentant la fréquence de mésappariements

La question du lien entre les DSB médiées par NucS et la plasticité du génome de *Streptomyces* est posée. Nos analyses de MA ont montré que le taux de mutation chez le mutant $\Delta nucS$ s'élève à $9,22 \times 10^{-7}$ mutations par cycle. Ainsi, le nombre de substrats potentiels de NucS, et donc de DSB, apparaît faible pour observer des phénomènes de remaniements génomiques. La construction d'un contexte génétique présentant un taux accru de mésappariements a donc été entrepris afin de faciliter l'observation d'éventuels réarrangements chromosomiques et donc d'appréhender le rôle de NucS dans la plasticité chromosomique de *Streptomyces*.

Une approche afin d'augmenter la fréquence des mésappariements dans l'ADN est d'utiliser des analogues de bases (Freese, 1959). Ces molécules chimiques possèdent une structure similaire aux bases azotées retrouvées naturellement dans l'ADN et peuvent, ainsi, être incorporées lors de la réplication. Les analogues de bases sont des agents mutagènes car ils induisent un appariement incorrect. Par exemple, le 5-bromouracil, un analogue structural de la thymine, s'apparie non seulement à l'adénine

mais aussi à la guanine (par tautomérisation, (Yu *et al.*, 1993)), conduisant, ainsi, à la mutation de type transition AT vers GC.

Une approche alternative consiste à créer un contexte génétique modifié pour l'activité de relecture d'épreuve (*proofreading*) de l'ADN polymérase, c'est-à-dire son activité exonucléase 3'-5'. Dans ce cas, les incorporations erronées générées par l'activité d'édition de la polymérase ne sont plus corrigées et engendrent des substitutions (de tous types) par réplication. C'est cette seconde approche qui a été mise en œuvre.

1.1. Ciblage de l'activité de relecture de l'ADN polymérase

L'activité de relecture d'épreuve (exonucléase 3'-5') de l'ADN polymérase III dans l'organisme modèle *E. coli* est détenue par la sous-unité ϵ , codée par le gène *dnaQ*. Chez d'autres bactéries telles que *Mycobacterium*, l'activité *proofreading* est assurée, non pas par DnaQ, mais par le domaine PHP (Polymerase and histidinol phosphatase) de DnaE qui correspond à la sous-unité α de l'ADN polymérase III (Rock *et al.*, 2015). Chez *S. lividans* TK24, il existe deux gènes de taille similaire (3540 bp vs 3558 pb) dont les produits présentent une identité de séquence avec DnaE (il s'agit des gènes *dnaE1* et *dnaE2*). Récemment, Sekurova et ses collègues (2021) ont mis en évidence par RNAseq que *dnaE1* est 90 fois plus exprimé que *dnaE2* suggérant que DnaE1 serait probablement l'acteur majeur la réplication et de l'activité de relecture qui lui est associée chez *S. lividans* TK24 (Sekurova *et al.*, 2021). Les études de mutagenèse dirigée montrent que la substitution du 44ème résidu (histidine) de DnaE1 en phénylalanine (H44F) engendre un phénotype mutateur avec une fréquence de mutation 500 fois supérieure à celle du WT.

Chez *S. ambofaciens* comme chez *S. lividans*, deux gènes dont les produits présentant une identité de séquence avec DnaE ont été identifiés : SAM23877_2151 et SAM23877_6470. L'alignement de la séquence protéique des produits de ces deux gènes révèle que les 9 résidus catalytiques importants du domaine PHP sont conservés (**Figure 15**).

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------------------------|-----|----|----|--------|----|-----|-----|----|---|---|---|---|----|---|---|---|---|---|-----|---|---|---|---|---|-----|---|---|---|---|---|---|---|---|---|---|---|-----|-----|---|---|---|---|---|
| <i>S. ambofaciens</i> | THL | HV | H | TQYSLL | D | GAA | MSD | H | G | N | L | I | G | I | E | A | Y | V | GYT | H | K | T | M | V | A | S | T | G | C | P | S | G | V | T | N | D | S | H | Y | | | | |
| <i>S. lividans</i> | THL | HV | H | TQYSLL | D | GAA | MSD | H | G | N | L | I | G | I | E | A | Y | V | GYT | H | K | T | M | V | A | S | T | G | C | P | S | G | V | T | N | D | S | H | Y | | | | |
| <i>C. glutamicum</i> | VHL | HN | H | TEFSML | D | GMA | I | T | D | H | G | N | M | I | G | I | E | T | Y | M | A | Y | L | H | Q | T | M | I | A | T | T | G | C | P | S | G | V | T | N | D | C | H | Y |
| <i>M. smegmatis</i> | VHL | HN | H | TEYSML | D | GAA | M | T | D | H | G | N | M | I | G | I | E | A | Y | I | A | Y | T | H | M | T | M | I | A | T | T | G | C | P | S | G | A | T | N | D | C | H | Y |
| <i>M. tuberculosis</i> | VHL | HN | H | TEYSML | D | GAA | M | T | D | H | G | N | M | I | G | V | E | A | Y | I | S | Y | T | H | L | T | M | I | I | T | T | G | C | P | S | G | A | T | N | D | C | H | Y |
| | | 10 | 12 | | 19 | | | 44 | | | | | 69 | | | | | | 103 | | | | | | 154 | | | | | | | | | | | | 222 | 224 | | | | | |

Figure 15. Alignement de la séquence protéique de DnaE de différentes actinobactéries. Les résidus essentiels à la coordination métallique (Rock *et al.*, 2015) sont surlignés en roses. La numérotation des résidus est relative à la séquence DnaE (SAM23877_2151) de *S. ambofaciens*.

Néanmoins, le produit de SAM23877_6470 est tronqué de la partie C-terminale et est de ce fait dépourvu du domaine OB-fold qui est impliqué dans la liaison de l'ADN simple-brin. Pour cette raison, nous avons poursuivi nos investigations avec le gène SAM23877_2151 que nous appellerons *dnaE* dans la suite du manuscrit.

Sur la base de ces connaissances, nous avons engagé l'inactivation de l'activité *proofreading* de l'ADN polymérase par mutagenèse dirigée en ciblant le résidu catalytique histidine 44 du domaine PHP de DnaE de *S. ambofaciens*.

➤ Utilisation de la technologie CRISPR-cBEST

Afin d'affecter l'activité *proofreading* de l'ADN polymérase DnaE, nous avons mis en œuvre l'édition de génome *via* la récente technologie CRISPR-Base Edition SysTem (CRISPR-BEST) (Tong *et al.*, 2019, 2020) en ciblant le domaine PHP de DnaE de *S. ambofaciens*. Les technologies d'édition de base médiées par CRISPR se sont développées ces dernières années ; elles sont aujourd'hui des outils puissants d'ingénierie génétique permettant une mutagenèse ciblée efficace et précise au nucléotide près. Le principe des outils d'édition de base repose sur la fusion d'une variante de l'endonucléase Cas9 avec une désaminase qui procède à la déamination de la base cible, conduisant à la substitution nucléotidique recherchée. Deux variantes du système CRISPR-BEST ont été développées pour *Streptomyces* : CRISPR-aBEST qui permet la conversion de l'adénine cible en guanine tandis que CRISPR-cBEST convertit la cytosine cible en thymine (Tong *et al.*, 2019). Cette spécificité est obtenue en ciblant une activité désaminase (adénosine désaminase ou cytosine désaminase) sur la séquence visée grâce à l'ARN guide (ADNg) et l'activité Cas9n (la Cas9n réalise une cassure simple-brin). Par déamination, ce système induit une transition adénosine vers guanine (A vers G) par CRISPR-aBEST, et cytosine vers thymine (C vers T) par CRISPR-cBEST. L'induction de cassures double-brin dans l'ADN par la Cas9 est abolie par mutation : Cas9n(D10A). Ce mécanisme présente l'avantage considérable de ne pas créer de DSB contrairement aux technologies CRISPR/Cas9 classiques. Les bactéries du genre *Streptomyces* sont connues pour leur instabilité génétique élevée associée à la réparation des DSB particulièrement. Par conséquent, la génération de DSB est à éviter lors des manipulations génétiques.

Afin de cibler le résidu catalytique histidine 44 du domaine PHP de DnaE de *S. ambofaciens*, nous avons utilisé le CRISPR-cBEST car il présente une plus grande efficacité comparé au CRISPR-aBEST (Tong *et al.*, 2019). A l'aide de l'outil CRISPR-cBEST, l'histidine 44 est modifiable en tyrosine. Pour ce faire, la cytosine en position 130 est substituée par une thymine ; ce qui mute le codon « CAC » codant pour l'histidine en un codon « TAC » codant pour la tyrosine. En somme, nous avons cherché à obtenir la substitution H44Y dans DnaE. Pour atteindre cet objectif, le *protospacer* ciblant la cytosine en position 130 de *dnaE* a été inséré dans le plasmide pCRISPR-cBEST natif, créant ainsi la construction appelée pCRISPR-

cBEST_ *dnaE*_{H44Y}. Au total, 48 transconjugants ont été comptabilisés suite au transfert du plasmide pCRISPR-cBEST_ *dnaE*_{H44Y} dans *S. ambofaciens* par conjugaison intergénérique. Parmi les 48 lignées, 32 ont été séquencées afin de vérifier la substitution de la cytosine 130 en thymine. Les résultats montrent que 15 des 32 lignées analysées présentent la mutation non ciblée C > T à la position 129 générant la mutation silencieuse GAC > GAT (l'acide aspartique en position 43 reste inchangé). Bien qu'aucune lignée ne présente exclusivement la substitution H44Y parmi les 32 analysées, nous avons identifié 5 lignées contenant à la fois la mutation recherchée conduisant la substitution H44Y et la mutation silencieuse C > T à la position 129. Ces lignées, désignées sous les noms *dnaE*_{H44Y_1}, *dnaE*_{H44Y_2}, *dnaE*_{H44Y_3}, *dnaE*_{H44Y_4}, *dnaE*_{H44Y_5}, ont été considérées comme portant une mutation silencieuse sans conséquences pour les investigations ultérieures. Par conséquent, nous avons entrepris de caractériser ces lignées *dnaE*_{H44Y}.

1.2. La mutation *dnaE*_{H44Y} confère un phénotype mutateur

Afin d'évaluer l'impact de la substitution H44Y dans DnaE sur la mutabilité de *S. ambofaciens*, nous avons réalisé des tests de fluctuations en suivant la méthode de Luria & Delbrück (1943), comme cela a été réalisé sur les mutants $\Delta nucS$ (chapitre 1). Nous avons mesuré l'apparition spontanée de résistants à la rifampicine comme indicateur du taux de mutation spontanée de la souche sauvage *S. ambofaciens* ATCC 23877 (WT) ainsi que de 3 mutants portant la substitution H44Y dans DnaE : *dnaE*_{H44Y_1}, *dnaE*_{H44Y_2} et *dnaE*_{H44Y_3}. Le taux de mutation spontanée chez la souche WT est de 1.24×10^{-8} tandis que les mutants *dnaE*_{H44Y_1}, *dnaE*_{H44Y_2} et *dnaE*_{H44Y_3} présentent des taux de mutation considérablement plus élevés, soit 2.74×10^{-6} , 2.16×10^{-6} et 1.45×10^{-6} respectivement (**Figure 16**). Ces résultats révèlent que la substitution H44Y dans DnaE est associée à une augmentation d'environ 100 fois du taux de mutation chez *S. ambofaciens*. Nous avons montré que la délétion de *nucS* provoque un accroissement d'un facteur 130 du taux de mutation (chapitre 1). Ainsi, la substitution H44Y dans DnaE et la délétion de *nucS* semblent affecter le taux de mutation chez *S. ambofaciens* de manière comparable. En comparaison, la mutation du domaine PHP de *M. tuberculosis* accroît le taux de mutation d'un facteur 3000 (Rock *et al.*, 2015).

En somme, les souches *dnaE*_{H44Y} présentent un phénotype mutateur qui semble être lié à la perturbation de l'activité *proofreading*. Il est donc raisonnable de supposer que la fréquence d'apparition de mésappariements substrats de NucS est augmentée dans ces mutants *dnaE*_{H44Y}. Pour cette raison, nous

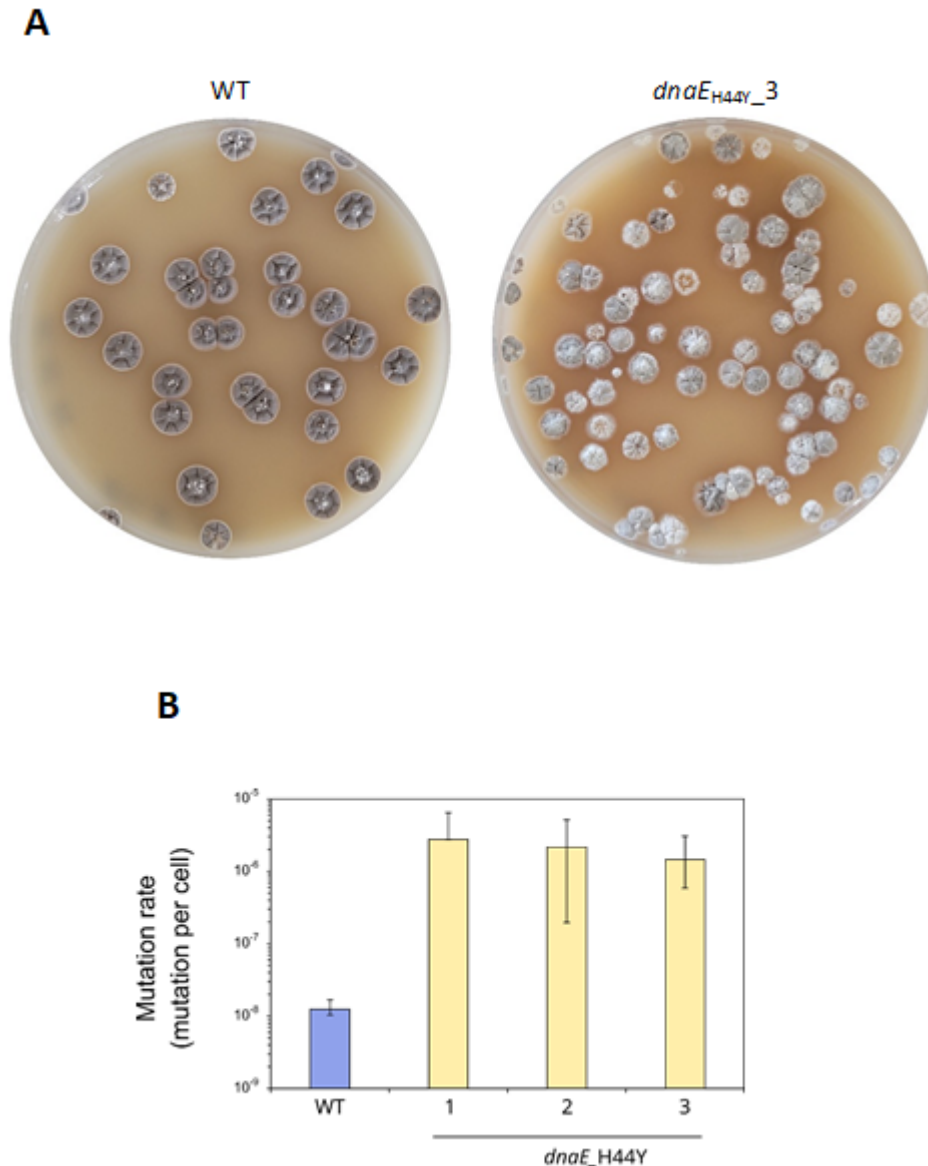


Figure 16. Phénotype mutateur des 3 souches *dnaE_{H44Y}* chez *S. ambofaciens*. La substitution H44Y dans l'ADN polymérase DnaE a été réalisée par CRISPR-cBEST (Tong *et al.*, 2019). **(A)** Hétérogénéité phénotypique sur SFM de la lignée *dnaE_{H44Y}_3* comparée au WT. **(B)** Les taux de mutations spontanées conférant la résistance à la rifampicine ont été estimés par des tests de fluctuations (Luria and Delbrück, 1943). Les barres d'erreurs représentent les intervalles de confiance à 95%.

avons choisi de les utiliser comme un fond génétique afin d'investiguer la relation entre l'activité de NucS et la dynamique du chromosome.

1.3. Instabilité génomique de l'hypermutateur *dnaE_{H44Y}*

L'augmentation du taux de mutation dans les souches *dnaE_{H44Y}* suggère une fréquence accrue d'apparition des mésappariements, qui sont les substrats de NucS. Afin de tester si les DSB NucS-dépendantes pourraient être impliquées dans la plasticité du génome de *Streptomyces*, nous avons

entrepris d'explorer la présence potentielle de remaniements chromosomiques dans la descendance de ces mutants. Pour cela, le génome de descendants de ces lignées *dnaE_{H44Y}* a été examiné par l'analyse de cartes de macro-restriction obtenues grâce à la séparation de larges fragments d'ADN par électrophorèse en champ pulsé (PFGE) (Schwartz and Cantor, 1984; Lopez-Canovas *et al.*, 2019). Lors d'une électrophorèse, les fragments de petite taille migrent plus vite que les grands fragments en raison de leur moindre rétention dans la matrice. Toutefois, au-delà d'une certaine taille (30 kb), tous les fragments, quelle que soit leur taille, migrent à la même vitesse. La PFGE résout ce problème en introduisant un paramètre essentiel : la modification de la direction du champ électrique. En effet, le changement périodique de la direction du champ perturbe la migration des fragments d'ADN et notamment celle des grands fragments. Ces derniers mettent plus de temps à se réorienter dans la direction opposée, tandis que les fragments de moindre taille réagissent plus rapidement au changement de direction du champ. La PFGE tire profit de cette différence de migration et offre une résolution élevée de séparation de grands fragments d'ADN. Cette technique fût largement utilisée dans les années 1990 notamment pour le génotypage des souches bactériennes et la cartographie physique des génomes. Le choix d'une enzyme de restriction digérant le chromosome en une dizaine (15 à 20) de fragments d'ADN est essentiel pour l'interprétation des profils génétiques et l'étude des grands réarrangements génomiques.

Le profil de macro-restriction du chromosome de *S. ambofaciens* ATCC23877 par *Asel* est connu (Leblond *et al.*, 1990, 1996), il présente 14 fragments notés de A à N (**Figure 17A**, panel gauche). Afin d'affiner l'analyse des profils génétiques et des remaniements chromosomiques, nous avons utilisé deux conditions électrophorétiques différentes, la première permettant de visualiser l'ensemble du profil (fragments notés A à N), la seconde permettant de mieux discriminer les bandes apparaissant en doublets dans la première condition et focalisant sur la séparation des fragments allant de G à K. Un exemple de gels PFGE est présenté en **Figure 17B**: le gel situé à gauche a migré dans les conditions 1, identiques à celles ayant permis la cartographie génomique de *S. ambofaciens* (Leblond *et al.*, 1990) alors que le gel de droite a migré dans les conditions 2 et il permet de distinguer les doublets H-I et J-K (Leblond *et al.*, 1996). Les deux gels PFGE montrent le profil génétique de 3 individus issus de la population *dnaE_{H44Y}_3* (puits 1 à 3) et d'un descendant du WT. Les gels de PFGE en conditions 1 et 2 sur le WT révèlent l'empreinte génétique de *S. ambofaciens* ATCC23877, telle que documentée dans la littérature (Leblond *et al.*, 1990, 1996). Pour les descendants de la lignée *dnaE_{H44Y}_3*, la PFGE en conditions 1 indique l'absence du fragment G chez l'individu 1, résultat corroboré par la PFGE en conditions 2 qui met en évidence, de surcroît, l'absence du fragment I. Cette observation met en lumière la complémentarité des conditions de migrations dans la détection de grands réarrangements. Pour l'individu 2, les fragments D et G sont manquants, tandis qu'une bande de taille comprise entre 820 kb

et 1100 kb apparaît sur la PFGE en conditions 1. La PFGE en conditions 2 confirme l'absence du fragment G. Enfin, l'individu 3 présente un profil identique à la souche WT, suggérant l'absence de remaniements d'ampleur dans ce clone.

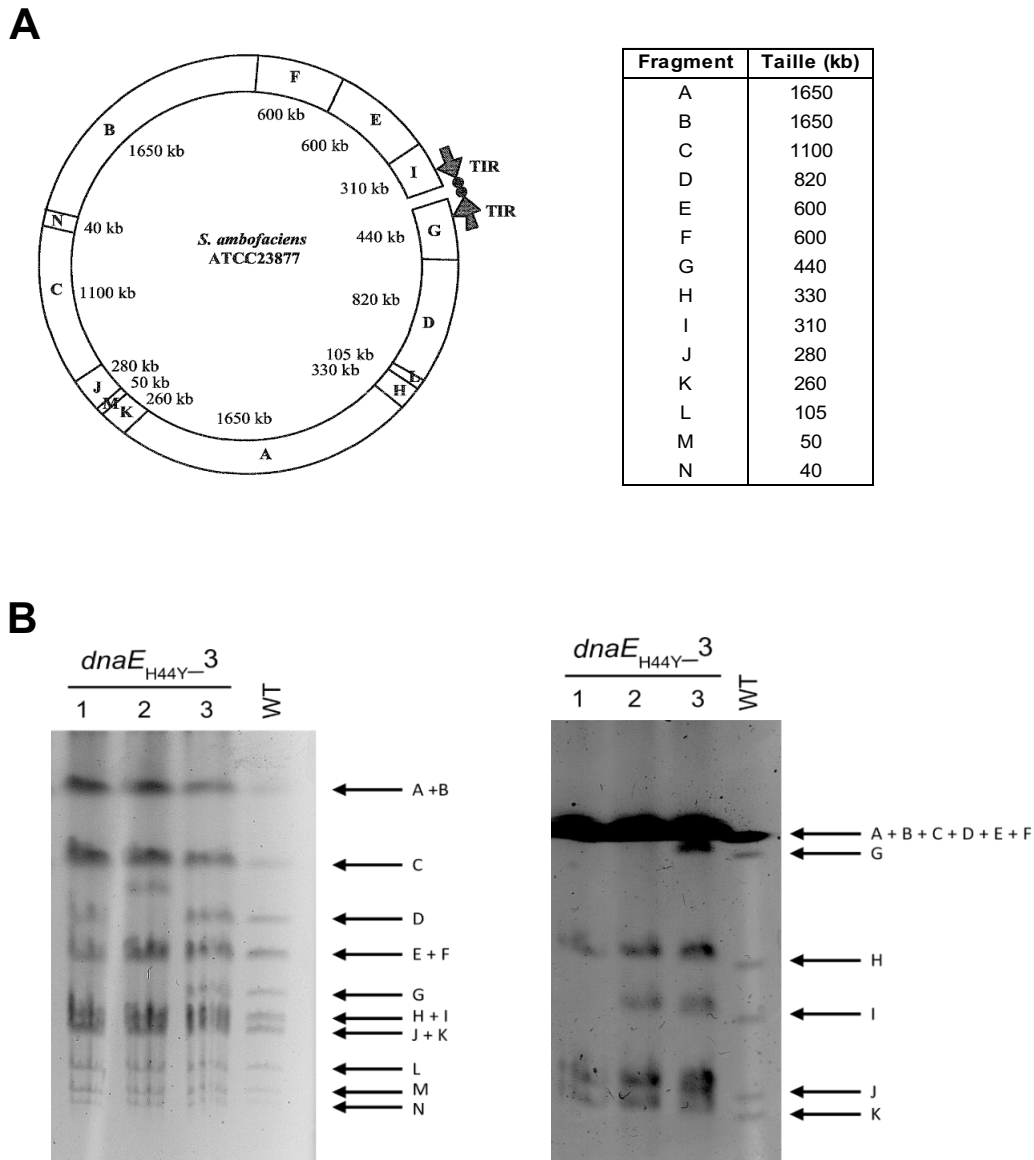


Figure 17. Profil de macro-restriction du chromosome *S. ambofaciens*. (A) Le panel gauche montre la cartographie du chromosome de *S. ambofaciens* ATCC23877 coupé par *Asel*. Le panel droit indique la taille des fragments (dans l'ordre décroissant) générés suite à la restriction du chromosome par *Asel*. (D'après Leblond *et al.*, 1996.). (B) Un exemple de gel PFGE en conditions de migration 1 et 2 (panel gauche et droite, respectivement) est présenté. Trois descendants de la souche *dnaE*_{H44Y-3} et un descendant du WT ont été testés. Les lettres à droite indiquent les fragments d'ADN attendus pour un profil sauvage suite à la restriction du chromosome par *Asel*.

Le **Tableau 1** résume les analyses de PFGE sur la population WT et sur les 3 mutants *dnaE*_{H44Y}. Au total, 34 descendants de la population WT ont été examinés et aucun réarrangement chromosomique majeur

n'a été observé dans ces derniers. En revanche, sur l'ensemble des mutants *dnaE_{H44Y}*, 8 génomes sur 51 analysés étaient remaniés, ce qui est statistiquement différent des résultats obtenus avec la souche WT. Parmi les 3 populations mutantes, *dnaE_{H44Y_3}* se démarque avec une fréquence de remaniements chromosomiques atteignant 33,3% (5 descendants au génome remanié sur 15 testés) qui est statistiquement différente de celle du WT ($p=0.002$, test exact de Fisher). En revanche, les populations *dnaE_{H44Y_1}* et *dnaE_{H44Y_2}* montrent des fréquences de réarrangement respectifs de 8% et 3,7% non significativement différent du WT. Il est nécessaire d'augmenter le nombre de descendants analysés pour ces deux derniers mutants, ainsi que pour la souche WT, afin de confirmer la tendance observée avec des statistiques plus robustes.

Tableau 1. Fréquence d'individus présentant un génome réarrangé dans la descendance du WT et des souches *dnaE_{H44Y}*. Les trois dernières colonnes désignent les 3 lignées *dnaE_{H44Y}* testées. La valeur de P a été obtenue en réalisant un test exact de Fisher comparant les souches *dnaE_{H44Y}* avec le WT.

| | WT | Tous les mutants <i>dnaE_{H44Y}</i> | <i>dnaE_{H44Y}</i> | | |
|---|-----|---|----------------------------|-------|-------|
| | | | 1 | 2 | 3 |
| Nombre total de descendants testés | 34 | 51 | 25 | 27 | 15 |
| Nombre de descendants remaniés | 0 | 8 | 2 | 1 | 5 |
| Nombre de descendants dépourvus du fragment D | 0 | 0 | 0 | 0 | 0 |
| Nombre de descendants dépourvus du fragment G | 0 | 0 | 0 | 0 | 0 |
| Nombre de descendants dépourvus du fragment I | 0 | 1 | 0 | 0 | 1 |
| Nombre de descendants dépourvus du fragment G + D | 0 | 4 | 2 | 0 | 2 |
| Nombre de descendants dépourvus du fragment G + I | 0 | 3 | 0 | 1 | 2 |
| Fréquence de descendants remaniés (%) | 0.0 | 15.7 | 8.0 | 3.7 | 33.3 |
| P | | 0.0193 | 0.175 | 0.443 | 0.002 |

Concernant les régions du chromosome qui ont été remaniées, nous n'avons observé des réarrangements localisés exclusivement aux extrémités du chromosome. Les réarrangements touchent les fragments D, G et I. Ces régions sont connues pour être des points chauds de remaniements chromosomiques comme décrit dans la littérature (Martin *et al.*, 1999). De manière intéressante, parmi les 8 descendants *dnaE_{H44Y}* totaux présentant des réarrangements chromosomiques, la moitié est affectée dans les deux bras du chromosome. En contraste, la fréquence de réarrangements du bras chromosomique droit est 3 fois plus élevée que celle du bras gauche (37,5% *versus* 12,5%). Cette observation est en accord avec une étude antérieure démontrant que l'instabilité génétique se concentre particulièrement sur les deux bras chromosomiques ou le bras droit seulement (Martin *et al.*, 1999).

Au regard de ces données, une tendance semble se dégager : les descendants des mutants *dnaE_{H44Y}* affichent une propension accrue à la réorganisation de leur matériel génétique par rapport aux descendants du WT. Il convient de rappeler que les mutants *dnaE_{H44Y}* sont affectés au niveau de l'activité

exonucléase 3'-5' de l'ADN polymérase DnaE ; ce qui se traduit par l'incapacité de corriger les erreurs d'appariement de bases survenues lors de la réplication. Ces mutants *dnaE_{H44Y}* présentent, de ce fait, un taux de mutations accru se traduisant par une augmentation de la fréquence d'apparition de mésappariement (et donc de substrats) reconnus par NucS. Par conséquent, ces données alimentent l'hypothèse selon laquelle l'augmentation du taux de réarrangement serait liée à l'apparition à haute fréquence de mésappariements et donc de DSB provoquées par NucS. Des investigations complémentaires corrélant directement l'activité de NucS à l'augmentation de la fréquence de réarrangements demeurent nécessaires pour étayer cette hypothèse. Dans cet objectif, nous avons entrepris d'étudier la stabilité génétique des mutants *dnaE_{H44Y}* dans un fond génétique affecté dans l'expression de *nucS*.

2. Implication de NucS dans l'augmentation de la fréquence des réarrangements génomiques

D'après nos précédents résultats, les remaniements sont plus fréquents dans les chromosomes des mutants *dnaE_{H44Y}* relativement à la souche sauvage. Cette propension à la plasticité génomique pourrait trouver source dans l'activité de NucS, qui génère (*in vitro*) une DSB au site d'un mésappariement. Nos travaux sur les mutants *dnaE_{H44Y}* ne permettent pas de lier directement l'augmentation de la fréquence des réarrangements à la protéine NucS. C'est pourquoi, afin de corréler l'activité de NucS avec la plasticité chromosomique accrue des mutants *dnaE_{H44Y}*, nous avons imaginé deux approches génétiques :

- (i) Abolir l'activité de NucS dans une souche mutatrice *dnaE_{H44Y}* en introduisant un codon STOP au début du gène *nucS*, résultant ainsi en une protéine tronquée inactive. En absence de NucS fonctionnelle, nous nous attendons à une baisse de la fréquence des remaniements liée à l'absence des DSB causées par cette dernière.
- (ii) Augmenter l'activité de NucS en surexprimant le gène dans un mutant *dnaE_{H44Y}*. Dans une telle souche, si effectivement l'activité de NucS génère des DSB recombinogènes, lors de la correction des mésappariements, cela se traduira par une augmentation de la fréquence des remaniements et ainsi de la plasticité du génome.

Les deux stratégies ont été testées ; les méthodes ainsi que les résultats associés sont détaillés dans les paragraphes ci-dessous.

2.1. Approche 1 : abolition de l'activité de NucS dans une souche hypermutatrice *dnaE_{H44Y}*

2.1.1. Mutants *nucS_{W49*}* dans le contexte sauvage

Comme dans le chapitre 1 de la partie « Résultats », nous aurions pu abolir l'activité de NucS en créant un mutant de délétion du gène *nucS* dans les mutants *dnaE_{H44Y}*. Cependant, cette méthode est longue et fastidieuse. Nous avons, de ce fait, choisi d'employer la technologie CRISPR-cBEST afin d'introduire un codon de terminaison en début de séquence du gène *nucS* et générer ainsi une protéine tronquée inactive. Le résidu 49 correspondant au tryptophane est le premier résidu modifiable en un codon STOP par CRISPR-cBEST.

Nous avons choisi, en premier lieu, de tester la faisabilité de la méthode sur *nucS* dans le contexte sauvage. Ainsi, l'ARNg permettant la modification W49* a été cloné dans le plasmide pCRISPR-cBEST puis le vecteur recombinant (pCRISPR-cBEST_*nucS*_W49*) a été transféré dans *S. ambifaciens* par conjugaison. Quatre transconjugants ont été analysés et tous portent la mutation générant un codon terminaison à la place du triplet codant pour le tryptophane en position 49. Après curage du plasmide pCRISPR-cBEST_*nucS*_W49*, ces 4 souches (considérés comme 4 mutants indépendants) sont nommées : *nucS_{W49*_1}*, *nucS_{W49*_2}*, *nucS_{W49*_3}*, *nucS_{W49*_4}*.

A l'instar des mutants de délétion *nucS* présentés dans le chapitre 1, les 4 lignées mutantes *nucS_{W49*}* montrent aussi une hétérogénéité phénotypique avec notamment la formation de secteurs, des bords dentelés, une dépigmentation des colonies, et la formation de colonies *bald* (dépourvu de mycélium aérien) (**Figure 18**). Bien que leur taux de mutation n'ait pas été déterminé, ces lignées *nucS_{W49*}* ont un phénotype compatible avec ceux des mutants Δ *nucS*, suggérant qu'elles produisent bien une protéine NucS non fonctionnelle.

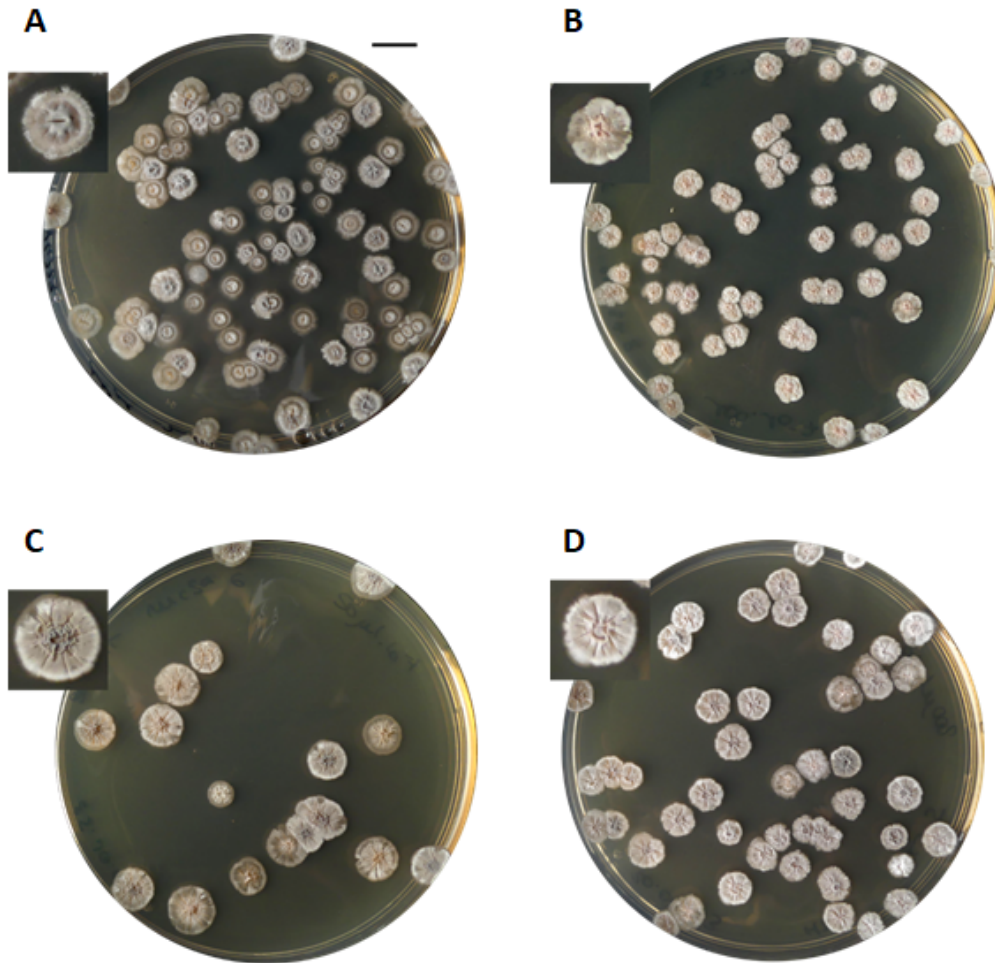


Figure 18. Hétérogénéité coloniale des mutants *nucSW49.** (A-D) Phénotype des colonies issues des lignées *nucSW49*_1*, *nucSW49*_2*, *nucSW49*_3*, *nucSW49*_4*, respectivement, après 7 jours d'incubation à 30°C sur milieu HT. Une colonie représentative de l'hétérogénéité coloniale est représentée en haut à gauche de chaque panel. La barre noire indique 1 cm.

2.1.2. Mutants *nucSW49** dans le contexte mutateur *dnaEH44Y*

La mutation W49* étant possible dans la souche sauvage, nous avons cherché à obtenir le double mutant *dnaEH44Y nucSW49**. Pour cela, le plasmide pCRISPR-cBEST_*nucS_W49** a été introduit dans le mutant *dnaEH44Y_1*. Les résultats de séquençage montrent que seulement 1 transconjugant sur 16 analysés porte la substitution W49* dans *nucS* tandis que 4/4 des transconjugants issus de fond génétique sauvage présentaient la modification cible. Ce résultat semble indiquer que l'abolition à la fois de l'activité de NucS et de l'activité de relecture de la polymérase DnaE affecte la survie des cellules.

Le plasmide pCRISPR-cBEST_*nucS_W49** a été curé et le double mutant fut renommé : *dnaEH44Y_1 nucSW49**. Le phénotype de ce dernier révèle une majorité de colonies *bald*, dépigmentées, de petite taille alors qu'elles ont poussé 7 jours à 30°C sur SFM (Figure 19). Ce phénotype très affecté impacte la suite

des travaux car il faut savoir que les PFGE sont réalisées en inoculant des spores des descendants des souches. Or, les descendants du double mutant *dnaE_{H44Y}_1 nucS_{W49*}* ne sporulent pas (ou très peu). Ainsi, les cultures liquides pour les PFGE ont été souvent infructueuses car aucune croissance n'était observée même après 3-5 jours. Par conséquent, l'analyse du génome de ces doubles mutants n'a pas pu être effectuée. Par ailleurs, au cours des différentes étapes de culture à la suite de l'obtention du double mutant (curage du plasmide pCRISPR-cBEST_*nucS_{W49*}*, mise en culture pour les PFGE), la conservation des deux mutations a été vérifiée par séquençage et des événements de réversions ont pu être mise en évidence (perte de la mutation H44Y dans *dnaE*). De fait, dans ce contexte hypermutateur, les révertants sont plus fréquents, et à cause du phénotype délétère du double mutant, ils sont avantagés. Nous avons ainsi eu des difficultés à stabiliser et à conserver les doubles mutants.

Toutefois, on peut retenir que la double mutation *dnaE_{H44Y} nucS_{W49*}* résulte en un phénotype que l'on pourrait qualifier de « délétère » puisque le taux de survie et la croissance des descendants est grandement affecté. Ce résultat rejoint celui évoqué plus haut lors de l'obtention du double mutant : la suppression concomitante de NucS et de l'activité *proofreading* de l'ADN polymérase DnaE semble conduire à une mortalité accrue des cellules. Cette observation peut être expliquée par une accumulation plus importante de mutations au cours du temps due à l'absence des corrections des mésappariements lors de la réplication.

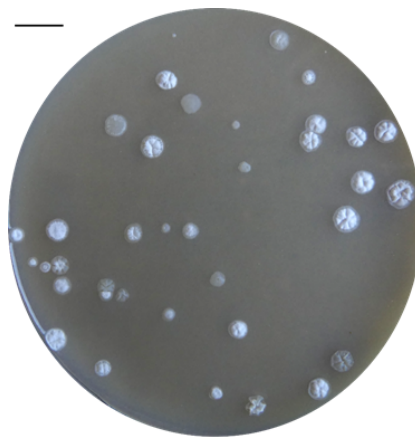


Figure 19. Phénotype colonial du mutant *dnaE_{H44Y}_1 nucS_{W49*}*. Phénotype des colonies issues du mutant *dnaE_{H44Y}_1 nucS_{W49*}* après 7 jours d'incubation à 30°C sur milieu SFM. La barre noire indique 1 cm.

2.2. Approche 2 : surexpression de *nucS* dans un contexte hypermutateur

2.2.1. Stratégie employée

D'après nos précédents résultats, les mutants *dnaE_{H44Y}* sont plus prompts à remanier leur chromosome relativement à la souche sauvage. Nous émettons l'hypothèse selon laquelle la capacité de NucS à produire des DSB au site de mésappariements serait une explication de cette hausse du taux de réarrangements. Afin de tester cette hypothèse, nous avons entrepris de créer un mutant surexprimant *nucS* dans un fond génétique *dnaE_{H44Y}*. Pour cela, nous avons utilisé le plasmide pDyn6902, un dérivé du plasmide pIJ6902 (Huang *et al.*, 2005), qui permet de cloner le gène d'intérêt derrière *PtipA*, un promoteur inductible au thiostrepton. Le vecteur pDyn6902 est intégratif chez *Streptomyces* et présente l'avantage d'être à faible nombre de copies chez *E. coli*, rendant possible le clonage de gènes dont l'expression pourrait être toxique chez cette dernière (Thibessard *et al.*, 2015). En raison du potentiel délétère de la surexpression de *nucS* chez *S. ambofaciens*, nous avons choisi de travailler avec le plasmide pDyn6902 pour pouvoir contrôler son expression et, ainsi, minimiser l'impact sur la viabilité des mutants. Toutefois, il est important de préciser qu'il existe une fuite du promoteur *PtipA* chez *Streptomyces*. En effet, malgré l'absence d'induction au thiostrepton, le gène cloné en aval de ce dernier demeure légèrement transcrit (Ali *et al.*, 2002).

Le gène codant pour *nucS* a été amplifié par PCR avec des amorces contenant le site de restriction de *NdeI* avant d'être ensuite cloné dans pDyn6902 par restriction-ligation. La construction plasmidique « pDyn6902::*nucS* » a été validée par séquençage : aucune mutation n'est présente ni dans le gène *nucS*, ni dans le promoteur *PtipA*. Rappelons que nous avons précédemment caractérisé les 3 lignées mutatrices *dnaE_{H44Y}* suivantes : *dnaE_{H44Y}_1*, *dnaE_{H44Y}_2*, *dnaE_{H44Y}_3*. Dans chaque lignée *dnaE_{H44Y}*, les vecteurs pDyn6902::*nucS* et pDyn6902 ont été introduits par conjugaison intergénérique afin de créer les sur-expressesurs *dnaE_{H44Y}* pDyn6902::*nucS* ainsi que les témoins associés (c'est-à-dire les mutants *dnaE_{H44Y}* contenant le plasmide pDyn6902). Pour chaque conjugaison, deux individus portant le plasmide de surexpression pDyn6902::*nucS* ont été sélectionnés pour la suite des travaux (ils ont été considérés comme des mutants indépendants) alors qu'un seul clone contenant le plasmide pDyn6902

vide (*i.e* souche témoin) a été choisi. Le **Tableau 2** récapitule l'ensemble des souches générées au cours de ce volet « surexpression de *nucS* dans un contexte hypermutateur ».

Tableau 2. Description des souches construites pour la surexpression de *nucS*.

| Nom de la souche | Description |
|--|--|
| <i>dnaE</i> _{H44Y_1} pDyn6902 | Mutant <i>dnaE</i> _{H44Y_1} contenant le plasmide intégratif pDyn6902 |
| <i>dnaE</i> _{H44Y_1} <i>nucS</i> _1 | Lignée 1 issue de l'introduction du plasmide pDyn6902:: <i>nucS</i> dans le mutant <i>dnaE</i> _{H44Y_1} |
| <i>dnaE</i> _{H44Y_1} <i>nucS</i> _2 | Lignée 2 issue de l'introduction du plasmide pDyn6902:: <i>nucS</i> dans le mutant <i>dnaE</i> _{H44Y_1} |
| <i>dnaE</i> _{H44Y_2} pDyn6902 | Mutant <i>dnaE</i> _{H44Y_2} contenant le plasmide intégratif pDyn6902 |
| <i>dnaE</i> _{H44Y_2} <i>nucS</i> _1 | Lignée 1 issue de l'introduction du plasmide pDyn6902:: <i>nucS</i> dans le mutant <i>dnaE</i> _{H44Y_2} |
| <i>dnaE</i> _{H44Y_2} <i>nucS</i> _2 | Lignée 2 issue de l'introduction du plasmide pDyn6902:: <i>nucS</i> dans le mutant <i>dnaE</i> _{H44Y_2} |
| <i>dnaE</i> _{H44Y_3} pDyn6902 | Mutant <i>dnaE</i> _{H44Y_3} contenant le plasmide intégratif pDyn6902 |
| <i>dnaE</i> _{H44Y_3} <i>nucS</i> _1 | Lignée 1 issue de l'introduction du plasmide pDyn6902:: <i>nucS</i> dans le mutant <i>dnaE</i> _{H44Y_3} |
| <i>dnaE</i> _{H44Y_3} <i>nucS</i> _2 | Lignée 2 issue de l'introduction du plasmide pDyn6902:: <i>nucS</i> dans le mutant <i>dnaE</i> _{H44Y_3} |

La surexpression effective du gène *nucS* dans les souches *dnaE*_{H44Y} pDyn6902::*nucS* a été quantifiée par PCR quantitative (qPCR). Au laboratoire, des travaux préliminaires ont permis de déterminer qu'une concentration de 2,5µg/ml de thiostrepton est suffisante pour induire l'expression du gène placé en aval de *PtipA* dans le plasmide pDYN6902. Pour cette raison, l'ARN total a été extrait des souches *dnaE*_{H44Y} pDyn6902 et *dnaE*_{H44Y} pDyn6902::*nucS*, après 48 heures d'incubation à 30°C sur un milieu solide Hickey Tresner (HT) supplémenté en thiostrepton (2.5 µg/mL), ou pas. Les ARN ont été extraits après une durée d'incubation de 2 jours. L'ADN complémentaire (ADNc), obtenu suite au traitement des ARN totaux par une DNase et une transcriptase inverse, a servi de matrice pour amplifier une région dans les gènes *nucS* et *hrdB*. Le gène *hrdB* code pour un facteur σ permettant l'expression de gènes de ménage et a été utilisé comme gène de référence pour la normalisation du niveau de l'expression de *nucS*.

2.2.2. Les mutants *dnaE_{H44Y}* surexpriment *nucS*

Les résultats de qPCR montrent un niveau d'expression de *nucS* similaire dans les souches *dnaE_{H44Y}* portant le plasmide pDyn6902, avec ou sans traitement au thiostrepton (**Figure 20**). Ce dernier ne semble donc pas avoir d'effet sur l'expression du gène chromosomique *nucS*. Concernant les souches *dnaE_{H44Y}* pDyn6902::*nucS*, le thiostrepton induit une augmentation de l'expression de *nucS* d'un facteur allant de 2 pour *dnaE_{H44Y}_3* pDyn6902::*nucS_2* à un facteur 37 pour *dnaE_{H44Y}_1* pDyn6902::*nucS_2*. Les données représentées en **Figure 20** montrent une des 3 répétitions biologiques. L'ensemble de ces résultats permettent d'établir la surexpression effective de *nucS* dans les mutants *dnaE_{H44Y}* *nucS* après induction au thiostrepton.

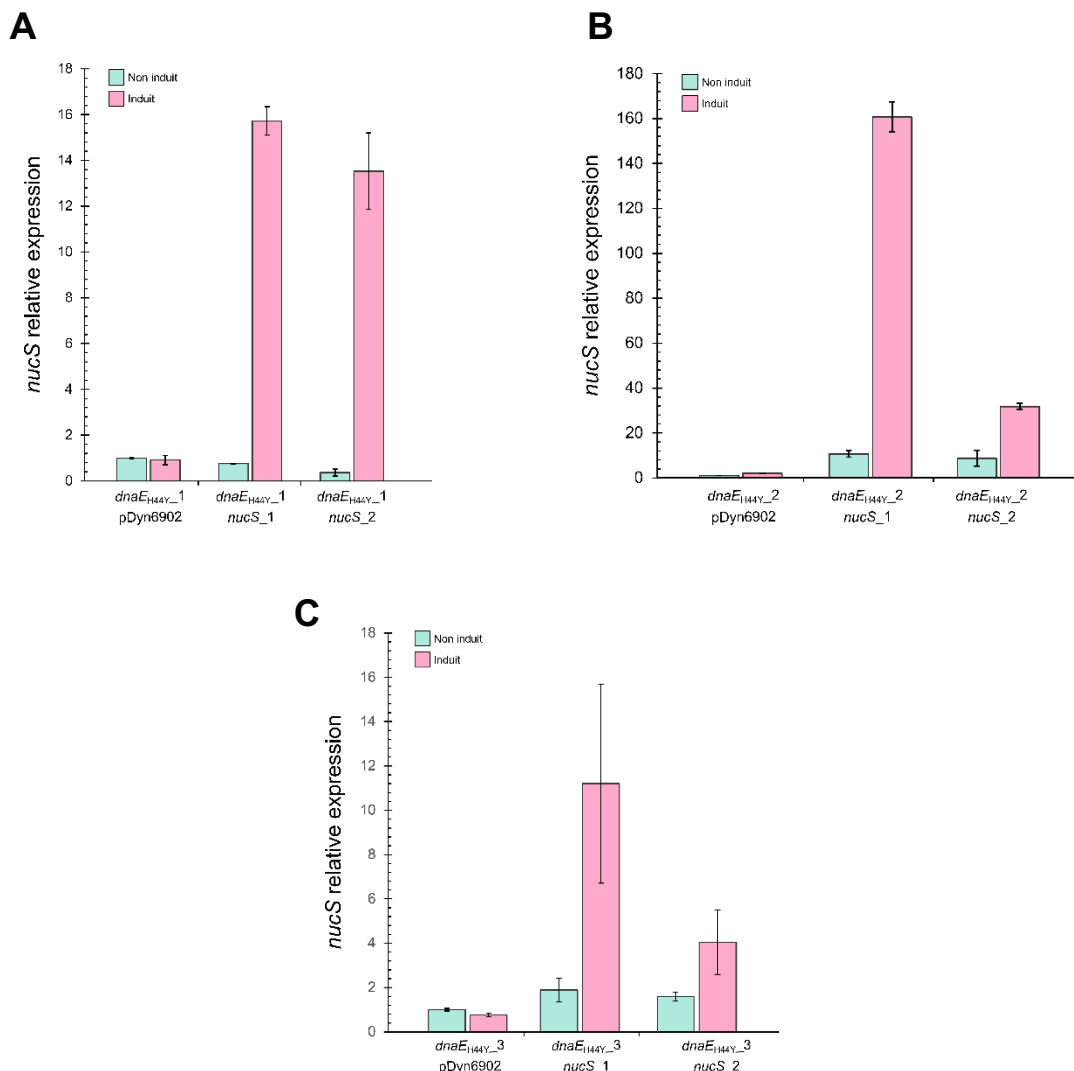


Figure 20. Expression relative de *nucS* dans les mutants *dnaE_{H44Y}*. Niveau d'expression, avec ou sans induction au thiostrepton, de *nucS* dans les souches **(A)** *dnaE_{H44Y}_1*, **(B)** *dnaE_{H44Y}_2*, **(C)** *dnaE_{H44Y}_3* contenant pDyn6902 ou pDyn6902::*nucS*. Deux individus contenant le vecteur pDyn6902::*nucS* ont été testés pour chaque mutant *dnaE_{H44Y}*. Une augmentation allant de 2 à 37 fois de l'expression de *nucS* est observée après induction au thiostrepton.

3. Perspectives à court terme

Dans le but de comprendre la corrélation potentielle entre l'activité de NucS et la plasticité génomique de *Streptomyces*, l'analyse du chromosome des descendants des mutants sur-expresses de *nucS* apparaît incontournable. A l'instar de la caractérisation des descendants mutants *dnaE_{H44Y}*, l'idée serait d'apprécier une éventuelle augmentation de la fréquence des réarrangements génomiques liée à la surexpression de *nucS*. Pour cela, des expériences de PFGE sont envisagées sur le chromosome des descendants des souches *dnaE_{H44Y} pDyn6902::nucS* après induction au thiostrepton. Les souches *dnaE_{H44Y} pDyn6902::nucS* non induits au thiostrepton constituent les témoins de l'expérience. Pour ces derniers, on peut s'attendre à un taux de remaniements similaire aux mutants *dnaE_{H44Y}*. Si NucS génère des DSB *in vivo* au site de mésappariements et si la réparation de ces dernières promeut la recombinaison, nous devrions observer un taux de réarrangements plus grand que celui déterminé dans les simples mutants *dnaE_{H44Y}*. Toutefois, il est important de souligner que la surexpression seule de *nucS* pourrait ne pas suffire à une hausse de la fréquence de recombinaison. En effet, l'accroissement de l'expression des partenaires moléculaires pourrait être nécessaire pour observer un tel effet.

4. Matériel et Méthodes

4.1. Mutagenèse dirigée avec CRISPR-cBEST

Les sites mutables par CRISPR-cBEST, dans une séquence nucléotidique donnée, sont définis par le site CRISPy-web (<https://crispy.secondarymetabolites.org/#/input>). La séquence des ARN guide et des amorces utilisées est listée en **Tableau 3**. Les étapes d'insertion et vérification de l'ARN guide dans le plasmide pCRISPR-cBEST (carte du plasmide en **Figure 21**) ont été réalisées comme décrit précédemment (Tong *et al.*, 2020). La transformation du plasmide pCRISPR-cBEST recombinant (qui porte un gène de résistance à l'apramycine) dans *E. coli* ET12567/pUZ8002 et la conjugaison intergénérique entre *E. coli* et *S. ambofaciens* ont été effectuées comme décrit dans le paragraphe 4.2. L'ADN génomique des transconjugants a été extrait et la mutation cible a été vérifiée par PCR puis séquençage. Les transconjugants portant la mutation cible ont poussé en milieu liquide YEME à 40°C pendant 3 jours afin d'éliminer le plasmide pCRISPR-cBEST. Les colonies issues de la culture étant sensibles à l'apramycine (50µg/mL) ont perdu le plasmide pCRISPR-cBEST.

Tableau 3. Oligonucléotides utilisés dans la partie chapitre 2 des résultats

| Nom de l'oligonucléotide | Séquence 5'-3' | Utilisation |
|--------------------------|--|---|
| sgRNA_dnaE_H44Y | CGGTTGGTAGGATCGACGGCTCCGACCACGGCAACCTGCAGTTTTAGAGCTAGAAATACC | ARN guide à cloner dans pCRISPR-BEST permettant de cibler la substitution H44Y dans DnaE |
| sgRNA-TEST-F | AATTGTACGCGGTCGATCTT | Vérification de l'intégration de la séquence spécifique à la mutation désirée dans le plasmide pCRISPR-BEST |
| sgRNA-TEST-R | TACGTAAAAAAGCACCGAC | |
| Cas9-C-terminal-TEST | GACCCTGATCCACCAGAGCA | Vérification de l'intégrité du plasmide pCRISPR-BEST |
| DnaE_H44Y_For | CACGTCCACACCCAGTACTC | Séquençage de la région portant la mutation H44 de DnaE chez <i>S. ambofaciens</i> |
| DnaE_H44Y_Rev | GATGCCCTCGGACCACTG | |
| sgRNA_nucS_W49* | CGGTTGGTAGGATCGACGGCCATCCAGTTGAGGGGCTGTGTTTTAGAGCTAGAAATAGC | ARN guide à cloner dans pCRISPR-BEST permettant de générer le codon STOP (W49*) au niveau de <i>nucS</i> |
| nucS_W49* | GAGCAGCTCCTGGAGGTGC | Vérification de la substitution W49* dans NucS chez <i>S. ambofaciens</i> |
| HrdB-F | CGCGGCATGCTCTTCCT | qPCR |
| HrdB-R | AGGTGGCGTACGTGGAGAAC | qPCR |
| nucS-F1 | AAGCCCCTGAACTGGATGTC | qPCR |
| nucS-R1 | GAGCAGCTCCTGGAGGTGC | qPCR |

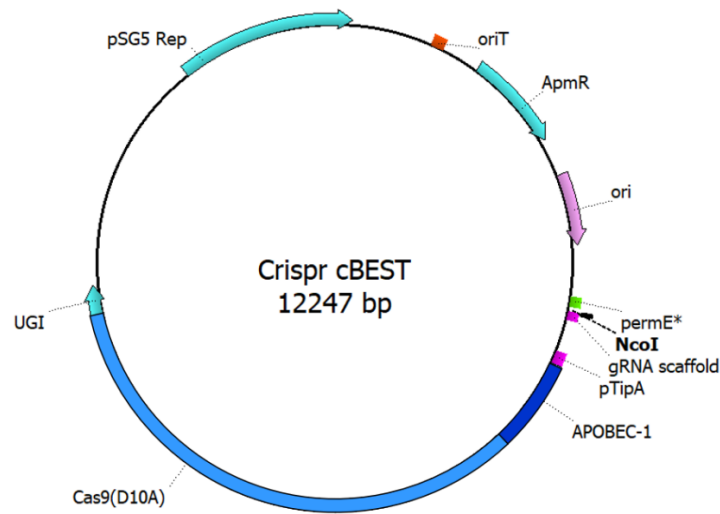


Figure 21. Carte du plasmide pCRISPR-cBEST. *pSG5 Rep* : gène codant protéine de réplication ; *oriT* : origine de transfert ; *ApmR* (*aac(3)IV*) : gène de résistance à l'apramycine ; *ori* : origine de réplication ; *permE** : promoteur fort chez *Streptomyces* ; *NcoI* : site de reconnaissance de l'enzyme de restriction *NcoI* ; gRNA scaffold : scaffold de l'ARN guide du système Crispr Cas9 de *Streptococcus pyogenes* ; *pTipA* : promoteur inductible au thiostrepton A : fusion de APOBEC-1 : cytidine désaminase, Cas9(D10A) : mutant nickase de l'endonucléase Cas9 du système CRISPR/Cas de type II de *Streptococcus pyogenes* ; UGI : inhibiteur d'uracil-ADN glycosylase provenant d'un bactériophage de *Bacillus subtilis*.

4.2. Conjugaison intergénérique

Dans un premier temps des souches d'*E. coli* ET12567/pUZ8002 ont été transformées avec le plasmide pDYN6902 avec ou sans *nucS* en vue de la conjugaison avec *S. ambofaciens*. Pour cela une culture d'*E. coli* ET12567/pUZ8002 a été lancée la veille de chaque transformation dans 3 mL de milieu LB avec de la kanamycine à 50 µg/mL et du chloramphénicol à 25 µg/mL.

Le lendemain une culture de 10 mL dans les mêmes conditions a été lancée à partir de la préculture de la veille diluée au millième. Cette culture a été stoppée quand la DO_{600nm} était comprise entre 0,4 et 0,6 puis centrifugée et lavée avec du glycérol 10%. Une nouvelle centrifugation s'en est suivie puis le surnageant a été retiré. Un volume de 50 µL de la culture et 1 µL de plasmides ont été déposés dans des cuves à électroporation. Après application du choc électrique, 1 mL de milieu riche SOC a été ajouté aux cellules. La culture est incubée 1h à 37°C. Un étalement de 100 µL de ces cellules a été réalisé (Non diluées, diluées à 10^{-1} , et un témoin non électroporé) sur des boîtes de LB avec chloramphénicol à 25 µg/mL, apramycine à 50 µg/mL et kanamycine à 50 µg/mL afin de sélectionner les transformants.

Concernant la conjugaison, des précultures des souches *E. coli* contenant ET12567/pUZ8002 pDYN6902 avec ou sans *nucS* sont tout d'abord réalisées dans 3 mL de milieu LB avec du

chloramphénicol à 25 µg/mL, de la kanamycine à 50 µg/mL et de l'apramycine à 50 µg/mL. Celles-ci sont incubées en agitation toute une nuit à 37°C. A partir de ces précultures, des cultures sont lancées dans les mêmes conditions et sont incubées à 37°C jusqu'à obtention d'une DO_{600nm} comprise entre 0,4 et 0,6. S'en suit une étape de lavage avec du LB, puis le culot est repris dans 1 mL de LB.

Cette culture est ensuite mise en contact avec une suspension de spores des mutants *dnaE_{H44Y}* de *S. ambifaciens*. Des dilutions en cascades sont ensuite effectuées et différentes dilutions sont étalées sur milieu SFM additionné de MgCl₂ (10 mM). L'application des antibiotiques de sélection est réalisée par « flooding » de 1 mL d'eau stérile additionnée d'acide nalidixique (25 µg/mL final) et d'apramycine (50 µg/mL final) et les boîtes sont ensuite incubées 5 à 7 jours à 30°C. Les transconjugants sont ensuite sous-clonés une première fois sur SFM (apramycine 50 µg/mL, acide nalidixique 12.5 µg/mL) puis sur SFM (apramycine 50 µg/mL). Des sporées sont ensuite récoltées après étalement en stries confluentes sur SFM (apramycine 50 µg/mL).

4.3. Préparation des lamelles, digestion enzymatique et PFGE

Les souches WT et ainsi que les mutants *dnaE_{H44Y_1}*, *dnaE_{H44Y_2}*, *dnaE_{H44Y_3}* ont été étalées sur SFM afin d'obtenir des colonies isolées après 7 jours d'incubation à 30°C. Les spores d'une colonie isolée des souches WT et des mutants *dnaE_{H44Y_1}*, *dnaE_{H44Y_2}*, *dnaE_{H44Y_3}* ont été inoculées à 30°C durant 3 jours dans 50 mL de milieu liquide YEME 34% + Glycine 10% (final) à 300 rpm. Les cellules ont été d'abord culotées et lavées une fois avec du TE-sucrose froid puis repris dans du TE-sucrose froid pour avoir une DO_{600nm} comprise entre 1,8 et 2. Ensuite, 300µL de ces cellules traitées a été mélangé avec 300 µL de 1,5% de Low Melting Point Agarose (LMPA) qui a été dissout dans du TE-sucrose. Notons que si les cellules doivent être maintenues à froid (4°C), la solution de LMPA 1,5% doit être à 50°C. Le mélange doit être homogène et il a été déposé dans un moule permettant la confection de lamelles dans lesquelles les cellules sont piégées dans l'agarose. Après 10 min, les lamelles ont été démoulées et transférées dans une solution de 5 mL de TE-sucrose supplémentée de lysozyme 2mg/mL. Après 6h d'incubation à 37°C, la solution a été remplacée par 5 mL de tampon contenant de la protéase 2,5 mg/mL, du SDS 1% et de l'EDTA 0,5 M. Après une incubation jusqu'au lendemain à 50°C, les lamelles ont été lavées toutes les 30 min avec du TE-sucrose préalablement mis à 50°C. Le 5^{ème} lavage a été réalisé avec du TE-sucrose froid afin de durcir les lamelles. Un morceau de lamelle a été digéré pendant 3h à 37°C avec *Asel* : 267 µL d'eau ultra pure + 30 µL de tampon de digestion + 3µL de l'enzyme restriction *Asel*. Le morceau de lamelle digéré a été placée dans un puit d'un grand gel 0,9% d'agarose SeaKem dissout dans du TBE 0,5X puis le puit a été bouché avec du LMPA 1,5%.

Les paramètres de la PFGE en conditions 1 sont les suivants : initial switch = 40s ; final switch = 160s ; voltage = 6V/cm ; durée de migration = 20h ; angle = 120°, ramp factor = linear.

La PFGE en conditions 2 requiert ces conditions de migration : initial switch = 20s ; final switch = 20s ; voltage = 6V/cm ; durée de migration = 40h ; angle = 120°.

Après migration, les gels sont incubés dans le BET puis révélés avec Uvidoc (Cambridge).

4.4. PCR quantitative

4.4.1. Extraction des ARN

Afin de faciliter les prélèvements, des membranes de cellophane ont été déposées sur le milieu gélosé (HT) additionné d'apramycine à 5 µg/mL et de thioestrepton (2,5 µg/mL) lors du traitement induction. Plusieurs stries de 2 µL de souches *S. ambofaciens* contenant pDYN6902 et pIB139 avec ou sans *nucS* ont été réalisées sur la membrane de cellophane et mises à incuber à 30°C. Au bout de 48h, les échantillons ont été récoltés délicatement avec une spatule (une demi-strie par répliquat), déposés dans des eppendorfs, plongés dans l'azote liquide puis conservés à -80°C.

Les ARN des prélèvements ont été extraits avec le kit BIO-RAD Aurum Total RNA Mini selon les recommandations du fournisseur. Les échantillons ont d'abord été repris dans des tubes de 2 mL avec 100 µL de TE pH 7,5 avec du lysozyme à 1 mg/mL. Le tout a été incubé 5 min à température ambiante puis 350 µL de solution de lyse a été ajoutée. Une étape de sonication s'en est suivi (3x30 secondes), avec l'appareil diagenode Bioruptor[®], afin de lyser les bactéries, suivi d'une centrifugation d'une minute à vitesse maximale. Suite à cette étape, le surnageant est transféré dans un eppendorf de 2mL et la manipulation reprend le protocole classique. L'élution est faite avec 20 µL de la solution d'élution.

Un dosage de la concentration des ARN dans les échantillons est réalisé avec le NanoDrop[®] ND-1000 spectrophotometer (Labtech) et grâce au logiciel ND-1000 (3.5). Les ARN ont ensuite été dosés sur gel d'agarose 0,8%, (1,5 µg d'ARN) afin de vérifier leur qualité et leur concentration.

4.4.2. Rétro-transcription et PCR quantitative

Un traitement à la DNase a été réalisé sur 2 µg des ARN extraits, avec 1,3 µL de tampon 10X, 0,5 µL d'inhibiteur à RNase, 1 µL de Dnase. Le volume est complété jusqu'à 10,2 µL avec de l'eau ultra pure RNase free puis incubé à 37°C pendant 30 minutes. Un volume de 3 µL d'EDTA à concentration finale de 5 mM est ensuite ajouté pour stopper la réaction et l'inactivation de la DNase se fait pendant 10 minutes à 75°C. Un volume de 6 µL a été prélevé à la suite de cette étape pour réaliser un contrôle

de l'efficacité des étapes de DNase (Vérification que les échantillons ne contiennent plus d'ADNg), par PCR sur le gène *hrdB*, puis électrophorèse sur gel d'agarose 1%. Une RT-PCR a ensuite été réalisée avec le kit « iScript™ Advanced cDNA Synthesis Kit » selon les recommandations du fournisseur dans 20 µL de volume réactionnel. Pour la PCR quantitative, l'amplification est réalisée à partir de 2 µL d'ADNc avec le kit « iQ™ SYBR® Green Supermix » selon les recommandations du fournisseur, dans le thermocycleur CFX96 (Bio-Rad). Le gène *hrdB* a servi de gène de référence pour la normalisation.

DISCUSSION ET PERSPECTIVES

DISCUSSION ET PERSPECTIVES

Dans cette partie seront plus particulièrement développés des éléments de discussion non abordés dans la discussion de l'article du chapitre 1 « Correction of non-random mutational biases along a linear bacterial chromosome by the mismatch repair endonuclease NucS » par Dagva *et al.* en révision dans Nucleic Acids Research.

I. Existence d'un MMR non-canonique chez *Streptomyces*

1. Phénotype mutateur lié à la délétion du gène *nucS*

De nombreux génomes complets de *Streptomyces* sont disponibles dans les bases de données. L'analyse d'un échantillon représentatif de la diversité du genre (Lorenzi *et al.*, 2021) a permis de montrer l'absence des protéines signatures du MMR canonique, MutS et MutL, chez les *Streptomyces*, comme chez l'ensemble des actinobactéries et des archées (Mizrahi and Andersen, 1998; Castañeda-García *et al.*, 2017). Ces organismes ne présentant pas des taux de mutation accrus comparés à ceux dotés du MMR canonique, il a été proposé qu'une voie alternative de correction des mésappariements existe chez ces organismes. En 2016, l'équipe du Pr. Y. Ishino a apporté, pour la première fois, des preuves génétiques quant au rôle de la protéine NucS dans le MMR non-canonique chez l'archée *T. kodakarensis*, en particulier à sa capacité à générer une DSB en clivant les deux brins de l'ADN au site d'un mésappariement. Cette découverte a suscité de nombreuses études visant, non seulement à préciser l'aspect mécanistique de cette voie, mais aussi à investiguer le rôle biologique de NucS dans l'évolution des génomes. Les travaux de cette thèse s'inscrivent précisément dans cet objectif.

Streptomyces, tout comme la majorité des actinobactéries, ne possède pas le gène *mutS* et *mutL* du MMR canonique. En revanche, nos analyses *in silico* révèlent que *nucS* appartient au génome « core » ; il est effectivement systématiquement retrouvé en une seule copie dans les 125 génomes testés (Lorenzi *et al.*, 2021) et sa séquence protéique est fortement conservée sein du genre. Pour explorer la fonction de NucS chez *Streptomyces*, nous avons construit 4 lignées $\Delta nucS$ qui, toutes, présentent un phénotype mutateur, caractérisé, d'une part, par l'apparition à haute fréquence de colonies à secteurs et d'autre part, par un taux de mutation 130 fois supérieur à la souche sauvage. L'ensemble de ces résultats est en accord avec les études antérieures sur le rôle de NucS comme garant de la stabilité du génome chez les actinobactéries (Castañeda-García *et al.*, 2017, 2020; Ishino *et al.*, 2018; Takemoto *et al.*, 2018).

2. NucS est une endonucléase spécifique des mésappariements

La caractérisation enzymatique de NucS a pour origine les travaux du groupe de H. Myllykallio (Ren *et al.*, 2009). Ainsi, chez *P. abyssi*, NucS a été caractérisée comme une *flap* endonucléase, c'est-à-dire ciblant les jonctions ADN double brin (ADNdb)/ADNss sur des substrats d'ADN ramifiés (Ren *et al.*, 2009). De manière surprenante, les travaux chez l'archée *T. kodakarensis* et l'actinobactérie *C. glutamicum* ont révélé une activité endonucléase spécifique des mésappariements. Contrairement au MMR canonique, la protéine coupe *in vitro* les deux brins de l'ADN au niveau des mésappariements G/T, G/G et T/T. Nos résultats de tests d'activité *in vitro* corroborent ces données : NucS de *S. ambofaciens* a la même spécificité de substrat que NucS de *T. kodakarensis* et *C. glutamicum* (Ishino *et al.*, 2016, 2018; Takemoto *et al.*, 2018) ; et il réalise un clivage double-brin. Ce résultat constitue une avancée majeure parce que d'importantes questions liées à la réparation des DSB médiées par NucS en découlent (voir le point 3. ci-dessous).

Très récemment, deux publications dans l'archive BioRxiv ont investigué la spécificité de substrat de NucS de *M. smegmatis in vivo* (Islam and Josephs, 2023; Rivera-Flores *et al.*, 2023). Les deux équipes ont utilisé une technique appelée « Oligo-mediated Recombineering » (OR) qui repose sur l'utilisation d'oligonucléotides pour générer des mésappariements au niveau des fourches de réplication. Les chercheurs ont examiné l'émergence de phénotype de résistance à la rifampicine afin de déterminer les mésappariements reconnus par NucS. Grâce à cette méthode de criblage, les chercheurs ont montré que, chez cette bactérie, l'activité de NucS est spécifique des mésappariements G/T, G/G et T/T, comme démontré *in vitro* chez les archées et l'actinobactérie *C. glutamicum*. Ces travaux apportent, pour la première fois, la preuve *in vivo* que NucS cible les mésappariements post-réplicatifs.

S'il est vrai que quelques études ont démontré l'activité de clivage de NucS au niveau des mésappariements *in vitro*, aucune publication ne relate une telle activité *in vivo* pour le moment. Cet aspect est au cœur de toutes les préoccupations des équipes travaillant sur le MMR non-canonique. L'une des approches pour répondre à cette question serait de quantifier les DSB, et également les MM, dans les cellules de *S. ambofaciens* en utilisant la microscopie optique au niveau d'une seule cellule (single-cell imaging). A cette fin, des fusions Gam-GFP (Shee *et al.*, 2013) ou KuA-GFP, se fixant sur les DSB, seront testées. De plus, différents fonds génétiques de *S. ambofaciens* exprimant des fusions de différentes protéines fluorescentes avec une protéine NucS catalytiquement inactive mais capable de se lier aux MM seront réalisées. Ces fusions rapportrices devraient se lier aux DSB ou aux MM dans les cellules vivantes et former des foyers de fluorescence stables qui peuvent être quantifiés en utilisant la microscopie de fluorescence à super-résolution (Lestini *et al.*, 2013; Delpech *et al.*, 2018). Pour ces études, il sera intéressant de se placer dans un contexte où la fréquence de mésappariements dans la cellule est

augmentée c'est-à-dire dans un mutant du domaine PHP de DnaE (*dnaE_{H44Y}*). Une autre solution serait d'estimer la quantité de DSB *in vivo* le long du chromosome dans différents contextes mutants (par exemple *dnaE_{H44Y}* et *dnaE_{H44Y}* surexprimant *nucS*) par la technique de i-BLESS (*immobilized-Breaks Labeling, Enrichment on Streptavidin and Next-Generation Sequencing*) qui permet de déterminer la distribution et la localisation exacte des DSB à l'aide de NGS (Crosetto *et al.*, 2013; Biernacka *et al.*, 2018). Cependant cette dernière n'a pour le moment été développée que pour les cellules eucaryotes, dont la levure.

3. Spectre mutationnel lié à la délétion de *nucS*

Les expériences d'accumulation de mutations chez le mutant Δ *nucS* révèlent un biais mutationnel vers les transitions. En effet, 97% des mutations accumulées au cours des 60 cycles de sporulations sont des transitions. Le même biais est observé pour le MMR canonique chez *E. coli* (Lee *et al.*, 2012). Il est connu de longue date que l'ADN polymérase corrige moins efficacement les mésappariements conduisant aux transitions que ceux conduisant aux transversions ; c'est pourquoi la machinerie MMR permettrait davantage l'évitement des transitions que transversions (Schaaper, 1993). Etant donné que les mismatches qui conduisent aux transitions sont G/T et A/C, et si on fait le postulat que NucS corrige uniquement les G/T *in vivo* comme c'est le cas *in vitro*, alors ces résultats laissent supposer que l'ADN polymérase serait plus encline à créer des mésappariements G/T, et/ou à moins bien les corriger *via* son activité de relecture « *proofreading* », une hypothèse déjà proposée par une équipe travaillant sur le MMR non-canonique de *C. glutamicum* (Takemoto *et al.*, 2018). Par ailleurs, nos résultats révèlent que les lignées Δ *nucS* accumulent davantage les transitions de type A:T>G:C que G:C>A:T alors que l'observation inverse est constatée chez les lignées sauvages. Au regard de ces données, et comme suggéré par Castañeda *et al.* (2020), le MMR non-canonique empêcherait l'augmentation du taux en GC dans les bactéries ayant déjà un contenu en GC important comme les mycobactéries ou les *Streptomyces*.

II. NucS et la réparation des DSB

1. Voies de réparation des cassures potentiellement médiées par NucS

A ce jour, l'activité enzymatique de NucS a été explorée uniquement *in vitro*. Si ces études tendent à montrer le clivage double-brin de l'ADN (donc la génération d'une DSB) au niveau d'un mésappariement, la formation de DSB dépendant de NucS *in vivo* n'est pas prouvée. Toutefois, plusieurs éléments impliqueraient la recombinaison homologe comme voie de réparation des DSB produites par NucS.

1.1. La recombinaison homologue

La recombinaison homologue pourrait favorablement prendre en charge les extrémités d'ADN générées par l'intervention de NucS aux sites des mésappariements post-réplicatifs. Plusieurs éléments ont amené les chercheurs à suspecter celle-ci. D'abord, le mécanisme de discrimination des brins néosynthétisé et parental par le MMR non-canonique n'est pas connu. Cependant, cette discrimination n'a pas lieu d'exister si la réparation des DSB est assurée par recombinaison homologue puisque la chromatide sœur servirait de matrice pour réparer la cassure et le mésappariement. De plus, on sait que NucS interagit avec le β -clamp (ou PCNA chez les archées et les eucaryotes). Ce dernier sert de plateforme universelle en recrutant de nombreuses protéines intervenant dans le métabolisme de l'ADN comme les enzymes impliquées dans la réplication et/ou dans la maintenance du génome. Récemment, le complexe MR, composé des protéines *Mre11* et *Rad50*, a été identifié comme partenaire de PCNA chez l'archée thermophile *P. abyssi* (Pluchon 2013, Hogrel 2018). Le complexe MR est retrouvé dans les trois branches de la vie ; il constitue un acteur majeur dans la maintenance du génome. En effet, de par sa capacité à détecter et à dégrader les extrémités endommagées (comme les DSB ; (Stracker, 2011)), il est en première ligne dans la réponse aux dommages à l'ADN. Si l'interaction entre PCNA et MRX/N (X pour *Xrs2* chez *Saccharomyces cerevisiae* ; N pour *Nbs1* chez l'Homme, le troisième acteur impliqué dans le complexe chez les eucaryotes) n'a pas été démontré chez les eucaryotes, des travaux ont révélé la co-localisation du complexe avec PCNA au niveau des fourches de réplication (Maser *et al.*, 2001). Ces résultats suggèrent l'implication du complexe MRN dans stabilité des fourches de réplication ; il pourrait initier la réparation, par recombinaison, des DSB apparaissant lors de la réplication.

Puisque NucS et le complexe MR sont deux acteurs interagissant avec PCNA chez *P. abyssi*, le complexe MR pourrait être recruté par PCNA afin d'intervenir dans la résection des extrémités des DSB induites par NucS, et promouvoir la réparation des cassures, et de fait des mésappariements, par recombinaison homologue.

Chez les bactéries, l'homologue du complexe MR est le complexe SbcCD qui intervient dans la réparation par recombinaison homologue des DSB chez *Bacillus subtilis* (Mascarenhas *et al.*, 2006). Chez *E. coli*, le complexe SbcCD est impliqué dans la résolution des intermédiaires lors la convergence des fourches de réplication (Yao and O'Donnell, 2018). Ainsi, le complexe SbcCD pourrait être impliqué dans la réparation des DSB au niveau des fourches de réplication. Chez *Streptomyces*, les homologues de SbcC et SbcD ont été identifiés (chez *S. ambofaciens*, loci SAM23877_1380 et 1381). L'effet de la mutation de ces gènes pourrait être révélatrice du rôle du complexe MR chez *Streptomyces* avec notamment un phénotype Δ nucS-like.

1.2. Autres mécanismes candidats

Selon les auteurs des deux publications sur BioRxiv (Islam and Josephs, 2023; Rivera-Flores *et al.*, 2023), il est peu probable qu'une résection des extrémités d'ADN ait lieu pour la correction des mésappariements par NucS. Si les deux études révèlent une activité exonucléasique 5'-3' au site de la cassure, la dégradation serait de courte longueur avec seulement une dizaine de nucléotides. Ce résultat serait incompatible avec une prise en charge des cassures par recombinaison homologue puisque la formation d'un ADN simple-brin d'au moins plusieurs kilobases (filament RecA) est nécessaire pour la recherche d'homologie (Chung *et al.*, 2010). Ainsi, le mécanisme proposé se rapprocherait du mécanisme *very short patch* (VSP) décrit chez les bactéries et les mammifères. Chez *E. coli* et les bactéries entériques apparentées, la réparation des mésappariements de bases est effectuée par deux processus biochimiques qui se chevauchent, la réparation des mésappariements réplcatifs dirigée par la méthylation (MMR) et la réparation des mésappariements courts (Very Short Patch repair) (Lieb and Bhagwat, 1996; Lieb, 1991). Alors que le MMR répare les erreurs de réplication, la réparation VSP corrige les mésappariements G/T créés par la désamination spontanée de la 5-méthylcytosine en T (Lieb and Bhagwat, 1996). Chez *E. coli*, la méthylation des cytosines résulte de l'action de la cytosine méthylase Dcm (au sein du motif $5'CC(A/T)GG^3'$). Le gène *dcm* est organisé en opéron avec le gène *vsr* dont le produit, l'endonucléase Vsr, est requis pour la voie VSP. La protéine Vsr reconnaît le mésappariement G/T et réalise une incision en 5' de la thymine à éliminer. Contrairement au MMR, un fragment de petite taille (<10 nucléotides) est dégradé puis resynthétisé par l'ADN polymérase I.

Dans le cas où le mécanisme dépendant de NucS impliquerait l'élimination d'un fragment d'ADN comme dans le mécanisme VSP, il est tout de même à noter que l'action de NucS est *a priori* indépendante de la méthylation, et que NucS induit une cassure double brin (*in vitro*) contrairement à Vsr.

2. Contrôle de la recombinaison par NucS

Depuis les travaux pionniers de M. Radman (Rayssiguier *et al.*, 1989) révélant l'impact du MMR dans la limitation de la recombinaison entre *Salmonella typhimurium* et *E. coli*, plusieurs études ont démontré le rôle de MutS dans le contrôle de la recombinaison (anti-recombinaison) de séquences faiblement divergentes (homéologues) comme chez la levure (Harfe and Jinks-Robertson, 2000) ou *Bacillus subtilis* (Carrasco *et al.*, 2019). Chez *M. smegmatis*, l'existence d'une barrière à la recombinaison fût suspectée dès 2000 (Springer *et al.*, 2004) puis l'implication de NucS et du MMR non-canonique dans l'anti-recombinaison résultent des recherches de l'équipe de J. Blasquez (Castañeda-García *et al.*, 2017). Ainsi, alors que la fréquence de recombinaison entre séquences identiques à 100% est similaire entre les

contextes sauvage et déficient pour NucS, elle est supérieure chez le mutant entre 3 et 10 fois pour des séquences respectivement divergentes à 95% et 90%.

Bien que l'implication du MMR ait été décrite tant chez les bactéries que chez les eucaryotes, les mécanismes restent à élucider (Tham *et al.*, 2013; Kunkel and Erie, 2015). En effet, des mésappariements peuvent être générés au cours du processus de recombinaison, et notamment homéologue, et peuvent servir de substrats à certaines protéines du MMR, comme MutS2, pour exercer leurs activités anti-recombinaison (Calmann *et al.*, 2005). Il a été démontré que chez *E. coli*, MutS et MutL régulent la recombinaison homologue en interrompant l'échange de brins médié par RecA (Worth *et al.*, 1994; Tham *et al.*, 2013, 2016). Cette interaction avec les protéines MMR et leur activité anti-recombinaison permettrait de favoriser la fidélité de la recombinaison homologue, et donc d'inhiber la recombinaison homéologue. Cette régulation permet le maintien de la stabilité du génome puisque la recombinaison entre séquences répétées non identiques peut générer des remaniements d'ampleur potentiellement source de maladies chez l'Homme (Lupski, 1998).

Dans le cadre de cette hypothèse, NucS serait capable de reconnaître et d'interagir avec des intermédiaires d'ADN branchés de la recombinaison, comme la boucle-D, et d'exercer une activité anti-recombinaison. En outre, les premières études sur la protéine NucS, chez l'archée *P. abyssi*, avaient démontré une affinité de NucS pour les substrats d'ADN branchés (Ren *et al.*, 2009).

3. NucS contribue-t-elle à la plasticité du chromosome des *Streptomyces* ?

Nous avons établi qu'en absence de NucS, la distribution des SNP (majoritairement de transitions) formait un gradient croissant de l'origine de répliation vers les extrémités du chromosome de *S. ambofaciens*. Cette observation suggère qu'en s'éloignant de l'*oriC*, l'ADN polymérase est plus encline à laisser échapper des erreurs d'incorporation de bases malgré son activité *proofreading*. D'après nos résultats, NucS intervient principalement dans l'évitement des transitions, la protéine ciblant davantage les mésappariements de type G/T et induisant une DSB *in vitro*. Il pourrait donc exister un gradient de DSB et conséquemment une fréquence de recombinaison croissante le long du chromosome de *S. ambofaciens*. Cette hypothèse est soutenue par des travaux de génomique comparée du laboratoire qui ont révélé un accroissement de la fréquence des insertions/délétions (indels) le long des bras chromosomiques chez les *Streptomyces* (Tidjani *et al.*, 2019b; Lorenzi *et al.*, 2021). Au sein d'une population (Tidjani *et al.*, 2019b, 2020) ou du genre (Lorenzi *et al.*, 2021), c'est-à-dire entre souches présentant des liens phylogénétiques allant d'étroits à distants, il est démontré que les bras chromosomiques évoluent plus rapidement que la région centrale : la fréquence de réarrangements est croissante (*shuffling*) ainsi que l'insertion de séquences variables (issues du transfert de gènes). Ainsi, la

remarquable plasticité génomique observée dans les régions subtélomériques du chromosome de *Streptomyces* pourrait trouver source dans la réparation des DSB générées par NucS.

Les résultats obtenus chez *S. ambofaciens* dans les contextes mutants pour l'activité *proofreading* de la polymérase (*dnaE_{H44Y}*, voir Résultats chapitre 2) soutiennent cette hypothèse : la fréquence des remaniements génomiques est significativement plus importante dans ce contexte que dans la souche sauvage. Dans ce contexte où les mésappariements répliatifs sont accrus, NucS interviendrait plus fréquemment, et pour une partie de son intervention, libérerait des extrémités d'ADN recombinogènes, favorisant les réarrangements du chromosome. Afin de relier NucS et la recombinaison, nous testons actuellement l'impact de la surexpression de *nucS* dans le contexte *dnaE_{H44Y}*. Si un accroissement de la fréquence de recombinaison est associé à la surexpression de *nucS* alors une corrélation entre l'activité de NucS et l'instabilité génétique pourra être établie. Cette expérience teste un impact dépendant de la seule quantité de NucS disponible ; on ne peut exclure que d'autres acteurs soient requis. Ainsi, parallèlement, la recherche de partenaires de NucS a été engagée par des expériences de purification de complexes cellulaires contenant NucS. A ce jour, hormis le β -clamp, les partenaires impliqués dans la machinerie du MMR non-canonique ne sont pas connus.

L'action différentielle de NucS le long du chromosome linéaire de *Streptomyces*, révélée par la densification des SNP, pourrait constituer un moteur de l'évolution rapide du chromosome de *Streptomyces* et être responsable de sa forte compartimentation. En effet, les bras chromosomiques sont des régions de faible expression, notamment en phase exponentielle de croissance, qui accumulent les mutations et la variabilité génétique issue des réarrangements d'ADN et du transfert horizontal. Ces phénomènes pourraient être liés par les mécanismes de réparation des cassures double brins favorisant la recombinaison (réarrangement, insertion de matériel exogène) et/ou sélection avec exclusion des gènes essentiels des régions terminales et enrichissement en gènes contingents. Ces régions pourraient également être le siège de phénomène d'évolution accélérée favorisant l'apparition de nouvelles fonctions. Par exemple, la fréquente recombinaison des voies de biosynthèse de métabolites spécialisés est vraisemblablement à l'origine de leur diversité ; leur distribution au sein des populations et espèces est également dépendante de la capacité à échanger et recombiner. Cette compartimentation remarquable est également superposable avec l'organisation *in vivo* déduite d'expériences de « capture de conformation chromosomique » ou 3C où la région centrale est plus structurée que les régions terminales (Lioy *et al.*, 2021; Szafran *et al.*, 2021). Chez *S. venezuelae*, la structuration spécifique des régions terminales est associée à un enrichissement en HupS, une protéine de structuration de l'ADN (Nucleic Associated Proteins, NAP) suggérant une composition de la chromatine distincte (Szafran *et al.*, 2021).

Ce phénomène de compartimentation n'est pas l'exclusivité du chromosome des *Streptomyces*. Ainsi, autour des 3 origines de réplication identifiées sur le chromosome circulaire de l'archée *Sulfolobus* (Badel *et al.*, 2022), des régions cumulant expression faible, variabilité et mutations, peuvent être observées. Ces régions sont par ailleurs superposables à une organisation chromosomique spécifique associée à la présence accrue d'une coalescence spécifique (NAP apparentée au système SMC). Des expériences d'ATAC-seq (Buenrostro *et al.*, 2013) montrent également que ces régions sont moins accessibles aux activités de recombinaison (Badel *et al.*, 2022).

Ainsi, la réplication et ses erreurs en lien avec la structure de la chromatine pourrait réguler la recombinaison le long du chromosome des *Streptomyces* ; ces mécanismes pourraient avoir été sélectionnés au cours de l'évolution pour aboutir à la structure contemporaine du chromosome ; une région centrale garante de la conservation et de l'expression des « gènes privés » nécessaires au fonctionnement cellulaire, et des régions terminales, laboratoire d'évolution rapide, favorisant l'apparition et la diversification de fonctions contingentes, et utiles à la population ou espèce (« gènes publics ») (Zhang *et al.*, 2022, 2020b).

III. Evolutivité et adaptation

L'impact de la mutation de différents gènes (*fitness*) qui composent un génome est à l'origine de la dynamique de l'adaptation évolutive. On sait peu de choses sur les facteurs qui influent sur cette diversité, en particulier lors de la diversification écologique et des changements dans la composition de la communauté. L'apparition de souches mutatrices est un facteur accélérant la diversification génétique. Associé à la déficience du système MMR, nous avons caractérisé un phénotype hypermutateur chez *Streptomyces* comme chez tous les autres organismes. Ces hypermutateurs émergent dans les écosystèmes complexes comme le microbiote humain (Barreto and Gordo, 2023) en réponse aux stress et détermine l'homéostasie de la symbiose avec l'hôte. Le déséquilibre du microbiote et la conversion d'un organisme commensal à pathogène peut aussi résulter de l'émergence de mutateurs dans une population ou communauté. Ces phénomènes ont également été étudiés au cours d'expériences d'évolution expérimentale (LTEE Long Term Evolution Experience) chez *E. coli* (Ascensao *et al.*, 2023; Lenski, 2023). La présence de mutateurs répond transitoirement à l'avantage de faire émerger une diversité à partir de laquelle des génotypes adaptés peuvent être sélectionnés ; réciproquement, lorsque la charge mutationnelle liée à la présence de telles lignées est forte, la présence de ces derniers s'estompe.

Les gènes impliqués dans le MMR présentent un degré plus élevé de mosaïcisme que les gènes de ménage (Denamur *et al.*, 2000) révélant des événements de recombinaison fréquents. Ce résultat révèle

que les gènes *mut* sont fréquemment perdus puis ré-acquis grâce aux transferts horizontaux. Ainsi, un hypermutateur ayant perdu un gène, tel que *mutS*, peut voir son taux de mutation revenir à sa valeur ancestrale *via* le transfert horizontal d'un homologue fonctionnel provenant d'une bactérie, même phylogénétiquement distante. Cette plasticité du MMR favoriserait la plasticité d'une population bactérienne.

L'émergence d'hypermutateurs est un enjeu clinique majeur. En effet, il a été montré que l'incidence des mutants parmi les isolats d'*E. coli* et de *Salmonella enterica* pathogènes est élevée (plus de 1 %) allant à l'encontre de la théorie fondée sur des études avec des souches atténuées en laboratoire, qui suggèrent que les hypermutateurs sont rares parmi les populations bactériennes (LeClerc *et al.*, 1996). De plus, chez les patients atteints de la mucoviscidose, la fréquence des isolats de *Pseudomonas aeruginosa* présentant un phénotype hypermutateur lié à un système MMR déficient peut atteindre 60% à 90% (Oliver and Mena, 2010). Un résultat similaire a été trouvé chez les isolats de *Staphylococcus aureus* et de *Haemophilus influenzae* accompagnant l'établissement de la mucoviscidose. En conséquence, ces pathogènes accroissent leurs chances d'acquérir des mutations leur conférant une résistance à des drogues et antibiotiques, rendant les infections plus difficiles à traiter (Oliver and Mena, 2010).

Chez les actinomycètes, une récente étude a examiné les conséquences de l'inactivation de NucS sur l'émergence de variants résistants à différents antibiotiques chez le pathogène *Mycobacterium abscessus* (Fressatti Cardoso *et al.*, 2022). La privation de NucS entraîne un phénotype mutateur avec une acquisition accrue de la résistance aux macrolides et aux aminoglycosides, les deux principaux groupes d'agents antimycobactériens pour le traitement de *M. abscessus*. L'inactivation du MMR non canonique chez *M. abscessus* multiplie par 10 à 22 de l'apparition de mutants spontanés résistants au macrolide clarithromycine et aux aminoglycosides amikacine, gentamicine et apramycine, par rapport à la souche de type sauvage. En outre, l'émergence d'une résistance aux fluoroquinolones (ciprofloxacine) a été détectée dans une souche déficiente en NucS, mais pas dans une souche de *M. abscessus* de type sauvage. Par ailleurs, il a été montré qu'il existe un polymorphisme des allèles *nucS* au sein des souches cliniques de *M. tuberculosis*. L'effet de ces polymorphismes sur l'activité de NucS affecte significativement à la hausse (jusqu'à 84 fois) le taux de mutation pour 5 des 9 variations détectées parmi 1.600 isolats. Ces résultats suggèrent l'existence de souches cliniques hypermutables de *M. tuberculosis* affectées par la modulation de l'activité de NucS (Castañeda-García *et al.*, 2017).

Au laboratoire DynAMic, nous étudions l'évolution génomique des bactéries *Streptomyces* provenant des sols forestiers, un environnement hautement compétitif et changeant. Dans ces conditions très sélectives, nous spéculons que des variants hypermutateurs de *nucS* pourraient émerger favorisant l'adaptabilité des populations. Nos résultats suggèrent que le phénotype de la déficience et de la surexpression de NucS serait vraisemblablement pléiotrope, avec un caractère mutateur marqué au

niveau génique (accumulation de mutations ponctuelles) mais également une dérégulation de la recombinaison (réparation des DSB). Aussi, l'émergence de variants pourrait 'booster' l'adaptabilité des souches et populations au sein de l'écosystème en favorisant les mutations et les transferts horizontaux. Cette perspective est particulièrement attractive en termes de recherche fondamentale pour expliquer l'évolution et la diversité du métabolisme spécialisé des *Streptomyces*. A noter également que les producteurs de molécules à activité antibiotique sont également des réservoirs de résistance aux antibiotiques. D'un point de vue biotechnologique, la formation de nouveaux gènes ou de nouvelles voies de biosynthèse (Tidjani *et al.*, 2019b) (Choufa *et al.*, en préparation) est hautement intéressante dans le cadre de la course à l'armement contre la dissémination de l'antibiorésistance.

Des expériences sont actuellement initiées dans l'équipe pour déterminer le rôle de NucS dans les transferts de gènes, que ce soit dans l'anti-recombinaison ou barrière aux transferts de gènes, comme l'est le mécanisme MMR canonique MutS/L (Matic *et al.*, 1995), ou dans la stimulation des échanges *via* les DSB qu'elle pourrait induire. Nous utilisons des souches isolées de l'environnement qui sont quasi-isogéniques mais suffisamment éloignées pour les distinguer par l'analyse de SNP (collection de l'Unité, (Tidjani *et al.*, 2019b). Des études sur l'importance des transferts entre ces souches ont déjà été réalisées dans l'équipe et nous disposons ainsi de couples de souches "parentales" marquées (chacune avec un gène de résistance à un antibiotique différent) pour lesquels la fréquence et la nature des transferts par conjugaison a été évaluée (Choufa *et al.*, en préparation). L'objectif est de muter le gène *nucS* (avec le système d'édition CRISPR-BEST développé pour les *Streptomyces*, dans ce cas introduction d'un codon stop) dans la réceptrice, ou la réceptrice et la donatrice simultanément, puis de déterminer ensuite si l'absence de NucS affecte les transferts d'ADN. L'importance des transferts dans les recombinants sera estimée par séquençage et analyse bio-informatique des génomes (l'origine parentale de la séquence du recombinant est définie par l'analyse des SNP). Cette approche permettra d'apporter des éléments de réponse sur le rôle du MMR non-canonique dans les échanges de gènes entre bactéries du sol.

REFERENCES

- Ahmad, S., Huang, Q., Ni, J., Xiao, Y., Yang, Y., and Shen, Y.** 2020. Functional Analysis of the NucS/EndoMS of the Hyperthermophilic Archaeon *Sulfolobus islandicus* REY15A. *Front Microbiol* **11**.
- Ali, N., Herron, P.R., Evans, M.C., and Dyson, P.J.** 2002. Osmotic regulation of the *Streptomyces lividans* thiostrepton-inducible promoter, *ptipA*. *Microbiology (Reading)* **148**, 381–390.
- Aniukwu, J., Glickman, M.S., and Shuman, S.** 2008. The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev* **22**, 512–527.
- Aravind, L., and Koonin, E.V.** 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* **11**, 1365–1374.
- Ascensao, J.A., Wetmore, K.M., Good, B.H., Arkin, A.P., and Hallatschek, O.** 2023. Quantifying the local adaptive landscape of a nascent bacterial community. *Nat Commun* **14**, 248.
- Badel, C., Samson, R.Y., and Bell, S.D.** 2022. Chromosome organization affects genome evolution in *Sulfolobus* archaea. *Nat Microbiol* **7**, 820–830.
- Bai, H., and Lu, A.-L.** 2007. Physical and functional interactions between *Escherichia coli* MutY glycosylase and mismatch repair protein MutS. *J Bacteriol* **189**, 902–910.
- Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y., and van Wezel, G.P.** 2015. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiology and Molecular Biology Reviews* **80**, 1–43.
- Barreto, H.C., and Gordo, I.** 2023. Intrahost evolution of the gut microbiota. *Nat Rev Microbiol* **21**, 590–603.
- Bertrand, C., Thibessard, A., Bruand, C., Lecoite, F., and Leblond, P.** 2019. Bacterial NHEJ: a never ending story. *Molecular Microbiology* **111**, 1139–1151.
- Biernacka, A. et al.** 2018. i-BLESS is an ultra-sensitive method for detection of DNA double-strand breaks. *Communications Biology* **1**, 1–9.
- Birch, A., Häusler, A., Vöggtli, M., Krek, W., and Hütter, R.** 1989. Extremely large chromosomal deletions are intimately involved in genetic instability and genomic rearrangements in *Streptomyces glaucescens*. *Molec. Gen. Genet.* **217**, 447–458.
- Bonura, T., and Smith, K.C.** 1975. Enzymatic production of deoxyribonucleic acid double-strand breaks after ultraviolet irradiation of *Escherichia coli* K-12. *J Bacteriol* **121**, 511–517.

- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J.** 2013. Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. *Nat Methods* **10**, 1213–1218.
- Burke, J., Schneider, D., and Westpheling, J.** 2001. Generalized transduction in *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A* **98**, 6289–6294.
- Bury-Moné, S., Thibessard, A., Lioy, V.S., and Leblond, P.** 2023. Dynamics of the *Streptomyces* chromosome: chance and necessity. *Trends in Genetics* **0**.
- Calmann, M.A., Evans, J.E., and Marinus, M.G.** 2005. MutS inhibits RecA-mediated strand transfer with methylated DNA substrates. *Nucleic Acids Research* **33**, 3591–3597.
- Carrasco, B., Serrano, E., Martín-González, A., Moreno-Herrero, F., and Alonso, J.C.** 2019. *Bacillus subtilis* MutS Modulates RecA-Mediated DNA Strand Exchange Between Divergent DNA Sequences. *Frontiers in Microbiology* **10**.
- Castañeda-García, A. et al.** 2017. A non-canonical mismatch repair pathway in prokaryotes. *Nat Commun* **8**.
- Castañeda-García, A., Martín-Blecua, I., Cebrián-Sastre, E., Chiner-Oms, A., Torres-Puente, M., Comas, I., and Blázquez, J.** 2020. Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies. *Sci. Adv.* **6**, eaay4453.
- Cebrián-Sastre, E., Martín-Blecua, I., Gullón, S., Blázquez, J., and Castañeda-García, A.** 2021. Control of Genome Stability by EndoMS/NucS-Mediated Non-Canonical Mismatch Repair.
- Chater, K.F.** 2006. *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philos Trans R Soc Lond B Biol Sci* **361**, 761–768.
- Choufa, C., Tidjani, A.-R., Gauthier, A., Harb, M., Lao, J., Leblond-Bourget, N., Vos, M., Leblond, P., and Bontemps, C.** 2022. Prevalence and mobility of integrative and conjugative elements within a *Streptomyces* natural population. *Frontiers in Microbiology* **13**.
- Choulet, F. et al.** 2006a. Evolution of the Terminal Regions of the *Streptomyces* Linear Chromosome. *Molecular Biology and Evolution* **23**, 2361–2369.
- Choulet, F. et al.** 2006b. Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambofaciens*. *J Bacteriol* **188**, 6599–6610.
- Chung, W.-H., Zhu, Z., Papusha, A., Malkova, A., and Ira, G.** 2010. Defective Resection at DNA Double-Strand Breaks Leads to De Novo Telomere Formation and Enhances Gene Targeting. *PLOS Genetics* **6**, e1000948.
- Cox, M.M.** 1999. Recombinational DNA repair in bacteria and the RecA protein. *Prog Nucleic Acid Res Mol Biol* **63**, 311–366.

- Creze, C., Ligabue, A., Laurent, S., Lestini, R., Laptенок, S.P., Khun, J., Vos, M.H., Czjzek, M., Myllykallio, H., and Flament, D.** 2012. Modulation of the *Pyrococcus abyssi* NucS Endonuclease Activity by Replication Clamp at Functional and Structural Levels *. *Journal of Biological Chemistry* **287**, 15648–15660.
- Crosetto, N. et al.** 2013. Nucleotide-resolution DNA double-strand breaks mapping by next-generation sequencing. *Nat Methods* **10**, 361–365.
- Datta, A., Adjiri, A., New, L., Crouse, G.F., and Jinks Robertson, S.** 1996. Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccaromyces cerevisiae*. *Mol Cell Biol* **16**, 1085–1093.
- Datta, A., Hendrix, M., Lipsitch, M., and Jinks-Robertson, S.** 1997. Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc Natl Acad Sci U S A* **94**, 9757–9762.
- David, S.S., O’Shea, V.L., and Kundu, S.** 2007. Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950.
- De Bragança, S., Dillingham, M.S., and Moreno-Herrero, F.** 2023. Recent insights into eukaryotic double-strand DNA break repair unveiled by single-molecule methods. *Trends Genet* **39**, 924–940.
- Dehé, P.-M., and Gaillard, P.-H.L.** 2017. Control of structure-specific endonucleases to maintain genome stability. *Nat Rev Mol Cell Biol* **18**, 315–330.
- Delpech, F., Collien, Y., Mahou, P., Beaurepaire, E., Myllykallio, H., and Lestini, R.** 2018. Snapshots of archaeal DNA replication and repair in living cells using super-resolution imaging. *Nucleic Acids Research* **46**, 10757–10770.
- Denamur, E. et al.** 2000. Evolutionary Implications of the Frequent Horizontal Transfer of Mismatch Repair Genes. *Cell* **103**, 711–721.
- Dillon, M.M., Sung, W., Sebra, R., Lynch, M., and Cooper, V.S.** 2017. Genome-Wide Biases in the Rate and Molecular Spectrum of Spontaneous Mutations in *Vibrio cholerae* and *Vibrio fischeri*. *Mol Biol Evol* **34**, 93–109.
- Dupuy, P., Sauviac, L., and Bruand, C.** 2019. Stress-inducible NHEJ in bacteria: function in DNA repair and acquisition of heterologous DNA. *Nucleic Acids Res* **47**, 1335–1349.
- Fichtinger-Schepman, A.M., van der Veer, J.L., den Hartog, J.H., Lohman, P.H., and Reedijk, J.** 1985. Adducts of the antitumor drug cis-diamminedichloroplatinum(II) with DNA: formation, identification, and quantitation. *Biochemistry* **24**, 707–713.
- Fischer, G., Holl, A.C., Volff, J.N., Vandewiele, D., Decaris, B., and Leblond, P.** 1998. Replication of the linear chromosomal DNA from the centrally located oriC of *Streptomyces ambofaciens* revealed by PFGE gene dosage analysis. *Research in Microbiology* **149**, 203–210.

- Formosa, T., and Alberts, B.M.** 1986. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell* **47**, 793–806.
- Freese, E.** 1959. The specific mutagenic effect of base analogues on Phage T4. *Journal of Molecular Biology* **1**, 87–105.
- Fressatti Cardoso, R., Martín-Bleuca, I., Pietrowski Baldin, V., Meneguello, J.E., Valverde, J.R., Blázquez, J., and Castañeda-García, A.** 2022. Noncanonical Mismatch Repair Protein NucS Modulates the Emergence of Antibiotic Resistance in Mycobacterium abscessus. *Microbiol Spectr* e02228-22.
- Fukui, K.** 2010. DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids* **2010**, 260512.
- Getino, M., and de la Cruz, F.** 2018. Natural and Artificial Strategies To Control the Conjugative Transmission of Plasmids. *Microbiol Spectr* **6**.
- Grogan, D.W.** 2015. Understanding DNA Repair in Hyperthermophilic Archaea: Persistent Gaps and Other Reasons to Focus on the Fork. *Archaea* **2015**, e942605.
- Gu, Y., Parker, A., Wilson, T.M., Bai, H., Chang, D.-Y., and Lu, A.-L.** 2002. Human MutY Homolog, a DNA Glycosylase Involved in Base Excision Repair, Physically and Functionally Interacts with Mismatch Repair Proteins Human MutS Homolog 2/Human MutS Homolog 6 *. *Journal of Biological Chemistry* **277**, 11135–11142.
- Guarné, A.** 2012. The functions of MutL in mismatch repair: the power of multitasking. *Prog Mol Biol Transl Sci* **110**, 41–70.
- Harfe, B.D., and Jinks-Robertson, S.** 2000. Mismatch repair proteins and mitotic genome stability. *Mutat Res* **451**, 151–167.
- Herman, G.E., and Modrich, P.** 1981. Escherichia coli K-12 clones that overproduce dam methylase are hypermutable. *J Bacteriol* **145**, 644–646.
- Hoff, G. et al.** 2016. Multiple and Variable NHEJ-Like Genes Are Involved in Resistance to DNA Damage in Streptomyces ambofaciens. *Front. Microbiol.* **7**.
- Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A., and Leblond, P.** 2018. Genome plasticity is governed by double strand break DNA repair in Streptomyces. *Scientific Reports* **8**, 5272.
- Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A., and Leblond, P.** 2017. Implication of RuvABC and RecG in homologous recombination in Streptomyces ambofaciens. *Research in Microbiology* **168**, 26–35.
- Honda, M., Okuno, Y., Hengel, S.R., Martín-López, J.V., Cook, C.P., Amunugama, R., Soukup, R.J., Subramanyam, S., Fishel, R., and Spies, M.** 2014. Mismatch repair protein hMSH2–hMSH6 recognizes mismatches and forms sliding clamps within a D-loop recombination intermediate. *Proc Natl Acad Sci U S A* **111**, E316–E325.

- Hopwood, D.A.** 2006. Soil To Genomics: The *Streptomyces* Chromosome. *Annual Review of Genetics* **40**, 1–23.
- Huang, J. et al.** 2005. Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol Microbiol* **58**, 1276–1287.
- Huang, T.-W., and Chen, C.W.** 2006. A *recA* Null Mutation May Be Generated in *Streptomyces coelicolor*. *J Bacteriol* **188**, 6771–6779.
- Ishino, S., Nishi, Y., Oda, S., Uemori, T., Sagara, T., Takatsu, N., Yamagami, T., Shirai, T., and Ishino, Y.** 2016. Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Res* **44**, 2977–2986.
- Ishino, S., Skouloubris, S., Kudo, H., l’Hermitte-Stead, C., Es-Sadik, A., Lambry, J.-C., Ishino, Y., and Myllykallio, H.** 2018. Activation of the mismatch-specific endonuclease EndoMS/NucS by the replication clamp is required for high fidelity DNA replication. *Nucleic Acids Research* **46**, 6206–6217.
- Islam, T., and Josephs, E.A.** 2023. Genome Editing Outcomes Reveal Mycobacterial NucS Participates in a Short-Patch Repair of DNA Mismatches.pdf.
- Jackson, S.P., and Jeggo, P.A.** 1995. DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK. *Trends in Biochemical Sciences* **20**, 412–415.
- Khushboo, Kumar, P., Dubey, K.K., Usmani, Z., Sharma, M., and Gupta, V.K.** 2022. Biotechnological and industrial applications of *Streptomyces* metabolites. *Biofuels, Bioproducts and Biorefining* **16**, 244–264.
- Krokan, H.E., and Bjørås, M.** 2013. Base excision repair. *Cold Spring Harb Perspect Biol* **5**, a012583.
- Kunkel, T.A., and Erie, D.A.** 2015. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annual Review of Genetics* **49**, 291–313.
- Leblond, P., Demuyter, P., Moutier, L., Laakel, M., Decaris, B., and Simonet, J.M.** 1989. Hypervariability, a new phenomenon of genetic instability, related to DNA amplification in *Streptomyces ambofaciens*. *J Bacteriol* **171**, 419–423.
- Leblond, P., Fischer, G., Francou, F.-X., Berger, F., Guérineau, M., and Decaris, B.** 1996. The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Molecular Microbiology* **19**, 261–271.
- Leblond, P., Francou, F.X., Simonet, J.-M., and Decaris, B.** 1990. Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofaciens* strains. *FEMS Microbiology Letters* **72**, 79–88.

- Leblond, P., Redenbach, M., and Cullum, J.** 1993. Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). *J Bacteriol* **175**, 3422–3429.
- LeClerc, J.E., Li, B., Payne, W.L., and Cebula, T.A.** 1996. High Mutation Frequencies Among *Escherichia coli* and *Salmonella* Pathogens. *Science* **274**, 1208–1211.
- Lederberg, J., and Tatum, E.L.** 1946. Gene Recombination in *Escherichia Coli*. *Nature* **158**, 558–558.
- Lee, H., Popodi, E., Tang, H., and Foster, P.L.** 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.
- Lenhart, J.S., Pillon, M.C., Guarné, A., Biteen, J.S., and Simmons, L.A.** 2016. Mismatch repair in Gram-positive bacteria. *Research in Microbiology* **167**, 4–12.
- Lenski, R.E.** 2023. Revisiting the Design of the Long-Term Evolution Experiment with *Escherichia coli*. *J Mol Evol* **91**, 241–253.
- Lestini, R., Laptinok, S.P., Kühn, J., Hink, M.A., Schanne-Klein, M.-C., Liebl, U., and Myllykallio, H.** 2013. Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication. *Nucleic Acids Research* **41**, 10358–10370.
- Lewis, J.S., Jergic, S., and Dixon, N.E.** 2016. The *E. coli* DNA Replication Fork. *Enzymes* **39**, 31–88.
- Lieb, M.** 1991. Spontaneous mutation at a 5-methylcytosine hotspot is prevented by very short patch (VSP) mismatch repair. *Genetics* **128**, 23–27.
- Lieb, M., and Bhagwat, A.S.** 1996. Very short patch repair: reducing the cost of cytosine methylation. *Mol Microbiol* **20**, 467–473.
- Lin, Y.-S., Kieser, H.M., Hopwood, D.A., and Chen, C.W.** 1993. The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Molecular Microbiology* **10**, 923–933.
- Lioy, V.S. et al.** 2021. Dynamics of the compartmentalized *Streptomyces* chromosome during metabolic differentiation. *Nat Commun* **12**, 5221.
- Liti, G., and Louis, E.J.** 2005. Yeast evolution and comparative genomics. *Annu Rev Microbiol* **59**, 135–153.
- Long, H., Sung, W., Miller, S.F., Ackerman, M.S., Doak, T.G., and Lynch, M.** 2014. Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair-deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol* **7**, 262–271.
- Lopez-Canovas, L., Martinez Benitez, M.B., Herrera Isidron, J.A., and Flores Soto, E.** 2019. Pulsed Field Gel Electrophoresis: Past, present, and future. *Analytical Biochemistry* **573**, 17–29.

- Lorenzi, J.-N., Lespinet, O., Leblond, P., and Thibessard, A.** 2021. Subtelomeres are fast-evolving regions of the *Streptomyces* linear chromosome. *Microbial Genomics* **7**.
- Lorenzi, J.-N., Thibessard, A., Lioy, V.S., Boccard, F., Leblond, P., Pernodet, J.-L., and Bury-Moné, S.** 2022. Ribosomal RNA operons define a central functional compartment in the *Streptomyces* chromosome. *Nucleic Acids Res* **50**, 11654–11669.
- Lupski, J.R.** 1998. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. **14**.
- Luria, S.E., and Delbrück, M.** 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**, 491–511.
- Majewski, J., and Cohan, F.M.** 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* **148**, 13–18.
- Malinge, J.-M., Giraud-Panis, M.-J., and Leng, M.** 1999. Interstrand cross-links of cisplatin induce striking distortions in DNA. *Journal of Inorganic Biochemistry* **77**, 23–29.
- Marschner, P., Crowley, D., and Rengel, Z.** 2011. Rhizosphere interactions between microorganisms and plants govern iron and phosphorus acquisition along the root axis – model and research methods. *Soil Biology and Biochemistry* **43**, 883–894.
- Marti, T.M., Kunz, C., and Fleck, O.** 2002. DNA mismatch repair and mutation avoidance pathways. *Journal of Cellular Physiology* **191**, 28–41.
- Martin, P., Dary, A., André, A., Fischer, G., Leblond, P., and Decaris, B.** 1999. Intracolonial polymorphism in the bacterium *Streptomyces ambofaciens* ATCC23877: evidence for a high degree of heterogeneity of the wild type clones. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **430**, 75–85.
- Mascarenhas, J., Sanchez, H., Tadesse, S., Kidane, D., Krisnamurthy, M., Alonso, J.C., and Graumann, P.L.** 2006. *Bacillus subtilis* SbcC protein plays an important role in DNA inter-strand cross-link repair. *BMC Molecular Biology* **7**, 20.
- Maser, R.S., Mirzoeva, O.K., Wells, J., Olivares, H., Williams, B.R., Zinkel, R.A., Farnham, P.J., and Petrini, J.H.J.** 2001. Mre11 Complex and DNA Replication: Linkage to E2F and Sites of DNA Synthesis. *Mol Cell Biol* **21**, 6006–6016.
- Matic, I., Rayssiguier, C., and Radman, M.** 1995. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* **80**, 507–515.
- McGovern, S. et al.** 2016. C-terminal region of bacterial Ku controls DNA bridging, DNA threading and recruitment of DNA ligase D for double strand breaks repair. *Nucleic Acids Res* **44**, 4785–4806.
- Meslet-Cladière, L., Norais, C., Kuhn, J., Briffotiaux, J., Sloostra, J.W., Ferrari, E., Hübscher, U., Flament, D., and Myllykallio, H.** 2007. A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen. *J Mol Biol* **372**, 1137–1148.

- Mizrahi, V., and Andersen, S.J.** 1998. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Molecular Microbiology* **29**, 1331–1339.
- Modrich, P.** 1987. Dna Mismatch Correction. *Annual Review of Biochemistry* **56**, 435–466.
- Nassif, N., Penney, J., Pal, S., Engels, W.R., and Gloor, G.B.** 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol Cell Biol* **14**, 1613–1625.
- Nudler, E.** 2023. Transcription-coupled global genomic repair in E. coli. *Trends in Biochemical Sciences* **48**, 873–882.
- Oliver, A., and Mena, A.** 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. *Clinical Microbiology and Infection* **16**, 798–808.
- On, Y.Y., and Welch, M.** 2021. The methylation-independent mismatch repair machinery in Pseudomonas aeruginosa. *Microbiology (Reading)* **167**, 001120.
- Palaniyandi, S.A., Yang, S.H., Zhang, L., and Suh, J.-W.** 2013. Effects of actinobacteria on plant disease suppression and growth promotion. *Appl Microbiol Biotechnol* **97**, 9621–9636.
- Pannunzio, N.R., Watanabe, G., and Lieber, M.R.** 2018. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry* **293**, 10512–10523.
- Paradkar, A., Trefzer, A., Chakraborty, R., and Stassi, D.** 2003. Streptomyces Genetics: A Genomic Perspective. *Critical Reviews in Biotechnology* **23**, 1–27.
- Pfeiffer, P., Goedecke, W., and Obe, G.** 2000. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis* **15**, 289–302.
- Pillon, M.C., Miller, J.H., and Guarné, A.** 2011. The endonuclease domain of MutL interacts with the β sliding clamp. *DNA Repair (Amst)* **10**, 87–93.
- Pitcher, R.S., Brissett, N.C., and Doherty, A.J.** 2007. Nonhomologous End-Joining in Bacteria: A Microbial Perspective. *Annual Review of Microbiology* **61**, 259–282.
- Polosina, Y.Y., and Cupples, C.G.** 2010. Wot the 'L—Does MutL do? *Mutation Research/Reviews in Mutation Research* **705**, 228–238.
- Possoz, C., Ribard, C., Gagnat, J., Pernodet, J.-L., and Guérineau, M.** 2001. The integrative element pSAM2 from Streptomyces: kinetics and mode of conjugal transfer. *Molecular Microbiology* **42**, 159–166.
- Putnam, C.D.** 2021. Strand discrimination in DNA mismatch repair. *DNA Repair* **105**, 103161.
- Rayssiguier, C., Thaler, D.S., and Radman, M.** 1989. The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. *Nature* **342**, 396–401.

- Ren, B., Kühn, J., Meslet-Cladiere, L., Briffotoux, J., Norais, C., Lavigne, R., Flament, D., Ladenstein, R., and Myllykallio, H.** 2009. Structure and function of a novel endonuclease acting on branched DNA substrates. *EMBO J* **28**, 2479–2489.
- Ricchetti, M., Fairhead, C., and Dujon, B.** 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402**, 96–100.
- Rivera-Flores, I.V., Wang, E.X., and Murphy, K.C.** 2023. Mycobacterium smegmatis NucS-promoted DNA Mismatch Repair involves limited resection by a 5'-3' exonuclease and is independent of homologous recombination and NHEJ.
- Robinson, A., and van Oijen, A.M.** 2013. Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat Rev Microbiol* **11**, 303–315.
- Rocha, E.P.C., Cornet, E., and Michel, B.** 2005. Comparative and Evolutionary Analysis of the Bacterial Homologous Recombination Systems. *PLOS Genetics* **1**, e15.
- Rock, J.M., Lang, U.F., Chase, M.R., Ford, C.B., Gerrick, E.R., Gawande, R., Coscolla, M., Gagneux, S., Fortune, S.M., and Lamers, M.H.** 2015. DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. *Nature Genetics* **47**, 677–681.
- Roelants, P., Konvalinkova, V., Mergeay, M., Lurquin, P.F., and Ledoux, L.** 1976. DNA uptake by Streptomyces species. *Biochim Biophys Acta* **442**, 117–122.
- Sachadyn, P.** 2010. Conservation and diversity of MutS proteins. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **694**, 20–30.
- Schaaper, R.M.** 1993. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. *Journal of Biological Chemistry* **268**, 23762–23765.
- Schwartz, D.C., and Cantor, C.R.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75.
- Sekurova, O.N., Sun, Y.-Q., Zehl, M., Rückert, C., Stich, A., Busche, T., Kalinowski, J., and Zotchev, S.B.** 2021. Coupling of the engineered DNA “mutator” to a biosensor as a new paradigm for activation of silent biosynthetic gene clusters in Streptomyces. *Nucleic Acids Research* **49**, 8396–8405.
- Sharda, M., Badrinarayanan, A., and Seshasayee, A.S.N.** 2020. Evolutionary and Comparative Analysis of Bacterial Nonhomologous End Joining Repair. *Genome Biol Evol* **12**, 2450–2466.
- Shee, C. et al.** 2013. Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. *eLife* **2**, e01222.
- Sinha, K.M., Unciuleac, M.-C., Glickman, M.S., and Shuman, S.** 2009. AdnAB: a new DSB-resecting motor-nuclease from mycobacteria. *Genes Dev* **23**, 1423–1437.

- So, A., Le Guen, T., Lopez, B.S., and Guirouilh-Barbat, J.** 2017. Genomic rearrangements induced by unscheduled DNA double strand breaks in somatic mammalian cells. *The FEBS Journal* **284**, 2324–2344.
- Spies, M., and Fishel, R.** 2015. Mismatch repair during homologous and homeologous recombination. *Cold Spring Harb Perspect Biol* **7**, a022657.
- Springer, B., Sander, P., Sedlacek, L., Hardt, W.-D., Mizrahi, V., Schär, P., and Böttger, E.C.** 2004. Lack of mismatch correction facilitates genome evolution in mycobacteria. *Molecular Microbiology* **53**, 1601–1609.
- Štambuk, S., and Radman, M.** 1998. Mechanism and Control of Interspecies Recombination in *Escherichia coli*. I. Mismatch Repair, Methylation, Recombination and Replication Functions. *Genetics* **150**, 533–542.
- Stephanou, N.C., Gao, F., Bongiorno, P., Ehrt, S., Schnappinger, D., Shuman, S., and Glickman, M.S.** 2007. Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks. *J Bacteriol* **189**, 5237–5246.
- Stracker, T.H.** 2011. The MRE11 complex: starting from the ends.
- Stuttard, C.** 1979. Transduction of Auxotrophic Markers in a Chloramphenicol-producing Strain of *Streptomyces*. *Microbiology* **110**, 479–482.
- Sung, W., Ackerman, M.S., Gout, J.-F., Miller, S.F., Williams, E., Foster, P.L., and Lynch, M.** 2015. Asymmetric Context-Dependent Mutation Patterns Revealed through Mutation–Accumulation Experiments. *Mol Biol Evol* **32**, 1672–1683.
- Suzuki, S., and Kurosawa, N.** 2019. Endonucleases responsible for DNA repair of helix-distorting DNA lesions in the thermophilic crenarchaeon *Sulfolobus acidocaldarius* in vivo. *Extremophiles* **23**, 613–624.
- Szafran, M.J., Małecki, T., Strzałka, A., Pawlikiewicz, K., Duława, J., Zarek, A., Kois-Ostrowska, A., Findlay, K.C., Le, T.B.K., and Jakimowicz, D.** 2021. Spatial rearrangement of the *Streptomyces venezuelae* linear chromosome during sporogenic development. *Nat Commun* **12**, 5222.
- Takemoto, N., Numata, I., Su’etsugu, M., and Miyoshi-Akiyama, T.** 2018. Bacterial EndoMS/NucS acts as a clamp-mediated mismatch endonuclease to prevent asymmetric accumulation of replication errors. *Nucleic Acids Research* **46**, 6152–6165.
- Tang, W., Zhou, H., and Li, W.** 2015. Silver and Cyanine Staining of Oligonucleotides in Polyacrylamide Gel. *PLoS One* **10**, e0144422.
- Tham, K.-C., Hermans, N., Winterwerp, H.H.K., Cox, M.M., Wyman, C., Kanaar, R., and Lebbink, J.H.G.** 2013. Mismatch repair inhibits homeologous recombination via coordinated directional unwinding of trapped DNA structures. *Mol Cell* **51**, 326–337.

- Tham, K.-C., Kanaar, R., and Lebbink, J.H.G.** 2016. Mismatch repair and homeologous recombination. *DNA Repair* **38**, 75–83.
- Thibessard, A., Bertrand, C., Bartlett, E.J., Doherty, A.J., Bruand, C., Leblond, P., and Lecointe, F.** 2021. DNA repair | Nonhomologous End-Joining in Bacteria, pp. 289–295. *Encyclopedia of Biological Chemistry III*. Elsevier.
- Thibessard, A., Bertrand, C., Hiblot, J., Piotrowski, E., and Leblond, P.** 2015. Construction of pDYN6902, a new *Streptomyces* integrative expression vector designed for cloning sequences interfering with *Escherichia coli* viability. *Plasmid* **82**, 43–49.
- Thibessard, A., and Leblond, P.** 2014. Subtelomere Plasticity in the Bacterium *Streptomyces*, pp. 243–258. *In* Louis, E.J. and Becker, M.M., eds, *Subtelomeres*. Springer, Berlin, Heidelberg.
- Thoma, L., Dobrowinski, H., Finger, C., Guezguez, J., Linke, D., Sepulveda, E., and Muth, G.** 2015. A Multiprotein DNA Translocation Complex Directs Intramyceial Plasmid Spreading during *Streptomyces* Conjugation. *mBio* **6**, e02559-14.
- Thoma, L., and Muth, G.** 2016. Conjugative DNA-transfer in *Streptomyces*, a mycelial organism. *Plasmid* **87–88**, 1–9.
- Thomas, C.M., and Nielsen, K.M.** 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**, 711–721.
- Tidjani, A.-R., Bontemps, C., and Leblond, P.** 2020. Telomeric and sub-telomeric regions undergo rapid turnover within a *Streptomyces* population. *Sci Rep* **10**, 7720.
- Tidjani, A.-R., Lorenzi, J.-N., Toussaint, M., van Dijk, E., Naquin, D., Lespinet, O., Bontemps, C., and Leblond, P.** 2019a. Genome Sequences of 11 Conspecific *Streptomyces* sp. Strains. *Microbiol Resour Announc* **8**.
- Tidjani, A.-R., Lorenzi, J.-N., Toussaint, M., van Dijk, E., Naquin, D., Lespinet, O., Bontemps, C., and Leblond, P.** 2019b. Massive Gene Flux Drives Genome Diversity between Sympatric *Streptomyces* Conspecifics. *mBio* **10**.
- Tong, Y., Whitford, C.M., Blin, K., Jørgensen, T.S., Weber, T., and Lee, S.Y.** 2020. CRISPR–Cas9, CRISPRi and CRISPR-BEST-mediated genetic manipulation in streptomycetes. *Nat Protoc* **15**, 2470–2502.
- Tong, Y., Whitford, C.M., Robertsen, H.L., Blin, K., Jørgensen, T.S., Klitgaard, A.K., Gren, T., Jiang, X., Weber, T., and Lee, S.Y.** 2019. Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proceedings of the National Academy of Sciences* **116**, 20366–20375.
- Uchida, T., Miyawaki, M., and Kinashi, H.** 2003. Chromosomal arm replacement in *Streptomyces griseus*. *J Bacteriol* **185**, 1120–1124.

- Viaene, T., Langendries, S., Beirinckx, S., Maes, M., and Goormachtig, S.** 2016. Streptomyces as a plant's best friend? *FEMS Microbiol Ecol* **92**, fiw119.
- Vogelmann, J., Ammelburg, M., Finger, C., Guezguez, J., Linke, D., Flötenmeyer, M., Stierhof, Y.-D., Wohlleben, W., and Muth, G.** 2011. Conjugal plasmid transfer in Streptomyces resembles bacterial chromosome segregation by FtsK/SpoIIIE. *EMBO J* **30**, 2246–2254.
- Volff, J.N., and Altenbuchner, J.** 1998. Genetic instability of the Streptomyces chromosome. *Mol Microbiol* **27**, 239–246.
- Vulić, M., Dionisio, F., Taddei, F., and Radman, M.** 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc Natl Acad Sci U S A* **94**, 9763–9767.
- Vurukonda, S.S.K.P., Giovanardi, D., and Stefani, E.** 2018. Plant Growth Promoting and Biocontrol Activity of Streptomyces spp. as Endophytes. *Int J Mol Sci* **19**, 952.
- Weller, G.R. et al.** 2002. Identification of a DNA Nonhomologous End-Joining Complex in Bacteria. *Science* **297**, 1686–1689.
- Wenner, T., Roth, V., Fischer, G., Fourier, C., Aigle, B., Decaris, B., and Leblond, P.** 2003. End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in Streptomyces ambifaciens. *Molecular Microbiology* **50**, 411–425.
- Widenbrant, E.M., Tsai, H.-H., Chen, C.W., and Kao, C.M.** 2007. Streptomyces coelicolor Undergoes Spontaneous Chromosomal End Replacement. *Journal of Bacteriology* **189**, 9117–9121.
- Worth, L., Clark, S., Radman, M., and Modrich, P.** 1994. Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proceedings of the National Academy of Sciences* **91**, 3238–3241.
- Yao, N.Y., and O'Donnell, M.E.** 2018. Replication fork convergence at termination: A multistep process. *Proceedings of the National Academy of Sciences* **115**, 237–239.
- Yao, N.Y., Schroeder, J.W., Yurieva, O., Simmons, L.A., and O'Donnell, M.E.** 2013. Cost of rNTP/dNTP pool imbalance at the replication fork. *Proc Natl Acad Sci U S A* **110**, 12942–12947.
- Yu, H., Eritja, R., Bloom, L.B., and Goodman, M.F.** 1993. Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. *Journal of Biological Chemistry* **268**, 15935–15943.
- Zahrt, T.C., Mora, G.C., and Maloy, S.** 1994. Inactivation of mismatch repair overcomes the barrier to transduction between Salmonella typhimurium and Salmonella typhi. *J Bacteriol* **176**, 1527–1529.

- Zhang, L., Nguyen, H.C., Chipot, L., Piotrowski, E., Bertrand, C., Thibessard, A., and Leblond, P.** 2014. The *adnAB* Locus, Encoding a Putative Helicase-Nuclease Activity, Is Essential in *Streptomyces*. *J Bacteriol* **196**, 2701–2708.
- Zhang, L., Shi, H., Gan, Q., Wang, Y., Wu, M., Yang, Z., Oger, P., and Zheng, J.** 2020a. An alternative pathway for repair of deaminated bases in DNA triggered by archaeal NucS endonuclease. *DNA Repair* **85**, 102734.
- Zhang, Z., Du, C., de Barsy, F., Liem, M., Liakopoulos, A., van Wezel, G.P., Choi, Y.H., Claessen, D., and Rozen, D.E.** 2020b. Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. *Science Advances* **6**, eaay5781.
- Zhang, Z., Shitut, S., Claushuis, B., Claessen, D., and Rozen, D.E.** 2022. Mutational meltdown of putative microbial altruists in *Streptomyces coelicolor* colonies. *Nat Commun* **13**, 2266.

Résumé

Les *Streptomyces*, bactéries ubiquistes des sols forestiers, présentent une grande plasticité génomique. Ces bactéries possèdent un chromosome linéaire hautement compartimenté avec, en son centre, une région conservée au travers du genre, bordée par des extrémités ou bras chromosomiques particulièrement variables et sujets à la recombinaison. La réparation des cassures double-brin (DSB), dommage le plus délétère pour une cellule vivante, constitue un moteur majeur de la plasticité et de l'évolution du génome chez les *Streptomyces*. Une source potentielle de DSB a été découverte récemment : l'endonucléase NucS. Cette dernière joue un rôle clé au sein du système de correction des mésappariements (MMR) non-canonique, présent chez les archées et plus sporadiquement chez les bactéries avec les actinobactéries.

L'objectif de ce travail est de comprendre si l'activité de NucS est impliquée dans la stimulation de la recombinaison chez les *Streptomyces*. Nos investigations ont révélé que la déplétion en NucS entraîne une augmentation significative du taux de mutation et un phénotype colonial marqué. Les tests d'activité *in vitro* montrent que NucS génère une DSB au site des mésappariements G/T, G/G et T/T. Son activité est augmentée en présence du facteur de processivité de la réplication β -clamp suggérant que NucS cible les mésappariements post-réplicatifs. Une expérience d'accumulation de mutations sur le long-terme (60 cycles de croissance) montre que NucS corrige plus favorablement les mésappariements conduisant à une transition qu'à une transversion. Une augmentation graduelle du nombre de transitions depuis l'origine de réplication vers les extrémités du chromosome a été caractérisée. Ce résultat original révèle que NucS est davantage sollicitée dans les régions terminales, et pourrait favoriser la recombinaison en générant des DSB. L'augmentation de la fréquence de réarrangements dans les régions terminales dans un contexte où la fréquence des mésappariements réplicatifs est accrue renforce cette hypothèse. En effet, les mésappariements sont le substrat de NucS et la cassure double brin résultant de son activité induirait la recombinaison.

Ces résultats ouvrent la voie de la compréhension de ce mécanisme paradoxal de maintien de l'intégrité du génome ; pourquoi casser le génome pour mieux le réparer ? La diversification génétique résultant de la plasticité du génome pourrait justifier de la sélection d'un tel mécanisme de réparation. Enfin, la diversification du métabolisme spécialisé chez ces organismes du sol revêt un intérêt majeur pour leur adaptation à l'écosystème, et aussi pour l'identification de nouvelles molécules d'intérêt biotechnologique.

Abstract

Streptomyces are ubiquitous bacteria in forest soils known for their remarkable genomic plasticity. These bacteria possess a highly compartmentalized linear chromosome, with a central region shared across the genus. This central region is bordered by variable chromosomal ends or arms that are prone to recombination. The repair of double-strand breaks (DSB), the most harmful cellular damage, is a major driver of genome plasticity and genome evolution in *Streptomyces*. Recently, a potential source of DSB has been identified: the endonuclease NucS, which plays a pivotal role in the non-canonical mismatch repair system (MMR) present in archaea and sporadically in bacteria, particularly in actinobacteria.

The objective of this study is to determine whether NucS activity is involved in stimulating recombination in *Streptomyces*. Our investigations unveil that NucS depletion results in a significant increase of the mutation rate and in a marked colonial phenotype. *In vitro* activity tests showed that NucS generates DSB at G/T, G/G, and T/T mismatches. Its activity is enhanced in the presence of the replication processivity factor β -clamp, indicating that NucS targets post-replicative mismatches. A long-term mutation accumulation experiment conducted over 60 sporulation cycles have demonstrated that NucS is more prone to correct mismatches leading to transitions than transversions. A crescent-increase in the number of transitions from the replication origin towards the chromosome ends is observed. This original finding suggests that NucS is more active in terminal regions and could promote recombination by generating DSB. The increased frequency of rearrangements in the terminal regions in a mutant exhibiting an elevated frequency of replicative mismatches supports this hypothesis.

These findings advance our understanding of this paradoxical mechanism for maintaining genome integrity: why break the genome in order to repair it better? The genetic diversification arising from genome plasticity could explain the selection of such a repair mechanism. Finally, the diversification of specialized metabolism in these soil-dwelling bacteria is of great interest both for their adaptation to the ecosystem and for the identification of new molecules with biotechnological interests.