



HAL
open science

CRISPR/Cas9 chez les mollicutes : Du système naturel aux outils d'ingénierie de génome

Thomas Ipoutcha

► **To cite this version:**

Thomas Ipoutcha. CRISPR/Cas9 chez les mollicutes : Du système naturel aux outils d'ingénierie de génome. Immunologie. Université de Bordeaux, 2022. Français. NNT : 2022BORD0029 . tel-03633779

HAL Id: tel-03633779

<https://theses.hal.science/tel-03633779>

Submitted on 7 Apr 2022

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.

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTE
SPÉCIALITÉ MICROBIOLOGIE – IMMUNOLOGIE

Par Thomas IPOUTCHA

**CRISPR/Cas9 chez les mollicutes : Du système naturel aux
outils d'ingénierie de génome**

Sous la direction de Pascal Sirand-Pugnet

Unité Mixte de Recherche 1332 Biologie du Fruit et Pathologie INRAE – Bordeaux

Soutenue le 03/02/2022

Membres du jury :

Mme BÉBÉAR, Cécile, PU-PH à l'Université de Bordeaux

M. BIKARD, David, Chargé de Recherche, Institut Pasteur, Paris

Mme CITTI, Christine, Directeur de Recherche à l'ENVT, Toulouse

M. FREMAUX, Christophe, Cadre scientifique à l'IFF, Dangé-Saint-Romain

Mme LARTIGUE, Carole, Chargée de Recherche, INRAE, Villenave d'Ornon

M. SIRAND-PUGNET, Pascal, Maître de conférences à l'Université de Bordeaux

Président

Rapporteur

Rapporteur

Examineur

Invité

Directeur de thèse

RÉSUMÉ

CRISPR/Cas9 chez les Mollicutes : Du système naturel aux outils d'ingénierie de génome

Les systèmes CRISPR/Cas sont largement représentés chez les bactéries et les archées et sont à la base de mécanismes de défense adaptatifs contre l'ADN et l'ARN exogène. De par son intérêt comme outil génétique, le système CRISPR de type II de *Streptococcus pyogenes* a été particulièrement étudié. Il est composé principalement d'une endonucléase Cas9 et d'un sgRNA qui a pour rôle de guider Cas9 jusqu'à une séquence cible, la spécificité de reconnaissance dépendant également d'une séquence NGG située en aval appelée PAM. Les mollicutes sont des bactéries minimales, dépourvues de paroi et qui utilisent, pour la plupart, un code génétique alternatif. Elles possèdent de petits génomes riches en AT (~70-75%) allant de 0,58 à 2,2 Mpb. La plupart d'entre elles ont un mode de vie parasitique et sont des pathogènes d'une grande diversité d'hôtes (plantes et animaux y compris humains). Chez la plupart des mollicutes, les possibilités d'ingénierie génomique restent limitées car la recombinaison homologue est peu efficace. Le but de ma thèse a été (1) de caractériser les systèmes CRISPR/Cas9 des mollicutes, (2) de développer un outil endogène CRISPR/Cas9 pour l'ingénierie de génome de mollicutes et (3) d'étendre cette boîte à outils chez certaines espèces d'intérêt pour lesquelles aucun outil génétique de précision n'était disponible.

La première partie de mon travail de thèse était l'étude globale des systèmes CRISPR chez les mollicutes. Une reconstruction phylogénétique a permis de mettre en évidence une origine commune pour tous les systèmes CRISPR de mycoplasmes et deux origines pour les systèmes de spiroplasmales. L'étude structurale de la protéine Cas9 des mollicutes montre une forte proximité structurale avec la protéine Cas9 de *Staphylococcus aureus* à l'exception du domaine PI qui intervient dans l'interaction avec la séquence PAM. Ceci suggère que les séquences PAM des systèmes de mollicutes sont sans doute très différentes. Le système CRISPR/Cas9 de *Mycoplasma gallisepticum* (*Mgal*), un pathogène d'oiseau, a ensuite été l'objet d'une étude approfondie. Le système naturel a été démontré comme fonctionnel *in vivo* et le système minimal composé de la *Mgal*Cas9 et d'un sgRNA a permis d'induire un clivage de l'ADN *in vitro*. Une différence de séquence PAM consensus a été observée entre deux souches de *Mgal* isolées de poulet et de pinson et a pu être mise en corrélation avec le récent changement d'hôte du pathogène et l'émergence de celui-ci dans les populations de pinsons d'Amérique. Un plasmide « all-in-one » contenant le système a été conçu, et la fonctionnalité du système comme outil précis de clivage a été démontré dans différentes espèces, notamment *Mgal*, *Mycoplasma capricolum* (*Mcap*), *Mycoplasma pulmonis* et *Mycoplasma mycoides*. Cet outil de contre sélection a été appliqué à *Mcap* afin de curer son génome d'un élément mobile intégré de 23 kpb (ICE). Finalement, pour pallier la faible efficacité de recombinaison des mycoplasmes, deux nouveaux outils d'ingénierie génomique ont été mis au point. Un outil « CRISPR Base Editor », composé d'une protéine Cas9 désactivée et fusionnée à une protéine de type déaminase, qui permet de faire des modifications de base de manière très ciblée. La fonctionnalité de cet outil, très efficace et simple d'utilisation, a été validée dans 3 espèces pathogènes majeures de mycoplasmes. Le second outil est basé sur l'importation d'un système de recombinaison RecET issu de phages. Chez *Mgal*, cet outil d'ingénierie de génome permet ainsi pour la première fois de réaliser des modifications précises à plus grande échelle (délétions, insertions et remplacements).

En conclusion, nos travaux ont concerné l'étude fonctionnelle des systèmes CRISPR des mycoplasmes et le développement d'outils d'ingénierie de génomes chez ces mêmes bactéries. Ce travail ouvre de manière significative les possibilités d'études et d'utilisation des mycoplasmes pour des applications biotechnologiques.

Mots clés : CRISPR, Ingénierie de génome, Mycoplasme, *Mycoplasma gallisepticum*

UMR 1332 Biologie du Fruit et Pathologie - Equipe Mollicutes

Centre INRA Bordeaux-Aquitaine, Bâtiment IBVM 71 Avenue Edouard Bourlaux, CS 20032
33882 Villenave d'Ornon Cedex, France

CRISPR/Cas9 in mollicutes: From natural systems towards tools for genome engineering

CRISPR/Cas systems are widely represented in bacteria and archaea and provide them an adaptative defense mechanism against exogenous DNA or RNA. Because of its wide use as a genome engineering tool, type II CRISPR system from *Streptococcus pyogenes* has been extensively studied. The core system includes the endonuclease Cas9 and one sgRNA that drives Cas9 on a target site with a recognition specificity depending on the NGG Protospacer Adjacent Motif (PAM). Mollicutes are minimal bacteria that have no cell wall and most of them use an alternative genetic code. Their genomes are AT rich (~70-75%) with a size ranging from 0.58 to 2.2 Mbp. Most of them have a parasitic lifestyle and are pathogens of a wide host diversity including humans, other animals and plants. For most mollicutes species, genome engineering currently remains challenging mainly because homologous recombination is poorly efficient. The aim of my thesis was to (1) characterize CRISPR/Cas9 systems of mollicutes, (2) develop a genome engineering tool based on endogenous CRISPR/Cas9 system and, (3) extend the toolbox for some mycoplasmas of veterinary interest for which any precise genetics tool was available.

The first part of my thesis work was a global survey of CRISPR systems in mollicutes. Complete or degraded CRISPR systems were found in 21 out of the 52 representatives mollicutes species included in the study. Phylogenetic reconstruction indicated a probable unique origin of mycoplasma CRISPR/Cas9 systems while two different origins were predicted for systems present in spiroplasmas. Comparison of the predicted structures Cas9 from mollicutes and *Staphylococcus aureus* showed a strong structural proximity except for the PI domain that is involved in the interaction with the PAM sequence. This suggests probable differences of PAM recognition specificity. We next focused on the CRISPR/Cas9 system of the bird pathogen *Mycoplasma gallisepticum* (*Mgal*). The functionality of the system was demonstrated *in vivo* and a minimal system including Cas9 and a sgRNA was designed and evaluated *in vitro*. An extensive evaluation of the PAM motifs recognized by two *Mgal*Cas9 from two distinct strains of *Mgal* isolated in chicken and house-finch was conducted using an *in vitro* cleavage assay coupled with a deep sequencing strategy. A difference of PAM consensus was observed between the two strains and was correlated with a recent host jump from chicken to American house-finch. Next, a set of "all in one" plasmids was constructed and *Mgal*Cas9 was successfully used as a tool for precise DNA cleavage in different mycoplasma species: *Mgal*, *Mycoplasma capricolum* subsp. *capricolum* (*Mcap*), *Mycoplasma pulmonis* et *Mycoplasma mycoides* subsp. *mycoides*. Using this counter selection tool, we were able to cure the *Mcap* genome from a 23 kbp auto-excisable genetic mobile element (ICE). In order to overcome the low efficiency of natural recombination in mycoplasmas that limits genome engineering possibilities, I developed two new generations tools. The first one is a "CRISPR Base Editor" consisting in an inactivated Cas9 protein that is fused to a deaminase protein (*pmcDA1*). This hybrid enzyme allowed very specific base modification in the mycoplasma genome. This simple tool was demonstrated to be highly efficient in three main pathogenic mycoplasmas, opening new possibilities for their study. The second tool is based on the phage recombination system RecET that was used to drive homologous recombination in mycoplasma. For the very first time in *Mgal*, this tool allowed us to perform large scale and precise modifications (Insertion, deletion and replacement).

In conclusion, my PhD thesis was dedicated to the studies of CRISPR system of mycoplasma and the development of new precise genome engineering tools for these bacteria. This work opened unprecedented possibilities for the study of pathogenic mycoplasmas and further biotechnological applications.

Key words : CRISPR, Genome engineering, Mycoplasma, *Mycoplasma gallisepticum*

UMR 1332 Biologie du Fruit et Pathologie - Equipe Mollicutes

Centre INRA Bordeaux-Aquitaine, Bâtiment IBVM 71 Avenue Edouard Bourlaux, CS 20032
33882 Villenave d'Ornon Cedex, France

REMERCIEMENTS

"The future belongs to those who believe in the beauty of their dreams."

- Eleanor Roosevelt

REMERCIEMENTS :

Je tiens tout d'abord à remercier Madame la **Professeure Cécile Bébéar**, Directrice de l'USC EA3671, Infections humaines à mycoplasmes et chlamydiae de l'Université de Bordeaux, de me faire l'honneur de présider mon jury de thèse. Je souhaite remercier Monsieur le **Docteur David Bikard**, Chargé de recherche à l'Institut Pasteur de Paris, et Madame la **Docteure Christine Citti**, Directrice de Recherche à l'INVT de Toulouse, pour avoir accepté d'examiner et de juger ce manuscrit. Je remercie également le **Docteur Christophe Fremaux**, cadre scientifique chez IFF d'avoir accepté de faire partie de ce jury de thèse.

Je remercie infiniment Monsieur le **Docteur Pascal Sirand-Pugnet**, Maître de conférences à l'université de Bordeaux, qui dirige également l'équipe Mollicutes au sein de l'Unité Mixte de Recherche 1332, Biologie du Fruit et Pathologie (INRAE). Merci de m'avoir permis de faire une thèse au sein de l'équipe, merci de m'avoir encadré, merci surtout de m'avoir fait confiance et laissé autant d'autonomie, d'avoir laissé libre cours à mes idées et de m'avoir permis de les approfondir cela a été tellement enrichissant. Merci également, malgré tout le travail que tu avais et l'emploi du temps horriblement surchargé, d'avoir consacré le temps nécessaire, pour m'aider et m'aiguiller, pour m'écouter et me conseiller, je t'en serai éternellement reconnaissant, Merci Pascal.

Je tiens ensuite à m'adresser à Madame la **Docteure Carole Lartigue**, Chargée de recherche à l'INRAE, ma compatriote Basque, sans qui je n'aurais très certainement jamais intégré l'équipe. Je te remercie du fond du cœur pour ton sourire constant et ta bienveillance, merci pour tous tes (très) nombreux conseils, pour toutes nos discussions et nos réflexions passionnées, c'est un honneur d'avoir pu te côtoyer et travailler à tes côtés. Tu es une personne brillante et l'on ne peut que s'améliorer à ton contact, j'espère qu'un jour nous pourrons développer ensemble des projets au sein de l'institut Basque de la recherche... Merci Carole.

Je tiens à remercier le **Professeur Alain Blanchard**, directeur de l'ISVV (Institut des sciences de la vigne et du vin), pour ses conseils avisés lors des data clubs et pour son aide très importante au moment de l'écriture des publications. Merci pour toutes tes histoires et anecdotes que tu nous racontes et qui sont toujours super à entendre. Merci Alain.

Tant que j'y suis avec les chefs, je remercie le **Docteur Yonathan Arfi**, Maître de conférences à l'université de Bordeaux, une personne qui m'a appris plus de bêtises et de Rap du sale que de connaissances scientifiques. Tu es un vrai puit de connaissances (Wikinathan/Wikipénible) certaines bien plus utiles que d'autres ... Je te remercie donc pour tout, et je te souhaite plein de bonnes choses pour la suite avec tes futurs quadruplés.

Je tiens également à remercier Mesdames la Docteure **Camille Bonneaud** et la **Docteure Patricia Thébault** pour leur aide respective et leur contribution dans la construction de mes travaux et papiers. Vous avez toujours été d'une grande aide et d'une grande sympathie.

Pour les autres je vais commencer par remercier tous les **Arosatar de la team Rugby-Piscine** avec qui j'ai beaucoup décompressé le weekend lors de soirées plus que mémorables, qui parfois se terminent dans une brouette... il fallait bien ça pour mieux travailler après... J'espère que l'on continuera encore pendant des années à se désaltérer allègrement !

Merci à tous mes amis de collège et lycée et même fac pour certains, je ne les citerais pas tous, mais particulièrement **Etienne, Elise, Marion, Imanol** et **Maxime** que je n'ai pas perdus de vu, même si forcément on se côtoie moins qu'avant, c'est toujours un bon moment de vous revoir !!

Merci la team MPG et les plus grands pipoteurs qui existent : **Adrien, Yadaly, Antonin, Romain, Karl, Gauthier, Toto et Fab**. Vous êtes de grands malades mais c'est pour ça que je vous aimeeee, de nombreuses soirées et week-ends mémorables pour quand même en arriver là, on se remémorera tout ça encore pendant longtemps et on en écrira de nombreuses autres !! **Adrien** et **Yadaly** surtout ne changez rien, vous êtes des vrais dingues et des vrais gentils avec qui on est sûre de toujours rigoler et pourtant des futurs Docteurs !

Docteur André, même si je t'ai déjà remercié, tu as été présent avec moi au Québec et mon colocataire dans 2 appartements, merci à toi pour toutes les barres que l'on a pu avoir en tant que colocataires, toutes les cuites à 2 auxquelles on ne s'attendait pas et à l'entraide dans les études qu'on a pu se fournir, sans toi je ne sais pas si les résultats auraient été les mêmes pour les examens et les concours! Donc je suis aussi là en partie grâce à toi, Mercii !

Merci **Carine** pour ta joie de vivre et ta façon d'être, tu es quelqu'un de formidable et je sais qu'on est loin d'en avoir fini nous deux !!

Merci encore **Romain**, une personne formidable, un parieur hors pair, le tombeur de ses dames, c'est avec toi que j'ai dû vivre le plus de moments formidables, et ce n'est pas fini...

Merci **Martin**, mon ami, avec ta longue carrière en fac de bio et tes belles performances, on a tellement vécu de chose et de soirées incroyables que je ne me lasserai jamais de remémorer ces braves souvenirs. Et là aussi ça va continuer j'en suis sûr !

Merci à Monsieur **Greg Delas**, l'Ingénieur en Cryptomonnaie, un très mauvais joueur à fifa et malgré tout un phénomène au basketball (tu as même réussi à casser le genou de Martin...), avec toi aussi on en a fait des belles et j'espère que cela va continuer encore longtemps ;

Et bien évidemment Merci à tous les autres, rencontrés pendant ces belles années fac à Bordeaux et qui sont devenus de très bons copains :

- Ceux qui font partie de la famille des fêtes d'Arrossa, notamment **Pierre, Melvin, Chloe D, Elise B**, vous qui m'avez vu dans de bons états comme dans les plus mauvais, avec qui j'ai toujours passé des moments excellents, Merciiii.

- Ceux que j'ai pu rencontrer grâce aux différents amis et qui sont devenus aussi des potes de bringues (la plupart des Croques Mais, mais pas que !)

J'espère n'avoir oublié personne sur cette période, si c'est le cas ne m'en voulez pas imaginez bien que je suis fatigué au moment où j'écris ça.... ;)

Je passe finalement aux personnes rencontrées lors de ma jeune carrière scientifique :

Dans un premier temps tous **les membres du laboratoire de la Professeure Cécile Bébéar** que j'ai rencontré lors de mon stage de M1, mais particulièrement **Chloé**, tu m'as en partie encadré, tu es une bonne personne, on a beaucoup parlé, je me suis souvent référé à toi pour des conseils, des corrections (d'orthographe ^^) et on a surtout beaucoup rigolé. Merci pour ta gentillesse et tes conseils.

Ensuite un paragraphe entier est nécessaire pour le phénomène, el mago, je parle bien évidemment du **Docteur Fabien Curtain**, qui aurait préféré une page entière de remerciement pour lui tout seul (Mais je fais plus court pour ton bien afin d'éviter de te faire saigner les yeux). Mon maitre de stage dans un premier temps, mon poteau alcoolique dans un second temps, mon collègue dans un troisième temps et pour finir mon ami (et futur marié aux yeux de Las Vegas), tu as eu une grosse influence sur mon parcours, tu m'as d'abord choisi pour le stage de M1, puis donné envie de faire de la recherche, tu m'as ensuite recommandé auprès de Carole pour le stage de M2 et la suite tu l'as connue : des bringues, des voyages, des escapes, des idées (oui on travaillait quand même !) et des branlés au squash. C'est grâce à toi si je suis autant épanoui dans cette vie de chercheur que ce soit côté travail ou personnel (bon d'accord je te le concède, c'est grâce à toi que j'ai trouvé ma chère et tendre....). Merci infiniment, je n'ai aucun doute sur le fait que notre route ne va pas se séparer, peu importe les kilomètres.

J'en arrive aux membres du laboratoire de l'INRAE, la meilleure équipe qu'il soit... Jamais on ne pourra retrouver une équipe comme celle-là, et pour beaucoup vous êtes maintenant plus que de simples collègues. Merci à toi, **Vincent Talenton**, pour toutes les soirées chez toi, pour ta bonne humeur et ton mauvais humour ;). Merci **Estelle** pour ta gentillesse constante, tes discussions (surtout celles sur les fleurs ;)) et ta capacité à me donner faim... Bien évidemment j'en arrive à ma dernière collègue de bureau, la personne la plus importante du labo, sans laquelle il ne tournerait pas, Merci **Géraldine** (première fois que je ne vais pas juste dire GG), la best, tu nous fait gagner à tous du temps par ton travail, ta rigueur et

ton efficacité, tu n'es pas seulement importante pour tes compétences tu es une personne gentille, drôle et attentionnée, tu es toujours là quand on a besoin, que ce soit pour le travail ou le perso, tu mérites toutes les bonnes choses que tu as dans ta vie et bien plus encore, en espérant que l'euro millions soit pour toi et ta magnifique famille un beau jour ! Mon passage dans ton bureau restera la meilleure chose du laboratoire et je me souviendrais toujours de tes/nos concerts et nos ragots dans le P2 !! Merci **Vincent Baby** pour ton humour (très) gras et ton alcoolisme chronique, pour ton bel accent et tes poutines et merci aussi à **Cat** pour ta gentillesse et tes expressions, à vous deux je n'ai qu'une dernière chose à dire, la plus belle expression canadienne : « Bouffe moi le *** ***** ». Merci **Laure bataille**, tu es un rayon de bonne humeur et la reine de la protéine, tu auras essayé de me convertir à la protéine à coup de « PETIT TOTO », tu es une personne formidable, ne change jamais ! Des bisouuuus. Merci aux nouveaux arrivés, **Djudju le BG, Franceskita, Bastos et Arthouritoze**, vous redonnez un souffle de rigolade au lab et cette année passée à votre côté aura été fort sympathique. Merci également aux autres membres passés au laboratoire, **Flavie, Mamadou, Paul, Alexis, Luis**, Merci à tous « ceux du hauts », **Christophe** (tu m'as quand même appris à jongler), **Mpi, Sybille, Pascal, Sandrine, Nathalie, Sylvie, Delphine, Xavier, Thierry, Denis, Jean Said, Laure, Jean-Paul** et Merci aux autres indispensables du lab, **Laurence, Patrice**, et bien sûr les trois bests **Annie, Isa et Martine** (avec notamment une soirée mémorable gravée dans ma mémoire).

Enfin pour finir au laboratoire, (non je ne t'ai pas oubliée...) Madame la future **Docteur Gabrielle Guesdon**, c'est à tes côtés que 3 ans de thèse se sont passés, un binôme hors pair pour faire des bêtises, rigoler et picoler. Ton amour pour le Patx et les produits locaux basques nous ont permis de vivre de nombreuses choses en trois ans et je suis très heureux que les étoiles aient pu s'aligner. Je te souhaite le meilleur et espère te mettre dans des cartons encore longtemps.

Je veux terminer par remercier les personnes les plus importantes dans ma vie, ma famille, **Aita et Ama** depuis le départ vous m'avez apporté votre amour, votre soutien, vous m'avez permis de faire mes propres choix dans les études et si j'en suis arrivé ici aujourd'hui c'est grâce à vous. Ensuite mes sœurs **Pauline et Charlotte**, qui ont cru en ce que je faisais et qui m'ont soutenu (vous étiez plus habituée à me voir bourré que sérieux, aujourd'hui jour de ma thèse j'espère vous démontrer que je ne suis pas seulement un Astoa). À tous les quatre, malgré qu'il soit difficile de comprendre mon travail, vous vous êtes intéressés à ce que je faisais (pour résumer des OGMs) et ça fait plaisir. Je remercie également les derniers arrivants dans la famille **Franck** et ma petite nièce et filleule **Rose Henriette** d'amener une vague de nouveautés dans notre petite famille.

LISTE DES ABRÉVIATIONS

LISTE DES ABRÉVIATIONS :

aacA-aphD	Gène de résistance à la gentamycine
AAV	Virus Adéno-Associé
A, C, T, G, U	Adénine, cytosine, thymine, guanine, uracile
Abi	Abortive infection system
ABEs	Systèmes Adenine Base Editor
Acr	Anti-CRISPR
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNt	ARN transfert
BIM	Bacteriophage insensitive mutants
BEs	Systèmes Base Editor
BER	Voie de réparation d'excision de base
BRD	Bovine respiratory disease
BS	Biologie de Synthèse (ou Biologie Synthétique)
CBEs	Systèmes Cytosine Base Editor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas	CRISPR-associated gene
nCas9	Cas9 nickase
dCas9	Cas9 dead
cat	Gène de résistance au chloramphénicol
crRNA	ARN du CRISPR array
DR	Direct Repeat
EDTA	Ethylène Diamine Tétra Acétate
gRNA (sgRNA)	Guide ARN
HBSS	Hank's Balanced Salt Solution
HF	House Finch
HGT	Horizontal Gene Transfer
HR (HDR)	Homology Direct Repair
Indel	Insertions-délétions
IL	Interleukine
IS	Insertion Sequence
kpb	Kilo paire de bases
Mpb	Méga paire de bases
ICE	Integrative and Conjugative Element
JCVI	John Craig Venter Institute
MGE	Eléments génétiques mobiles
<i>Mgal</i>	<i>Mycoplasma gallisepticum</i>
<i>Mmc</i>	<i>Mycoplasma mycoides subsp. capri</i>
<i>Mmm</i>	<i>Mycoplasma mycoides subsp. Mycoides</i>
<i>Mbov</i>	<i>Mycoplasma bovis</i>
NHEJ	Non Homologous End Joining
Nt	Nucléotide
OriC	Origine de réplication chromosomique

pac	Gène de résistance à la puromycine
PAM	Protospacer Adjacent Motif
PBS	Phosphate Buffered Saline
Pb	Paire de bases
PCR	Polymerase Chain Reaction (Réaction en chaîne par polymérase)
PEG	Polyéthylène glycol
Pgl	Phage growth limitation
pegRNA	ARN prime editing guide
PT	Phosphorothioation
PPLO	Pleuropneumonia-Like Organisms
PPCB	Péripneumonie contagieuse bovine
RT	Reverse transcriptase
R-M	Restriction-Modification
ROS	Dérivé réactif de l'oxygène
SDS	Dodécylsulfate de sodium
SNV	Single Nucléotide Variation
SSAPs	Phage-encoded ssDNA annealing proteins
SSR	Recombinases à Site Spécifique
TA	Toxines-Antitoxines
tet(M)	Gène de résistance à la tétracycline
tracrARN	Trans-activating crARN
Tris	Tris-(hydroxyméthyl)-aminoéthane
UFC	Unité Formant Colonie

TABLE DES MATIÈRES

TABLE DES MATIERES :

RÉSUME.....	1
REMERCIEMENTS	4
LISTE DES ABRÉVIATIONS	8
TABLE DES MATIÈRES	11
LISTE DES TABLES ET DES FIGURES	15
INTRODUCTION ET OBJECTIFS.....	17
ÉTUDE BIBLIOGRAPHIQUE.....	23
1. GENERALITES SUR LES MOLLICUTES ET LES MYCOPLASMES.....	24
1.1. Classification taxonomique et phylogénétique	24
1.2. Caractéristiques générales.....	26
1.2.1. Epidémiologie et habitat	26
1.2.2. Génomes de mycoplasme, une évolution régressive	27
1.2.3. ... accompagnée de nombreuses acquisitions.....	29
1.3. Objets d'études.....	30
1.3.1. Etude d'organismes minimaux et biologie de synthèse.....	31
1.3.2. Relation hôte-pathogène et virulence.....	32
1.3.2.1. Facteurs de virulence.....	33
1.3.2.2. Interaction avec le système immunitaire.....	35
1.4. <i>Mycoplasma gallisepticum</i>	36
1.4.1. Généralités sur la bactérie.....	36
1.4.2. <i>Mycoplasma gallisepticum</i> chez la volaille.....	37
1.4.3. Émergence chez les oiseaux sauvages.....	38
1.4.4. Virulence	40
1.4.4.1. Adhérence	40
1.4.4.2. Défense contre le système immunitaire.....	41
1.4.4.3. Production de molécules cytotoxiques	42
1.4.4.4. Autres facteurs de virulence	42
1.4.5. Interactions avec le système immunitaire.....	43
1.4.6. Perspectives d'études	44
1.5. <i>Mycoplasma mycoides subsp. mycoides</i>	45
1.5.1. Généralités sur la bactérie.....	45
1.5.2. PPCB.....	46
1.5.3. Virulence	48
1.5.3.1. Adhérence	48
1.5.3.2. Défense contre le système immunitaire.....	48
1.5.3.3. Production de molécules cytotoxiques	49
1.5.3.4. Autres facteurs.....	50
1.5.4. Perspectives d'études	50
1.6. <i>Mycoplasma bovis</i>	51
1.6.1. Généralités sur la bactérie.....	51
1.6.2. Virulence	52
1.6.2.1. Adhérence	52
1.6.2.2. Défense contre le système immunitaire.....	52
1.6.2.3. Autres facteurs.....	53
1.6.3. Pathologie.....	53
1.6.4. Perspectives d'études	55
2. SYSTEMES DE DEFENSE BACTERIENS : FONCTIONS ET APPLICATIONS.....	56

2.1.	<i>Systèmes immunitaires procaryotes</i>	57
2.1.1.	Immunité innée bactérienne.....	57
2.1.2.	Dormance et mort programmée.....	61
2.1.3.	Diversité des systèmes.....	62
2.2.	<i>CRISPR : systèmes de défense adaptatifs</i>	64
2.2.1.	Découverte du système.....	64
2.2.2.	Cibles principales.....	66
2.2.3.	Classification évolutive.....	69
2.2.4.	Etapas du système CRISPR.....	73
2.2.4.1.	Etape initiale d'adaptation.....	73
2.2.4.2.	Expression et mise en place du système.....	75
2.2.4.3.	Elimination de la menace par l'étape d'interférence.....	76
2.2.4.4.	Systèmes atypiques.....	78
2.2.5.	Balance bénéfiques-coûts du système adaptatif.....	79
2.2.6.	Modèle dynamique du combat bactérie-phage.....	81
2.3.	<i>Détournement du système CRISPR pour l'édition génétique</i>	84
2.3.1.	Premières applications d'un outil universel.....	85
2.3.2.	Outil génétique pour les bactéries.....	85
2.4.	<i>Limites et expansion de la boîte à outils</i>	87
2.4.1.	Entrée du système dans la cellule.....	87
2.4.2.	Clivage de l'ADN.....	89
2.4.2.1.	Toxicité cellulaire : Cytotoxicité et off-target.....	89
2.4.2.2.	Contrôle de l'expression dans le temps et dans l'espace.....	91
2.4.2.3.	Limitation du nombre de cibles.....	92
2.4.2.4.	Utilisation de systèmes hétérologues.....	92
2.4.2.5.	Evolution dirigée de l'effecteur.....	94
2.4.3.	Réparation de l'ADN : voies et alternatives.....	95
2.4.3.1.	Voies endogènes de réparation.....	96
2.4.3.2.	Systèmes exogènes phagiques.....	97
2.4.3.2.1.	Systèmes SSR.....	97
2.4.3.2.2.	Systèmes RecET et lambda Red.....	98
2.4.3.3.	Outils dérivés : de puissantes alternatives.....	99
2.4.3.3.1.	Modulation de la transcription.....	99
2.4.3.3.2.	Systèmes Base Editor (BE) : mutagenèse dirigée.....	101
2.4.3.3.3.	Autres outils d'application.....	103
2.5.	<i>Applications et éthiques : quelles limites ?</i>	104
3.	OUTILS GENETIQUES CHEZ LES MOLLICUTES	106
3.1.	<i>Transformation et vecteurs</i>	107
3.1.1.	Marqueurs de sélection et promoteurs.....	108
3.1.2.	Plasmide suicide.....	110
3.1.3.	Plasmides oriC.....	111
3.1.4.	Transposons.....	112
3.1.5.	Systèmes de recombinaison importés.....	114
3.2.	<i>Approches de biologie de synthèse chez les mycoplasmes</i>	115
3.3.	<i>Outils génétiques pour Mycoplasma gallisepticum</i>	117
3.4.	<i>Outils génétiques pour Mycoplasma mycoides subsp. mycoides et Mycoplasma bovis</i>	118
3.5.	<i>Conclusion</i>	118

CHAPITRE 1 : ETUDE DES SYSTEMES CRISPR CHEZ LES MOLLICUTES120

1.	<i>Principaux résultats</i>	121
2.	<i>Article publié dans Frontier in Microbiology</i>	122
3.	<i>Données supplémentaires</i>	137

CHAPITRE 2 : LE SYSTEME CRISPR-CAS9 DE MYCOPLASMA GALLISEPTICUM : DE LA CARACTERISATION DU SYSTEME NATUREL A L'OUTIL D'INGENIERIE DES GENOMES146

1. Principaux résultats	148
2. Article en finalisation pour soumission à CRISPR journal	149
3. Données et discussions supplémentaires non publiées	169
CHAPITRE 3 : UTILISATION D'UN SYSTEME DE RECOMBINAISON EXOGENE CHEZ MYCOPLASMA GALLISEPTICUM COMME OUTIL D'INGENIERIE DE GENOME	171
1. Principaux résultats	172
2. Article soumis à ACS Synthetic Biology (27/10/2021, en cours de révision)	173
3. Données supplémentaires de la publication	182
4. Données et discussions supplémentaires non publiées	186
CHAPITRE 4 : CONSTRUCTION D'UN OUTIL UNIVERSEL POUR DES BACTERIES A FAIBLE EFFICACITE DE RECOMBINAISON : LES SYSTEMES BASE-EDITOR	188
1. Principaux résultats	189
2. Article en finalisation pour soumission à PNAS – mBio	190
3. Données supplémentaires de la publication	210
4. Données et discussions supplémentaires non publiées	213
DISCUSSION GÉNÉRALE ET CONCLUSION	214
1. DISCUSSION GENERALE	215
1. Les systèmes de défense chez les mycoplasmes	215
2. La boîte à outils : développement de nouveaux outils et de nouveaux modèles	222
2.1. Les outils développés	222
2.2. Amélioration des outils construits	223
2.3. Guide pour l'ingénierie d'une espèce de mycoplasme	224
2. CONCLUSION	227
RÉFÉRENCES BIBLIOGRAPHIQUES	228

LISTE DES TABLES ET DES FIGURES

LISTES DES TABLES ET DES FIGURES :

Figure 1 : Arbre phylogénétique des *Mollicutes*.

Table 1 : Caractéristiques générales des génomes de mycoplasmes.

Figure 2 : Détermination des gènes essentiels chez *Mesoplasma florum*.

Figure 3 : Représentation schématique des principaux facteurs de virulence identifiés chez *Mycoplasma gallisepticum*, *Mycoplasma mycoides* subsp. *mycoides* et *Mycoplasma bovis*.

Figure 4 : *Mycoplasma gallisepticum*.

Figure 5 : *Mycoplasma mycoides*.

Figure 6 : *Mycoplasma bovis*.

Figure 7 : Représentation schématique des différents mécanismes procaryotes de défenses contre l'ADN exogène.

Figure 8 : Historique des découvertes sur les systèmes CRISPR.

Figure 9 : Représentation du locus CRISPR de *S. pyogenes*.

Table 2 : Tableau récapitulatif de la classification des systèmes CRISPR de classe 1.

Table 3 : Tableau récapitulatif de la classification des systèmes CRISPR de classe 2.

Figure 10 : Représentation schématique de la réponse CRISPR pour les systèmes de classe 1 et 2.

Figure 11 : Représentation détaillée de la phase d'adaptation du système de défense CRISPR.

Figure 12 : Représentation schématique de l'étape d'interférence.

Figure 13 : Représentation schématique des domaines de la protéine Cas9.

Figure 14 : Balance coût-bénéfice du système CRISPR.

Figure 15 : Représentation des stratégies d'adaptation adoptées par les phages en réponse au système CRISPR.

Figure 16 : Les outils CRISPR et leurs dérivés.

Figure 17 : Représentation des opportunités d'ingénierie par le système Cre-lox.

Figure 18 : Représentation d'un modèle d'ingénierie génétique médié par le système RecET.

Table 4 : Ingénierie génétique publiées chez les mollicutes.

Table 5 : Récapitulatif des différents noms de plasmides utilisés dans la littérature.

Figure 19 : Représentation schématique des principaux outils disponibles chez les mycoplasmes.

Figure 20 : Représentation de la méthode RAGE.

Figure 21 : Cycle d'ingénierie de génome de mycoplasme dans la levure.

Figure 22 : Ilot de défense chez *Mycoplasma gallisepticum*.

Figure 23 : Modèle d'évolution du système CRISPR dans le contexte de l'émergence et l'adaptation de *Mgal* chez le Roselin.

Figure 24 : L'ingénierie des mycoplasmes pour les Nuls.

INTRODUCTION ET OBJECTIFS

Les *Mollicutes* représentent une classe de petites bactéries particulièrement fascinantes. Elles sont considérées comme les plus petites bactéries capables de se multiplier en milieu axénique. Ce sont des bactéries ayant évolué principalement de manière régressive, c'est-à-dire par perte de gènes. Les mollicutes ne possèdent pas de paroi cellulaire et ont de très petits génomes (de 0,58 Mb à 2,2 Mb). La majorité des espèces présente un mode de vie parasitaire et est retrouvée chez une grande diversité d'organismes vivants comme les animaux (humains, mammifères variés, insectes, reptiles, oiseaux, *etc.*) ou les végétaux (vignes, citronniers, pins *etc.*). Ces bactéries peuvent être commensales et non pathogènes, mais malgré leur caractère minimal, elles sont souvent responsables ou associées à de nombreuses maladies (Pleuropneumonie contagieuse bovine, Flavescence dorée de la vigne, *etc.*) ayant des répercussions sur les productions agricoles mais également sur la santé animale et humaine. Malgré leur simplicité, les mollicutes sont également capables d'émerger dans de nouveaux hôtes ce qui explique en partie qu'ils soient aussi répandus dans la nature. Un évènement récent de changement d'hôte a notamment eu lieu en 1994 pour le mycoplasme pathogène d'oiseaux *Mycoplasma gallisepticum (Mgal)*, ce qui fait de cette bactérie, une des rares espèces à pouvoir être suivie en temps réel lors d'un tel évènement. Il est donc important d'étudier ces organismes pour comprendre l'interaction de ces bactéries avec les différents hôtes, identifier des facteurs de virulence et finalement développer des méthodes de contrôle efficaces des maladies qu'elles provoquent chez leurs différents hôtes. Les mycoplasmes, qui sont un groupe particulier de mollicutes, intéressent de nombreux chercheurs avec notamment, l'avènement récent de la biologie des systèmes et la biologie de synthèse. Les procédés complexes tels que la compréhension des mécanismes cellulaires et l'étude de certaines voies métaboliques étant difficiles à étudier dans des systèmes complexes, l'utilisation d'organismes manipulables minimaux tels que les mycoplasmes, est une option intéressante pour servir de modèle à la compréhension de tous ces mécanismes et des fonctions essentielles à la vie cellulaire.

Pour pouvoir manipuler génétiquement ces bactéries et en faire l'étude fonctionnelle, des outils précis pour l'ingénierie de génomes sont nécessaires. Des plasmides réplikatifs de type *oriC* et des plasmides intégratifs de types transposons sont disponibles pour certaines espèces, mais la recombinaison homologue n'étant pas assez efficace chez les mollicutes, il est difficile de faire de l'ingénierie précise sur leurs génomes. Les récentes avancées en biologie de synthèse ont apporté des méthodes très

puissantes permettant par exemple la création de souches de mycoplasme dont le génome a été reconstruit à partir d'oligonucléotides. Malheureusement, ces approches restent longues et laborieuses à mettre en place pour de nouvelles espèces, la limite étant la transplantation de génome qui pour le moment n'est au point seulement que pour 4 des 5 espèces pathogènes de ruminants du cluster Mycoides et trois espèces du groupe phylogénétique Spiroplasma.

Les mycoplasmes sont importants sur les plans sanitaire et économique mais malgré tout de manière générale, aucuns traitements ou vaccins efficaces sont disponibles. Ce développement doit passer par la compréhension de l'interaction hôte – bactérie de ces bactéries minimales qui est limité par le manque d'outils génétiques. Ainsi ma thèse a pour objectif de développer des outils d'ingénierie de génomes précis, pour de nombreuses espèces de mollicutes, particulièrement pour les Mycoplasmes afin de comprendre un évènement de changement d'hôte sur le modèle de *Mgal*.

L'histoire évolutive des systèmes de défense bactérien CRISPR corrélée au changement d'hôte de *Mgal* et l'avènement de ces systèmes comme outil d'ingénierie de génome a été le point de départ dans la construction de mon projet de thèse qui finalement est découpé en quatre parties, développées chacune dans mes chapitres 1, 2, 3 et 4 respectivement.

La première partie de mon travail de thèse était l'étude globale des systèmes CRISPR chez les mollicutes par des approches de génomique comparée. Sur une sélection de 52 espèces représentatives de la classe *Mollicutes*, 21 espèces présentent un locus CRISPR ou des vestiges de locus. Deux systèmes distincts ont été identifiés, grâce à des analyses computationnelles. Le premier est retrouvé chez certains spiroplasmales et est caractérisé comme un système CRISPR de type II-C ; le second système présente les caractéristiques d'un système CRISPR de type II-A. Une reconstruction phylogénétique a permis de mettre évidence une origine commune pour tous les systèmes de mycoplasmes et deux origines pour les systèmes de spiroplasmales. La présence de nombreux systèmes tronqués, l'absence du système CRISPR dans le cluster mycoides (malgré de nombreux transferts horizontaux entre les espèces présentes chez les ruminants par exemple) et la présence de nombreux codons tryptophane UGA dans certaines séquences protéiques du système, tendent à démontrer que le système est arrivé il y a longtemps chez les mycoplasmes et qu'il était présent chez l'ancêtre commun aux mycoplasmes. Malgré la

réduction drastique des génomes caractéristique de l'évolution des mycoplasmes, le système a été maintenu dans différentes espèces, ce qui suggère qu'il est important et certainement actif. La protéine Cas9 qui est conservée dans l'ensemble des systèmes de mycoplasme, la principale différence se situe sur les résidus du domaine PI qui a un rôle dans la reconnaissance de la séquence PAM, pour la différenciation du soi et du non-soi par le système CRISPR. Etant donné l'isolement génétique des mycoplasmes (les transferts horizontaux vers d'autres espèces étant vraisemblablement défavorisés à cause du codon UGA codant un tryptophane et reconnu comme un codon stop dans les autres espèces bactériennes), le système CRISPR pourrait être unique et avoir évolué de manière intéressante. D'autres indices tendent à montrer qu'il est encore actif, au moins chez certaines espèces, comme chez *Mgal*, où le système CRISPR montre une évolution rapide suite à un changement d'hôte, suggérant son importance dans la persistance de la bactérie. Cette partie du travail a été initiée par un précédent doctorant de l'équipe, Iason Tsarmopoulos (Thèse soutenue en 2017) et par mon encadrant de thèse le docteur Pascal Sirand-Pugnet et j'ai complété ce travail pendant la première année de ma thèse. Ceci a conduit à une première publication (**Ipoutcha et al. 2019**) dans la revue scientifique **Frontiers in Microbiology**.

Ce travail a introduit ma deuxième partie de thèse qui est la caractérisation fonctionnelle d'un système CRISPR-Cas9 de mycoplasme et son adaptation comme outil génétique pour un large panel d'espèces de mollicutes. Dans un premier temps, il a été question de démontrer la fonctionnalité du système CRISPR de *Mgal* S6. Grâce aux expériences préliminaires réalisées par Iason Tsarmopoulos, le système de défense de la bactérie est apparu comme fonctionnel chez *Mgal*. Par la suite, après des analyses de génomique comparée, nous avons cherché à caractériser la séquence PAM de deux Cas9 provenant de deux souches de *Mgal* à l'aide d'une méthode de clivage *in vitro* couplée à du séquençage haut-débit. Ceci a permis de mettre en évidence une corrélation entre l'évolution de la spécificité de la Cas9 et l'émergence du pathogène dans les populations de pinsons d'Amérique. Des études approfondies laissent même imaginer un rôle des systèmes CRISPR et autres systèmes de défense pour l'émergence dans un nouveau contexte écologique tel qu'un changement d'hôte, et dans l'adaptation de la bactérie à de nouvelles contraintes environnementales telle que l'apparition d'une population d'hôte plus résistante à la bactérie. Une fois la caractérisation fonctionnelle effectuée, l'objectif de la transformation du système naturel en outil s'est concrétisé et un système minimal,

composé de la protéine MgalCas9 et d'un ARN guide reconstitué (fusion des crRNA et tracrRNA), a été testé à l'aide de multiples plasmides *oriC* « all in one » pour des applications de contre-sélection dans des espèces de mycoplasmes éloignées phylogénétiquement les unes des autres : *Mgal*, *Mycoplasma capricolum* subsp. *capricolum* (*Mcap*), *Mycoplasma pulmonis*, *Mycoplasma pneumoniae* et *Mycoplasma mycoides* subsp. *mycoides* (*Mmm*). L'outil a finalement été mis en application comme preuve de concept pour la sélection de souche de *Mcap* curées de leur élément conjugatif et intégratif (ICE). Ceci a permis d'ajouter ce système à la boîte à outils pour l'ingénierie génomique des mycoplasmes. Cette seconde partie du travail va permettre une deuxième publication (**Ipoutcha et al. 2022**) qui est en cours de préparation.

Enfin les troisième et quatrième parties de ma thèse se sont développées autour de la problématique liée au manque d'efficacité de la recombinaison homologue chez les mycoplasmes, déjà décrit dans la littérature, mais également vérifiée par l'impossibilité à obtenir des recombinants à l'aide du système CRISPR développé dans la partie précédente. Cela concerne l'extension de la boîte à outils par la création d'outils d'ingénierie des mycoplasmes, notamment pour des espèces vétérinaires majeures pour lesquelles aucun outil de mutagenèse ciblée n'était disponible : *Mgal*, *Mmm* et *Mycoplasma bovis* (*Mbov*). Ces trois pathogènes sont responsables de pathologies importantes d'un point de vue économique et écologique chez les oiseaux (*Mgal*) d'une part et chez les ruminants (*Mmm* et *Mbov*) d'autre part. Cette partie a été développée par le docteur Pascal Sirand-Pugnet et moi-même avec les conseils du docteur Carole Lartigue. Deux types d'approches très différentes ont été évaluées en parallèle. La première approche est basée sur l'utilisation d'un système de recombinaison exogène phagique (type RecET) pour réaliser des recombinaisons homologues et des modifications génomiques de tous types (insertion, délétion et remplacement). Plusieurs systèmes ont été identifiés et testés chez *Mgal*. La fonctionnalité d'un de ces systèmes nous a permis d'obtenir les premiers mutants de manière ciblée chez *Mgal*. Ces travaux font l'objet d'une troisième publication (**Ipoutcha et al. 2022**) qui vient d'être soumise au journal ACS Synthetic Biology. La deuxième approche est basée sur le système CRISPR et l'utilisation d'une version désactivée de la protéine Cas9 de *S. pyogenes* fusionnée à une désaminase. Ceci permet l'obtention de mutations ponctuelles ciblées de type Cytosine en Thymine et Adénine en Guanine. Ces systèmes sont décrits comme simples et efficaces et permettent l'inactivation ou la modification de gènes sans avoir recours à des mécanismes de

INTRODUCTION ET OBJECTIFS

recombinaison. Nous avons testé ici aussi différents types de constructions dans *Mgal*, *Mmm* et *Mbov* et un de ces systèmes s'est montré fonctionnel dans les trois espèces. Le développement de cet outil et son efficacité dans différentes espèces de mycoplasmes éloignées phylogénétiquement, nous a permis de réaliser une preuve de concept solide qui suggère le côté universel de cet outil pour la génomique fonctionnelle des mollicutes. Cette dernière partie clôturera finalement l'ensemble de mes objectifs de thèse, et sera valorisée dans une quatrième publication (**Ipoutcha et al. 2022**) qui est en cours de rédaction.

L'ensemble de ce travail vise, d'une part, à acquérir des connaissances sur l'émergence de la bactérie *Mgal* dans un nouvel hôte et d'autre part, à développer des outils génétiques utilisables chez plusieurs espèces de mycoplasmes dans le but final de pouvoir les modifier, les étudier et développer des traitements et vaccins efficaces.

ÉTUDE BIBLIOGRAPHIQUE

1. GENERALITES SUR LES MOLLICUTES ET LES MYCOPLASMES

1.1. Classification taxonomique et phylogénétique

Le terme *Mollicutes* vient du latin « *mollis* » signifiant peau et « *cutis* » signifiant molle. Les *Mollicutes* sont donc une classe de bactéries ayant la « peau molle », en référence à l'absence de paroi cellulaire caractéristique des bactéries appartenant à cette classe (Razin, Yogev and Naot, 1998). Ce terme est apparu pour la première fois en 1967 et a finalement été adopté par la communauté scientifique (Edward and Freundt, 1967). Cette classe de bactéries a longtemps été assimilée aux mycoplasmes, qui ne représentent en fait qu'une partie des mollicutes. La classe *Mollicutes* contient quatre ordres et a été divisée sur des bases phylogénétiques et phénotypiques : les Mycoplasmatales, les Entomoplasmatales, les Acholeplasmatales et les Anaeroplasmatales (Tully *et al.*, 1993). L'ordre I des *Mycoplasmatales* est un ordre avec des espèces infectant principalement les humains et les animaux, il est notamment composé des genres *Ureaplasma* et *Mycoplasma* (May, Balish and Blanchard, 2014). L'ordre II des *Entomoplasmatales* est quant à lui composé des espèces associées majoritairement aux insectes et aux plantes. Il est composé en partie par les genres *Mesoplasma* (seules espèces des ordres I et II ne nécessitant pas de cholestérol) et *Spiroplasma* (espèces caractérisées par une structure hélicoïdale). L'ordre III des *Acholeplasmatales* regroupe des espèces trouvées principalement associées aux arthropodes et aux plantes. Il est caractérisé par des espèces ne nécessitant pas de cholestérol, mais également par des espèces utilisant le code génétique universel contrairement aux autres mollicutes ayant un codon UGA alternatif codant le tryptophane (Yamao *et al.*, 1985; Blanchard, 1990; Bové, 1993). Finalement l'ordre IV des *Anaeroplasmatales* est un ordre regroupant des espèces anaérobiques obligatoires (Tully *et al.*, 1993).

D'un point de vue évolutif, les mollicutes ont d'abord été considérés comme des organismes primitifs descendants de l'ancêtre bactérien sans paroi qui a ensuite donné toutes les autres bactéries (Morowitz and Wallace, 1973), mais cette théorie a rapidement été réfutée par l'analyse des ADNr 16s (Woese, Maniloff and Zablen, 1980) qui suggère, au contraire, que ces bactéries proviennent d'une bactérie à GRAM positif et qu'elles ont suivi une évolution dite régressive, c'est-à-dire s'accompagnant de pertes massives en termes de répertoire de gènes et de capacités métaboliques (Neimark, 1983). La

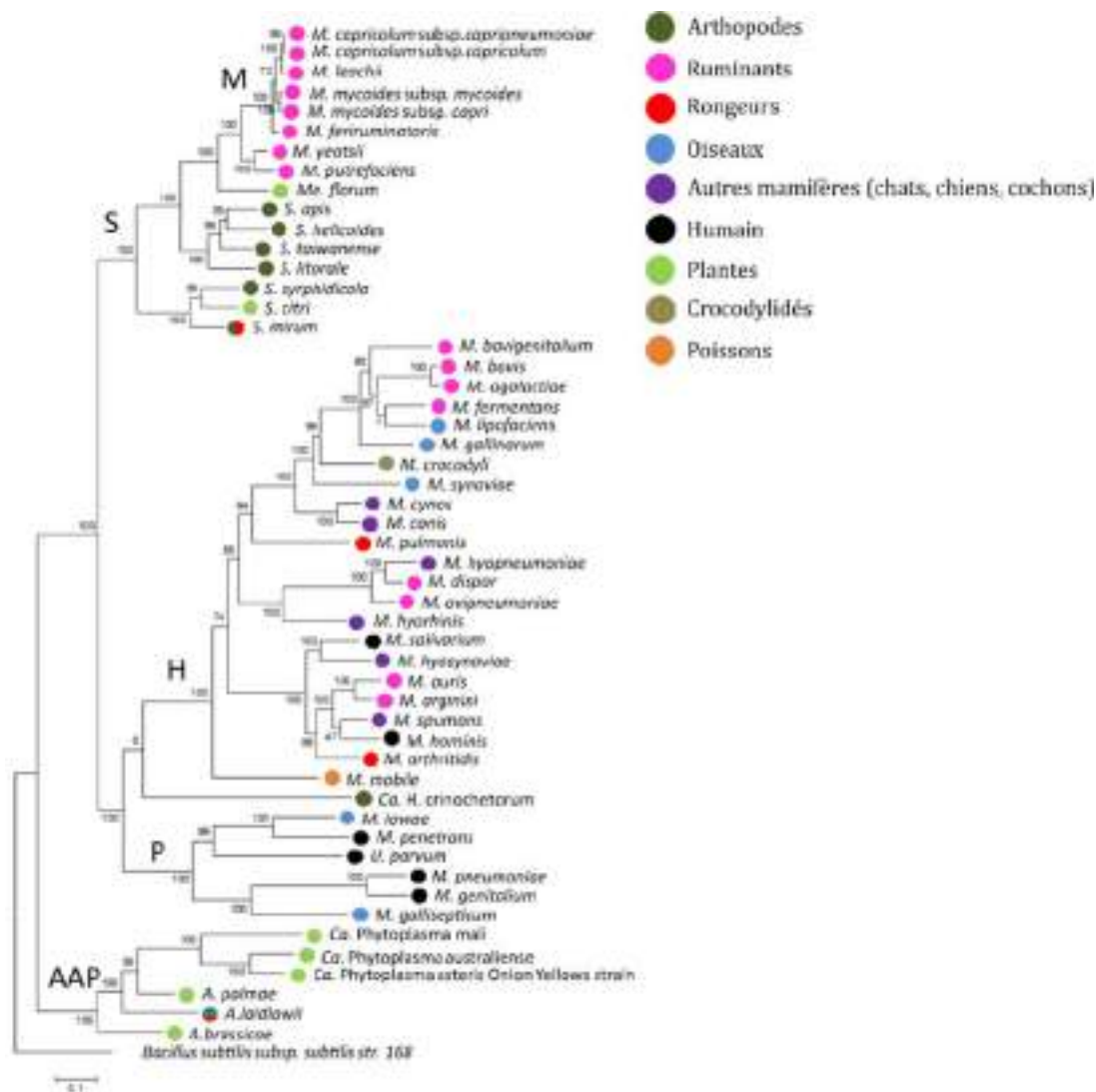


Figure 1 : Arbre phylogénétique des *Mollicutes*. L'arbre provient de Ipoutcha *et al.* 2019 et a été généré à partir de la comparaison de 62 séquences protéiques orthologues impliquées dans la traduction en utilisant la méthode maximum likelihood. Les principaux groupes phylogénétiques sont indiqués, S, Spiroplasma, H, Hominis, P, Pneumoniae, AAP, Acholeplasma/Phytoplasma. Le cluster Mycoides (M) appartenant au groupe phylognétique Spiroplasma est indiqué. *Bacillus subtilis* a été utilisé comme outgroup. Les valeurs statistiques sont indiquées sur les branches. Les hôtes sont indiqués par des ronds colorés : Arthropodes en verts foncés, Ruminants en rose, Rongeurs en rouge, Oiseaux en bleu, Autres mammifères en violet, Humain en noir, Plantes en vert clair, Crocodylidés en marron et Poissons en orange.

ÉTUDE BIBLIOGRAPHIQUE : Généralités sur les mollicutes et les mycoplasmes

comparaison des ADN ribosomiques a même permis de déterminer un taux d'évolution très élevé chez les mycoplasmes (Woese, Stackebrandt and Ludwig, 1985; Delaney *et al.*, 2012) caractéristique non pas d'organismes primitifs comme énoncé dans la première théorie, mais de bactéries très évoluées comme cela est illustré par la longueur des branches correspondant à ces bactéries dans les arbres phylogéniques (Ciccarelli *et al.*, 2006). Finalement, d'un point de vue phylogénétique, la branche des mollicutes a dérivé de celle des *Firmicutes* il y a environ 605 MA, une évolution régressive s'en est suivie (Woese, Maniloff and Zablen, 1980; Weisburg *et al.*, 1989; Maniloff, 2002; Sirand-Pugnet, Citti, *et al.*, 2007). Cette branche se serait ensuite séparée en deux il y a 470 Ma pour donner la branche AAP (*Acholeplasmataceae*, *Anaeroplasmataceae* et *Phytoplasma*) et la branche SEM (*Spiroplasmataceae*, *Entomoplasmataceae* et *Mycoplasmataceae*) (Figure 1). Dans cette dernière, les *Spiroplasmataceae* - *Entomoplasmataceae* se seraient séparés des *Mycoplasmataceae* il y a 410 Ma pour finalement donner les trois branches distinctes il y a 170 Ma. Les *Mycoplasmataceae* ont également été séparés sur des critères phylogénétiques dans les groupes Pneumoniae (P), Hominis (H), Spiroplasma (S) (Weisburg *et al.*, 1989; Maniloff, 2002; Sirand-Pugnet, Citti, *et al.*, 2007). Paradoxalement, le groupe Spiroplasma regroupe les mollicutes des genres *Spiroplasma*, *Mesoplasma*, *Entomoplasma* et une partie des mycoplasmes. Ces derniers sont tous associés à des ruminants, les plus importants d'un point de vue vétérinaire étant regroupés dans le cluster Mycoïdes (M). Les mycoplasmes du cluster Mycoïdes sont peu différenciables en comparant les ADNr 16S, et il a fallu attendre des approches moléculaires plus poussées, par exemple la MLST (multi locus sequence typing) (Manso-Silván, Perrier and Thiaucourt, 2007) pour clarifier les relations taxonomiques et phylogénétiques au sein de ce dernier groupe (Manso-Silván *et al.*, 2009). Dernièrement la place du cluster mycoïdes au sein de la branche des spiroplasma a suscité des divergences d'opinions entre quelques chercheurs proposant une reclassification taxonomique (Gupta *et al.*, 2018). Celle-ci a majoritairement été rejeté par la communauté des mycoplasmodistes qui prône la simplicité afin de ne pas amener de la confusion dans la taxonomie de la classe *Mollicutes* (Balish *et al.*, 2019).

Au cours de ce manuscrit, les mycoplasmes qui sont l'objet principal de ma thèse seront particulièrement détaillés.

1.2. Caractéristiques générales

1.2.1. Epidémiologie et habitat

Les mollicutes sont des organismes considérés comme minimaux de par leurs morphologies et la taille de leurs génomes. Les mycoplasmes sont notamment considérés comme les plus petites bactéries capables d'autoréplication (Fraser *et al.*, 1995; Peterson and Fraser, 2001; Sirand-Pugnet, Lartigue, *et al.*, 2007). Malgré tout, les mollicutes sont associés à de très nombreux hôtes et habitats différents.

La branche des AAP qui regroupe les acholeplasmales, les anaéroplasmes et les phytoplasmes, sont tout particulièrement retrouvés au sein de plantes dans lesquelles ils peuvent causer différents types de symptômes comme des perturbations hormonales, une virescence ou une stérilité (Gasparich, 2010). Ils sont disséminés de plantes à plantes par des insectes vecteurs dans lesquels ils ne produisent pas de symptômes et sont souvent des commensaux de l'estomac ou des glandes salivaires (Razin, Yogev and Naot, 1998). Par exemple, le phytoplasme de la flavescente dorée (FD) qui est étudié dans l'équipe est un phytoplasme de la vigne transmis principalement par un insecte vecteur, la cicadelle *Scaphoideus titanus*. Il est responsable de la FD qui est considérée comme une maladie de quarantaine par des directives européennes et peut entraîner des pertes considérables dans les vignobles (Malembic-Maher *et al.*, 2020).

Les spiroplasmales sont des mollicutes commensaux, pathogènes ou endosymbiotes d'arthropodes (Perilla-Henao and Casteel, 2016). Contrairement aux phytoplasmes, seuls trois espèces sont des pathogènes de plantes (*S. citri*, *S. kunkelii*, and *S. phoeniceum*). C'est lorsqu'ils vont passer la barrière de l'hémolymphe qu'ils vont être disséminés et se retrouver dans divers habitats, dans lesquels ils peuvent être responsables de maladies. *S. citri*, responsable de la maladie du stubborn chez le citronnier, est transmis par la cicadelle *Circulifer haematoceps* (Cole *et al.*, 1973; Gasparich, 2010). Il est responsable chez les agrumes de symptômes tels qu'un nanisme accru des plantes additionné d'une perturbation de la floraison et de la fructification. De nombreuses espèces de spiroplasmales sont retrouvées dans les crustacés (*Spiroplasma panaei*) ou d'autres animaux (*Spiroplasma mirum*).

Enfin les mycoplasmes sont des mollicutes infectant une grande diversité de métazoaires. On les retrouve notamment chez les poissons (*Mycoplasma mobile*), les reptiles

Espèces	Groupes phylogénétiques	Tailles génomes (Mb)	G + C (%)	Hôte
<i>Mycoplasma mycoides subsp. capri</i>	Spiroplasma	1,08	24	Ruminant
<i>Mycoplasma mycoides subsp. mycoides</i>	Spiroplasma	1,20	24	Ruminant
<i>Mesoplasma florum</i>	Spiroplasma	0,79	27	Insecte
<i>Spiroplasma citrii</i>	Spiroplasma	1,64-2,16	25	Plante - Insecte
<i>Mycoplasma mobile</i>	Hominis	0,78	25	Poisson
<i>Mycoplasma hyopneumoniae</i>	Hominis	0,89	28	Cochon
<i>Mycoplasma pulmonis</i>	Hominis	0,96	27	Rongeur
<i>Mycoplasma synoviae</i>	Hominis	0,80	28	Oiseau
<i>Mycoplasma agalactiae</i>	Hominis	0,88	30	Ruminant
<i>Mycoplasma penetrans</i>	Pneumoniae	1,36	26	Humain
<i>Mycoplasma gallisepticum</i>	Pneumoniae	0,92-1,01	31	Oiseau
<i>Mycoplasma pneumoniae</i>	Pneumoniae	0,82	40	Humain
<i>Mycoplasma genitalium</i>	Pneumoniae	0,58	32	Humain
<i>Ureaplasma urealyticum</i>	Pneumoniae	0,75	25	Humain
<i>Phytoplasma Aster Yellow</i> s	Phytoplasma	0,71	27	Plante - Insecte
<i>Acholeplasma laidlawii</i>	Acholeplasma	1,50	32	Animaux

Table 1 : Caractéristiques générales des génomes de mycoplasmes.
Adapté de Sirand-Pugnet et al. 2007

ÉTUDE BIBLIOGRAPHIQUE : Généralités sur les mollicutes et les mycoplasmes

(*Mycoplasma crocodyli*), les oiseaux (*Mgal*) ainsi que les mammifères (ruminants, cochons, humains...). Les mycoplasmes peuvent être commensaux ou non et peuvent provoquer des maladies. Ils sont la plupart du temps tissus et organes spécifiques ce qui est lié à leurs modes de vie parasitique (Razin, Yogev and Naot, 1998). Par exemple, *Mycoplasma hominis* est un mycoplasme humain, commensal de la flore vaginale qui, lors d'un déséquilibre de la flore vaginale peut être associé à une vaginose bactérienne (Waites and Talkington, 2005).

Comme on peut le voir dans la Figure 1 et la Table 1, les mollicutes sont retrouvés dans un très large panel d'hôtes, et on peut penser que de nombreuses espèces restent encore inconnues. Actuellement, plus de 300 espèces de mollicutes sont recensées au NCBI et parmi elles, plus de 150 sont des espèces de mycoplasmes. Outre les habitats dans lesquels on les retrouve naturellement, les cultures cellulaires se trouvent également être un habitat fréquent de certains mycoplasmes. En effet de nombreuses espèces sont retrouvées comme contaminants de culture, c'est notamment le cas de *Mycoplasma hyorhinis* et *Mycoplasma arginini* (Razin, Yogev and Naot, 1998).

Les espèces de mollicutes ne sont pas toutes cultivables en laboratoire *in vitro*. Les milieux utilisés sont riches et la vitesse de croissance est généralement plutôt lente (entre 0,5 et 6 h de temps de génération). La source d'énergie utilisée est le plus souvent le glucose chez la majeure partie des espèces, mais parfois certaines espèces peuvent utiliser l'arginine comme *M. hominis* ou l'urée comme *Ureaplasma urealyticum* (Razin, Yogev and Naot, 1998).

1.2.2. Génomes de mycoplasme, une évolution régressive...

Les mollicutes sont des organismes avec des génomes très petits allant de 564 kb pour *Mycoplasma parvum* (NC_022575.1) à 2,22 Mb pour *Spiroplasma ixodetis* (Tully *et al.*, 1995). Les génomes de *Mollicutes* et de mycoplasmes ont un taux très faible de GC : allant de 22,4 % pour *Mycoplasma collis* (GCA_000701825.1) à 40 % pour *M. pneumoniae* avec un % médian de 27,7 % (Table 1) (Sirand-Pugnet, Citti, *et al.*, 2007). Cet effet serait le résultat d'une pression de sélection AT pendant l'évolution (Muto and Osawa, 1987), qui aurait alors favorisé l'apparition de codons riches en AT et provoqué un biais de codon important chez de nombreuses espèces de mycoplasmes. Ainsi, peu de codons de type GGN, CCN, GCN et CGN et donc peu d'acides aminés de type Glycine, Proline, Alanine et Arginine sont retrouvés au sein des séquences protéiques de mycoplasmes (Razin, Yogev

and Naot, 1998). Le codon CGG aurait même quasi-disparu chez *Mcap* (Krishnakumar *et al.*, 2013). Cette pression de sélection AT est probablement à l'origine d'une autre caractéristique des génomes de mollicutes : l'utilisation d'un code génétique atypique où le codon UGA code un tryptophane au lieu d'un codon-stop (Yamao *et al.*, 1985; Blanchard, 1990), l'exception vient de l'ordre *Acholeplasmatales* avec les genre *Acholeplasma* et *Phytoplasma* (Bové, 1993; Dybvig and Voelker, 1996). La perte de nombreux gènes au cours de l'évolution des mycoplasmes a également eu de nombreuses conséquences sur ces bactéries. En effet, de nombreuses voies métaboliques classiques non essentielles ont disparu, par exemple les enzymes nécessaires à la production d'une paroi (Dybvig and Voelker, 1996). D'autres pertes comme celles d'enzymes ayant un rôle dans la synthèse d'acides aminés, de petits métabolites, de nucléotides ou de lipides sont observées (Grosjean *et al.*, 2014) et ceci peut être mis en lien avec le mode de vie parasitique des mycoplasmes ou les milieux de cultures très riches qui sont nécessaires pour leur croissance en laboratoire. Certains mycoplasmes sont intracellulaires et peuvent être retrouvés dans des cellules eucaryotes c'est le cas de *Mgal*, *Mycoplasma genitalium* ou *Mycoplasma penetrans* (Lo *et al.*, 1992; Jensen, Blom and Lind, 1994; Winner, Rosengarten and Citti, 2000), ce qui facilite certainement l'accès à de nombreux facteurs nécessaires à leur croissance. L'ancêtre des mycoplasmes a également perdu plusieurs enzymes nécessaires à la réparation de l'ADN et notamment pour la recombinaison homologue. En effet, chez les différentes espèces de mycoplasmes, entre 15 et 24 enzymes sont impliquées dans différentes voies de réparation de l'ADN, contre 73 chez *Escherichia coli* ou 45 chez *Lactobacillus plantarum* (Carvalho *et al.*, 2005). La recombinaison médiée par RecA serait bien présente dans plusieurs espèces de mycoplasmes, mais il semblerait que son rôle soit lié à la variation antigénique et non à la réparation de l'ADN comme cela a pu être montré chez *M. genitalium* (Burgos *et al.*, 2012). Les autres voies de réparation sont également toutes réduites, seule la voie de réparation par excision de nucléotide semble intacte. Quelques protéines de la voie de réparation d'excision de base (BER) peuvent également être identifiées et pourraient compléter la voie NER. Cette faible capacité à réparer l'ADN entraîne une accumulation de mutations et par la suite, l'élimination de gènes sans utilité. Ceci peut expliquer l'évolution dégressive et rapide des génomes de mycoplasmes (Carvalho *et al.*, 2005; Sirand-Pugnet, Lartigue, *et al.*, 2007).

1.2.3. ... accompagnée de nombreuses acquisitions

Contrairement à la vision qui a prévalu pendant des années, l'évolution des mollicutes ne peut être réduite à une réduction inexorable de leur génome et des phénomènes d'acquisition de gènes par transfert horizontal de gènes (HGT) ont été identifiés. Les HGT sont connus comme étant le principal mécanisme d'évolution chez les bactéries (Koonin, 2019) mais les transferts horizontaux de mycoplasmes vers d'autres bactéries ne sont pour le moment pas décrits dans la littérature. Les gènes de mycoplasmes et plus largement de mollicutes, au cours de l'évolution ont massivement été transformés. Suite au biais de codon important, le codon tryptophane TGG a par exemple été remplacé par un codon TGA qui est reconnu chez la majorité des autres espèces bactériennes comme un codon stop. Ainsi les transferts de gène de mycoplasmes vers d'autres espèces bactériennes sont quasi-impossibles puisque l'espèce receveuse ne devrait pas avoir d'avantages conséquents pour le maintien d'un gène de mycoplasme probablement tronqué par des codons TGA. Les mollicutes sont susceptibles de recevoir des éléments génétiques mobiles, mais il est peu probable que d'autres espèces bactériennes conservent en retour ce qui pourrait provenir de mollicutes. En revanche, des HGT provenant d'autres bactéries vers les mycoplasmes (Omelchenko *et al.*, 2003; Ganter *et al.*, 2019) ou entre mycoplasmes ont été identifiés par analyses comparées des génomes. Les HGT sont particulièrement suspectés entre bactéries partageant une même niche écologique par exemple entre les espèces *Mgal* et *Mycoplasma synoviae* chez l'oiseau (Vasconcelos *et al.*, 2005), également chez *Mycoplasma agalactiae* dont le génome serait composé à 16% de morceaux d'autres espèces de mycoplasmes de ruminants (Sirand-Pugnet, Citti, *et al.*, 2007; Citti *et al.*, 2018) ou même entre *M. hominis* et *Ureaplasma parvum* chez l'homme (Pereyre *et al.*, 2009). Ces échanges d'ADN ne sont pas forcément associés à des éléments mobiles comme cela peut être le cas chez *M. genitalium* (Torres-Puig *et al.*, 2018) ou chez *Mbov* (García-Galán *et al.*, 2021). En revanche pour des échanges de grands fragments d'ADN, ils sont associés à divers éléments mobiles (Sirand-Pugnet, Lartigue, *et al.*, 2007; Dordet-Frisoni *et al.*, 2019). En effet au sein des génomes de mycoplasmes on trouve différents éléments mobiles. Les éléments intégratifs et conjugatifs (ICE) sont les principaux éléments mobiles retrouvés chez les mycoplasmes. Les ICE sont des éléments mobiles qui codent pour leur propre excision, leur transfert et l'intégration dans un autre génome (Wozniak and Waldor, 2010). Le premier ICE identifié

chez les mycoplasmes est celui de *Mycoplasma fermentans* présent en 4 copies ; il représente près de 8% de son génome (Calcutt, Lewis and Wise, 2002). De nombreuses études bio-informatiques mais aussi fonctionnelles ont permis de montrer que les ICE sont répandus chez les mycoplasmes et on les retrouve dans les différents groupes phylogénétiques (Tardy *et al.*, 2015). Ils peuvent être entiers et actifs ou tronqués et ont des tailles comprises entre 20 et 30 kb (Marenda *et al.*, 2006; Sirand-Pugnet, Citti, *et al.*, 2007; Nouvel *et al.*, 2010; Dordet Frisoni *et al.*, 2013; Dordet-Frisoni *et al.*, 2014; Citti *et al.*, 2018). Ces éléments permettent ainsi, le transfert de gènes qui sont compris dans l'ICE d'une espèce à l'autre. La présence d'une ICE peut également permettre l'échange de fragments d'ADN provenant de n'importe quelle partie du génome, cela a récemment été montré chez *M. agalactiae* et *Mbov* (Dordet-Frisoni *et al.*, 2014; García-Galán *et al.*, 2021). D'autres éléments mobiles sont retrouvés chez les mycoplasmes comme des IS (séquence d'insertion) et des prophages. Peu de phages de mycoplasmes ont pour le moment été identifiés malgré la grande diversité d'hôtes infectés par ces bactéries. Seuls quatre phages ont été identifiés : le phage Br1 retrouvé chez *Mycoplasma bovirhinis* ; les prophages MAV1/MFV1 qui ont été identifiés chez *Mycoplasma arthritidis* et *M. fermentans* qui sont très proches, et qui sont distribués chez différentes espèces de mycoplasmes (MHoV1 chez *M. hominis*) ; les prophages MAgV1 identifiés chez *M. agalactiae* mais présents dans le génome d'autres espèces ; et finalement le prophage P1 de *M. pulmonis* (Citti *et al.*, 2020).

Pour conclure sur ces aspects, il est vrai que les mollicutes sont dotés de petits génomes qui résultent d'une évolution par perte massive de gènes. Cependant, on se rend compte de plus en plus que de nombreux éléments mobiles sont présents au sein de leurs génomes, ce qui leur permet de réaliser de nombreux échanges d'ADN et d'évoluer rapidement. L'ensemble de ces échanges peut aboutir à des génomes dits mosaïques composés d'éléments génétiques provenant de plusieurs espèces ou souches différentes (García-Galán *et al.*, 2021). Ceci peut rendre les analyses phylogénétiques de mycoplasmes plus complexes

1.3. Objets d'études

Les mollicutes, et particulièrement les mycoplasmes, sont des espèces pouvant servir de modèles pour l'étude de nombreux phénomènes biologiques. En tant qu'organismes minimaux, ils peuvent notamment être utilisés pour comprendre les principes même de

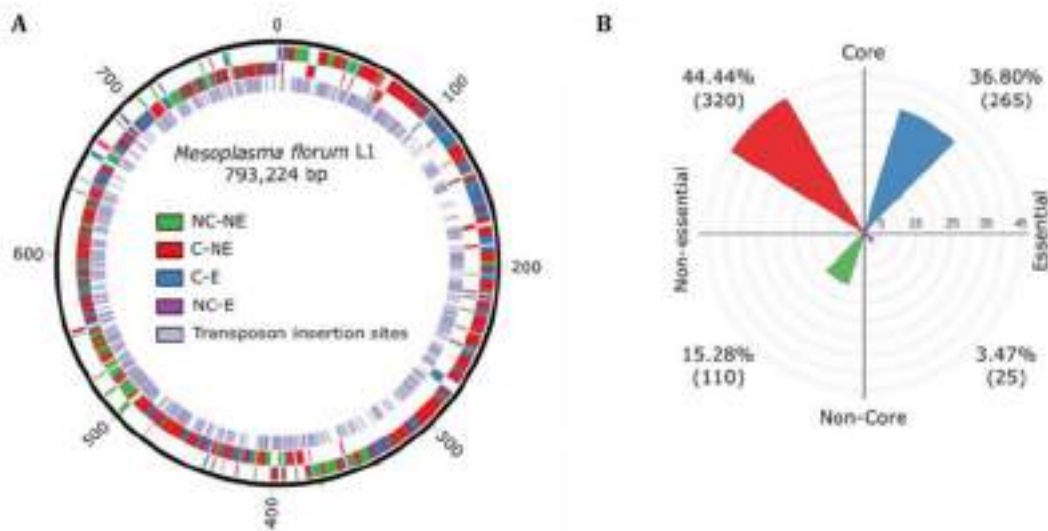


Figure 2 : Détermination des gènes essentiels chez *Mesoplasma florum*. Adapté de Baby *et al.* 2018 ; NC : Non-Core ; C : Core ; NE : Non-essentiel ; E : Essentiel. A- Représentation des résultats d'essentialité des gènes par mutagenèse à saturation chez *Mesoplasma florum*. B- Distribution de l'ensemble des gènes selon leur classification.

la vie. Les connaissances fondamentales acquises pourront à terme servir à une meilleure compréhension de la relation hôte-pathogène et ainsi à combattre les espèces les plus pathogènes.

1.3.1. Etude d'organismes minimaux et biologie de synthèse

Comme expliqué précédemment, les mycoplasmes sont des organismes aux génomes réduits naturellement et malgré une perte massive de nombreux gènes non essentiels, les mycoplasmes ont conservé les éléments nécessaires pour leur survie ainsi que leur capacité à infecter le large panel d'hôte que l'on connaît. On estime qu'au minimum 256 et au maximum 422 gènes, seraient essentiels chez les mycoplasmes (respectivement chez *M. genitalium* et *M. pulmonis*) (Mushegian and Koonin, 1996; Dybvig *et al.*, 2010; Grosjean *et al.*, 2014). Chez *Mesoplasma florum*, 290 gènes sont essentiels à la croissance en laboratoire (Figure 2) et parmi ceux-là, 25 sont spécifiques à *Mesoplasma florum* et n'appartiennent pas au core génome des mollicutes (Matteau *et al.*, 2017). En comparant le nombre de CDS aux gènes essentiels estimés, il est admis que les mycoplasmes ont tout de même conservé de nombreux gènes ayant certainement des rôles dans d'autres conditions de croissance notamment en milieu naturel : 482 CDS sont estimés chez *M. genitalium* et 783 CDS chez *M. pulmonis* (Dybvig *et al.*, 2010; Grosjean *et al.*, 2014). Ces génomes de mycoplasmes dits minimaux peuvent être toutefois encore réduits en laboratoire et cela a par exemple conduit à la construction de la bactérie au génome synthétique JCVI-Syn3.0 (530kb), à partir du génome de *Mycoplasma mycoides* subsp. *capri* de plus de 1Mb (Hutchison *et al.*, 2016). Le groupe des mycoplasmes est une référence en terme de simplicité pour l'étude des différents mécanismes cellulaires les plus répandus dans le monde du vivant comme l'étude du set de gènes minimal pour la traduction chez *M. genitalium* qui aurait 144 CDS contre 228 pour *Escherichia coli* et 210 pour *Bacillus subtilis* (Grosjean *et al.*, 2014). Chez *M. pneumoniae*, une étude approfondie démontre la présence d'ARN non codants dans ces bactéries minimales ainsi que l'importance particulière des petites ORF de moins de 100 résidus qui sont proportionnellement plus essentielles que les autres ORF (Lluch-Senar *et al.*, 2015).

Outre l'aspect fondamental de « compréhension de la vie », ces organismes minimaux intéressent aussi de nombreuses équipes pour des applications biotechnologiques nécessitant des approches de biologie de synthèse. Un axe majeur de la biologie de synthèse est la construction de châssis bactériens possédant un génome minimal

nécessaire à sa survie, dans lequel on va pouvoir insérer des modules génétiques afin d'ajouter une fonction particulière à cet organisme et qui sera finalement comparable à une usine de production. La construction de toutes les pièces d'un châssis bactérien fait partie des objectifs à long terme, mais actuellement, les connaissances nécessaires pour la construction d'une cellule synthétique ne sont pas suffisantes. Des stratégies plus classiques, reposant sur des bactéries non réduites sont en cours de développement avec l'utilisation d'espèces non pathogènes comme, par exemple, *Staphylococcus epidermidis*, utilisée comme vecteur pour une thérapie contre certains cancers (Charbonneau *et al.*, 2020). En revanche, utiliser une bactérie pourvue de toutes ses « capacités génétiques » naturelles peut poser des problèmes d'une part éthique, avec la possibilité de libérer un organisme vivant non-naturel, mais aussi des problèmes d'interférence entre les voies métaboliques ou les fonctions ajoutées et les voies déjà présentes. Les mycoplasmes et les mollicutes en général, possèdent naturellement des génomes réduits qui utilisent également un code génétique alternatif limitant les risques de transferts horizontaux de gènes vers d'autres espèces bactériennes hors mollicutes. Ils sont donc potentiellement des « quasi-châssis » naturels très intéressants. Plusieurs groupes travaillent sur la compréhension approfondie d'une espèce particulière pour des applications dans le domaine de biologie de synthèse et au-delà, comme le groupe de Sébastien Rodrigue (Université de Sherbrooke, Canada) sur *Mesoplasma florum* ou le groupe de Luis Serrano (Center for Genomic Regulation, Barcelone, Espagne) sur *M. pneumoniae*. Récemment, les nombreux développements et travaux fondamentaux réalisés sur ce mycoplasme ont permis de mettre en place une preuve de concept pour son utilisation comme châssis de sécrétion de molécules pour combattre les biofilms bactériens présents lors d'infection à *Staphylococcus aureus*, un problème clinique important (Garrido *et al.*, 2021).

1.3.2. Relation hôte-pathogène et virulence

Paradoxalement à leur génome dit minimal, les mycoplasmes sont capables d'infecter, de s'adapter et de persister chez de très nombreux hôtes et niches écologiques, provoquant ou non des maladies. Ceci nécessite la présence de facteurs de virulence et d'adaptation à l'hôte, pour s'y installer et pour y persister. Les facteurs de virulence tels que des toxines ne sont qu'exceptionnellement décrits chez les mycoplasmes, et ce sont le plus souvent les réponses immunitaires exacerbées qui sont à l'origine des dommages liés aux infections à mycoplasmes (Razin, Yogev and Naot, 1998; Yiwen *et al.*, 2021).

1.3.2.1. Facteurs de virulence

Les principaux facteurs de virulence de ces bactéries sont associés à leurs capacités à infecter et persister au sein d'un hôte. L'adhérence est un des processus les plus importants à l'établissement de la bactérie chez son hôte. De nombreuses adhésines et protéines de surface différentes selon les espèces sont impliquées dans l'adhérence des mycoplasmes aux cellules épithéliales de l'organe qu'il infecte, mais également aux cellules immunitaires. Le cytosquelette présent chez certaines espèces comme *M. pneumoniae* joue un rôle dans cette adhérence via la polarisation de la localisation des adhésines, mais la plupart des autres ne possédant pas de cytosquelette sont tout de même capables d'adhérer aux cellules (Razin, Yogev and Naot, 1998; Balish, 2014). L'adhésion est la première étape importante permettant pour le développement de l'infection. Ensuite, différents mécanismes d'évasion et modulation du système immunitaire, bien connus chez les mycoplasmes, vont être à l'origine de la persistance des bactéries et donc de la chronicité des infections dans l'organe cible et même dans différents organes de l'hôte.

La production de biofilm par exemple (McAuliffe *et al.*, 2006), est caractérisée chez plusieurs espèces et permet une résistance accrue à de nombreux facteurs environnementaux.

Certains mycoplasmes peuvent être trouvés dans des cellules eucaryotes, phagocytaires ou non, ce sont des bactéries intracellulaires facultatives, c'est par exemple le cas de *M. penetrans*, *M. genitalium*, *Mycoplasma hyopneumoniae* et de *Mgal* (Lo *et al.*, 1992; Jensen, Blom and Lind, 1994; Razin, Yogev and Naot, 1998; Winner, Rosengarten and Citti, 2000; Razin and Hayflick, 2010; Raymond *et al.*, 2018). Ce phénomène d'invasion cellulaire permet le passage de la barrière épithéliale pour se disséminer dans l'organisme, mais également d'établir une niche écologique pour l'organisme qui n'est plus directement menacé par le système immunitaire. Il va subsister de cette manière et ceci peut également être à l'origine d'infections chroniques.

La variation antigénique de surface est un des mécanismes majeurs permettant la diversification de la population de mycoplasmes afin d'échapper à une réponse spécifique du système immunitaire. Celle-ci est notamment due à la présence, de nombreuses copies variantes de gènes codant ces protéines dans les génomes de mycoplasmes. Plusieurs mécanismes régissent celle-ci (Citti, Nouvel and Baranowski, 2010; Arfi *et al.*, 2021) : (i) le premier est un changement par insertion ou délétion qui se produirait au niveau des

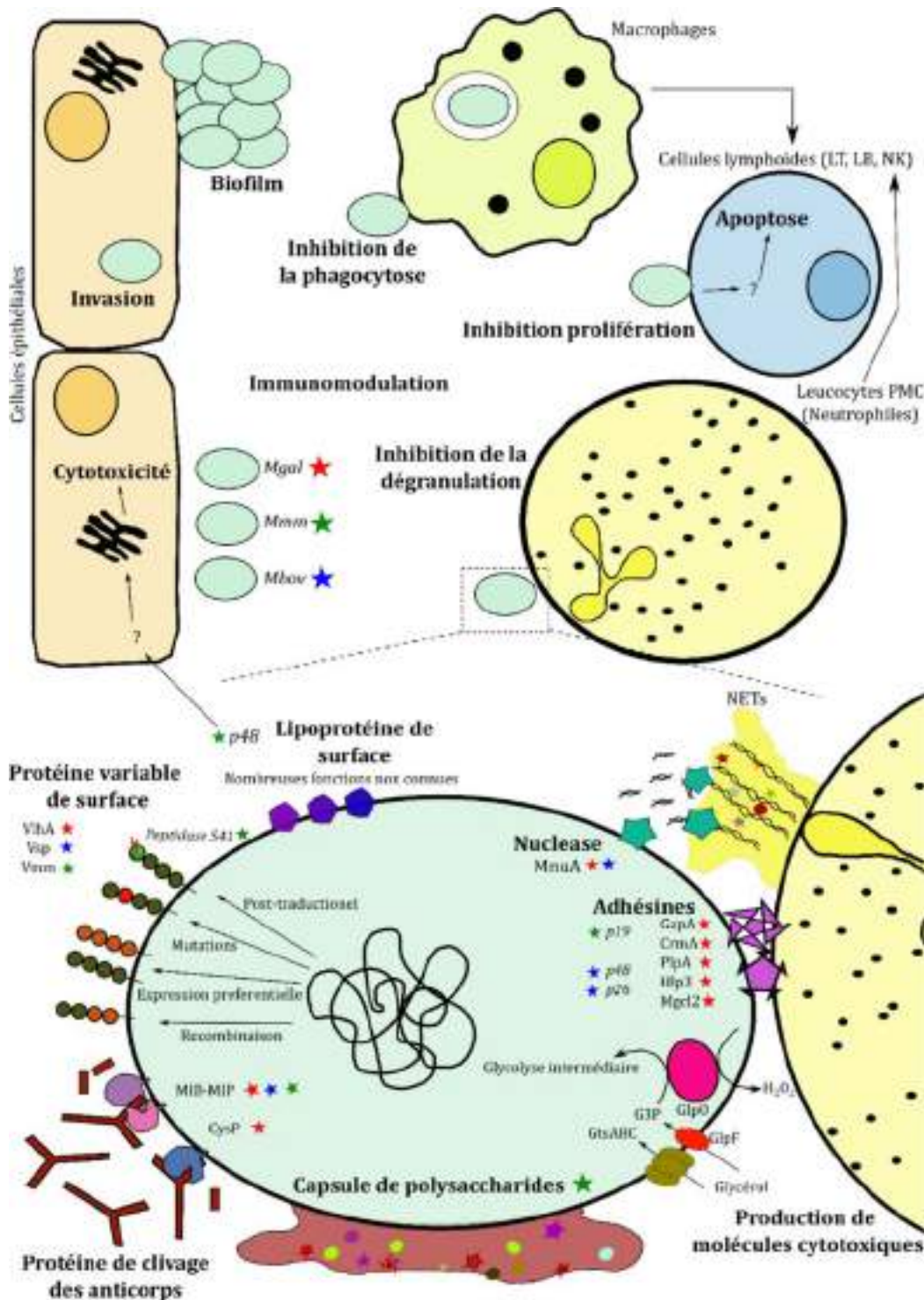


Figure 3 : Représentation schématique des principaux facteurs de virulence identifiés chez *Mycoplasma gallisepticum*, *Mycoplasma mycoides* subsp. *mycoides* et *Mycoplasma bovis*. Les principaux mécanismes sont représentés en gras : L'inhibition de la phagocytose, l'invasion cellulaire, l'immunomodulation, l'induction de l'apoptose, l'inhibition de la dégranulation, l'induction de la cytotoxicité par le réticulum endoplasmique, la formation de biofilm, la production de protéines variables de surface, la production de nucléases, d'adhésines, de lipoprotéines, le clivage des anticorps, la production d'une capsule de polysaccharides et la production de molécules cytotoxiques. La bactérie est représentée en vert, les cellules épithéliales en orange, les cellules immunitaires innées en jaune et les cellules immunitaires adaptatives en bleu.

ÉTUDE BIBLIOGRAPHIQUE : Généralités sur les mollicutes et les mycoplasmes

promoteurs des différentes copies de gènes et qui permet l'activité ou non du promoteur et donc la transcription ou non de la copie ; (ii) le second mécanisme est basé sur la recombinaison site-spécifique médiée par exemple par la protéine Xer1 chez *M. pulmonis*, où les copies sont inactives et recombinent sur un site ou un promoteur unique qui permet l'expression de la copie présente ; (iii) le dernier mécanisme a été décrit chez *M. genitalium*, et est médié par la recombinaison de différents morceaux de copies du gène afin de générer des variations et des gènes mosaïques (Citti, Nouvel and Baranowski, 2010; Arfi *et al.*, 2021). Plus récemment, il a été montré que d'autres mécanismes, n'agissant pas au niveau génomique, pouvaient également être responsables de la diversification des protéines de surfaces. Chez les mycoplasmes, l'absence de aminoacyl-tRNA synthetases entraîne des erreurs fréquentes sur l'acylation des acides aminés sur les tRNA et par conséquent des erreurs de traduction plus fréquentes (L. Li *et al.*, 2013). Des modifications protéolytiques sont également caractérisées ce qui permet là encore de diversifier la population de protéines de surfaces comme cela a pu être montré avec la serine protéase S41 chez *Mcap* (Ganter *et al.*, 2019).

Outres les protéines d'adhésion et les protéines variables de surface, des facteurs de virulence permettant de piéger et/ou de cliver les anticorps ont récemment été caractérisés chez les mycoplasmes : par exemple chez *M. genitalium* la protéine M (Glover *et al.*, 2014) ou de façon plus répandue chez plusieurs espèces de mycoplasmes le système MIB-MIP (Arfi *et al.*, 2016; Nottelet *et al.*, 2021). Ce dernier est notamment capable de lier grâce à la protéine MIB un anticorps déjà fixé à son antigène, de le détacher et de le cliver à l'aide de la protéine MIP.

De nombreux autres facteurs de virulence plus ou moins spécifiques de certaines espèces sont connus. La production d' H_2O_2 en est un, même si son mode d'action et son rôle dans la virulence n'est pas toujours bien claire. La cytotoxicité sur les cellules de l'immunité a parfois été démontrée par le biais de la production d' H_2O_2 (Pilo *et al.*, 2005; Szczepanek *et al.*, 2014). D'autres facteurs plus ou moins connus sont capables d'induire de la cytotoxicité allant jusqu'à l'apoptose des cellules eucaryotes.

Les mycoplasmes ont finalement un large panel de facteurs de virulence qui leur permet d'infecter et se maintenir dans l'hôte. Tous ne sont pas encore connus ou résolus et feront l'objet d'études dans les prochaines années. Les facteurs retrouvés chez nos trois espèces d'intérêts sont décrits par la suite et représentés Figure 3.

1.3.2.2. Interaction avec le système immunitaire

Tout d'abord l'immunité innée est composée majoritairement par la réponse des macrophages, des neutrophiles et des Lymphocytes NK (natural killers). Ces cellules peuvent reconnaître les mycoplasmes grâce aux récepteurs de type TLR (Toll like receptor) qui reconnaissent des structures conservées entre microbes (Shimizu, Kida and Kuwano, 2005). Une fois reconnues, elles sont en capacité de produire des signaux pour détruire les micro-organismes par endocytose et production de molécules cytotoxiques. Les neutrophiles peuvent également produire des NETs (Neutrophil Extracellular Traps) lors de la nétose, ou un amas de longs fragments d'ADN et de protéines diversifiées produit par les neutrophiles vont piéger et détruire les agents pathogènes extracellulaires. Ces signaux peuvent également participer à la signalisation à l'échelle de l'organisme pour déclencher par exemple la réponse inflammatoire avec la production de TNF- α et autres cytokines pro-inflammatoire (IL-1 β , IL-8 etc.) puis la réponse adaptative spécifique médiée principalement par les lymphocytes B et lymphocytes T (Qin, Chen and You, 2019; Di Federico *et al.*, 2020; Baquero *et al.*, 2021). L'H₂O₂ produit par les mycoplasmes serait également à l'origine de l'expression de TNF α et du facteur de transcription NF κ B qui induit l'inhibition de l'apoptose et l'activation des macrophages (Blötz and Stülke, 2017). L'ATPase F1F0 ferait aussi l'objet de la reconnaissance par certains TLR présents sur les cellules de l'immunité innée et entrenerait l'activation du NF κ B (Shimizu, Kida and Kuwano, 2005). Malgré tout cela, les mycoplasmes parviennent à survivre et se propager et ceci va engendrer une augmentation de la destruction des tissus (Razin, Yogev and Naot, 1998; Qin, Chen and You, 2019). Comme on a pu le voir précédemment, plusieurs facteurs de virulence de mycoplasmes permettent d'échapper au système immunitaire de l'hôte et de persister. L'invasion cellulaire, la production de biofilm, les protéines variables de surface sont, en partie, responsables mais de nombreuses interactions entre les mycoplasmes et le système immunitaire de l'hôte (dont le clivage des anticorps par MIB-MIP) sont désormais caractérisées ce qui participe grandement à leur pathogénicité (Razin, Yogev and Naot, 1998; Yiwen *et al.*, 2021). En effet, la plupart des espèces de mycoplasmes interagissent avec les cellules du système immunitaire, conduisant à une modulation de la réponse de l'hôte et une augmentation des chances de persistance de la bactérie. Ces différents mécanismes immunomodulateurs sont loin d'être entièrement compris tant la relation entre la bactérie et le système immunitaire est complexe. Plusieurs facteurs sont notamment

responsables de la régulation négative de la réponse de l'hôte, par exemple, *M. hyorhinis* possède une protéase capable de dégrader l'IL-2 qui est un stimulateur de la prolifération lymphocytaire (Razin, Yogev and Naot, 1998). La lipoprotéine de surface p68 chez *M. hyopneumoniae* ou p48 chez *M. fermentans* sont également capables d'induire l'apoptose de nombreuses cellules de l'immunités tout en maintenant l'induction de la production de cytokines pro-inflammatoires telles que l'IL-1 β et le TNF α (Qin, Chen and You, 2019; W. Liu *et al.*, 2019). Certaines espèces de mycoplasmes peuvent induire la sécrétion par les cellules dendritiques d'IL-10 qui est une cytokine anti-inflammatoire, c'est notamment le cas de *M. hominis* (Goret *et al.*, 2017). Un mécanisme d'inhibition de l'activation des L_{T/B} est également décrit pour plusieurs espèces de mycoplasmes (Qin, Chen and You, 2019).

Cette interaction est à l'origine de nombreuses recherches chez les mycoplasmes pour comprendre comment ces organismes minimaux peuvent se retrouver et infecter une gamme d'hôtes aussi large.

Finalement, les traitements antibiotiques étant limités et entraînant l'apparition de nombreuses résistances dans le temps chez les mycoplasmes (Pereyre and Tardy, 2021), la compréhension de la virulence et de la relation entre l'hôte et le mycoplasme permettra le développement de traitements contre ces infections et particulièrement le développement de souches vaccinales qui pour le moment ne sont que pas ou peu efficaces pour la majorité des espèces pathogènes importantes.

1.4. *Mycoplasma gallisepticum*

Mgal a été décrit pour la première fois en 1905 en Angleterre, associé à une maladie observée chez la dinde. Ce n'est qu'en 1943 que cette bactérie a été cultivée à partir d'isolats de poulet infecté (Ley, 2013). C'est aujourd'hui une des espèces de mycoplasmes les plus étudiées et ayant un impact important dans le monde agro-industriel, mais pas seulement.

1.4.1. Généralités sur la bactérie

Mgal est une espèce de mollicutes appartenant au groupe phylogénétique Pneumoniae (Figure 1). Cette espèce est capable d'infecter les oiseaux notamment les volailles et certains oiseaux migrateurs. C'est une bactérie qui croit rapidement à 37 °C en milieu Hayflick, son temps de génération est estimé à 1,2 heure en milieu liquide (Beaman and Pollack, 1983). Sur milieu gélosé, les colonies apparaissent entre 3 à 15 jours, sont de

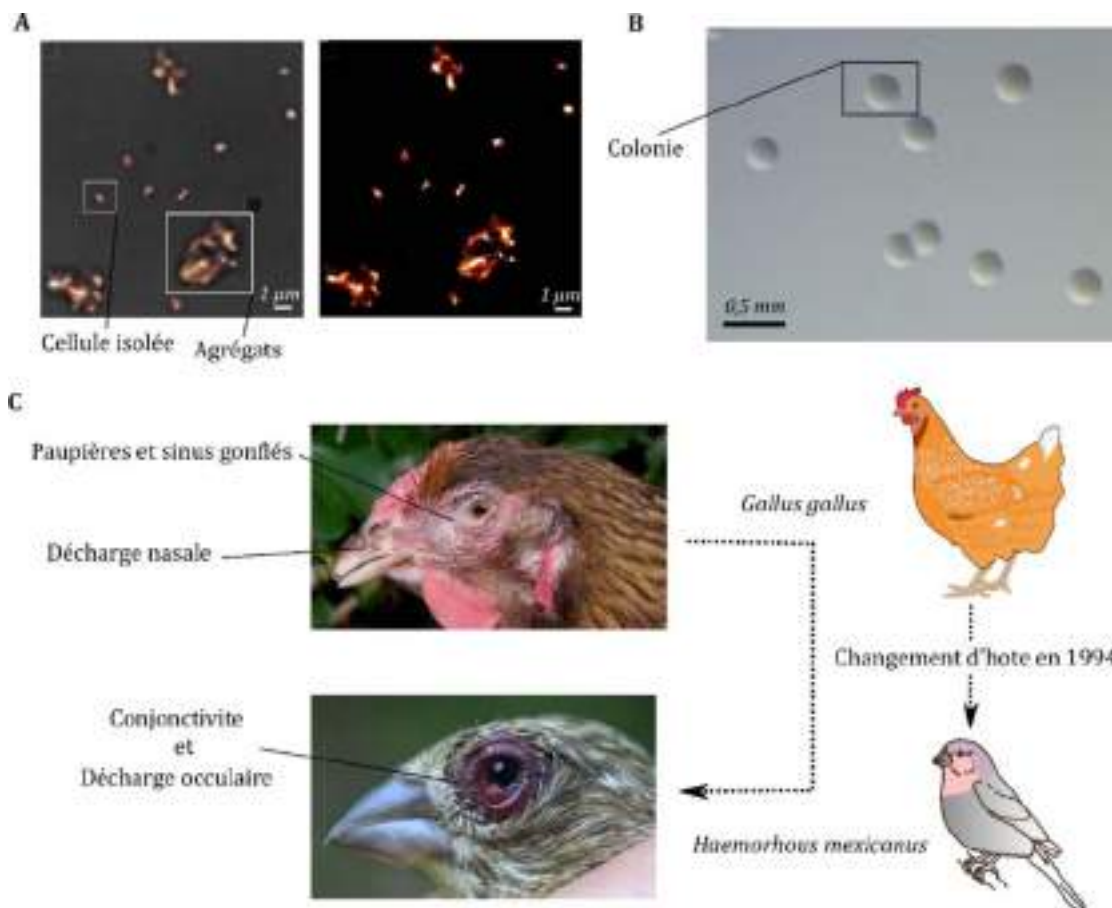


Figure 4 : *Mycoplasma gallisepticum*. **A-** Observation de *Mgal* exprimant la mEos3.2 par microscopie à super résolution (images obtenues par le Docteur Fabien Rideau). **B-** Photo d'une colonie de *Mgal* obtenue sur boîte et observée à la loupe binoculaire. **C-** Images des symptômes provoqués par *Mgal* sur ses deux hôtes principaux. (Fournies par le Docteur Camille Bonneaud)

petite taille (0,2-0,3 µm), et ont un aspect d'œuf sur le plat, typique des mycoplasmes. Les cellules individuelles quant à elles ont une taille variant de 0,2 à 0,5 µm (Figure 4) ; elles sont généralement de forme conchoïdales avec parfois apparition d'une organelle qui polarise la cellule avant la division. Ce sont des bactéries qui fermentent le glucose et le maltose, ce qui provoque une acidification du milieu de culture qui peut être suivie par l'ajout de rouge de phénol qui vire à l'orange lors de la croissance de la bactérie (Ley, 2013). Au total, 27 génomes de *Mgal* sont disponibles dans GenBank dont 18 assemblés entièrement. Les tailles de génome de ces souches varient de 926 kb à 1 010 kb, soit une variation de 8,3 %. Le pourcentage GC de ces génomes est de l'ordre de 31,5 %. La souche S6 qui contient 694 CDS prédits a été plus particulièrement utilisée dans le cadre de ce travail de thèse. Cette souche a été décrite pour la première fois en 1955 lors du 4^{ème} congrès « Western Poultry Disease Conference » (Zander, 1961) et a été isolée dans une volaille. Actuellement aucun élément mobile autre que des IS n'a été mis en évidence chez *Mgal*, ceci peut sûrement être dû à la présence d'un système CRISPR chez les souches ancestrales telles que la souche S6.

1.4.2. *Mycoplasma gallisepticum* chez la volaille

Mgal est le mycoplasme pathogène d'oiseaux, le plus important connu à ce jour (Ley, 2013). Il serait responsable de grandes pertes économiques dans l'industrie de la volaille, à l'échelle mondiale (Levisohn and Kleven, 2000; Dhondt *et al.*, 2005; Ley, 2013). Aux USA, l'impact économique sur la production d'œuf est estimé à 150 millions de dollars par an (Hennigan *et al.*, 2012). Il est à l'origine d'une maladie respiratoire chronique chez les volailles, souvent associée à une perte de poids (particulièrement chez la dinde) et une diminution de la production d'œufs (10-20% chez les poulets)(Yadav *et al.*, 2021). Chez les jeunes poulets, la mortalité peut atteindre 5-10%, Des complications sévères peuvent également subvenir lors de coïnfections avec d'autres pathogènes, ce qui augmente l'impact sur les élevages. (Ley, 2013; Wijesurendra *et al.*, 2015). Étant donné l'impact économique de *Mgal*, des moyens de lutte (curatif et préventif) ont été mis en place : (i) sur des animaux malades, l'antibiothérapie reste la principale stratégie de lutte ; elle permet d'éviter, sur du court terme, la perte d'une trop grosse partie de l'élevage (Kleven, 2008). Cependant, l'utilisation des antibiotiques chez les animaux est de plus en plus controversée avec l'apparition et la possible dissémination de résistances ; (ii) en prévention, la seconde solution mise en place est la vaccination. Pour le moment trois

ÉTUDE BIBLIOGRAPHIQUE : Généralités sur les mollicutes et les mycoplasmes

souches vaccinales sont disponibles : Souche F, F6-85 et ts-11 (Levisohn and Kleven, 2000; Ley, 2013; Wijesurendra *et al.*, 2015). Elles ont montré une certaine efficacité dans certains cas (notamment sur le poulet), mais pour le moment aucune ne l'est vraiment sur la dinde. De plus, ces souches étant utilisées comme vaccins vivants atténués, l'apparition de symptômes voire même de révertants est possible ainsi que la transmission aux autres espèces comme la dinde (Whithear, 1996; Peebles and Branton, 2012; Ishfaq *et al.*, 2020). Une alternative vaccinale a été initiée en 2007, avec l'expression de l'antigène majeur de *Mgal*, VlhA, à la surface de *Bacillus thuringiensis*, malheureusement aucun essai complémentaire n'a été publié (Liu *et al.*, 2008).

Mgal se propage par transfert vertical *in ovo* (poule à œuf) ou par transfert horizontal avec un contact direct (Levisohn and Kleven, 2000). Une transmission par gouttelettes et aérosols a également été mise en évidence. *Mgal* est un parasite obligatoire, mais il a une forte capacité à survivre hors de l'hôte (dépendant du pH, de la température et de l'humidité). Il peut par exemple survivre jusqu'à 3 j dans les fèces de poulet, 24 h dans les nasaux humains, 18 semaines dans le blanc d'œuf à 37 °C (Ley, 2013).

1.4.3. Émergence chez les oiseaux sauvages

En 1994, aux USA, *Mgal* a effectué un changement d'hôte des volailles vers les oiseaux sauvages notamment le Roselin (House finch : oiseaux migrateurs) et les oscines (songbirds). Il est responsable dans ces nouveaux hôtes d'infection sur les voies respiratoires supérieures et de conjonctivites sévères entraînant une forte mortalité liée à la diminution de l'alimentation et de la capacité d'échappement aux prédateurs (Dhondt, Tessaglia and Slothower, 1998; Dhondt *et al.*, 2005; Bonneaud *et al.*, 2011; Ley, 2013). Sur les quatre premières années d'infection, il est estimé que plus de 228 millions d'oiseaux seraient morts (Nolan, Hill and Stoehr, 1998). La bactérie est ensuite passée d'une vague épizootique à un niveau enzootique (Staley and Bonneaud, 2015). Des événements de changement d'hôte pour divers pathogènes ont été décrits (exemple du virus SARS-CoV2 récemment, ou de la bactérie *Staphylococcus aureus* de l'humain vers d'autres animaux) (Viana *et al.*, 2015), mais peu de cas permettent un suivi en temps réel de l'interaction hôte-bactérie depuis le début d'une infection bactérienne. De nombreuses équipes se sont alors intéressées à *Mgal* et un groupe de surveillance (<http://feederwatch.org/>) des Roselins a été mis en place (Dhondt, Tessaglia and Slothower, 1998). Des études évolutives ont démontré une adaptation importante de l'hôte qui résiste de plus en plus à

Mgal et qui tolère plus facilement la bactérie. Mais une évolution de la virulence de *Mgal* est également observée contre ce même hôte. Cette coévolution continue a pu être démontrée à l'aide de souches de *Mgal* ancestrales au changement d'hôte (avant 1994) et de souches récupérées annuellement depuis 1994 (au moment de l'infection saisonnière) mais aussi grâce à la possibilité de réaliser des tests chez les animaux. En effet, des oiseaux dits « non exposés » provenant d'Arizona et n'ayant jamais été mis au contact de *Mgal* (provenant de l'Ouest américain non touché) et des oiseaux dits « adaptés » provenant de l'Alabama (présents dans les zones infectées) sont disponibles et peuvent servir comme matériel d'étude (Bonneaud *et al.*, 2011, 2018; Staley *et al.*, 2018). Cette adaptation bactérienne rapide peut s'expliquer notamment grâce au fort taux de substitutions observées dans le génome de *Mgal* : 10^{-5} sites/ans. En comparaison, le plus rapide taux jusqu'alors caractérisé était de l'ordre de 10^{-6} sites/ans chez *Helicobacter pylori* (Delaney *et al.*, 2012).

La caractérisation de nombreuses souches a permis des travaux de génomique comparée afin de comprendre et observer les différents facteurs changeants entre les souches de *Mgal* de volailles dites ancestrales et les souches de roselins. Différents traits évolutifs ont pu être relevés, l'un d'eux est la perte très rapide dans le génome de certaines souches du locus CRISPR présent à l'origine dans toutes les souches ancestrales (avant 1994). En effet, il a été observé que le taux d'évolution du locus a été très important et ceci, jusqu'à la perte complète des protéines codant le système CRISPR dans 50 % des souches à partir de 2007. Il s'agit là d'une évolution très rapide en comparaison avec d'autres organismes ; A titre d'exemple, il aura fallu plusieurs centaines voire milliers d'années pour qu'*Escherichia coli* et *Yersinia pestis* perdent le système (Delaney *et al.*, 2012; Tulman *et al.*, 2012). Cette évolution pourrait être en lien à la perte d'utilité du système CRISPR lors du changement d'hôte qui aurait donc été inactivé, puis progressivement éliminé. Cette hypothèse peut notamment s'appuyer sur l'observation que les souches de *Mgal* isolées chez le Roselin en 1994/1995/1996 (dates du changement d'hôte) ont une capacité accrue à envahir les cellules eucaryotes (Dowling, Hill and Bonneaud, 2020). Etant ainsi protégées des phages à l'intérieur des cellules eucaryotes, le système CRISPR pourrait alors ne plus avoir d'utilité.

En plus de la perte de locus CRISPR, d'autres différences ont été observées entre les souches infectant la volaille et les souches infectant les roselins, notamment dans les

gènes codant pour les lipoprotéines de surface V_{lha}, mais aussi dans les gènes *hlp3*, *plpA*, *gapA*, *mgc2* et *crmB*-like ayant tous un rôle dans la cytoadhérence et portant tous des mutations ponctuelles (Szczepanek *et al.*, 2010; Tulman *et al.*, 2012).

1.4.4. Virulence

Mgal est une bactérie appartenant à la classe *Mollicutes*, c'est donc une bactérie dite minimale, elle est pourtant retrouvée comme pathogène d'oiseaux domestiques notamment chez le poulet et la dinde. Cette bactérie a même été capable d'effectuer un changement d'hôte et émerger chez des oiseaux sauvages : les Roselins. Au même titre que les autres mollicutes, on retrouve chez *Mgal*, les facteurs de virulence « classiques » que sont l'adhérence (1), la protection active contre le système immunitaire (2) et la production de molécules cytotoxiques (3).

1.4.4.1. Adhérence

L'adhérence, comme énoncé précédemment, est un processus majeur participant à la virulence des mycoplasmes. *Mgal* possède une structure ayant un rôle dans la chimiotaxie (Lam, 2005), la cytoadhérence et la pathogénicité (Ley, 2013). Cette organelle polaire appelée Bleb chez *Mgal*, a particulièrement été étudiée chez *M. pneumoniae* une espèce proche de *Mgal*. À la pointe de cette structure sont trouvées de nombreuses adhésines qui vont être responsables de la cytoadhérence de la bactérie. Les protéines GapA (adhésine) ou CrmA ayant un rôle dans la cytoadhérence ont été identifiées comme des facteurs de virulence importants pour *Mgal* (Goh *et al.*, 1998; Papazisi *et al.*, 2000, 2002; Indiková *et al.*, 2013; Tseng *et al.*, 2013). Cette structure étant similaire à celle de *M. pneumoniae*, il a été inféré que les protéines GapA, CrmA, PlpA, Hlp3 et Mgc2 étaient les constituants majeurs de cette structure chez *Mgal* (Nakane and Miyata, 2009). Ce Bleb aurait également un rôle dans la motilité de la bactérie : le gliding. L'ensemble de ces fonctions sont contrôlées par cette organelle et il est difficile d'étudier l'une de ces fonctions sans en modifier une autre (Balish, 2006). Ces protéines sont associées à différentes fonctions qui ont été décrites comme essentielles à l'établissement de l'infection chez le poulet et à la colonisation de l'hôte (Papazisi *et al.*, 2002; Indiková *et al.*, 2013; Ley, 2013; Tseng *et al.*, 2013; Indikova, Vronka and Szostak, 2014).

1.4.4.2. Défense contre le système immunitaire

Mgal possède de nombreuses stratégies pour se protéger du système immunitaire. Malgré l'absence de paroi, on retrouve des lipoprotéines de surface de la famille des pMGA (aussi nommées VlhA) ayant un rôle dans l'adhérence et l'hémagglutination, mais également dans l'échappement au système immunitaire. En effet, les gènes de la famille des VlhA (~2.3 kb) peuvent représenter jusqu'à 10 % du génome (43 gènes) chez la souche Rlow de *Mgal* (Pflaum *et al.*, 2015). De nombreuses études montrent que *Mgal* module grandement l'expression phénotypique de VlhA par des variations d'expression au cours du temps. Ce phénomène a tout d'abord été montré *in vitro* (Glew *et al.*, 1998) mais également *in vivo* au cours de l'infection (Pflaum *et al.*, 2015). Le mécanisme régissant cette variation est lié à la présence d'un motif portant des répétitions GAA dans le promoteur du gène. La présence de très exactement 12 répétitions permet l'expression du gène (Glew *et al.*, 2000; Liu *et al.*, 2000). Le nombre de génomes et de souches disponibles étant croissant, des analyses de génomique comparée ont permis de démontrer qu'il y avait de nombreuses différences au niveau des gènes codant les protéines VlhA selon les souches de *Mgal* (Szczepanek *et al.*, 2010). En revanche, il a été montré que les copies exprimées ne sont pas spécifiques d'une virulence forte ou faible notamment sur les souches GT5 et Mg7 utilisées parfois comme vaccins (Pflaum *et al.*, 2018). Chez *Mgal*, l'hypothèse préalable d'une sélection naturelle des variants par pression de sélection du système immunitaire a été réfutée dernièrement puisque des variants apparaissent très rapidement après infection, avant même l'apparition de la réponse immunitaire adaptative (Pflaum *et al.*, 2015). Il semblerait également que plusieurs variants soient exprimés au même moment dans la cellule et que la répétition GAA au niveau du promoteur ne soit pas le seul système de régulation (Pflaum *et al.*, 2015). Dernièrement il a été observé que selon la niche écologique de la bactérie (la trachée du poulet ou les sinus du roselin), le profil d'expression d'une même souche de *Mgal* est différent, ce qui pourrait montrer un nouveau rôle des protéines variables de surface (Pflaum *et al.*, 2020). La compréhension de ce mécanisme de variation d'antigène de surface grâce à un changement rapide et réversible des lipoprotéines de surface exprimées nécessite d'être encore approfondi pour comprendre son fonctionnement dans l'optique de la production de souches vaccinales.

ÉTUDE BIBLIOGRAPHIQUE : Généralités sur les mollicutes et les mycoplasmes

La présence de protéines variables à la surface de *Mgal* compromet la mise en place d'une réponse immunitaire efficace et l'apparition d'anticorps spécifiques de l'infection. A cela s'ajoute, un second système de contournement de l'immunité : le système MIB-MIP responsable du clivage des anticorps (Arfi *et al.*, 2016). Une autre protéase ayant un rôle similaire a été identifiée et semble fonctionnelle chez *Mgal*, il s'agit de la protéine CysP (Cizelj *et al.*, 2011).

Mgal fait également partie des mycoplasmes capables d'envahir des cellules non phagocytaires notamment les cellules épithéliales eucaryotes (HeLa et CEF). Cette capacité pourrait avoir deux rôles majeurs, notamment le franchissement de la barrière épithéliale pour établir l'infection dans d'autres organes, mais également la protection contre le système immunitaire de l'hôte pouvant entraîner alors une infection chronique. (Winner, Rosengarten and Citti, 2000; Dowling, Hill and Bonneaud, 2020). Là encore, il a été montré que des protéines ayant un rôle dans l'adhérence telle que GapA, jouaient un rôle majeur dans la colonisation de ces organes (Indiková *et al.*, 2013).

1.4.4.3. Production de molécules cytotoxiques

Tout comme dans le groupe *Mycooides*, l'opéron *glpOKF* responsable de la production d'H₂O₂ en présence de glycérol est retrouvé chez *Mgal*. Le rôle exact du peroxyde hydrogène au cours de l'infection n'a pas encore été démontrée. En revanche, il a été montré *in vitro* que celui-ci était cytotoxique sur les cellules eucaryotes lors de co-cultures (Szczepanek *et al.*, 2014). Les souches vaccinales ts-11 et 6/85 produiraient moins ou pas d'H₂O₂ contrairement à la souche vaccinale F ou la souche virulente R. L'H₂O₂ est considéré chez les mollicutes comme un facteur de virulence, car chez certaines espèces de spiroplasmales, il a été démontré que les protéines GlpO-K-F sont associées à la virulence chez l'insecte vecteur (Lo *et al.*, 2013; Blötz and Stülke, 2017).

1.4.4.4. Autres facteurs de virulence

Finalement, de nombreux autres facteurs de *Mgal* ont été identifiés comme essentiels/importants à la persistance et/ou à la colonisation de l'hôte dans certaines souches de *Mgal*. Parmi ceux-là, le gène *lpd* (codant pour la dihydrolipoamine dehydrogenase) aurait un impact sur la colonisation de l'hôte (Hudson *et al.*, 2006), le gène *malF* (codant pour une perméase transporteur de sucre) ayant un rôle essentiel dans la persistance de la bactérie (Tseng *et al.*, 2013) et le gène *oppD* (de l'opéron *OppABCDF*

qui est un gène présent en deux copies dans des souches virulentes dont une acquise récemment par transfert horizontal de gène depuis *M. synoviae*), qui a un rôle non essentiel, mais important dans la capacité de persistance dans la trachée se traduisant chez un mutant, par des lésions réduites (Tseng *et al.*, 2017). De plus un facteur comme la protéine codée par le gène *mslA*, aurait un rôle dans le recyclage d'acide nucléique de la cellule hôte pour l'utilisation des nucléotides par *Mgal* (Masukagami *et al.*, 2013). On peut également noter la présence de la nucléase de surface MnuA ayant été caractérisée chez *Mbov* comme facteur de virulence important (cf 1,6), qui n'a pour le moment pas été décrite chez *Mgal* comme fonctionnelle, mais qui l'est au moins dans la souche S6 (données non publiées).

Mgal fait également partie des espèces de mycoplasmes capables de former des biofilms (McAuliffe *et al.*, 2006; Chen *et al.*, 2012), ceci permettant la résistance accrue aux antibiotiques et autres stress environnementaux. Récemment, 10 gènes ont été identifiés dans la souche R-Low comme ayant un rôle dans la diminution de la production de biofilm lorsqu'ils sont inactivés (Y. Wang *et al.*, 2017). Cette étude a été réalisée par mutagenèse aléatoire, à l'aide de transposons.

Il semblerait donc que de nombreux facteurs aient un rôle plus ou moins essentiel dans la virulence de *Mgal*. Il semblerait également que cela soit différent selon les souches, selon les niches écologiques (intra ou extra cellulaire) et selon l'hôte lui-même. Une compréhension accrue de ces différents facteurs est nécessaire dans l'optique de produire des souches vaccinales.

1.4.5. Interactions avec le système immunitaire

Les mycoplasmes sont connus pour leur capacité à échapper au système immunitaire et à le moduler. C'est notamment pour cela qu'ils sont responsables d'infections chroniques ou même capables de devenir des organismes commensaux. La réponse immunitaire chez le poulet et le roselin a été étudiée lors de la réponse contre *Mgal* (Staley and Bonneaud, 2015). Lorsque *Mgal* adhère aux cellules épithéliales de la trachée de poulet, la bactérie établit un foyer infectieux et se multiplie. La réponse inflammatoire peut alors provoquer la rupture de la barrière épithéliale et le recrutement de macrophages et de leucocytes dans un premier temps puis de lymphocytes (NK, LT et LB) par la suite. Mais cette inflammation entraîne le relargage de protéines inflammatoires telles que le TNF- α et l'Interleukine 6 (IL-6). Cela va finalement amplifier la destruction de tissus locaux et

permettre à la bactérie de se disséminer dans l'organisme. Les bactéries vont finalement persister à l'aide de leurs nombreux mécanismes de virulence et vont également inhiber les autres réponses immunitaires via la sous-régulation de la production d'IL-8, IL1- β et IL-12 ce qui entraîne une réponse spécifique LT et LB plus faible. Cette immunomodulation est liée aux facteurs de virulence (connus ou non) de la bactérie et permet la persistance de celle-ci dans l'organisme. De manière intéressante, chez le nouvel hôte, il a été montré qu'une forte réponse inflammatoire est présente chez les rosélins provenant de la population non-exposée à la bactérie. Au contraire chez les populations d'oiseaux ayant été au contact de la bactérie, il est observé une diminution de la réponse inflammatoire via une diminution de l'IL-1 β et une augmentation de l'IL-10. Tout ceci est corrélé à la plus grande sévérité des symptômes dans la population ancestrale. Finalement la résistance du roselin a été augmentée au cours du temps puisque la nouvelle population semble capable de résister à la sous-régulation de la réponse spécifique par rapport à la population ancestrale, ce qui permet de maîtriser la bactérie. Cette interaction entre la réponse immunitaire de l'hôte et la bactérie est complexe et de nombreux mécanismes sont encore à étudier.

1.4.6. Perspectives d'études

Mgal est un pathogène important, pourtant, de nombreux mécanismes liés à sa pathogénicité ainsi qu'à l'interaction avec la réponse de l'hôte reste inexplorés. La compréhension de l'émergence chez un nouvel hôte est également un point majeur qui continuera de faire l'objet d'étude dans les prochaines années. Ce pathogène offre la double opportunité de travailler sur des souches de *Mgal* qui sont antérieures et postérieures au changement d'hôte, mais aussi des modèles expérimentaux animaux qui n'ont pas été confrontés à *Mgal* (Amérique de l'ouest) ou des hôtes adaptés (Amérique de l'est) à la bactérie. Celle-ci peut servir de modèle à la compréhension de l'émergence d'un pathogène et à l'adaptation de l'hôte en retour (Staley *et al.*, 2018). L'émergence de la bactérie chez un oiseau sauvage migrateur tel que le Roselin a des conséquences importantes puisque la bactérie peut se disséminer bien plus facilement vers de nouveaux hôtes que les volailles de la ferme. Il a par exemple été montré que sur 53 espèces d'oiseaux provenant de 19 familles distinctes capturés aux USA, 27 espèces montraient la présence d'anticorps dirigés contre *Mgal* même si seules 4 espèces présentent pour le moment des conjonctivites (Dhondt *et al.*, 2014; Staley and Bonneaud, 2015). Une souche

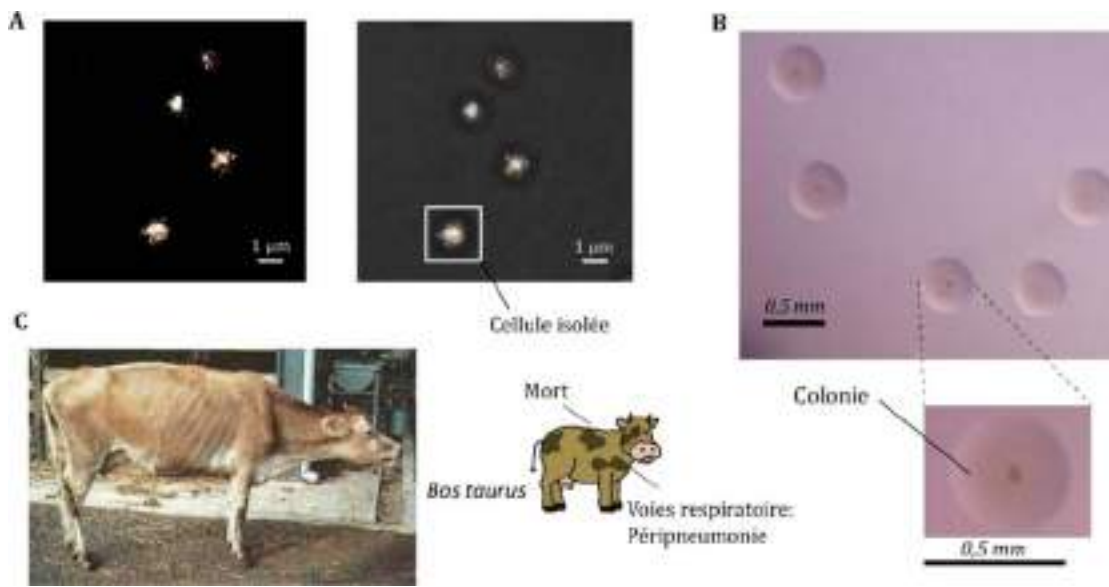


Figure 5 : *Mycoplasma mycoides*. A- Observation de *Mmc* (une espèce voisine de *Mmm*) exprimant la mEOS3.2 par microscopie à super résolution (images obtenues par le Docteur Fabien Rideau). B- Photo d'une colonie de *Mmm* obtenue sur boîte et observée à la loupe binoculaire. C- Image des symptômes provoqués par *Mmm* sur les bovins (provient du site de la FAO (Food and Agriculture Organization) : <https://www.fao.org/>)

vaccinale efficace et sécurisée serait une avancée importante pour l'industrie de la volaille, mais la variation d'antigène de surface médiée par *Mgal* et le peu d'outils d'ingénierie de génome rendent ces développements difficiles.

Que ce soit pour la compréhension des mécanismes permettant un changement d'hôte, l'étude des interactions hôte/pathogène, l'étude évolutive d'une bactérie au sein d'un hôte, ou le développement de souches vaccinales pour pallier les problèmes engendrés par *Mgal*, il est important de pouvoir étudier ce pathogène.

1.5. *Mycoplasma mycoides* subsp. *mycoides*

Mmm est une bactérie classée par l'OIE (Organisation mondiale de la santé animale) comme un pathogène important à surveiller. Ce pathogène a été décrit pour la première fois en Europe au milieu du 16^{ème} siècle (Di Teodoro *et al.*, 2020). Deux rapports de 1765 et 1773 en France et en Suisse en font également référence. La PPCB s'est ensuite mondialement répandue avec les échanges de bétail en Afrique en 1853, puis en Australie en 1858 et aux USA en 1871 (Jores *et al.*, 2020). *Mmm* a finalement été isolé pour la première fois par Nocard et Roux en 1898. Cette bactérie est encore aujourd'hui un problème majeur en Afrique, mais également ailleurs dans le monde où quelques émergences sont parfois observées.

1.5.1. Généralités sur la bactérie

Mmm est une espèce de mycoplasmes appartenant au cluster mycoides (Figure 1) composé également de *Mycoplasma mycoides* subsp. *capri* (*Mmc*), *Mycoplasma capricolum* subsp. *capricolum* (*Mcap*), *Mycoplasma capricolum* subsp. *capripneumoniae* (*Mccp*) et *Mycoplasma leachii*. Ces espèces sont capables d'infecter les ruminants de manière générale (caprins et bovins). *Mmm* est responsable de maladies chez les bovins (yak, bisons et vaches) même si on la retrouve parfois chez les brebis et les chèvres (OIE, 2009). Cette espèce est très proche de la sous-espèce *Mmc* (pathogène de caprins) ; Elle aurait intégré le nouvel hôte bovin assez récemment, dans les années 1700 (Thiaucourt *et al.*, 2011; Dupuy *et al.*, 2012). Cette bactérie fait partie des espèces de mycoplasmes qui poussent de plutôt rapidement (24 h) à 37 °C avec ou sans CO₂ en milieu riche tels que le PPLO ou le SP4. Sur milieu gélosé, les colonies apparaissent entre 3 à 10 jours et ont un aspect d'œuf sur le plat, typique des mycoplasmes (Figure 5). Environ 17 génomes sont disponibles sur NCBI et parmi eux, 9 sont circularisés. Les tailles de génomes varient entre

1,09 et 1,21 Mb soit une variation de 9 %, cela correspond au nombre de pseudogènes qui sont caractérisés chez cette espèce, ce qui est important, mais caractéristique d'une espèce en pleine adaptation à un nouvel hôte (Thiaucourt *et al.*, 2011). Le pourcentage GC du génome est de l'ordre de 23,9 %. La souche sur laquelle nous avons particulièrement travaillé est la souche *Mmm* T1-44. Préalablement nommée *Mycoplasma mycoides* subsp. *mycoides* Small Colony (SC), cette souche a été isolée à partir d'une vache en Tanzanie en 1951 et a été atténuée par 44 passages dans des œufs. Elle est ainsi utilisée comme souche vaccinale et donne une protection partielle à l'hôte (Gourgues *et al.*, 2016).

1.5.2. PPCB

La Péripleumonie contagieuse bovine (PPCB) est une maladie classée par l'OIE pouvant être responsable de la mort de bovins (Figure 5). Au milieu du 20^{ème} siècle, grâce à une large campagne de vaccination et d'abattage sanitaire, la maladie a été éradiquée d'Europe et d'Asie, mais des re-émergences ont parfois été observées comme en 1935, 1953 et 1993 au Portugal, en 1967 et 1982 en Espagne et en 1993 en Italie (Yaya *et al.*, 2008; Di Teodoro *et al.*, 2020). En Afrique, la maladie est endémique encore de nos jours malgré une campagne de vaccination réussie jusqu'aux années 1980. Des réémergences sont ensuite apparues et sont liées à plusieurs facteurs comme le mouvement incontrôlé des animaux, la faible couverture vaccinale et l'immunité qui est lente à acquérir (Thiaucourt *et al.*, 1998). Quelques succès notoires comme au Botswana sont tout de même à noter où malgré la réapparition de la maladie en 1995, 320 000 animaux ont été abattus afin de rendre la région à nouveau exempt de maladie (Amanfu *et al.*, 2006). Des analyses génotypiques ont mis en évidence des différences importantes entre les souches des différents continents, mettant en relief que les réémergences européennes de la PPCB sont dues à un réservoir non identifié présent en Europe et au regain de virulence de certaines souches (Yaya *et al.*, 2008).

Selon l'OIE, le pathogène se transmet principalement par inhalation ou gouttelettes et parfois par la salive ou les urines. La maladie se transmet donc par contact rapproché entre animaux malades et animaux sains. Le temps d'incubation peut être très long entre 1 et 4 mois.

La PPCB est une maladie qui se présente sous trois formes et dépend énormément de la virulence de la souche (Di Teodoro *et al.*, 2020) :

ÉTUDE BIBLIOGRAPHIQUE : Généralités sur les mollicutes et les mycoplasmes

- La forme hyperaiguë, rare, mais caractérisée par une mort rapide du bovin.
- La forme aiguë observée au début d'une épidémie, qui se traduit par de la fièvre et un stress respiratoire important. L'émergence d'une souche épidémique de *Mmm* se traduit le plus souvent par une de ces deux formes et entraîne une mortalité importante (jusqu'à 90 %).
- La forme subaiguë ou chronique, qui se développe chez un hôte ayant déjà eu une forme aiguë. Cette forme est retrouvée dans les zones endémiques et entraîne une mortalité de l'ordre des 10 %.

Les signes cliniques dépendent de nombreux paramètres comme la virulence de la souche, par exemple les souches européennes sont souvent associées à des taux de morbidités et mortalités plus faibles que les souches africaines ; mais également à la santé et le bien être des vaches qui sont parfois en meilleures conditions nutritionnelles par exemple en Europe qu'en Afrique (Di Teodoro *et al.*, 2020). En tout cas, les symptômes peuvent être des dépressions, des pertes d'appétit, de la fièvre, de la toux, des douleurs thoraciques et une respiration accrue. Des lésions au niveau d'un poumon sont principalement retrouvées chez les animaux atteints. Il a été démontré que tout comme *Mgal*, l'inflammation des poumons est en partie due à la réponse immunitaire de l'hôte et que l'augmentation de TNF- α , un activateur de la réponse immunitaire, corrèle avec la sévérité de la maladie (Sacchini *et al.*, 2012; Di Teodoro *et al.*, 2020).

Le contrôle de la PPCB est basé sur des mesures sanitaires et prophylactiques. Le contrôle des mouvements des animaux, additionné de tests sérologiques dans les zones réputées indemnes sont des mesures qui se sont montrées efficaces, mais en Afrique ces méthodes n'ont pas abouti. Trois souches vaccinales sont recommandées par l'OIE, la souche T1-44 qui a été isolée en Tanzanie en 1951, la souche Ben181 et la souche T1sr. Cette dernière est responsable d'une immunité courte et peut entraîner de nombreux symptômes, ce qui limite son utilisation. La souche T1-44 qui est la plus utilisée résulte de 44 passages de culture dans des œufs afin de l'atténuer (Jores *et al.*, 2020). Malheureusement, des symptômes sont souvent rapportés et des cas sévères de PPCB se sont même déclarés lors d'une étude au cours de laquelle la transmission de la souche à une vache saine non vaccinée a été décrite (Mbulu *et al.*, 2004). Cela rend difficile l'adoption de ce vaccin par les populations locales dans les pays où la PPCB est endémique. De plus, le vaccin T1-44 a montré une protection limitée en temps et en intensité selon les études de la virulence

ÉTUDE BIBLIOGRAPHIQUE : Généralités sur les mollicutes et les mycoplasmes

des souches (Jores, Mariner and Naessens, 2013). Plus récemment en Chine, la souche Ben181, atténuée à partir de la souche Ben1 par 181 passages dans le lapin, a été construite et a permis l'éradication de la maladie dans le pays et sa classification en zone libre de PPCB par l'OIE (Y. Li *et al.*, 2016). Ce dernier semble efficace à 95 % pendant deux ans faisant de ce vaccin le plus efficace actuellement, mais la méthode de production de ce vaccin ne rentre pas dans les normes actuelles des standards de bien-être animal et l'efficacité contre des souches africaines ne semble pas supérieure à la souche T1-44 (Jores *et al.*, 2020).

1.5.3. Virulence

Au sein du cluster mycoïdes, les mêmes facteurs de virulence sont retrouvés. L'adhérence est là encore un processus majeur, la défense contre le système immunitaire un second ainsi que la production de molécules cytotoxiques.

1.5.3.1. Adhérence

L'adhérence est la première étape permettant la colonisation de l'hôte. *Mmm* est capable d'adhérer aux cellules non ciliées des poumons (Di Teodoro *et al.*, 2020). De nombreuses lipoprotéines de surface telles que LppA, LppB, LppC, LppQ et Vmm sont antigéniques chez *Mmm* et sont probablement liées à l'adhérence de la bactérie, mais aucune démonstration n'a pour le moment été effectuée. Récemment la protéine p19 a été caractérisée comme adhérente aux cellules EBL qui sont des cellules de poumon bovin (Zhou *et al.*, 2016). L'adhérence permet à la bactérie de pénétrer dans les vaisseaux lymphatiques et sanguins ce qui peut expliquer les symptômes de thrombose parfois observés et qui permet à la bactérie de se retrouver dans d'autres organes (Di Teodoro *et al.*, 2020).

1.5.3.2. Défense contre le système immunitaire

Chez les mycoplasmes, l'adhérence est souvent à l'origine de la capacité à rentrer dans les cellules non phagocytaires, ce qui lui permet de se cacher du système immunitaire. Cette caractéristique n'est pas très répandue dans le cluster mycoïdes mais il a été montré que *Mmm* peut être trouvé à l'intérieur des cellules (Di Teodoro *et al.*, 2018). La bactérie est tout de même capable de survivre de façon importante hors de ces cellules puisque les espèces du cluster mycoïdes, tout comme *Mgal*, possèdent de nombreuses lipoprotéines variables de surface appelées Vmm pour *Mmm* (Persson *et al.*, 2002; Wise *et al.*, 2006). Le

mécanisme de régulation ON/OFF principalement retrouvé chez *Mmm* est lié au mécanisme d'insertion ou délétion dans les promoteurs des différentes copies, par exemple au nombre de répétitions du dinucléotide TA (contrairement à GAA chez *Mgal*), ce qui permet leur expression de manière plus ou moins aléatoire et variée (Persson *et al.*, 2002). Comme expliqué précédemment, cette oscillation des protéines exposées à la surface de la cellule est une première stratégie d'échappement au système immunitaire. De plus, on retrouve des copies de gènes codant pour le système de surface MIB-MIP capable de cliver les anticorps de l'hôte. Ce système est fonctionnel chez *Mmc* qui est très proche de *Mmm* (Arfi *et al.*, 2016).

Finalement il a également été montré dans une étude que *Mmm* entraîne d'une certaine manière l'apoptose des leucocytes et que cela pourrait être lié à la sécrétion d'un effecteur particulier qui cherche encore à être caractérisé (Dedieu, Chapey and Balcer-Rodrigues, 2005).

1.5.3.3. Production de molécules cytotoxiques

La production d' H_2O_2 en réponse au métabolisme du glycérol est particulièrement étudiée dans le cluster mycoïdes. Chez *Mmm*, des corrélations entre l'absence du transporteur de glycérol GtsABC, la perte de cytotoxicité des souches en présence de glycérol et le défaut de production d' H_2O_2 ont été démontrées dans plusieurs expériences (Bischof *et al.*, 2008). GlpO (codée par l'opéron *glpOKF*) est l'enzyme responsable de la dégradation du glycérol après son entrée sous forme de glycerol 3 phosphate (G3P) dans la cellule (via les transporteurs *gstABC* et *glpKF*). Dans le cytoplasme, le G3P est oxydé en glycéraldéhyde-3-phosphate et dihydroxyacétone phosphate par la L- α -glycérophosphate oxydase. L'étape d'oxydation du G3P se fait en présence de NAD^+ et d'oxygène et ceci libère, en fin de réaction, des quantités massives de peroxyde d'hydrogène (H_2O_2). Or la production d' H_2O_2 comme dérivé réactif de l'oxygène (ROS), induit des lésions tissulaires (Pilo *et al.*, 2005) et la synthèse de cytokines pro-inflammatoires (Di Teodoro *et al.*, 2020). Ce facteur pourrait expliquer la différence de virulence entre les souches africaines et européennes, mais il ne peut être le seul facteur étant donné que la souche vaccinale *Mmm* T1-44, moins virulente que les autres souches, produit malgré tout une importante quantité d' H_2O_2 (Bischof *et al.*, 2008; Di Teodoro *et al.*, 2020; Jores *et al.*, 2020).

1.5.3.4. Autres facteurs

Outre les facteurs de virulence caractérisés, de nombreux autres sont certainement présents, mais le faible nombre d'études limite leur caractérisation et la compréhension de leur mécanisme. La compréhension des rôles que doivent jouer les différentes lipoprotéines de surface reste encore un challenge comme les protéines DUF285 qui sont probablement disséminés entre les différentes espèces par HGT et qui n'ont pas encore de rôle caractérisé chez les mycoplasmes (Röske *et al.*, 2010). Une capsule composée de polymères de polysaccharides est présente chez la majorité des espèces du cluster mycoïdes (Bertin *et al.*, 2013). Celle-ci est décrite comme pouvant provoquer des effets cytopathiques *in vitro* (Schieck *et al.*, 2016) mais également ayant un rôle dans la virulence *in vivo* chez *Mmc* (Jores, Schieck, *et al.*, 2019). Malgré la présence de cette capsule qui diminue la capacité d'adhérence de la bactérie (Schieck *et al.*, 2016), il a été montré que *Mmm* est capable de former des biofilms (McAuliffe *et al.*, 2008), lui conférant une résistance accrue à de nombreux stress environnementaux.

D'autres facteurs ont été caractérisés notamment la protéase S41 qui a été acquise par HGT au sein du cluster mycoïdes et qui a une activité caseinolytique et qui pourrait avoir un rôle sur la modification post-traductionnelle des protéines de surface (Ganter *et al.*, 2019). Récemment lors d'une expérience dans laquelle la souche Ben1 a subi 468 passages dans le lapin, 18 gènes ont été associés à l'adaptation dans le lapin, 6 associés à la virulence chez les bovins et 35 associés à la protection de la bactérie contre le système immunitaire chez les bovins (Y. Li *et al.*, 2016). Le rôle de ces gènes candidats reste tout de même à préciser, mais plusieurs lipoprotéines de surface semblent avoir des fonctions importantes.

1.5.4. Perspectives d'études

Le modèle de *Mmm* est un modèle très intéressant pour la compréhension de sa virulence, la construction d'une souche vaccinale ou même par son aspect évolutif. Récemment, une revue scientifique composée de nombreux auteurs travaillant sur le pathogène de la PPCB, décrit les enjeux majeurs des prochaines années sur *Mmm* avec la nécessité d'un vaccin efficace qui passe par la compréhension de la réponse immunitaire de l'hôte à *Mmm*, la compréhension de la virulence et donc par le développement d'outils d'ingénierie génétique pour cette bactérie (Jores, Mariner and Naessens, 2013; Jores *et al.*, 2020). La production d'une souche vaccinale devrait se concentrer sur un vaccin

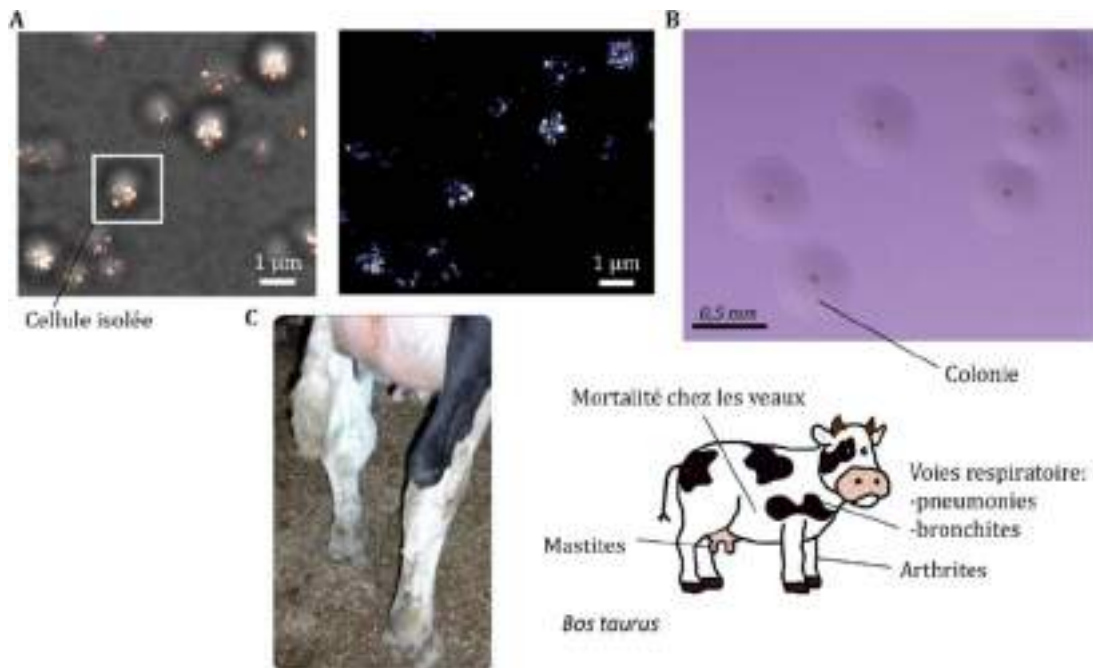


Figure 6 : *Mycoplasma bovis*. **A-** Observation de *Mbov* exprimant la mEOS3.2 par microscopie à super résolution (images obtenues par le Docteur Fabien Rideau). **B-** Photo d'une colonie de *Mbov* obtenue sur boîte et observée à la loupe binoculaire. **C-** Image des symptômes provoqués par *Mbov* sur les bovins (Nicholas *et al.* 2006)

répondant aux différentes souches des différents continents et qui devra s'appuyer sur la réponse à l'IFN gamma médié par les LT_{CD4} de l'hôte (Sacchini *et al.*, 2012).

Sur un aspect plus fondamental, la recherche sur *Mmm* pourrait s'intéresser à la capacité de cette bactérie à s'adapter à un nouvel hôte vu sa récente adaptation aux bovins et aux études qui montrent son gain de virulence chez le lapin par évolution forcée (Thiaucourt *et al.*, 2011; Y. Li *et al.*, 2016). La capacité d'adaptation aux bovins de cette bactérie est liée à la présence de nombreux éléments mobiles dans le génome de cette bactérie, aux nombreux pseudogènes présents, aux nombreux échanges avec d'autres bactéries et à la plasticité du génome de cette bactérie (Thiaucourt *et al.*, 2011).

1.6. *Mycoplasma bovis*

Mbov est depuis quelques années le mycoplasme pathogène de bovins le plus important dans le monde. Il a été isolé pour la première fois aux USA en 1961 sur des vaches atteintes de mastites sévères (Nicholas, 2011). Tout comme *Mmm*, ce pathogène s'est ensuite répandu dans le monde avec le commerce et les échanges de bétails : Israël en 1964, l'Australie en 1970, la France en 1977, le Maroc en 1988. Pas à pas, *Mbov* s'est propagé pour toucher la quasi-totalité des pays du monde. Deux des derniers pays qui n'étaient pas encore concernés, ont à leur tour été touchés, la Finlande en 2012 puis la Nouvelle-Zélande en 2017 (Dudek and Szacawa, 2020).

1.6.1. Généralités sur la bactérie

Comme son nom l'indique, *Mbov* est une bactérie capable d'infecter les bovins. Cette bactérie appartient au groupe phylogénétique Hominis (Figure 1). Elle fait partie des espèces de mycoplasmes qui poussent de manière rapide (24 h) à 37 °C avec ou sans CO₂ en milieu riche tel que le SP4. Sur milieu gélosé, les colonies apparaissent entre 3 à 8 jours et ont un aspect d'œuf sur le plat, typique des mycoplasmes (Figure 6). Les génomes de 234 souches ont été récemment séquencés et sont disponibles dans les bases de données internationales. Une telle représentativité reste rare pour les mycoplasmes, ceci étant lié à l'importance actuelle qui est accordée à cette espèce. Les tailles de génome varient entre 0,94 et 1,11 Mb soit une variation de taille de 15 %. Cette variation est importante et correspond majoritairement à la présence de nombreux éléments mobiles tels que des IS, des ICEs et des prophages (García-Galán *et al.*, 2021). Le pourcentage GC du génome est de l'ordre de 29 %. La souche sur laquelle nous avons particulièrement travaillé est la

souche de référence PG45. Cette souche a été publiée en 1962 sous le nom de *Mycoplasma agalactiae* subsp. *bovis* (Hale *et al.*, 1962) et a été isolée à partir d'un bovin.

1.6.2. Virulence

Tout comme *Mmm* et *Mgal* avant lui, *Mbov* n'est pas pathogène via le relargage de toxine ou autres facteurs de mêmes types. L'adhérence et la résistance au système immunitaire sont, là encore, les principales causes de la pathogénicité de cette bactérie.

1.6.2.1. Adhérence

Tout comme *Mmm* et contrairement à *Mgal*, aucune organelle n'est associée à l'adhérence chez *Mbov* et pourtant, les différentes souches de la bactérie sont capables d'adhérer à de nombreux types de cellules : EBL, EBTr, MDBK, RK (Thomas *et al.*, 2003). Certaines souches non pathogènes ont également été caractérisées comme moins adhérentes. Différents gènes et protéines sont associés à la capacité d'adhérence de la cellule comme la protéine p26, les protéines variables de *Mbov*, les Vsp (Thomas *et al.*, 2003), la protéine p48 (P. Fu *et al.*, 2014), mais également la cyto-adhésine Mbov_0553 qui faciliterait la translocation de la barrière épithéliale (Zhang *et al.*, 2020). Il faut noter aussi que les gènes codant l' α -enolase, la NADH oxydase et le facteur TrmFO-like sont impliqués dans la liaison à la fibronectine et au plasminogène qui sont retrouvés dans la matrice extracellulaire des cellules eucaryotes (Askar *et al.*, 2021).

1.6.2.2. Défense contre le système immunitaire

Comme de nombreuses bactéries et mycoplasmes pouvant établir une infection chronique, la bactérie est capable de moduler et se défendre contre le système immunitaire. Comme pour *Mmm* et *Mgal*, les protéines de surface variables classiquement retrouvées chez les mycoplasmes sont également présentes chez *Mbov*, il s'agit des Vsp. Elles participent à l'adhérence, mais permettent également d'échapper à la réponse immunitaire (Perez-Casal, 2020). Le système de clivage des anticorps MIB-MIP est également retrouvé chez *Mbov* même si, pour le moment, aucune expérience n'atteste de sa fonctionnalité (Arfi *et al.*, 2016). Une nucléase de surface, MnuA, est décrite comme un facteur de virulence important chez *Mbov*, puisqu'elle est capable de dégrader de manière très efficace l'ADN *in vitro* (Sharma, Tivendale, *et al.*, 2015). De par cette activité, il a également pu être démontré que MnuA dégrade de manière très efficace les NETs (Neutrophil Extracellular Traps) (Zhang *et al.*, 2016; Mitiku *et al.*, 2018). La production

d'H₂O₂ chez *Mbov* est possible, mais son association à la virulence n'a pour le moment pas été démontrée (Maunsell and Chase, 2019).

Finalement d'autres mécanismes, plus ou moins caractérisés au niveau génétique, sont décrits chez *Mbov*, par exemple la capacité à adhérer et inhiber la prolifération des PBMC sans inhiber sa production d'IFN γ , la capacité à envahir certaines cellules et s'y développer (Perez-Casal, 2020; Askar *et al.*, 2021), la capacité à adhérer et induire l'apoptose des cellules notamment des neutrophiles (Jimbo *et al.*, 2017; Perez-Casal, 2020; Askar *et al.*, 2021) et à moduler la production de ROS. Il a par exemple été démontré sur les cellules EBL que la protéine p48 de *Mbov* est capable d'induire l'apoptose via un stress du réticulum endoplasmique (Wu *et al.*, 2021), tout comme la protéine p280 sécrétée par *Mbov* pour les macrophages (Zhao *et al.*, 2021).

1.6.2.3. Autres facteurs

De nombreuses études n'ont impliqué que la souche de référence PG45, or, le nombre de passages *in vitro* que cette souche a subi est important et il est maintenant admis qu'une perte de virulence est associée aux nombreux passages *in vitro* (Thomas *et al.*, 2003; Askar *et al.*, 2021). Ainsi on peut imaginer que plusieurs facteurs de virulence n'ont pour le moment pas été détectés dans les très nombreuses souches de *Mbov* qui existent dans les différents pays. De plus, plusieurs études font état de protéines sécrétées qui pourraient avoir un lien avec la virulence de la bactérie (Zubair *et al.*, 2020; Zhao *et al.*, 2021).

Il a également été montré que l'expression de certaines de ces Vsp (VspO et VspB), est associée à la formation de biofilms par la bactérie, ce qui contribue à la résistance aux antibiotiques, à la dessiccation ou à l'échappement au système immunitaire (à l'opsonisation) (McAuliffe *et al.*, 2006; Perez-Casal, 2020).

1.6.3. Pathologie

Mbov est un pathogène de bovin qui est responsable d'une infection chronique particulièrement chez les vaches. Cette bactérie est également le principal pathogène lié à la maladie respiratoire du bovin (BRD : bovine respiratory disease) même si la maladie se déclare sur des animaux préalablement infectés par des bactéries comme *Mannheimia haemolytica*, *Pasteurella multocida* ou des virus comme le parainfluenza virus de type 3 (Nicholas, 2011; Dudek and Szacawa, 2020; Perez-Casal, 2020) ou lors de stress extérieur tel que des chocs de température ou une baisse de l'immunité. La bactérie et la BRD vont

provoquer une forte mortalité chez les veaux, une perte de poids importante chez les vaches et une diminution de la production de lait. Tout ceci a des répercussions économiques importantes : (i) aux Etats-Unis, les pertes économiques liées à *Mbov* sont estimées à 55 millions de dollars par an en 2015 (Perez-Casal, 2020) ; (ii) en Europe, les pertes liées à la mortalité chez les veaux notamment sont estimées à 140 millions d'euros par an (Zubair *et al.*, 2020) ; (iii) enfin, l'émergence de *Mbov* en Nouvelle-Zélande a entraîné des pertes économiques de plus de 200 millions de dollars (Dudek and Szacawa, 2020). Les pathologies engendrées par cette espèce de mycoplasme sont diverses et dépendent de la souche et de l'hôte. Le plus souvent, une hyperplasie péribronchiolaire lymphoïde et une bronchopneumonie caséo-nécrotique (Nicholas, 2011) sont observées au niveau des poumons, tout comme des mastites et des arthrites sévères (Dudek and Szacawa, 2020). Plus rarement, les yeux, l'ouïe et le cerveau peuvent également être atteints puisque des otites et des méningites sont parfois associées lors d'analyses post-mortem (Dudek and Szacawa, 2020). Une large population bactérienne est requise pour déclencher la réponse immunitaire qui se manifeste principalement par une production de cytokine pro-inflammatoire comme le TNF- α , l'IFN γ et l'IL4 et l'IL12 et par l'action des LT_{CD4} et des L_B (Perez-Casal, 2020; Askar *et al.*, 2021). *In vivo*, *Mbov* semble persister via la dérégulation de cette réponse par la production d'IL10 anti-inflammatoire par les macrophages et la sous régulation de la réponse adaptative (Maunsell and Chase, 2019). Cette réponse immunitaire provoquée et modulée par *Mbov* nécessite encore des recherches puisqu'elle semble donner des résultats divergents *in vitro* et *in vivo*, selon l'endroit de l'infection et selon la souche (Maunsell and Chase, 2019). La transmission de la bactérie se fait principalement par aérosols ou par contact proche, mais également par le lait ou le colostrum (Dudek and Szacawa, 2020).

Les mesures prises contre ces bactéries sont du même ordre que pour *Mmm* et *Mgal*. Des traitements antibiotiques sont possibles, mais sont limités, une augmentation de la résistance aux antibiotiques des souches est observée dans le temps de manière préoccupante (Dudek and Szacawa, 2020). De nos jours, aucun vaccin commercial n'est disponible en Europe (Dudek and Szacawa, 2020). Un vaccin inactivé a été utilisé dans le monde, mais son efficacité est limitée dans le temps et efficace seulement s'il est donné à de jeunes animaux, avant un fort taux d'infection (Nicholas, 2011). La prophylaxie, via le contrôle des mouvements des animaux et le dépistage massif sont les principales armes face à cette bactérie. La Nouvelle-Zélande, malgré la non-importation de bovins, a

finalement subi l'émergence de la bactérie en 2017. Mais la mise en place d'un programme d'éradication important basé sur la détection tend à montrer que la stratégie peut être payante (Jordan *et al.*, 2021).

1.6.4. Perspectives d'études

Mbov est un modèle d'étude très intéressant pour de nombreuses raisons. La principale est la compréhension de l'ensemble des facteurs de virulence et la création d'un vaccin efficace dans le but d'éradiquer la maladie. L'infection peut avoir des conséquences dramatiques, comme en 2007 dans un troupeau du nord de l'Angleterre où 24 % des veaux sont décédés (Nicholas, 2011). Les récentes émergences en Nouvelle-Zélande par exemple sont également à l'étude pour comprendre l'apparition de la bactérie et pour développer des stratégies pour la combattre (Dudek and Szacawa, 2020).

Sur un autre aspect, *Mbov* (tout comme *Mgal*) a été capable de franchir la barrière de l'espèce en infectant le cochon (Spargser *et al.*, 2013; Perez-Casal, 2020). Il semblerait que ce soit lié à une souche en particulier et que les répercussions n'aient pas été aussi importantes que pour *Mgal* sur les populations de roselin, mais cela met encore une fois en évidence la capacité d'une espèce de mycoplasme à s'adapter à de nouveaux hôtes et à leurs réponses immunitaires.

D'un point de vue génomique, *Mbov* est un très bon modèle d'étude pour la compréhension de l'évolution d'espèces minimales telles que les mycoplasmes. En effet l'évolution dégressive des mycoplasmes a permis de créer naturellement des espèces bactériennes pourvues de génomes réduits. A l'heure actuelle, les mycoplasmes sont des modèles pour l'étude du génome minimal afin de comprendre les fonctions nécessaires à la vie. Or, il est apparu, lors de la construction de banques de mutants obtenues par transposition, que 16 gènes essentiels chez certaines espèces de mycoplasmes, ne l'étaient pas chez *Mbov* et que 5 appartenaient même au core génome des mycoplasmes (Sharma, Markham and Browning, 2014). Ceci met en avant le fait qu'il n'existe pas un génome minimal unique et que chaque bactérie a son set de gènes essentiels. D'un point de vue fondamental, il serait très intéressant de regarder et de comprendre ces différences afin d'agrandir notre compréhension du vivant. D'autre part, *Mbov* est une espèce de mycoplasmes dans laquelle de nombreux HGT ont été caractérisés, entre souches mais aussi entre différentes espèces de mycoplasmes partageant une même niche écologique. Ainsi l'étude du moteur évolutif le plus important chez les bactéries (les HGTs)

peut également être effectuée au sein de ces organismes minimaux. Chez *Mbov*, ces évènements sont si prégnants, que des souches portant des génomes mosaïques ont récemment été mises en évidence (García-Galán *et al.*, 2021); les échanges de fragments d'ADN de tailles variées se faisant via des éléments mobiles tels que les ICEs, les phages et les IS.

2. SYSTEMES DE DEFENSE BACTERIENS : FONCTIONS ET APPLICATIONS

Le système immunitaire est un champ de recherche important notamment chez l'homme. Il assure la défense contre des agents pathogènes divers, qu'ils soient d'origine virale, bactérienne ou fongique, à l'aide de deux dispositifs complémentaires. L'immunité innée va répondre de manière immédiate contre différents types d'agressions via la reconnaissance de caractéristiques communes à de nombreux agresseurs telles que la paroi bactérienne. L'immunité adaptative va elle se mettre en place lors de la première rencontre avec l'agent pathogène et va permettre l'élaboration d'une défense ciblée très spécifique et efficace après quelques jours d'infection. Cette dernière va favoriser la reconnaissance de signaux très spécifiques de l'agresseurs tels qu'une protéine de surface caractéristique de l'assaillant. Depuis quelques années, la dimension de défense immunitaire procaryote a pris de l'ampleur. On comprend de plus en plus qu'eux aussi, dans leurs environnements naturels, sont confrontés à des virus (bactériophages) et une large diversité d'éléments mobiles susceptibles de « parasiter » les cellules ou de s'intégrer dans leurs génomes. Des mécanismes de défense bactérienne, plus ou moins répandus, ont été découverts récemment par analyses comparées des génomes, dont le nombre a considérablement augmenté depuis l'apparition des nouvelles technologies de séquençage. Aujourd'hui encore, nous sommes malgré tout, loin de connaître et comprendre l'ensemble de ces mécanismes de défense. Tout comme les défenses immunitaires humaines, il apparaît maintenant clairement que si des mécanismes de réponse innée tels que les systèmes de Restrictions-Modifications (RM) sont utilisés de manière rapide pour la défense des cellules procaryotiques, les systèmes CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), s'apparentent à des systèmes de défense adaptatifs spécifiques et très efficaces. A l'image des enzymes de RM, l'intérêt biotechnologique des systèmes CRISPR a rapidement été décrit (Barrangou and Horvath, 2012). Ceci a lancé la recherche des systèmes de défense bactériens pour la

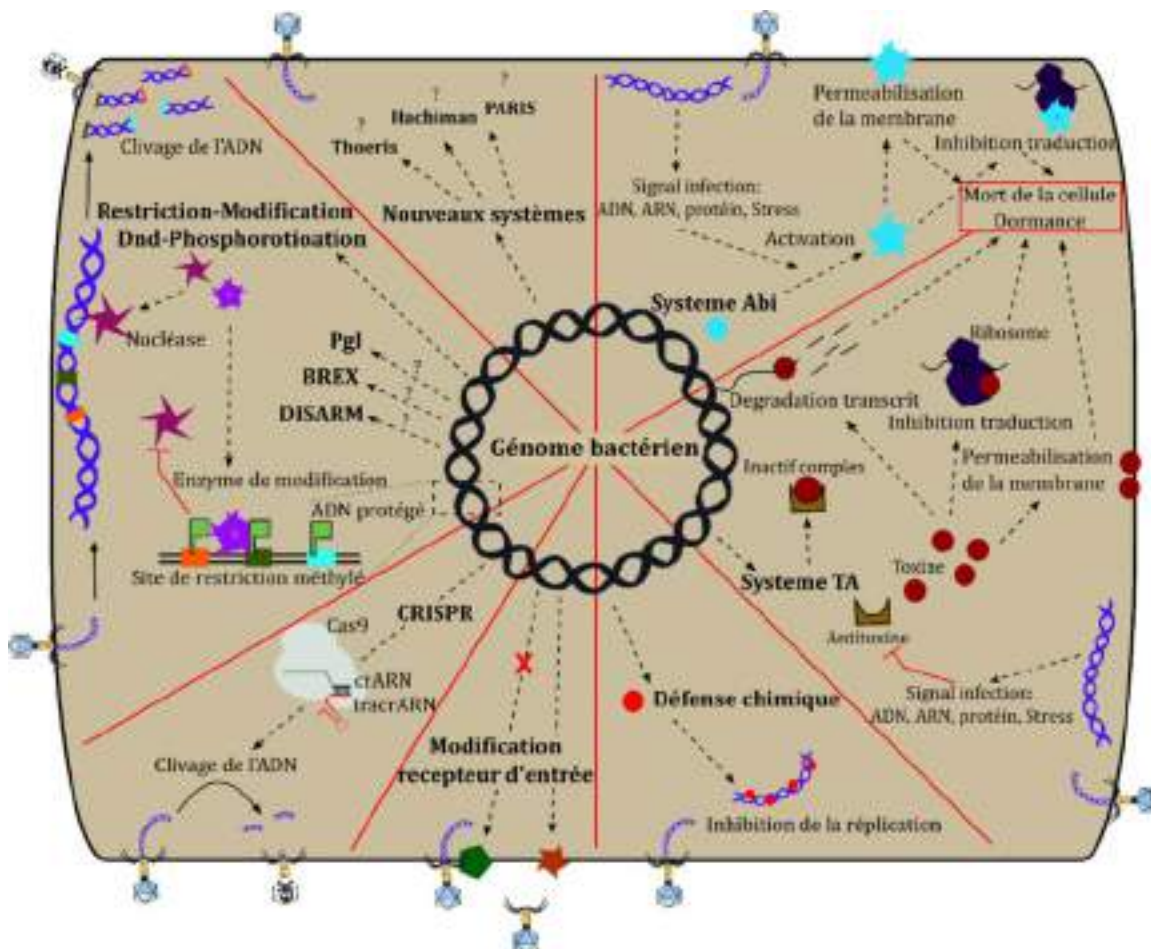


Figure 7 : Représentation schématique des différents mécanismes procaryotes de défenses contre l'ADN exogène. La cellule est séparée en différentes parties qui représentent la défense innée (Restriction-Modification, Dnd-Phosphorotioation, BREX, DISARM, Pgl), adaptative (CRISPR) ou de dernier recours par la dormance ou le suicide (Système Toxine-Antitoxine [TA] et Abi [Abortive]). Ainsi que les systèmes classés dans autres types de défenses (Modification du récepteur d'entrée, Défense chimique et les nouveaux systèmes).

découverte des « nouveaux CRISPR ». C'est devenu l'axe principal de certains groupes de recherche qui tend à caractériser les systèmes de défense bactériens.

2.1. Systèmes immunitaires procaryotes

Certains organismes, comme *Chlamydia trachomatis*, ont adopté un mode de vie parasitaire intracellulaire (Murray and McKay, 2021), s'éloignant ainsi des dangers extérieurs et ne nécessitent donc pas de nombreux mécanismes de défense contre les AN exogènes. Dans ces organismes, les nécessités ne sont pas les mêmes que dans ceux exposés à l'environnement et peu/aucun système de défense ne sont retrouvés. Pour l'autre majorité des bactéries, trois lignes de défense ont été identifiées dans les modèles actuels de défense bactérienne (Koonin *et al.*, 2017) :

- L'immunité innée, via des défenses non spécifiques (Restrictions-Modifications, Phosphorotioation, « Bactériophage Exclusion », Argonautes)
- La dormance ou la mort programmée de la cellule (Abi, Toxine-Antitoxine)
- L'immunité adaptative avec les systèmes CRISPR

Ce modèle composé de trois types de défense pourrait bien évoluer dans les prochaines années avec les nombreux systèmes putatifs de défenses bactériens en cours de caractérisation et les nombreux autres encore non identifiés (Figure 7).

Pour comprendre la dynamique des systèmes de défense procaryotes contre leurs envahisseurs, nous allons approfondir ces systèmes, des plus anciens comme les systèmes de Restrictions-Modification, aux plus récents tel que le système PARIS. Ceci va nous permettre de comprendre le modèle dynamique de l'immunité bactérienne qui peut agir de concert pour l'élimination efficace d'une menace extérieure (Figure 7) (Dupuis *et al.*, 2013; Koonin, 2019; Bernheim and Sorek, 2020).

2.1.1. Immunité innée bactérienne

L'immunité innée, comme chez les organismes eucaryotes complexes tels que les humains, va concerner des systèmes de protection n'étant pas spécifiques d'un envahisseur particulier, mais qui vont reconnaître une menace ou un changement environnemental. Les plus anciens connus chez les procaryotes sont les systèmes de Restrictions-Modification (RM) considérés très tôt comme des systèmes de défense de l'immunité innée (Arber and N, 1969) et qui sont devenus des outils standards en biologie

moléculaires depuis la caractérisation des systèmes de type II . Cette découverte a été réalisée par Hamilton Smith, Werner Arber et Daniel Nathans sur la bactérie *Haemophilus influenzae*, ce qui leur a valu le Prix Nobel en 1978.

Les RM sont des systèmes de défense très répandus, on les retrouve dans 75 %-90 % des génomes de procaryotes (Oliveira, Touchon and Rocha, 2014; Isaev, Musharova and Severinov, 2021). Ces systèmes sont composés de deux ou trois protéines et utilisent la méthylation de l'ADN, pour discriminer l'ADN soi, de l'ADN non-soi et cliver ce dernier par une endonucléase (Makarova, Wolf and Koonin, 2013). La protéine méthyltransférase va ajouter un groupement méthyle à une base spécifique de l'ADN immédiatement après réplication et la protéine restrictase n'aura alors pas accès au motif nucléotidique palindromique qui stimule son activité endonucléase. Par contre, lors d'une invasion par un ADN exogène, les sites de restriction non méthylés, seront reconnus et clivés par le système et l'ADN exogène sera ainsi éliminé par la bactérie. Il existe différentes types de RM et la nomenclature actuelle repose sur un consensus défini en 2003 (Roberts *et al.*, 2003).

- Les systèmes de Type I vont reconnaître l'ADN et le cliver à des positions variables par rapport au site de restriction (Roberts *et al.*, 2003). Ce sont des complexes protéiques de 5 sous-unités qui dépendent de l'expression de 3 gènes codant la sous unité spécificité, deux sous unités restrictases et deux sous unités méthylases (Oliveira, Touchon and Rocha, 2014; Isaev, Musharova and Severinov, 2021).
- Les systèmes de Type II sont les plus précis, car leur site de restriction et l'endroit de la coupure sont proches (Roberts *et al.*, 2003). Le premier découvert est celui d'*Haemophilus influenzae* (HindII) par Hamilton Smith en 1978 (Pingoud, Wilson and Wende, 2014). Depuis, grâce aux avancées liées au séquençage de génomes complets, la base de données « Rebase » répertoriant l'ensemble des systèmes RM connus (Roberts *et al.*, 2010) est passée de 4.000 enzymes connus de type II en 2011 (Stern and Sorek, 2011), à plus de 63 000 (même si tous ne sont pas caractérisés). Le système est encodé par deux gènes qui codent une restrictase homodimérique/tétramerique et une méthylase monomérique qui sont capables de travailler de manière indépendante (Oliveira, Touchon and Rocha, 2014; Isaev, Musharova and Severinov, 2021). Plusieurs sous-types existent pour discriminer

ÉTUDE BIBLIOGRAPHIQUE : Systèmes de défense bactériens : fonctions et applications

d'autres aspects particuliers, par exemple le type IIS qui coupe à l'extérieur du motif et non dans le motif comme ceux du type II (Roberts *et al.*, 2003).

- Les systèmes de Type III sont des systèmes intermédiaires qui vont couper à environ 30 pb du site de restriction (Isaev, Musharova and Severinov, 2021). Comme les systèmes de type I, ils peuvent avoir des rôles divers dans la régulation épigénétique de la bactérie (« Phase variable restriction-modification systems ») (Seib *et al.*, 2020).
- Les systèmes de Type IV sont particuliers, ils ont un rôle inversé. Ils ne nécessitent pas de méthylase et vont cliver l'ADN méthylé (Roberts *et al.*, 2003; Oliveira, Touchon and Rocha, 2014).

Les systèmes les plus retrouvés sont ceux de type II (42 %) suivis du type I (20 %) et du type III (8 %), et certaines espèces comme *Helicobacter pylori* en ont en moyenne 12 par génomes (Oliveira, Touchon and Rocha, 2014).

Des mécanismes d'adaptation ont été sélectionnés par les phages pour contrer ces systèmes de défense, notamment les anti-RM tels que des protéines mimant l'ADN pour lier le RM et le bloquer, mais également d'autres mécanismes non résolus avec des actions indirectes comme deux systèmes KclA et ArdB qui inhibent les systèmes de type I (Serfiotis-mitsa *et al.*, 2010). On peut aussi trouver des phages codants une méthylase afin d'arriver déjà protégés dans la cellule suivante (Pinilla-redondo *et al.*, 2020). Ces gènes codants des anti-systèmes de défense ont récemment été retrouvés dans des ilots « anti-défense bactérienne » dans les génomes de phages ; de nombreux autres sont certainement encore à caractériser (Pinilla-redondo *et al.*, 2020).

Pour survivre, les bactéries doivent répondre à ces inhibiteurs, cela peut passer par des variants de ces systèmes, elles ont développé par exemple le système de type IV des RM, mais également d'autres mécanismes de défense, comme d'autres systèmes basés sur la modification de l'ADN tels que les BREX, Pgl, DISARM et Dnd que l'on retrouve également chez les bactéries et les archées.

Les systèmes BREX pour « bactériophage exclusion », sont présents chez pas moins de 10 % des espèces bactériennes (Doron *et al.*, 2018). Ils ont été découverts chez *Bacillus cereus* pour la première fois où 6 gènes se sont avérés conférer une résistance à des bactériophages (Goldfarb *et al.*, 2014). Ce système n'est pas résolu, mais serait basé sur la

méthylation d'un site non palindromique comme système de reconnaissance du soi. L'hypothèse principale serait qu'un complexe soit formé autour du site de restriction, ce qui bloquerait la réplication, l'initiation à la transcription ou l'initiation à la traduction (Goldfarb *et al.*, 2014; Barrangou and Oost, 2015). Différents types de BREX sont connus et ils seraient tous porteurs d'une méthyltransferase et de la protéine pglZ associées à d'autres protéines moins conservés (Isaev, Musharova and Severinov, 2021).

Les systèmes Pgl (phage growth limitation) sont des systèmes multiprotéiques ressemblant au BREX avec qui il partage deux gènes (PglX et PglW). Une méthyltransferase est présente ainsi que différentes protéines présentant des domaines Duf262, HNH, HEPN, souvent associés à des systèmes de défenses (Makarova, Wolf and Koonin, 2013). Cette fois-ci, c'est le phage qui serait ciblé par la méthyltransferase, ce qui entrainerait la production de phages modifiés qui eux seraient reconnus par la cellule suivante puis détruit par un mécanisme là aussi non résolu (Koonin *et al.*, 2017; Isaev, Musharova and Severinov, 2021).

Le système DISARM, est un cluster de 5 gènes souvent associés aux systèmes RM. La présence d'une méthyltransferase et de son site associé CCWGG montre bien une discrimination du soi et non-soi par ce mécanisme, mais ici encore le mécanisme n'est pas résolu (Ofir *et al.*, 2018).

Les systèmes Dnd sont des systèmes basés sur la modification de type Phosphorothioation (PT), qui sont des modifications de sulfure dans l'ADN (Makarova, Wolf and Koonin, 2013). Le système de modification est composé de 5 protéines (DndABCDE) et celui de restriction de 3 protéines (DndFGH) (L. Wang *et al.*, 2018; Isaev, Musharova and Severinov, 2021). Tout le mécanisme n'est pas entièrement compris, mais on sait que DndA (Cystéine Desulfurase) est un donneur de l'atome sulfure qui va être transféré à DndC (PAPS reductase) pour l'ajout du groupement sulfure à l'ADN double brin par DndD (ATPase nicking endonucléase) possédant une activité de nickase de l'ADN. La spécificité des sites PT n'est pour le moment pas résolue, il se pourrait que cela soit dépendant d'une géométrie et non d'une séquence (Isaev, Musharova and Severinov, 2021). De la même manière que pour les systèmes RM, la différenciation du soi et du non-soi, permet à la bactérie de reconnaître et cliver l'ADN exogène par DndFGH. Récemment, le système dérivé SspABCD-SspE, a été découvert dans plusieurs espèces bactériennes où cette fois-ci, c'est l'ADN simple brin du phage qui est reconnu par

SSpABCD, puis modifié avec un PT notamment grâce à SspB et c'est SspE qui détecte et coupe l'ADN du phage évitant ainsi sa réplication (Xiong *et al.*, 2020). C'est encore un exemple d'une nouvelle diversification des systèmes de défense de la bactérie.

Outre les systèmes basés sur la modification de l'ADN, il existe également d'autres systèmes tels que les protéines Argonautes. Celles-ci, sont à la base de l'interférence à l'ARN chez les eucaryotes, mais elles sont également retrouvées chez les procaryotes (Koonin *et al.*, 2017) et notamment dans 10 % des génomes bactériens (Doron *et al.*, 2018). Ce sont de grosses protéines (de 800 à 1200 acides aminés) avec des activités nucléases qui sont guidées par des ADN (ou des ARN plus faiblement) (Makarova, Wolf and Koonin, 2013; Koonin *et al.*, 2017), et qui inhibent la propagation de l'ADN exogène (Swarts *et al.*, 2014). Leurs mécanismes ne sont pas résolus *in vivo* mais seulement démontrés *in vitro*. Une hypothèse est mise en avant, il s'agit de la destruction d'un ADN exogène qui permettrait la production de petits morceaux d'ADN exogènes qui seraient ensuite utilisés par la nucléase comme des guides pour dégrader d'autres copies de cet ADN exogène (Koonin *et al.*, 2017).

Finalement la plupart des systèmes innés sont tous basés sur des mécanismes agissant sur l'ADN. Mais le plus simple des mécanismes dont je n'ai pas encore mentionné le rôle dans l'immunité innée est la variation des protéines de surfaces. En effet, pour son entrée dans la cellule, le bactériophage nécessite un récepteur d'entrée pour sa liaison et l'injection son ADN dans la cellule. Or, la simple modification de ce récepteur par arrêt d'expression ou modification est une stratégie utilisée par la bactérie afin de ne plus être sensible à ce phage (Chevallereau *et al.*, 2021).

2.1.2. Dormance et mort programmée

Une stratégie d'adaptation originale est déployée par la bactérie, elle concerne l'activation de la dormance ou de la mort programmée pour éviter la dissémination du virus au sein de la population bactérienne. C'est une preuve de la coopération de la communauté bactérienne avec des individus seuls qui se sacrifient pour le bien de la population. Deux types de systèmes sont connus actuellement : les systèmes Abi (abortive infection system) et les systèmes Toxines-Antitoxines (TA) qui sont parfois considérés comme appartenant aux systèmes Abi. Les systèmes TA sont les plus diversifiés (Koonin *et al.*, 2017). Ils sont composés d'une toxine qui, à une certaine concentration, entraîne la mort de la cellule et d'une antitoxine maintenant la toxine à une activité faible ou nulle. Lors

d'une réponse à un stress tel que l'infection par un phage, l'antitoxine va être inactivée ou dégradée, la toxine va s'accumuler et va induire par différents mécanismes la dormance ou la mort de la cellule (Makarova, Wolf and Koonin, 2013). Les TA ont des rôles divers plus ou moins caractérisés, certains ont un rôle de défense contre des phages, mais ce n'est que la partie émergée de l'iceberg. Les TA sont classifiés en six types de systèmes (Page and Peti, 2016), différents selon la nature de l'antitoxine, un ARN pour le type I et III et une protéine pour le type II, mais également selon la nature de l'inhibition, l'ARN interfère avec l'ARNm de la toxine pour le type I tandis que pour le type III, c'est une interaction directe entre l'ARN et la protéine qui va inhiber la toxine. Pour les types II, IV, V et VI, les deux acteurs sont des protéines, l'antitoxine va inhiber la toxine par contact direct pour le type II, par antagonisme pour le type IV, par dégradation de l'ARNm de la toxine pour le V et par la promotion de la dégradation protéolytique pour le VI. Récemment, un groupe VII a fait son apparition avec l'antitoxine qui est une enzyme capable de régulation post-traductionnelle sur la toxine l'inactivant ainsi (Wang *et al.*, 2020). Certains de ces systèmes ont un rôle dans la défense contre les phages, par exemple le système de type III ToxIN-ToxN qui va dégrader les transcrits lorsque l'antitoxine ToxN n'est plus transcrite lors d'une infection (Guegler and Laub, 2021). Certaines toxines vont également perméabiliser la membrane ou attaquer des procédés cellulaires essentiels tels que le système de traduction, la division, etc (Koonin *et al.*, 2017). Ce mode d'action est associé à celui des systèmes Abi (abortive infection), qui sont des systèmes de suicide bactérien en réponse à un signal (souvent provenant du phage ou de la perturbation qu'il entraîne dans la cellule). Très peu sont caractérisés fonctionnellement, mais de nombreux systèmes existent (Makarova, Wolf and Koonin, 2013) et sont continuellement découverts. Certains peuvent être extrêmement complexes et être composé d'une voie de signalisation comme le système CBASS l'ancêtre d'une voie de signalisation antivirale (cGAS-STING) retrouvée chez les animaux (Cohen *et al.*, 2019).

2.1.3. Diversité des systèmes

La diversité de système de défense bactériens décrits dans la littérature est depuis une dizaine d'années en train d'exploser. Ceci est dû à l'intérêt croissant de nombreux groupes dans la découverte et la compréhension de ces mécanismes, mais également en grande partie grâce aux nouvelles technologies de séquençage haut débit et d'analyses de données. Plusieurs études à très grandes échelles ont permis d'identifier de nombreux

nouveaux systèmes putatifs et plusieurs ont même été caractérisés comme fonctionnels. Le plus souvent, ces systèmes sont très souvent clustérisés afin de faciliter les échanges (Doron *et al.*, 2018). Ainsi à partir de recherches informatiques sur des génomes procaryotes couplés à la recherche d'ilots de défenses, de nombreux systèmes putatifs ont été déterminés. Certains ont pu même montrer une fonctionnalité dans des tests d'infection de *E. coli* par une large gamme de phages. Neuf systèmes supplémentaires ont alors montré qu'ils étaient aptes à protéger contre au moins un phage en 2018 (Doron *et al.*, 2018), 29 systèmes en 2020 (Gao *et al.*, 2020), 14 systèmes de plus en 2021 (Rousset *et al.*, 2021) et 6 très récemment (Payne *et al.*, 2021). On trouve alors un très large éventail de systèmes, comme le système Thoeris (Doron *et al.*, 2018) qui semble dégrader le NAD (Nicotinamide Adénine Dinucléotide) de la cellule pour induire la mort de la cellule après activation (Ka *et al.*, 2020), mais dont le mécanisme d'activation reste encore inconnu. Le système PARIS s'active lors de la reconnaissance d'une protéine anti-restriction et entraîne l'arrêt de la croissance de la cellule (Rousset *et al.*, 2021). De la même manière, il a été montré qu'en réponse à ces systèmes anti-anti défense, le phage pouvait encore évoluer et s'adapter. Par exemple, le phage T4, par une de ses protéines virales, est capable d'inhiber le système de R-M bactérien. La bactérie active à son tour un système suicide via le clivage de l'ARNt, et finalement, une seconde protéine du phage T4 permet la réparation de ce clivage. Ce dernier correspond finalement à un anti-anti-anti système de défense (Rocha and Bikard, 2021). D'autres systèmes ont également été caractérisés dernièrement après des dizaines d'années de recherche, comme les systèmes de type Retron (Millman *et al.*, 2020). Ces systèmes codent une reverse transcriptase associée à un ARN non codant et couplé à une protéine effectrice. Ils sont associés à la mort de la cellule comme seconde ligne de défense. Ces systèmes sont activés lorsque RecB est inhibé lors de l'infection par la protéine « Gam » qui est caractéristique de certains phages.

Outre les systèmes dernièrement cités, des systèmes de défense chimiques ont également été mis en évidence ces dernières années, notamment dans *Streptomyces* qui produit des molécules de la famille des anthracyclines (daunorubicin et doxorubicin) qui sont des agents intercalant de l'ADN. L'hypothèse est que le système va préférentiellement s'attaquer au génome du phage qui est différent d'un point de vue topologique (linéaire, superenroulé) (Kronheim *et al.*, 2018). C'est aussi ce qui a été retrouvé chez d'autres procaryotes avec la viperine (retrouvé également chez les eucaryotes) qui modifie les

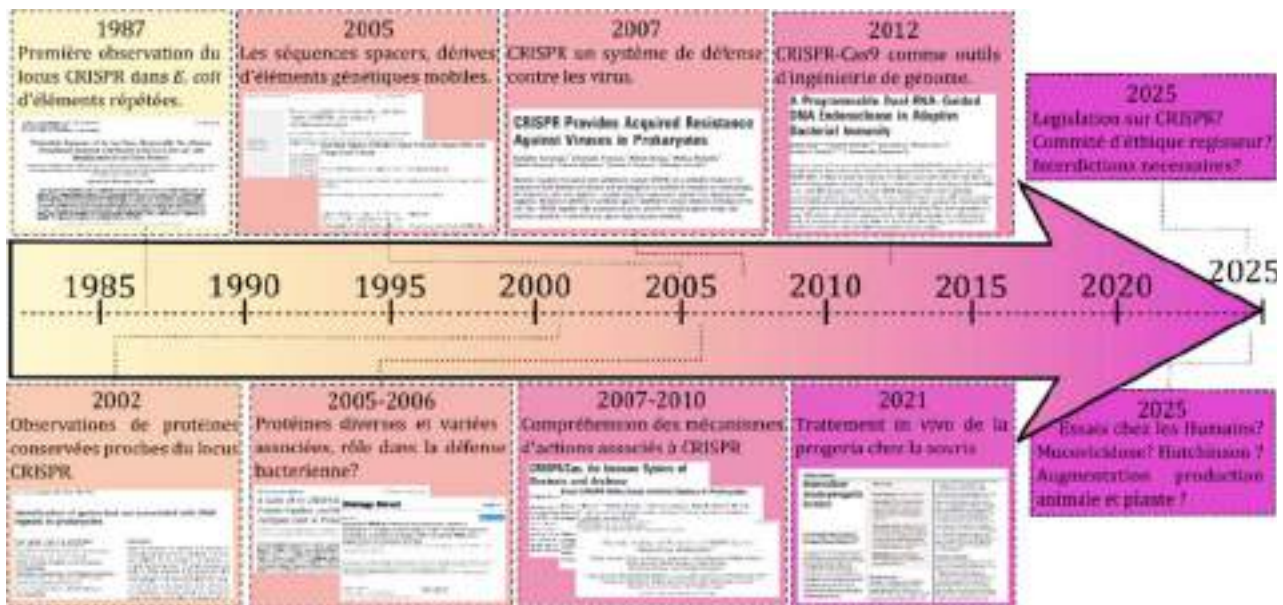


Figure 8 : Historique des découvertes sur les systèmes CRISPR. Les grandes dates et les découvertes majeures dans le domaine CRISPR correspondantes sont représentées dans chaque encadré et sont associées à l'image de l'article associé.

ribonucléotides ddhCTP, ddhGTP et ddhUTP, inhibant ainsi la polymérase virale (Bernheim *et al.*, 2021).

2.2. CRISPR : systèmes de défense adaptatifs

Les systèmes CRISPR sont des systèmes adaptatifs naturels de défense contre l'invasion par de l'ADN exogène retrouvé chez de nombreuses espèces de procaryote. La révolution technologique de 2012 via l'utilisation du système naturel CRISPR-Cas9 comme outil d'ingénierie de génome, a considérablement accéléré leur étude et ils représentent un des sujets les plus étudiés de la décennie (plus de 27 500 publications référencées dans PubMed). Comprendre leurs mécanismes fondamentaux, leur évolution, leur diversité, leurs rôles comme système de défense et leurs applications comme outil d'ingénierie génomique sont actuellement les principaux défis. Dans cette partie seront détaillés les systèmes CRISPR dans leurs environnements naturels.

2.2.1. Découverte du système

Historiquement, les systèmes CRISPR ont été décrits dans *E. coli* en 1987 avec l'observation de séquences répétées inter-espacées par des séquences uniques. Leurs fonctions n'étaient alors pas connues (Ishino *et al.*, 1987) (Figure 8). Il a ensuite fallu attendre les années 2000 pour voir apparaître de nouvelles données. Des études bio-informatiques visant à identifier ces séquences répétées sur les génomes ont alors permis de mettre en évidence ces séquences répétées chez de nombreuses bactéries et archées. Elles sont le plus souvent à proximité de protéines conservées chez de nombreuses espèces telles que Cas3 et Cas4 (Jansen *et al.*, 2002). Finalement, 45 protéines Cas ayant des domaines protéiques impliqués dans l'interaction à l'ADN, avec par exemple des domaines nucléases et hélicases, ont été identifiées. Il a très vite été mis en évidence une certaine hétérogénéité des systèmes CRISPR, et il est apparu dès le milieu des années 2000 que les systèmes CRISPR étaient de plusieurs types ou familles et qu'ils pourraient avoir un rôle dans la défense bactérienne (Haft *et al.*, 2005; Makarova *et al.*, 2006). Rapidement des études mettant en évidence ce rôle ont alors vu le jour. Les séquences uniques appelées « spacers », retrouvées entre les séquences répétées correspondent à des séquences complémentaires présentes dans des génomes de phages et autres éléments génétiques mobiles et plus rarement dans des génomes bactériens (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Makarova *et al.*, 2006; Horvath *et al.*, 2008). Chez *Streptococcus thermophilus* notamment, plusieurs loci possédant de 21 à 48 paires de

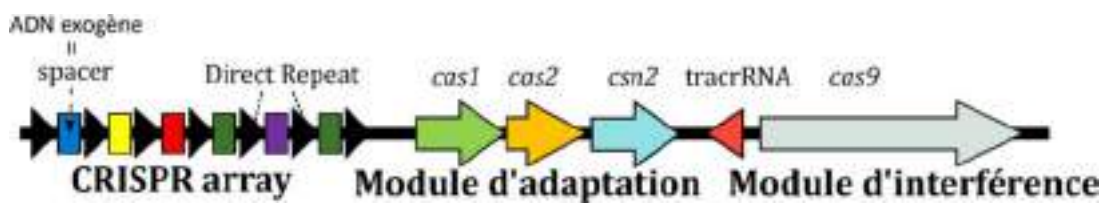


Figure 9 : Représentation du locus CRISPR de *S. pyogenes*. Le locus comporte différents modules, un module ARN le CRISPR array, et deux modules protéiques celui permettant l'acquisition de spacers (adaptation) et le module conférant une défense spécifique (l'interférence).

bases de séquences répétées sont espacées de séquences uniques de même taille (Deveau *et al.*, 2008). La présence ou non de ces séquences, associée à une protéine, Cas5 (appelée maintenant Cas9), détermine une résistance ou une sensibilité aux bactériophages chez *S. thermophilus* (Barrangou *et al.*, 2007). Ces deux acteurs jouent également un rôle en inhibant la conjugaison bactérienne et la transformation chez *S. epidermidis* (Marraffini and Sontheimer, 2008). Ces premières démonstrations de l'existence d'un mécanisme de défense bactérien ont alors intéressé un nombre croissant de chercheurs qui ont essayé de décrypter leurs mécanismes d'action menant à l'élimination d'un élément génétique mobile.

Le locus CRISPR est finalement composé de manière générale, de protéines Cas qui vont avoir un rôle dans les trois étapes de l'immunité bactérienne : l'adaptation, l'expression et l'interférence. L'autre partie de ce locus est le CRISPR array qui est composé de séquences répétées (DR = direct repeat) de 21 à 48 pb selon les systèmes, espacées par les séquences uniques (spacers) de 26 à 72 pb (Deveau, Garneau and Moineau, 2010) (Figure 9). Ce locus est apparu comme étant très dynamique et évoluant rapidement : de nouveaux spacers sont acquis dans le locus de manière polarisée après infection par des phages. Cette étape appelée l'acquisition, est médiée par des protéines n'ayant aucun rôle dans l'étape de clivage spécifique de l'ADN appelé l'interférence (Barrangou *et al.*, 2007). Le rôle essentiel de Cas1 et Cas2 dans l'intégration de nouveaux spacers a été mis en évidence dès la fin des années 2000 (Beloglazova *et al.*, 2008; Marraffini and Sontheimer, 2009). Ce système dynamique d'adaptation procaryote a pour conséquence l'application d'une pression de sélection sur les phages ciblés par le système CRISPR. Des variants (BIM= bacteriophage insensitives mutants), présentant des mutations au niveau des séquences complémentaires des spacers (dites protospacers) ont alors été identifiés comme résistants au système (Barrangou *et al.*, 2007; Deveau, Garneau and Moineau, 2010; Sun *et al.*, 2013). Toutes ces données expérimentales ont permis une compréhension de plus en plus précise du système CRISPR et notamment de l'implication de spacers comme supports de l'immunité acquise par les bactéries et leur conférant une résistance aux phages possédant une séquence identique dans leur génome. Mais deux éléments majeurs étaient encore indéterminés, tous deux liés aux mécanismes d'interférence. Il a été décrit tout d'abord dans *E. coli* que l'ARN du CRISPR array (crRNA) est à l'origine de la cascade d'interférence rendant la bactérie/archée immunisée contre le phage et qu'une maturation du crRNA est essentielle au phénomène d'interférence et à

sa spécificité (Brouns *et al.*, 2008). C'est la question de l'auto-immunité et de la discrimination du soi et du non-soi qui s'est alors posée, puisque la séquence correspondant au spacer est également présente dans le locus CRISPR de la bactérie. Pour des systèmes de défense classiques de types Restriction-Modification, la méthylation de l'ADN permet cette discrimination, mais comment le système CRISPR peut-il couper la séquence phagique tout en épargnant son propre locus CRISPR (Marraffini and Sontheimer, 2008) ? Des analyses bio-informatiques de l'ensemble des protospacers de différentes souches de *S. thermophilus* additionné d'analyses phénotypiques, ont conduit à l'identification d'un motif nucléotidique spécifique adjacent aux protospacers : le motif PAM (Brouns *et al.*, 2008; Deveau *et al.*, 2008; Horvath *et al.*, 2008). Ce motif semble être présent sur toutes les cibles phagiques, mais pas sur le locus CRISPR. Ainsi le motif PAM semble être la pièce du puzzle manquante pour comprendre l'ensemble du système. Ce qui a été confirmé par des expériences dans lesquelles un système CRISPR de *S. thermophilus* a été importé chez *E. coli*, l'immunité spécifique contre les phages est bien conservée et dépend toujours du motif PAM (Saprunauskas *et al.*, 2011). De plus, il a été démontré que cette séquence PAM est différente selon les types de CRISPR ou la provenance des protéines utilisées pour l'interférence (Mojica *et al.*, 2009; Sinkunas *et al.*, 2013).

C'est à l'aide de toute cette recherche pionnière que les systèmes CRISPR ont été décrits comme des systèmes de défense adaptatifs bactériens. Ces systèmes dépendent d'un ARN issu du locus CRISPR, comprenant une séquence répétée et d'un spacer associé, qui permet de cibler de manière très spécifique une séquence complémentaire présente sur un phage ou un élément mobile. Le tout apporte une immunisation spécifique contre l'élément invasif grâce à des protéines associées appelées « Cas ». Ce système de défense adaptatif original a intéressé très rapidement la communauté scientifique. Au début des années 2010, les études se sont multipliées pour décrire de nombreux types de systèmes et les détails des mécanismes sous-jacents.

2.2.2. Cibles principales

Les cibles des systèmes CRISPR sont les éléments mobiles tels que les phages, mais il est difficile d'identifier ces cibles puisque seulement 7 % des spacers connus en 2017 trouvent des cibles dans les différentes banques génomiques sur les 363 000 uniques spacers disponibles (Shmakov *et al.*, 2017). Dans la plupart des exemples précédents, j'ai

mis l'accent sur l'interaction CRISPR - phages, car ce sont les principales cibles (Shmakov *et al.*, 2017). Mais comme déjà cité brièvement, les systèmes CRISPR cibles toute sortes d'autres sources d'ADN invasifs communément appelés : les éléments génétiques mobiles (MGE). Les MGE peuvent être classés en deux familles, ceux qui sont transmissibles d'une cellule à une autre comme les phages, mais également les plasmides et les éléments intégratifs conjugatifs (ICE), et ceux qui sont intracellulaires comme les transposons, les intégrons, les introns et les séquences d'insertions (IS). Ensemble, ils forment le mobilome d'une bactérie (Siguier *et al.*, 2006).

Les phages également appelés bactériophages, sont des virus infectant les bactéries. Ils sont nombreux et retrouvés de manière ubiquitaire dans tout environnement. Ils sont reconnus comme les agents biologiques les plus abondant sur Terre. Ils sont composés d'un génome ADN ou ARN de 3 kb à 30 kb et sont encapsulés dans une capsid protectrice (Hatfull and Hendrix, 2011; Kasman and Porter, 2021). Une fois dans la bactérie, les phages adoptent deux stratégies de répllication : le cycle lysogénique ou lytique. Cette stratégie consiste à l'intégration du génome de phage dans le génome de la bactérie, à cette étape le prophage est alors en dormance et est transmis seulement d'une bactérie « mère » à une bactérie « fille », sans tuer la cellule hôte. Lors d'un changement environnemental, ce phage peut se réactiver pour passer en cycle lytique. Le cycle lytique permet le détournement de l'ensemble des ressources d'une bactérie par le phage afin de produire à partir d'un génome phagique, de nombreuses copies de celui-ci ainsi que les protéines phagiques. La bactérie va ensuite mourir et libérer de nouvelles particules phagiques dans l'environnement qui infecteront les bactéries environnantes (Kasman and Porter, 2021).

Les IS sont des éléments génétiques de 0,7 à 2,5 kb présents dans de nombreux génomes bactériens, qui codent une transposase ainsi qu'une séquence terminale répétée. La transposase va permettre le clivage de l'ADN et le transfert de l'IS dans une autre localisation, ce qui peut notamment entraîner l'élimination ou le déplacement de gènes environnants. Leurs rôles dans l'élimination ou l'insertion des gènes dans un génome, le transfert de gènes d'un génome à un plasmide ou d'un plasmide à un autre par exemple a été démontré (Siguier *et al.*, 2006).

Le transposon quant à lui à une structure plus complexe, il est composé lui aussi d'une transposase mais possède un agencement plus complexe avec de nombreux gènes qui

vont le suivre à chaque transposition. Les gènes présents dans les transposons vont le plus souvent coder des facteurs importants favorisant le maintien de cet élément mobile dans le génome bactérien. De manière similaire, les intégrons vont se différencier du transposon par la présence d'une intégrase site-spécifique et non d'une transposase (Piotrowska and Popowska, 2015). Tous ces éléments génétiques mobiles ne sont pas capables de se répliquer eux-mêmes et vont être intégrés au génome bactérien. Ils vont être à l'origine d'évolution importante avec acquisition et délétion de gènes et vont avoir un rôle important dans l'écologie bactérienne particulièrement en réponse à des changements environnementaux.

Le transfert de ces éléments d'une cellule à l'autre peut se faire lors d'événements de transformation de l'ADN, mais aussi de conjugaison pour certains transposons ayant acquis la machinerie de transposition. Certains de ces éléments, comme les IS, s'intègrent dans d'autres éléments mobiles pour profiter de leurs machineries de transport intercellulaire (Piotrowska and Popowska, 2015). Ces autres éléments mobiles, qui possèdent le plus souvent un moyen intrinsèque de passer d'une cellule à une autre, sont les phages, certains plasmides et les ICE (Siguier *et al.*, 2006).

Les ICE, comme précédemment décrit dans la première partie, sont des éléments mobiles plutôt grands (20 à 30 kb) qui codent les facteurs nécessaires à leurs excision, transfert et intégration dans un autre génome. Ils peuvent être retrouvés intégrés aux génomes bactériens ou excisés et circulaires (Wozniak and Waldor, 2010).

Les plasmides conjugatifs sont des éléments génétiques qui sont extrachromosomiques, que l'on retrouve dans la plupart des classes bactériennes. Ils sont capables de se répliquer de manière indépendante au chromosome.

Ces deux MGE se transmettent par conjugaison d'une cellule à une autre grâce au système de sécrétion de type IV qu'ils transportent eux même (Virolle *et al.*, 2020). Ils véhiculent également le plus souvent, des éléments avantageux pour la bactérie tels que des facteurs de virulence ou de résistance aux antibiotiques (Piotrowska and Popowska, 2015), ce qui peut expliquer leur persistance puisqu'ils sont conservés dans une bactérie si le bénéfice qu'ils apportent est supérieur au coût.

De manière étonnante, une étude récente a démontré la présence de nombreux systèmes CRISPR sur des plasmides, avec une moyenne de 3 % des plasmides qui possèderait un

système. Paradoxalement, ces plasmides participeraient à la dissémination des systèmes CRISPR et possèderaient bien plus de spacers ciblant d'autres plasmides (66 % des 1100 spacers identifiés cibleraient un plasmide contre 24 % sur les 11 000 spacers identifiés dans les systèmes CRISPR présents sur les chromosomes) (Pinilla-Redondo *et al.*, 2021). Tout comme les plasmides, les transposons, les IS, les introns et les virus sont les cibles principales des systèmes CRISPR, et codent malgré tout pour des morceaux de systèmes CRISPR voire des systèmes entiers (Faure *et al.*, 2019). Il en est de même pour d'autres systèmes de défense comme les RM que l'on retrouve aussi fréquemment portés par des MGE (Oliveira, Touchon and Rocha, 2014). Toutes ces données ne font que confirmer que l'interaction bactéries-MGE est une interaction très complexe, régie par des conditions écologiques spéciales et que le changement de ces conditions entraîne des adaptations importantes du côté bactérien, mais également du côté du MGE. Cette relation étroite entre bactérie et élément mobile permet à l'un et à l'autre d'évoluer vers des entités de plus en plus robustes.

2.2.3. Classification évolutive

Comme cela a pu être remarqué dès les années 2000, les systèmes CRISPR sont divers (Haft *et al.*, 2005). La découverte de nouveaux systèmes CRISPR par l'exploration des génomes et leur classification est en continuelle évolution et de nouveaux systèmes CRISPR sont régulièrement ajoutés. Lors de la dernière analyse à grande échelle (13 116 génomes analysés), 85 % des espèces d'archées (dont 97 % des hyperthermophiles) et 42 % des bactéries possèderaient un type de système CRISPR plus ou moins complet (Makarova *et al.*, 2020). Au départ, les gènes *cas1* et *cas2* étant présents de manière quasi systématique dans les systèmes CRISPR, un arbre des *cas1* était couramment employé pour discriminer les systèmes (Horvath *et al.*, 2009). Mais la grande diversité ainsi que l'implication de transferts horizontaux dans l'évolution de ces systèmes CRISPR ont conduit à proposer un modèle de classification plus sophistiqué, prenant en compte de nombreux paramètres. La comparaison des similarités de séquence, l'analyse des voisins les plus proches, l'analyse phylogénétique, les éléments composant le système et leurs combinaisons de gènes sont dorénavant pris en compte (Makarova *et al.*, 2020). Seuls trois types de systèmes CRISPR étaient observés avant 2011, puis 5 types et 16 sous-types en 2015 et ce sont 2 classes, 6 types et 33 sous-types qui sont répertoriés à l'heure actuelle (Makarova *et al.*, 2011, 2015, 2020). Le descriptif des différents types de systèmes

Classe	Types	Gène	Espèce hôtes	Répétition CRISPR	Adaptation				Expression		Interférence			
					Cas1	Cas2	Cas3	PE	Cas6	Cas5	Cas7	Cas8	Cas9	Cas10
1	1A	ADN	<i>Archaeoglobus fulgidus</i>	Oui	1	1	2		1	1	1	1	2	
	1B	ADN	<i>Clostridium kluyveri</i>	Oui	1	1	1		1	1	1	1	1	
	1C	ADN	<i>Bacillus halodurans</i>	Oui	1	1	1			1	1	1	1	
	1D	ADN	<i>Cyanothece</i> sp.	Oui	1	1	1		1	1	1	1	1	1
	1E	ADN	<i>Escherichia coli</i> K12	Oui	1	1			1	1	1	1	1	
	1F1	ADN	<i>Fraxinus pseudo-nubigena</i>	Oui	1	Fusion Cas1			1	1	1	1	1	
	1F2	ADN	<i>Shewanella putrefaciens</i>	Oui	1	Fusion Cas1			1	1	1	1	1	
	1F3	NE	<i>Vibrio cholerae</i>	Oui					1	1	1	1	1	
	1G	ADN	<i>Geobacter sulfurreducens</i>	Oui	1	1	Fusion Cas1		Fusion Cas5	1	1	1	1	
	1H-A	ADN-ARN	<i>Staphylococcus epidermidis</i>	Oui	0-1	0-1			1	1	2			1
	1H-B	ADN-ARN	<i>Styrococcus lentus</i>	Non					1	1	1			1
	1H-C	ADN-ARN	<i>Microparvum rubeum</i> thermotogastrum	Non					1	1	1			1
	1H-D	ARN?	<i>Syngnathus</i> sp.	Non					1	1	1			1
	1H-E	ARN?	<i>Candida lusitana</i> Endler	Oui	0-1	0-1		Fusion Cas1			Fusion PE			
	1H-F	ADN?	<i>Thermotoga litoralis</i>	Non					1	1	1			1
	1H-A	NE	<i>Thiobacillus</i> sp.	Oui					0-1	1	1	1	1	
	1H-B	NE	<i>Thiobacillus</i> sp. PHT2	Non					1	1	1	1	1	
1H-C	ADN?	<i>Thermotoga doerrieri</i>	Non					1	1	1	1	1		

Table 2 : Tableau récapitulatif de la classification des systèmes CRISPR de classe 1. Un exemple pour chaque type de CRISPR est donné avec les protéines associées. Le nombre de protéines associées à ces systèmes est indiqué par un chiffre, un - indique différents nombres possibles selon les bactéries possédant ce système. Certaines protéines sont identifiées comme fusionnées et sont mentionnées dans le tableau.

CRISPR présenté ci-dessous est basé principalement sur plusieurs revues faisant le bilan de ceux-ci.

Les systèmes CRISPR de classe 1 regroupent les types I, III et IV (Table 2). Ils sont caractérisés par la présence d'effecteurs multiples ayant un rôle dans la maturation du crRNA et l'interférence à l'ADN médié par le système CRISPR. Ces protéines entraînent une réaction en chaîne appelé cascade ou le plus souvent, la protéine Cas6 est responsable de la maturation du crRNA, Cas5 et Cas7 ont un rôle central dans la reconnaissance du crRNA et l'interaction avec les effecteurs responsables du clivage Cas8-Cas3 ou Cas10. Des modules avec diverses combinaisons d'effecteurs, des protéines de fusion ou d'autres effecteurs pour la cascade sont ensuite les principales différences dans les types de CRISPR de classe 1 (Koonin, Makarova and Zhang, 2017; Makarova *et al.*, 2020).

Le type I est composé de 9 sous-types (A, B, C, D, E, F1, F2, F3 et G). La signature de ce système est la présence de Cas3 qui est une protéine parfois fusionnée à l'endonuclease ciblant des ADN exogènes (Makarova *et al.*, 2015). Les différentes protéines modulant l'acquisition sont présentes dans la majorité de ces systèmes.

Le type III est composé de 6 sous-types (A, B, C, D, E, F) et est quant à lui caractérisé par la présence de Cas10 qui est l'effecteur qui va permettre le clivage de l'ADN cible. Ces systèmes sont caractérisés par de nombreux gènes qui ont été dupliqués puis qui ont évolué (notamment Cas7) (Makarova *et al.*, 2020). L'adaptation et l'intégration des spacers sont souvent dépendantes des autres systèmes présents dans le génome puisque les protéines Cas1 et Cas2 sont peu retrouvées dans les CRISPR de type III. Ces systèmes sont capables, à l'aide de Cas10, de cliver l'ADN double brin, mais également l'ADN simple brin pendant la transcription, et l'ARN nouvellement transcrit (Koonin, Makarova and Zhang, 2017; Liu, Iavarone and Doudna, 2017; Gleditsch *et al.*, 2019). Par exemple, pour le type III-A de *S. thermophilus*, il a été montré que le système via la protéine Csm6 qui est capable de cliver l'ARNm après activation par Cas10 (Kazlauskienė *et al.*, 2017). Pour ces systèmes, la séquence PAM n'est pas nécessairement présente et la reconnaissance du non-soi s'effectue à l'inverse, par une non-complémentarité de base sur les nucléotides adjacents au protospacer (Koonin, 2019).

Le type IV est composé de trois sous-types (A, B, et C). Il est considéré comme un système minimal puisque Cas1 et Cas2 ne sont pas présents la plupart du temps dans le

Classe	Types	Clon	Exemple espèces	Répétition CRISPR	Adaptation				Maturation	Interférence	Maturation-Interférence		Autres	
					Cas1	Cas2	Cas3	Cas4	RNAseIII	Cas9	Cas13	Cas18		
II	V-A	ADN	<i>Streptococcus thermophilus</i>	Oui	1	1		1		1			tracARN	
	V-B	ADN	<i>Legionella pneumophila</i>	Oui	1	1	1		1	1			tracARN	
	F-C1	ADN	<i>Neisseria lactamica</i>	Oui	1	1			1	1			tracARN	
	F-C2	ADN	<i>Mycobacterium abscessus</i>	Oui	1	1	1		1	1			tracARN	
	V-A	ADN	<i>Francisella tularensis</i>	Oui	1	1	1				1			
	V-B1	ADN	<i>Acetivibrio acetivorans</i>	Oui	1	1		Fusion Cas1			1		trac ARN	
	V-B2	ADN	<i>Neisseria meningitidis</i>	Oui	1	1		Fusion Cas1			1		trac ARN	
	V-C	ADN	<i>Chlamydia sp.</i>	Oui	1						1			
	V-D	ADN	<i>Saccharum sp.</i>	Oui	0-1						1			
	V-F	ADN	<i>Bacteroides fragilis</i>	Oui	1	1	1				1		tracARN	
	V-F1	ADN	<i>Urethralyase</i>	Oui	1	1	1				1		tracARN	
	V-F2	ADN	<i>Neisseria meningitidis</i>	Oui							1			
	V-F3	ADN	<i>Neisseria meningitidis</i>	Oui							1			
	V-F4	ADN	<i>Urethralyase</i>	Oui	1	1	1				1			
	V-F5	ADN	<i>Neisseria meningitidis</i>	Oui	1	1	1				1			
	V-G	ARN	<i>Haemophilus influenzae</i>	Oui							1			tracARN
	V-H	ADN	<i>Hyperthermus aerophilus</i>	Oui							1			
	V-I	ADN	<i>Neisseria meningitidis</i>	Oui							1			
	V-K-1	ADN	<i>Cyanobacterium</i>	Oui							1			tracARN-cas3-cas4
	V-K-2	ADN	<i>Gordonia</i>	Oui							1			
V-L1	ADN	<i>Cyanobacterium</i>	Oui							1				
V-L2	ADN	<i>Cyanobacterium</i>	Oui							1				
V-L3	ADN	<i>Rhodospirillum rubrum</i>	Oui							1				
V-L4	ADN	<i>Leptothrix</i>	Oui	0-1	0-1							1		
V-M1	ADN	<i>Prevotella</i>	Oui									1	Cas2	
V-M2	ADN	<i>Streptococcus</i>	Oui									1	Cas2	
V-N	ADN	<i>Streptococcus</i>	Oui									1		
V-O	ADN	<i>Streptococcus</i>	Oui	0-1	0-1							1	WFL	

Table 3 : Tableau récapitulatif de la classification des systèmes CRISPR de classe 2. Un exemple pour chaque type de CRISPR est donné avec les protéines associées. Le nombre des protéines associées à ces systèmes est indiqué par un chiffre, un - indique différents nombres possibles selon les bactéries possédant ce système. Certaines protéines sont identifiées comme fusionnées et sont mentionnées dans le tableau.

locus CRISPR, mais également parce que les effecteurs sont peu nombreux (Csf1/Cas5/Cas7) (Makarova *et al.*, 2015).

Les systèmes CRISPR de classe 2 regroupent les types II, V et VI (Table 3). Ils sont caractérisés par un effecteur unique remplaçant la cascade protéique, avec de nombreux domaines pour réaliser l'ensemble de l'activité d'interférence. Les différences au sein de cette classe se situent principalement sur la maturation du crRNA et la présence de certaines protéines (Koonin, Makarova and Zhang, 2017; Koonin, 2019; Makarova *et al.*, 2020). Leur relative simplicité explique que ce sont les systèmes qui ont été les plus étudiés et les plus utilisés pour en dériver des outils d'ingénierie génomique.

Les systèmes de type II sont répartis en 4 sous-types (A, B, C-1 et C-2). La principale signature est la protéine Cas9 qui va cliver l'ADN exogène. Mais le type II est également caractérisé par la présence d'un second ARN, le tracr ARN qui est partiellement complémentaire au crRNA et qui s'hybride pour former un guide RNA conduisant Cas9 à sa cible (Jinek *et al.*, 2012). Ce système est retrouvé seulement chez les bactéries puisque la maturation du crRNA qui est nécessaire est réalisée par la RNase III endogène de la bactérie qui n'est pas présente chez les archées (Makarova *et al.*, 2015). Les différents sous-types sont ensuite différenciés par l'organisation du locus CRISPR et la présence de protéines du module d'adaptation : Cas1 et Cas2 seules (II-C), ou associées à Csn2 (II-A) ou Cas4 (II-B) (Chylinski *et al.*, 2014; Makarova *et al.*, 2020).

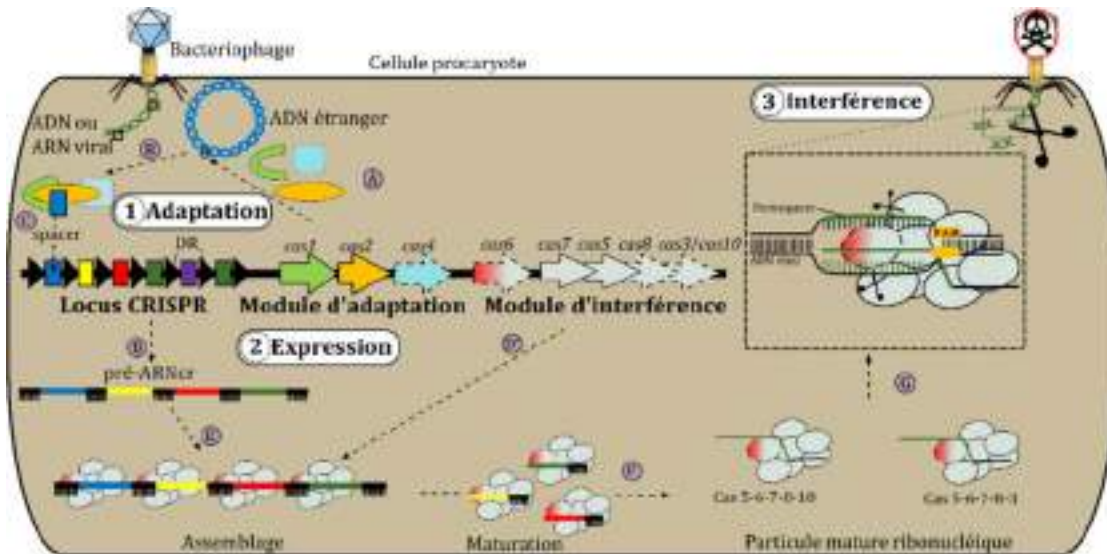
Le type V est composé de 17 sous-types (A, B1, B2, C, D, E, F1, F1-U3, F2, F3, G, H, I, K, U1, U2 et U4). C'est le plus large groupe car il se trouve dans un transposon de type TnpB ; de nombreux variants sont retrouvés dans la nature (Makarova *et al.*, 2020). La signature de ce groupe est la protéine effectrice Cas12. A la différence de Cas9 qui possède un domaine de clivage HNH et RuvC, Cas12 ne possède qu'un seul domaine catalytique RuvC pour le clivage de l'ADN (Makarova *et al.*, 2015). À l'inverse de Cas9, Cas12 ne nécessite pas de tracrRNA et mature le crRNA de manière autonome (Koonin, Makarova and Zhang, 2017). Les différences au sein du type V proviennent essentiellement des variants de Cas12 (Cpf1, Cas12b, C2C1, etc...) et des protéines du module d'adaptation Cas1, Cas2 et Cas4 qui sont présentes ou non (Makarova *et al.*, 2020).

ÉTUDE BIBLIOGRAPHIQUE : Systèmes de défense bactériens : fonctions et applications

Le type VI est le dernier à avoir été ajouté à la classification et il est composé de 5 sous-types (A, B1, B2, C et D). Cette fois-ci, la signature est la protéine Cas13 qui cible uniquement les ARN transcrits à partir des génomes invasifs, à l'aide de deux domaines HEPN. Les systèmes sont différenciés sur la base des variations de la protéine Cas13 ainsi que des protéines accessoires pouvant avoir des rôles dans la régulation de l'effecteur Cas13 (Koonin, Makarova and Zhang, 2017; Makarova *et al.*, 2020). Ce type de système a développé le moyen de discriminer l'ARN endogène de l'ARN étranger par une séquence dite PFS (Protospacers flanking site). Pour la protéine Cas13a chez *Leptotrichia shahii*, il a été montré que la protéine différencie les ARN endogènes par rapport aux protospacers ciblés, via la présence sur la séquence endogène d'une base Cytosine au niveau de la partie répétée du crRNA (Gleditsch *et al.*, 2019).

Définir l'origine de l'ensemble de ces systèmes n'est pas chose facile. Définir des liens de parenté entre certains d'entre eux est plus aisé, avec notamment le système II-A qui dérive du II-C par acquisition de Csn2 et qui donne un gain d'adaptation important (Chylinski *et al.*, 2014) ou le type IV qui provient probablement du type I ou III avec la perte de nombreux effecteurs (Makarova *et al.*, 2020). Mais dernièrement la découverte de relations évolutives entre certaines protéines Cas et des éléments mobiles, a permis de proposer un modèle d'évolution des systèmes CRISPR. Selon ce modèle, la protéine Cas1 proviendrait d'un transposon nommé Casposon, et la protéine Cas4 d'un bactériophage ou d'un plasmide. De plus, les protéines Cas2 et Csm6 auraient pour origine des systèmes de type Toxine-Antitoxine (Koonin, 2019). Le système de type III est celui proposé comme candidat pour être le système ancestral des CRISPR de classe I. Il est très répandu que ce soit chez les bactéries ou chez les archées, mais aussi, car il est le plus complexe et que de nombreux variants de l'effecteur Cas10 existent. Des validations expérimentales restent toutefois nécessaires. Les systèmes de classe I contemporains proviendraient donc d'un type III ancestral et de nouveaux effecteurs seraient alors apparus grâce à des duplications ou acquisitions de gènes. Pour les systèmes de classe 2, les type II et type V auraient évolué lors d'évènements multiples indépendants à partir d'une nucléase encodée par un transposon qui se serait inséré à proximité d'un locus CRISPR (Makarova *et al.*, 2020). Plus récemment, une nouvelle étude à grande échelle a permis d'identifier une nucléase, IscB, provenant des transposons IS200/IS605, comme ancêtre potentiel de

Classe 1



Classe 2

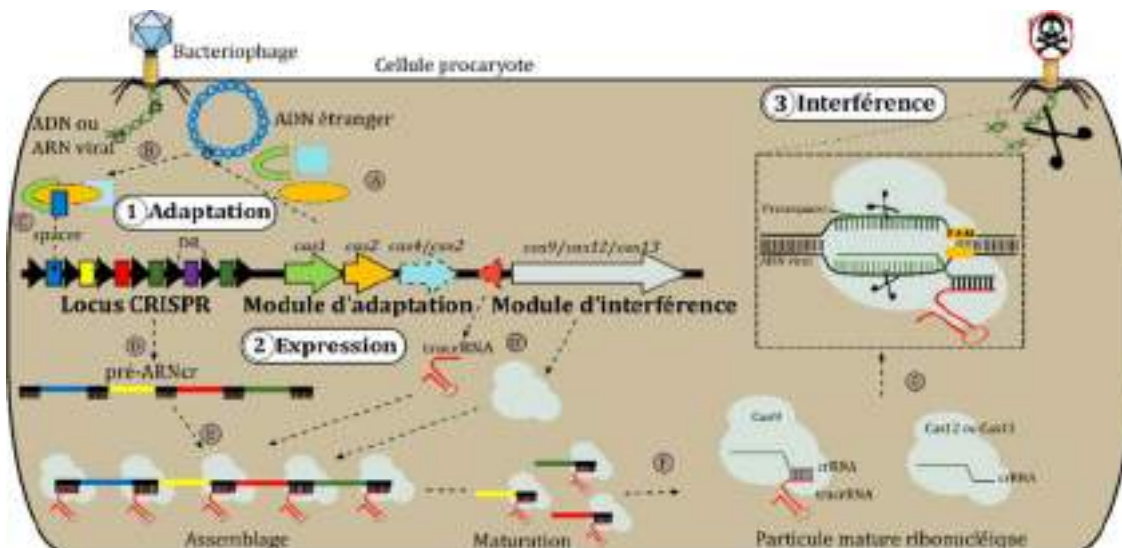


Figure 10 : Représentation schématique de la réponse CRISPR pour les systèmes de classe 1 et 2. La réponse est composée de trois étapes, l'adaptation, l'expression et l'interférence. Le locus CRISPR est représenté en noir et est composé du locus CRISPR, du module d'adaptation et du module d'interférence. Les éléments génomiques présents dans la grande majorité des systèmes CRISPR sont représentés par des flèches au contour continu. Les noms des éléments peuvent être séparés d'un "/" et cela signifie que selon le système CRISPR, le gène Cas4 ou Csn2 peut être présent. Les éléments présents dans certains systèmes seulement sont représentés par des flèches au contour pointillé. DR= direct repeat, représente les séquences répétées présentes dans le locus CRISPR. Le schéma représente 3 étapes 1, 2 et 3 subdivisées en sous étapes A, B, C, D/D', E, F et G. **L'étape 1** représente l'étape d'adaptation avec A, l'expression des gènes nécessaires à cette étape, B la réponse à l'infection d'une cellule procaryote par un élément viral (ADN ou ARN) ou un élément exogène tel qu'un plasmide, C, l'acquisition d'un petit morceau de l'ADN/ARN envahisseur et est inséré dans le locus CRISPR. **L'étape 2** représente l'étape d'expression du locus CRISPR (D) et des gènes codants les effecteurs protéiques d'interférence (D'). E représente l'assemblage de l'ARN CRISPR (crRNA) avec les protéines d'interférence ainsi que le tracrRNA pour les systèmes de types 2. F représente le résultat de la maturation permettant l'obtention des particules ribonucléiques matures composées des protéines interférentes, du crRNA mature avec une séquence nucléotidique complémentaire d'un élément exogène et du tracrRNA pour les systèmes de type II. **L'étape 3** représente la défense adaptative du système : l'interférence. G représente le clivage spécifique de l'ADN exogène lors d'une nouvelle apparition de l'élément invasif exogène.

Cas9. Ceci a mis en valeur l'existence d'une large classe de nucléases fonctionnelles guidées par un ARN (Altae-Tran *et al.*, 2021). Le type VI quant à lui, aurait pour origine un système de type ABI (abortive infection system) possédant un domaine HEPN typique pour l'interférence.

De nombreux éléments mobiles auraient permis une dissémination importante des systèmes CRISPR et la production de nombreux systèmes dérivés (Makarova *et al.*, 2020). Encore plus récemment, une autre étude a mis en évidence la présence de systèmes CRISPR plus ou moins complets et de tous les types (excepté type VI) sur de nombreux plasmides de bactéries ou d'archées. Au total, 3 % des plasmides possèderaient des systèmes CRISPR, ce qui a dû largement favoriser la propagation de ces systèmes et rend ainsi la découverte des origines très difficile (Pinilla-Redondo *et al.*, 2021).

2.2.4. Etapes du système CRISPR

Selon le type, les systèmes CRISPR ne fonctionnent pas exactement de la même manière, mais trois phases communes sont décrites dans la littérature pour expliquer le mécanisme de cette immunité acquise: (1) la phase d'immunisation communément appelée adaptation, (2) la mise en place de l'immunité ou phase d'expression et enfin (3) le clivage spécifique de l'ADN aussi appelée phase d'interférence (Figure 10) (Deveau, Garneau and Moineau, 2010; Horvath and Barrangou, 2010; Barrangou and Horvath, 2012; Doudna and Charpentier, 2014).

2.2.4.1. Etape initiale d'adaptation

Au moment d'une réponse primaire contre un phage (immunité innée), mais également après une réponse secondaire liée à l'interférence par le système CRISPR, une étape d'adaptation est à l'œuvre. Les principaux acteurs de celle-ci sont les protéines Cas1 et Cas2 qui sont essentielles à cette phase (Figure 11). Cas1 et Cas2 vont permettre à partir d'ADN exogène, une intégration de nouveaux spacers dans le CRISPR array. Ceci est le fondement de l'immunité adaptative bactérienne par les systèmes CRISPR. Cas1 et Cas2 sont des protéines métallo-dépendantes, leurs mécanismes moléculaires ont pu être résolues à l'aide de structures cristallines. C'est notamment ainsi que le complexe multimérique de quatre Cas1 externes et deux Cas2 centrales, de *E. coli* K12 a pu être décrit (système de type IE) (Nuñez *et al.*, 2014). Le complexe servirait à découper un fragment d'ADN exogène à l'aide de Cas1 qui possède une activité nucléase. Ce fragment

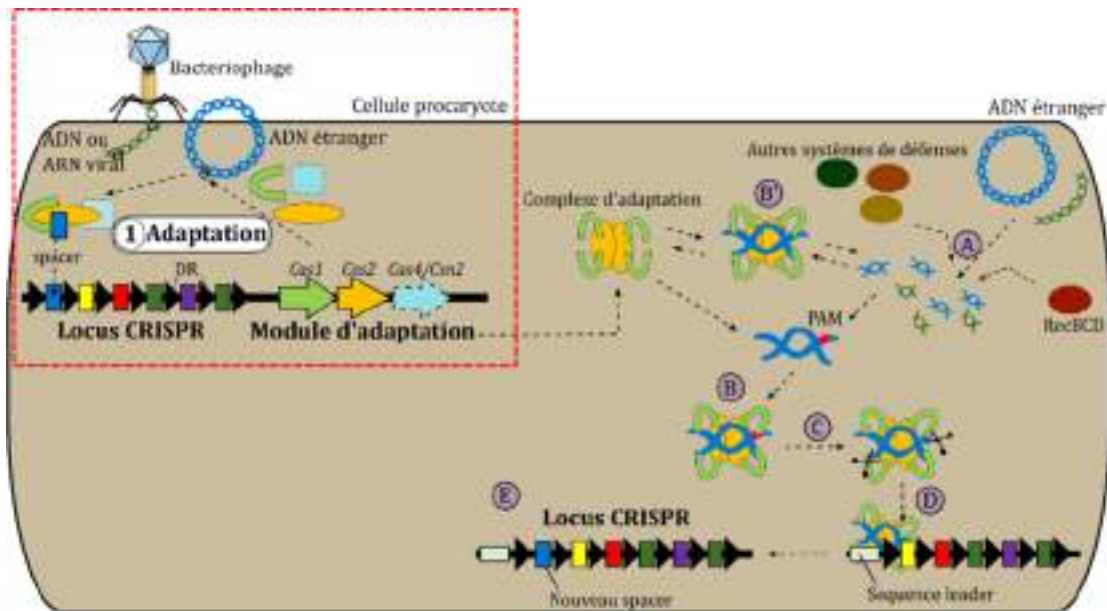


Figure 11 : Représentation détaillée de la phase d'adaptation du système de défense CRISPR. Les gènes codant les protéines sont représentés par des flèches dont le contour est plein si leur présence est obligatoire au système ou pointillé si alternatif. Le nom du gène possède un "/" si dans certains systèmes il s'agit du gène Cas4 ou du gène Csn2. L'ordre des étapes est représenté par les lettres A-B/B'-C-D-E et correspond au clivage de l'ADN exogène, à la reconnaissance d'un fragment de celui-ci possédant ou non une séquence PAM adéquate, à la préparation de ce fragment et à son intégration au niveau de la séquence leader du locus CRISPR afin d'obtenir un locus CRISPR composé d'un nouveau spacer.

est par la suite intégré à l'aide de Cas2 dans le CRISPR array près de la séquence leader (séquence la plus en amont du CRISPR array). Ce complexe fonctionne par paire et dans certains systèmes une protéine de fusion peut être directement retrouvée (Nuñez *et al.*, 2014). Une méthode de FRET (Transfert d'énergie entre molécules fluorescentes) en temps réel sur le complexe Cas1-Cas2 de *Enterococcus faecalis* (système de type II) a également permis de reconstituer de manière cinétique un modèle de base de l'intégration du spacer dans le locus CRISPR (Budhathoki *et al.*, 2020).

Le mécanisme d'adaptation décrit précédemment est le modèle le plus simple, mais de nombreuses études prouvent que selon les systèmes CRISPR, des divergences et spécificités sont retrouvées. Certaines équipes ont montré une spécificité dans le processus d'acquisition des spacers en fonction des séquences PAM qui seront nécessaires à l'effecteur pour l'interférence. Notamment, de nouvelles structures du complexe Cas1/Cas2 (*E. coli* K12, type IE) lié à l'ADN montrent que le complexe forme une fourche à deux branches qui change de conformation lorsqu'il interagit avec un spacer. La séquence PAM est reconnue par la sous-unité Cas1a par interaction spécifique de la séquence PAM et il y a alors clivage du morceau d'ADN par le complexe qui régit également la taille du spacer (J. Wang *et al.*, 2015). De plus, des simulations de dynamiques moléculaires basées sur des modèles mathématiques montrent que la stabilité du complexe Cas1/Cas2 augmente lors d'une liaison au protospacer suivi de la séquence PAM adaptée au système (Wan *et al.*, 2019). Chez d'autres systèmes, certains acteurs semblent importants à l'acquisition, pour les systèmes CRISPR ciblant l'ARN par exemple, une protéine reverse transcriptase (RT) serait nécessaire à l'intégration de nouveaux spacers (Koonin, 2019). Chez un certain nombre de systèmes CRISPR de type II-C, la présence de Cas4 ou Csn2 est également essentielle (Mir *et al.*, 2018). De plus, la protéine Cas9 qui est l'effecteur, semble dicter à Cas1 et Cas2, les spacers qu'il faut acquérir de manière directe ou indirecte puisqu'il a été montré qu'une Cas9 reconnaissant une PAM différente entraîne l'acquisition de spacers avec une PAM différente (Heler *et al.*, 2015; Gleditzsch *et al.*, 2019).

L'étape d'adaptation nécessite également une discrimination du soi et du non-soi sans laquelle, de nombreux spacers provenant de leur propre ADN seraient intégrés au locus CRISPR. Il a été montré que, pour le système CRISPR de type II-A chez *S. thermophilus*, lorsque la Cas9 est inactivée, les spacers intégrés sont peu discriminés. De nombreux

spacers ciblant le génome de la bactérie elle-même sont alors acquis. Ceci peut être expliqué par la pression de sélection qui, en contexte naturel, aurait éliminé la bactérie ayant intégré un spacer ciblant son propre génome. Cela impliquerait un taux de suicide important et expliquerait que l'on ne retrouve pas naturellement de souches avec une accumulation de spacers dirigés contre le génome (Wei, Terns and Terns, 2015). Dans cette étude les auteurs soulignent l'acquisition de spacers de manière aléatoire et non spécifique d'un ADN exogène et une régulation par Cas9 qui élimine les cellules ayant acquis des spacers provenant de son propre génome.

A contrario, pour le système de type IE de *E. coli*, il a été montré une acquisition préférentielle de spacers ciblant des ADN exogènes. En effet, 100 à 1000 fois plus d'acquisitions auraient pour origine de petites molécules d'ADN (plasmides) plutôt que l'ADN de l'hôte (Levy *et al.*, 2015). Cela a été également observé chez *Sulfolobus islandicus* (Erdmann *et al.*, 2014). L'hypothèse la plus admise provient du taux de réplication plus élevé des éléments mobiles. Or, lors de la réplication en général, des cassures de l'ADN peuvent survenir et sont prises en charge par le complexe endogène de réparation RecBCD qui va digérer l'ADN jusqu'à une séquence d'ADN précise appelé Chi-site. Un grand nombre de spacers au niveau de ces séquences Chi-site ont été observés. Cas1 qui lie l'ADN simple brin, aurait alors une facilité à lier ces morceaux d'ADN digérés préalablement par RecBCD et ceci favoriserait l'acquisition de spacers (Levy *et al.*, 2015; Wimmer and Beisel, 2020). Ceci peut être également relié au fait qu'un ADN exogène d'ores et déjà ciblé par le système CRISPR, génère des morceaux d'ADN double brins qui sont inactifs mais qui peuvent par la suite être pris en charge par le complexe RecBCD. C'est alors que de nouveaux spacers provenant d'une position proche de la précédente cible sont acquis préférentiellement (Westra *et al.*, 2015).

Finalement, le mécanisme de base d'adaptation qui repose sur Cas1 et Cas2 est globalement le même pour l'ensemble des systèmes CRISPR et ce sont quelques différences sur la spécificité ou l'efficacité de l'acquisition qui divergent entre les systèmes. Il s'agit là de la première phase immunitaire, essentielle à la réponse adaptative.

2.2.4.2. Expression et mise en place du système

L'expression du locus/système est la seconde étape, il s'agit là de la mise en place à proprement parler de la phase immunitaire adaptative. De manière générale et chez la plupart des systèmes, le CRISPR array est transcrit, les protéines Cas et autres protéines

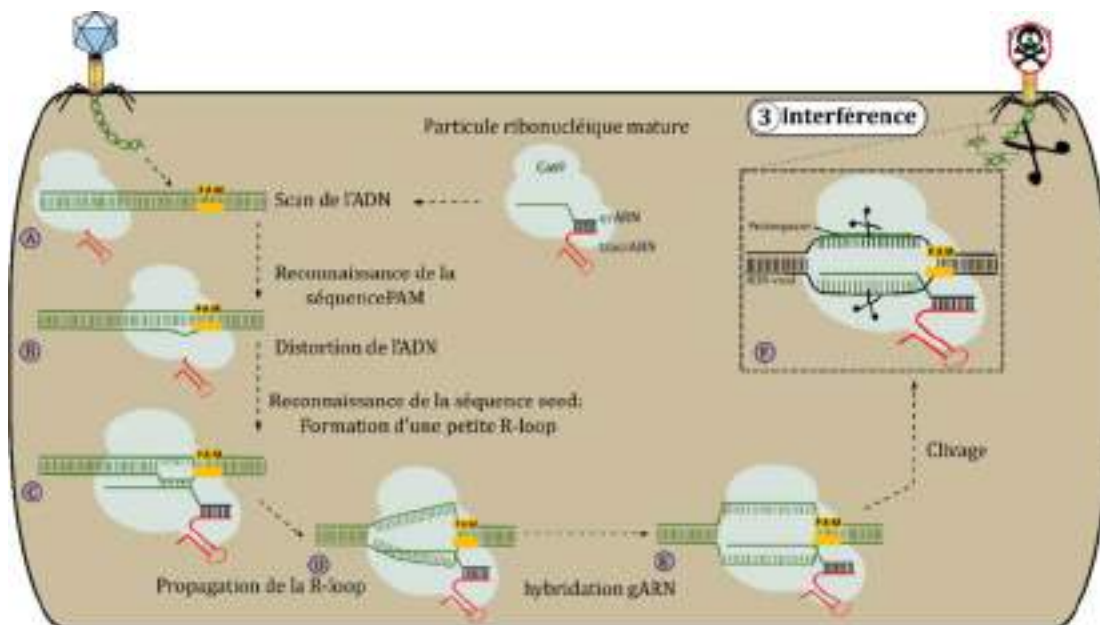


Figure 12 : Représentation schématique de l'étape d'interférence. Le schéma est découpé en 6 parties de A à F. La partie A, représente le scan de l'ADN effectué par la particule ribonucléique, B représente l'étape de reconnaissance d'une séquence PAM et la distorsion de l'ADN associé, l'étape C'est l'étape de reconnaissance et hybridation de la séquence seed nécessaire à la formation d'une pseudo R-loop qui permet par la suite l'étape de propagation de la R-loop lors de l'étape D. Finalement, l'étape E représente l'hybridation totale de l'ARN guide (gRNA), la stabilisation de la R-loop qui entraîne finalement le clivage de l'ADN lors de l'étape finale F.

associées sont exprimées et le pre-crRNA, est mûri en crRNA par un clivage qui adapte notamment la taille du guide ARN. Ce dernier est alors pris en charge par un ou plusieurs effecteurs qui vont former une particule ribonucléique mature prête à défendre la bactérie ou l'archée (Barrangou and Horvath, 2017; Koonin, 2019). Concernant cette phase, la différence majeure entre les systèmes se situe au niveau de la maturation du système puisque pour les CRISPR de type II, Cas9 est incapable de mûrir elle-même le crRNA et c'est la RNase III endogène de la bactérie qui va avoir ce rôle (Fonfara *et al.*, 2014). C'est également pour cela que ce système est retrouvé uniquement chez les bactéries, les archées ne possédant pas cette activité (Makarova *et al.*, 2020). A contrario, pour les systèmes de type IV et V, c'est l'effecteur lui-même qui va mûrir le crRNA, et chez l'ensemble des systèmes CRISPR de classe 1 (type I, III et IV), c'est la protéine Cas6 qui réalise ce processus (Makarova *et al.*, 2020). Une fois la particule assemblée et mûrie, il faut attendre une nouvelle infection pour alors avoir la phase immunitaire adaptative : l'interférence.

2.2.4.3. *Élimination de la menace par l'étape d'interférence*

L'interférence est la dernière étape de défense (Figure 12). Elle correspond à l'action de la particule ribonucléique mature qui va être guidée par son crRNA sur une cible précise de l'ADN exogène. La reconnaissance de la séquence complémentaire au spacer, le protospacer, va permettre la formation d'une structure R-loop stable et l'enclenchement de l'action d'interférence qui aboutit au clivage spécifique de l'ADN exogène et l'inhibition de l'invasion (Barrangou and Horvath, 2017; Koonin, 2019).

Le mécanisme général décrit précédemment est cette fois encore, un modèle qui présente des variantes en fonction des différents types de système CRISPR. Le mécanisme le plus résolu est celui des systèmes de types II et IV qui sont les systèmes mono-effecteurs et d'intérêt biotechnologique. Pour les systèmes les plus étudiés, différents modèles fonctionnels très précis ont été obtenus à l'aide notamment de nombreuses structures cristallines : (i) Cas9 de type II-A de *S. pyogenes* et *S. aureus* (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014, 2015) ; (ii) Cas9 de type II-C de *Actinomyces naeslundii* (Jinek *et al.*, 2014), *Campylobacter jejuni* (Yamada *et al.*, 2017) et *Corynebacterium diphtheriae* (Hirano *et al.*, 2019) ; (iii) Structures des complexes liés à leur ADN cible pendant la R-loop (Jiang *et al.*, 2016) ; (iv) des structures de la Cas12a de *Francisella novicida* (Swarts *et al.*, 2017). L'ensemble de celles-ci ainsi que des technologies innovantes telles que la high-speed

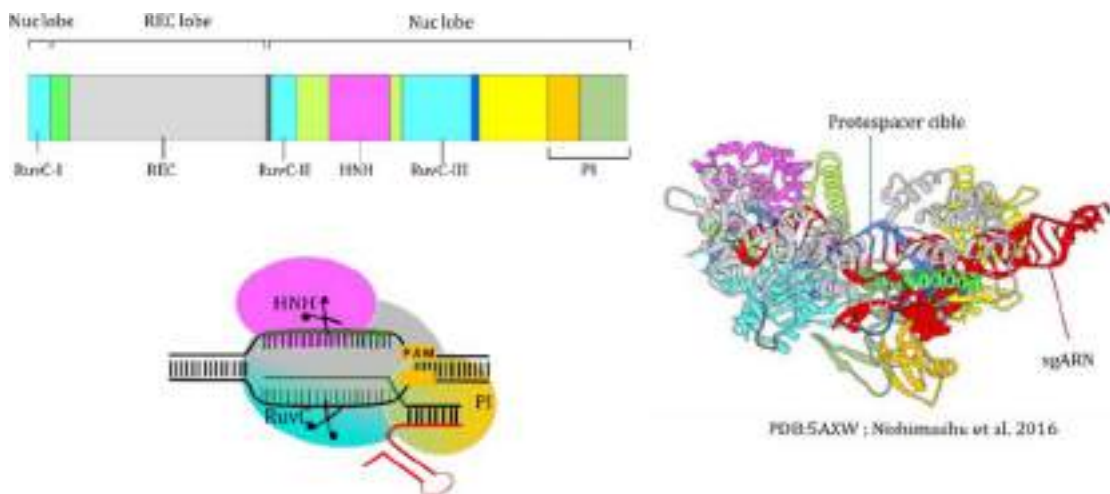


Figure 13 : Représentation schématique des domaines de la protéine Cas9. Représentation des principaux domaines de la protéine Cas9 et de la structure du complexe ADN-sgRNA-Cas9 de *Staphylococcus aureus*.

atomic force microscopy (HS-AFM), qui permet de regarder en temps réel la dynamique d'action de Cas9 (Shibata *et al.*, 2017) ou la single-molecule fluorescence energy transfer (smFRET) pour observer la dynamique de la conformation de la R-loop (Zeng *et al.*, 2018), ont permis de préciser le modèle de base de l'interférence (Figure 12).

L'étape initiale du processus est celle de la reconnaissance de la séquence PAM par le complexe ribonucléique. Il a été montré que l'absence de séquence PAM adaptée affecte la production de la R-loop par manque de stabilité (Szczelkun *et al.*, 2014; Swarts *et al.*, 2017) et que c'est la reconnaissance de la séquence PAM qui produit une distorsion de l'ADN de 30° rendant alors possible l'accès à l'ADN pour l'ARNg (Jiang *et al.*, 2016). La plupart du temps, la séquence PAM n'est reconnue que sur un brin de l'ADN, mais pour quelques systèmes de type IIC comme *C. jejuni* et *C. diphtheriae* (Yamada *et al.*, 2017; Hirano *et al.*, 2019), la séquence PAM interagit avec la protéine Cas9 sur les deux brins de l'ADN. La structure du complexe dans sa conformation finale étant résolu pour de nombreuses Cas9, cette étape est apparue particulièrement similaire pour toutes ces Cas9. Le domaine structurel le plus variable de Cas9 est le domaine PI qui a pour rôle la reconnaissance et l'interaction avec la séquence PAM (Figure 13) (Nishimasu *et al.*, 2015). Le domaine PI interagit de manière spécifique avec les bases correspondantes à la séquence PAM. Ces interactions constituent la base moléculaire de la reconnaissance de la séquence PAM canonique qui peut aller de trois nucléotides comme pour spCas9 (NGG) à dix nucléotides pour la Cas9 de *Clostridium perfringens* (Karvelis *et al.*, 2015; Anders, Bargsten and Jinek, 2016). Des différences structurelles du domaine PI, liées à des séquences protéiques différentes, sont corrélées à des variations de motif PAM canonique. Ceci explique la diversité très importante de séquences PAM selon les espèces bactériennes d'où provient la Cas9 (Gasiunas *et al.*, 2020).

La seconde étape est la reconnaissance de la séquence « seed ». Cette séquence est la partie proximale du guide ARN par rapport à la séquence PAM. La taille de celle-ci peut varier selon les systèmes, de 7 nucléotides décrits dans *E. coli* (Semenova *et al.*, 2011) jusqu'à 11 chez *S. thermophilus* (Szczelkun *et al.*, 2014). Cette interaction entre la séquence « seed » et la séquence ciblée va stabiliser la R-loop primaire, sans laquelle il n'y a pas d'interférence. Si la stabilité est suffisante, la R-loop se propage et le gRNA s'hybride à l'ADN sur toute sa longueur (Swarts *et al.*, 2017; Zeng *et al.*, 2018) formant ainsi un complexe très stable. Une fois la conformation adoptée, il y a clivage de chaque brin de

l'ADN par les deux domaines catalytiques RuvC et HNH pour Cas9 et seulement par le domaine RuvC pour Cas12 (Swarts *et al.*, 2017; Zeng *et al.*, 2018).

L'interférence a été décrite de manière très approfondie pour les systèmes à effecteur unique, mais les variations liées au type de système ne permettent pas pour le moment de généraliser les détails du mécanisme. Il a par exemple été remarqué avec Cas12a et Cas9 que de plus nombreux mésappariements étaient tolérés avec un plasmide super-enroulé négativement et qu'une accélération de la cinétique de clivage était également observée (Aelst *et al.*, 2019; Ivanov *et al.*, 2020). Cela a-t-il un rôle biologique ou est-ce simplement une coïncidence structurale ? Concernant l'interférence par les systèmes à multiples effecteurs, peu de données sont disponibles, notamment à cause de la complexité d'assembler des cascades effectrices avec de nombreuses protéines. Le complexe de type IC, ne possédant que 3 protéines, a pu être résolu par structure cristalline (O'Brien *et al.*, 2020), mais peu de données sont disponibles sur la grande diversité des systèmes à multiples effecteurs.

2.2.4.4. Systèmes atypiques

Les différents types de CRISPR précédemment décrits donnent une vision globale de leurs modes d'action. Mais ces systèmes CRISPR qui sont, à l'origine, des systèmes de défense, ont pu évoluer vers de nouvelles fonctions. C'est ainsi que certains systèmes atypiques ont été décrits dans la littérature ces dernières années.

Par exemple, chez *Francisella novicida*, la présence naturelle d'un petit ARN non codant supplémentaire a été décrite. Celui-ci, à l'aide de Cas9, cible l'ARN messager d'une lipoprotéine de surface de la bactérie, induisant une baisse de la quantité de la protéine à la surface de la bactérie, ceci entraînant une baisse de détection par le système immunitaire de son hôte (Sampson and Weiss, 2013).

Pour *Campylobacter jejuni*, deux rôles distincts ont été caractérisés pour la protéine Cas9 chez certaines souches : le premier est le clivage de l'ARNm ciblé par le crRNA (Dugar *et al.*, 2018), le second est la sécrétion de la protéine Cas9 seule dans le cytoplasme de cellule humaine, entraînant alors la mort de la cellule (Saha *et al.*, 2020).

Un autre système de type II, pour lequel un tracrRNA exceptionnellement grand a été identifié, entraîne l'autorépression du promoteur de Cas9 et l'inactivation de cette répression entraîne un taux d'immunisation multiplié par 3000 (Workman *et al.*, 2020).

Ces trois exemples imagent parfaitement les différences entre systèmes qui sont nombreuses et peuvent avoir des répercussions importantes sur leurs fonctions. De tels systèmes atypiques sont certainement nombreux dans la nature puisque chacun évolue individuellement dans chaque individu de chaque espèce procaryote. Des nouveaux devraient encore être identifiés dans les prochaines années.

2.2.5. Balance bénéfices-coûts du système adaptatif

La question du « fitness costs », ou coût d'adaptation, des systèmes CRISPR est abordée dès 2010 (Horvath and Barrangou, 2010). Environ 50 % des génomes qui contiennent un système CRISPR, possèdent plus d'un locus (Deveau, Garneau and Moineau, 2010). Une souche *Streptococcus thermophilus* possède quatre loci CRISPR différents (Horvath and Barrangou, 2010), *Tistrella mobilis* et *Azospirillum* B510 possèdent même trois types de CRISPR différent (Chylinski *et al.*, 2014) et ceci peut monter à 20 loci CRISPR chez *Methanococcus jannaschii*. Ces données semblent montrer que les systèmes CRISPR ont un faible coût de fitness, mais pourtant, de fréquents exemples de pertes de locus CRISPR suggèrent, au contraire, que ces systèmes sont coûteux (Delaney *et al.*, 2012; Koonin, 2019). Certains modèles prédisent même que 10^{-4} cellules au sein d'une population de 10^{10} cellules, seraient déficientes ou auraient perdu leurs systèmes CRISPR (Jiang, Maniv, *et al.*, 2013). Pourtant le gain de fitness de la bactérie ayant éliminé le système ne serait pas forcément considérable et perceptible au niveau de la croissance *in vitro* (Jiang, Maniv, *et al.*, 2013). Des études bio-informatiques ont permis de montrer que lorsqu'une bactérie a une faible diversité de parasites (éléments mobiles), le coût de l'immunité adaptative est trop important et un simple système de l'immunité innée est suffisant. Il en est de même pour des taux très élevés de diversité virales, le cout généré par l'acquisition de nombreux spacers n'est pas non plus bénéfique et agit sur la fitness de l'organisme. D'autres systèmes de défenses sont alors favorisés comme la variation des récepteurs d'entrée à la surface des bactéries (Broniewski *et al.*, 2020). En revanche, c'est dans une fenêtre de diversité phagique intermédiaire que la balance (coût généré par la présence de systèmes de l'immunité adaptative / fitness) est le plus favorable. (Weinberger *et al.*, 2012; Koonin, 2019), ce qui a été confirmé de manière expérimentale (Houte *et al.*, 2016; Morley *et al.*, 2016). Il en est de même pour l'abondance de phages qui augmente le poids du bénéfice dans la balance et favoriserait le maintien et l'utilisation d'une défense adaptative (Meaden *et al.*, 2021).

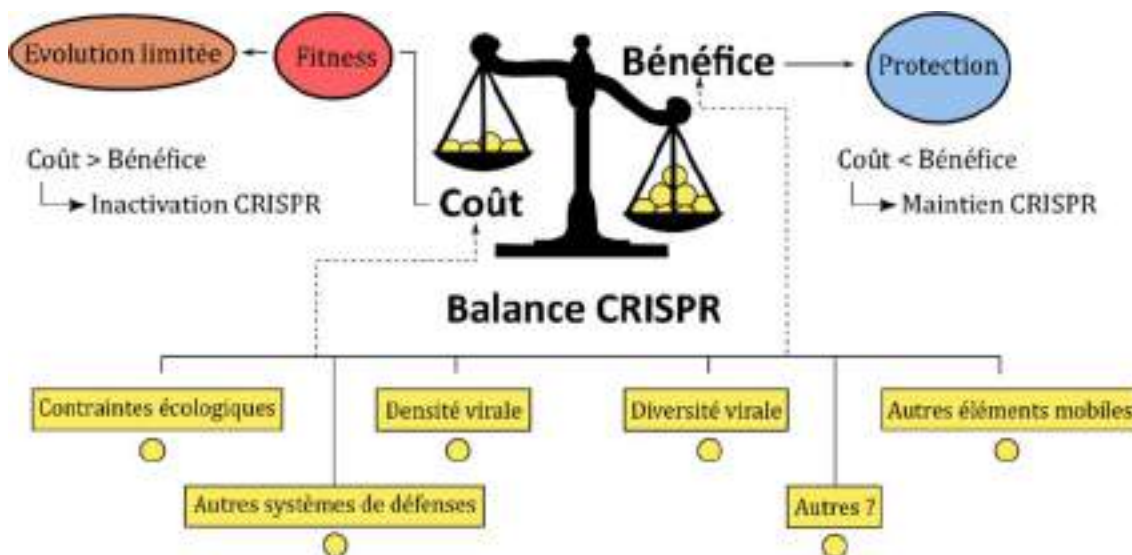


Figure 14 : Balance coût-bénéfice du système CRISPR. Les différents facteurs identifiés pesant dans la balance sont représentés en jaunes. Les coups principalement identifiés sont représentés en rouge et orange, le bénéfice principal en bleu.

Le coût sur la fitness de l'organisme peut être en grande partie expliqué par des effets délétères du système de défense. Par exemple, la balance entre l'insertion de nouveaux spacers dans le locus CRISPR et l'auto-immunité que cela pourrait engendrer (Iranzo *et al.*, 2013). En effet, il a été montré que l'insertion de nouveaux spacers pouvait amener une population bactérienne à ajouter des spacers ciblant son propre génome. L'accumulation de spacers représente donc fondamentalement un enrichissement de l'arsenal cellulaire pour lutter contre les éléments invasifs, mais aussi un risque accru d'effets délétères sur la survie de la population bactérienne (Wei, Terns and Terns, 2015). Ainsi, l'augmentation du nombre de spacers augmente le risque de potentiels clivages off-target liés à des séquences seed/PAM correctes présentes sur le génome et suffisantes pour induire le clivage du génome bactérien à faible fréquence (Szczelkun *et al.*, 2014). Pourtant, dans certains génomes, on peut retrouver le CRISPR array pouvant représenter jusqu'à 1 % du génome total (Deveau, Garneau and Moineau, 2010).

Une autre explication peut provenir du coût évolutif important pouvant être engendré par les systèmes CRISPR qui limitent les transferts horizontaux (Horizontal Gene Transfer, HGT). Le transfert horizontal de gènes est le moteur majeur de l'évolution des procaryotes et il est essentiel à la survie à long terme des populations microbiennes (Koonin, 2019). Étant donné que les ADNs exogènes apportés par transduction, par conjugaison (Marraffini and Sontheimer, 2008; Mahendra *et al.*, 2020) et transformation (Bikard *et al.*, 2012) sont la cible des systèmes CRISPR, on peut se demander quel est le coût évolutif relatif et le « compromis évolutif » mis en place entre ces différents processus. La trace fréquente de HGT dans les génomes procaryotes, malgré la présence de systèmes CRISPR, doit impliquer un avantage important pour la bactérie ou un mécanisme permettant de ne pas être repéré par le système immunitaire bactérien. En effet un transfert horizontal présente, tout comme les systèmes de défense, un coût chez la bactérie, ce coût étant le plus souvent compensé par un avantage qu'il confère (Rocha and Bikard, 2021). Mais entre le coût du système CRISPR pour son évolution et sa maintenance et le coût des transferts horizontaux avec les avantages de fitness ou non qu'il confère, un équilibre semble être présent dans le milieu naturel des bactéries. L'étude des conditions écologiques des bactéries est vraisemblablement une clef dans la compréhension de cet équilibre et pour comprendre quand le système peut être conservé ou non et quand la bactérie nécessite l'acquisition d'ADN exogène pour évoluer (Figure 14) (Westra and Levin, 2020).

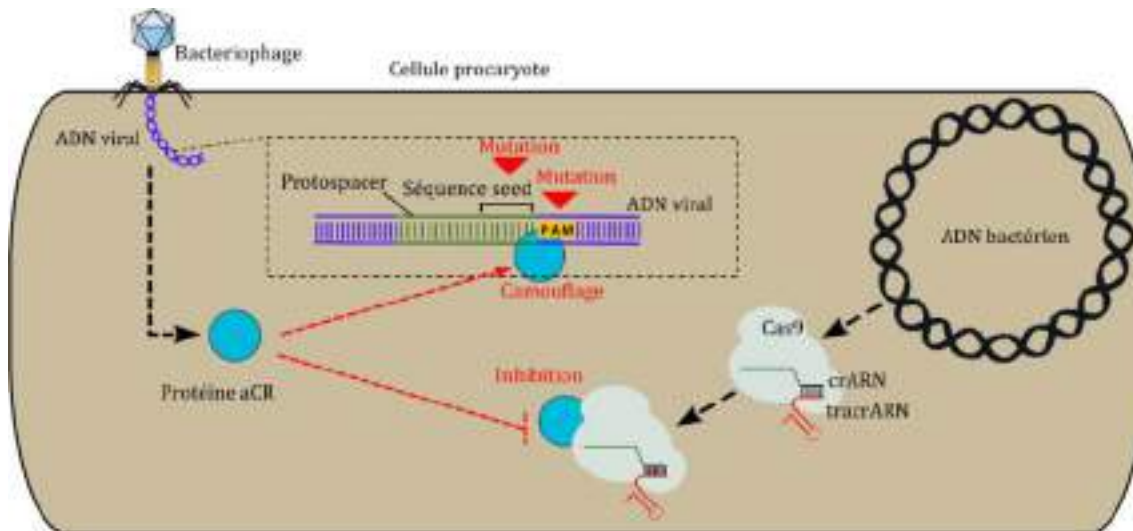


Figure 15 : Représentation des stratégies d'adaptation adoptées par les phages en réponse au système CRISPR. La séquence d'ADN viral est représentée en bleu à l'exception de la séquence ciblée par le système CRISPR, le protospacer est en vert. Les protéines codées par l'ADN viral, les aCR (anti CRISPR) sont représentés en bleu.

2.2.6. Modèle dynamique du combat bactérie-phage

Comme nous l'avons vu précédemment, les systèmes CRISPR sont très divers et possèdent de nombreuses caractéristiques spécifiques, mais l'ensemble de ces systèmes montre une dynamique évolutive très importante qui va être régie par la balance avantages versus inconvénients et le coût évolutif (Figure 14). Dans un contexte naturel, les bactéries et les éléments mobiles agissent tels le « chat et la souris ». Par le système CRISPR, les bactéries cherchent à se protéger des éléments mobiles qui pourraient interférer avec leur bon fonctionnement. Ces bactéries vont alors intégrer de nouveaux spacers dans leurs loci CRISPR. À l'inverse, les pressions de sélection exercées par les systèmes de défense entraînent l'adaptation des bactériophages qui vont évoluer pour échapper ou inhiber la machine à tuer CRISPR pour ainsi à nouveau se développer. Ce duel est constant et de nombreuses études ont mis en évidence divers mécanismes d'adaptations mis en place de part et d'autre (Figure 15).

Plusieurs analyses informatiques, mais également des expériences de co-évolution ont mis en évidence un système d'échappement naturel pour les bactériophages: l'évolution sous pression de sélection (Horvath and Barrangou, 2010; Sun *et al.*, 2013; Paez-Espino *et al.*, 2015). En effet, après infection d'une culture bactérienne par un phage, le phage prend le contrôle des cellules bactériennes pour se multiplier de manière dynamique. Mais très rapidement, de nouveaux spacers ciblant ces phages sont intégrés au sein de la population bactérienne comme cela a été vu par exemple chez *S. thermophilus* (Sun *et al.*, 2013). Ceci permet à la population bactérienne de reprendre le dessus et d'éliminer la quasi-totalité de la population phagique.

Lors d'expériences d'infection de bactérie par des phages, après mise en place de la défense adaptative par la bactérie, la population virale diminue. En revanche il est parfois possible d'observer l'apparition de phages résistants (BIM) aux systèmes CRISPR et la population virale peut reprendre sa croissance. Des mutations au niveau des spacers ou de la séquence PAM ont été caractérisées chez ces BIM. Ceci entraîne l'évolution du phage afin de ne plus être sensible au système CRISPR (Horvath and Barrangou, 2010; Sun *et al.*, 2013). Il a été montré que ces SNV (single nucleotide variation) sont surreprésentés dans les régions du phage ciblées par le système (Paez-Espino *et al.*, 2015). Cette adaptation est contournée par la bactérie à l'aide de l'intégration de nouveaux spacers mais pas seulement. Chez certaines espèces, une alternative est trouvée par la modulation de la

spécificité des étapes de l'interférence, telle une forme d'adaptation à plus long terme de la bactérie.

- Par exemple, une mutation dans la séquence seed signifie la plupart du temps l'inhibition de l'interférence, contrairement à une mutation sur la partie distale de le gRNA (Szczelkun *et al.*, 2014). Cette tolérance est variable selon les systèmes et les plus flexibles vont alors pouvoir répondre plus facilement contre certains BIM.
- Il a également été décrit qu'outre la séquence PAM dite naturelle ou canonique, d'autres motifs PAM non-canoniques peuvent permettre un clivage de l'ADN (Anders, Bargsten and Jinek, 2016). Ces motifs sont reconnus grâce à une accommodation stérique lié à une distorsion de l'ADN ou à une compensation énergétique améliorant la thermodynamique de la liaison à l'ADN (Anders, Bargsten and Jinek, 2016). Ceci a notamment été identifié chez certaines structures de protéines Cas liées avec un motif PAM canonique ou non (Yamano *et al.*, 2017), mais également observés dans de nombreux cas, par exemple pour la très étudiée SpCas9 qui reconnaît un motif NGG : une certaine flexibilité est observée pour les motifs NAG et NGA (Jiang, Bikard, *et al.*, 2013), mais aussi pour NAGG et NYGG (Collias *et al.*, 2020). Ces complexes seraient moins stables mais suffisant pour le clivage de l'ADN (Yamano *et al.*, 2017).

Ces différents types de flexibilité vont limiter l'évasion des bactériophages qui auraient muté dans le protospacer (Semenova *et al.*, 2011) ou sur la séquence PAM.

Un autre type adaptation a été mis en place par les bactériophages. Cette hypothèse avait d'ores et déjà été évoquée comme une possibilité en 2010 sans avoir été démontrée, il s'agit de la présence de protéines qui pourraient interférer avec la machinerie CRISPR (Horvath and Barrangou, 2010). Aujourd'hui, 32 sous-types de systèmes CRISPR possèdent des protéines dites Anti-CRISPR (Acr), capables d'inhiber la cascade d'interférence (Marino *et al.*, 2020). De nombreuses études les ont mises en évidence et différentes étapes de l'interférence sont ciblées. Par exemple, les protéines AcrF2 et F10 ou ACRIIA2 et A4 inhibent l'accès à la PAM en liant partiellement ou totalement le site de liaison à l'ADN (Chowdhury *et al.*, 2017; L. Liu *et al.*, 2019), AcrVA1 qui clive le crRNA de Cas12a, AcrV5A qui va acétyler Cas12a pour l'inhiber ou également des protéines qui vont inhiber le clivage comme ACRIIC1 qui lie le domaine HNH (Marino *et al.*, 2020).

Ainsi ce jeu du chat et de la souris entraîne un modèle dynamique important des systèmes CRISPR par des adaptations plus ou moins rapides (mutations, acquisitions de spacers, de gènes). De nombreux autres mécanismes pourraient être découverts dans les prochaines années concernant les systèmes CRISPR, mais pas seulement. Depuis l'avènement des systèmes CRISPR et leurs compréhensions, la recherche d'autres mécanismes de défense bactériens s'est développée, et de nombreux mécanismes sont à l'étude. On se rend compte de plus en plus que les bactéries subissent des pressions de sélections fortes de leurs environnements et notamment des phages (Jiang, Maniv, *et al.*, 2013; Bernheim and Sorek, 2020; Chevallereau *et al.*, 2021). Cette situation écologique est un moteur important de l'évolution des bactéries. En réponse à ces agressions de phages, mais également aux facteurs environnementaux, les systèmes de défense évoluent et restent dynamiques et ceci est favorisé par l'agencement de ces systèmes se regroupant le plus souvent en îlots de défense afin de favoriser les échanges intra-bactéries (Makarova, Wolf and Koonin, 2013; Doron *et al.*, 2018; Bernheim and Sorek, 2020; Rousset *et al.*, 2021). En effet, des flux constants d'échanges et de pertes de systèmes de défense sont caractérisés au sein de populations bactériennes (Jiang, Maniv, *et al.*, 2013; Bernheim and Sorek, 2020) ce qui leur confère une protection contre les diverses attaques, cela a par exemple été montré expérimentalement qu'une population avec une plus grande diversité génétique limite l'émergence de virus qui échappe au système immunitaire des bactéries (Houte *et al.*, 2016) et qu'une fois la menace éloignée, ces bactéries vont alors rediversifier leurs « PAN immune system » par l'échange et la perte de ces systèmes de défense (Bernheim and Sorek, 2020).

La recherche sur les systèmes de défense bactériens est en plein essor aussi bien sur les systèmes CRISPR que sur les nouveaux systèmes en cours de caractérisation. Il y a fort à parier que les prochaines décennies de recherches vont encore nous apporter de nombreuses connaissances dans la caractérisation fonctionnelle de tous ces systèmes nouvellement identifiés. Ceci élargira la vision de la dynamique des interactions procaryotes-phages et la compréhension de l'écologie de ces organismes. D'autres systèmes vont encore être découverts du côté des bactéries dans les prochaines années à l'aide d'outils de détection de plus en plus performants comme le logiciel PADLOCK (Payne *et al.*, 2021) ou DefenseFinder (Tesson *et al.*, 2021) pour l'identification des

systèmes connues dans des génomes, mais aussi du côté des phages avec l'identification de nouveaux mécanismes anti-défenses. À l'aide de la compréhension de leurs mécanismes, on pourra alors améliorer la compréhension de cette histoire évolutive intéressante entre MGE et bactérie. Une évolution continue entre défense contre ces éléments mobiles et maintien de la fitness et acquisition de ces éléments pour une évolution plus rapide avec l'acquisition de phénotypes avantageux. Quelques groupes de recherches essaient d'élargir la connaissance non plus à l'échelle d'un seul système ou d'un seul point de vue, mais en considérant ces systèmes à une échelle plus large, en prenant en compte le contexte écologique (Bernheim and Sorek, 2020; Westra and Levin, 2020; Chevallereau *et al.*, 2021; Rocha and Bikard, 2021).

Outre la compréhension de ces systèmes et l'aspect fondamental, l'intérêt biotechnologique des systèmes CRISPR a largement contribué au développement des recherches fondamentales sur ce sujet. La partie suivante sera consacrée aux outils d'ingénierie de génome développés à partir des systèmes CRISPR.

2.3. Détournement du système CRISPR pour l'édition génétique

La caractérisation des systèmes de défense CRISPR, et notamment la démonstration de l'utilisation en système hétérologue de Cas9 pour protéger *E. coli* contre certains phages (Sapranaukas *et al.*, 2011) a été un tournant dans les développements des outils CRISPR. Des applications à visée biotechnologique ont été imaginées, comme l'idée d'utiliser les systèmes CRISPR dans l'industrie pour conserver un système de production bactérien résistant aux bactériophages (Barrangou and Horvath, 2012). Mais finalement c'est en 2012 que l'application la plus impactante a fait son apparition. Les systèmes CRISPR ont révolutionné le domaine de l'ingénierie de génome chez les eucaryotes comme chez les procaryotes. Le système Cas9 est particulièrement concerné, il est le premier à avoir été utilisé comme outil pour le clivage double brin de l'ADN à un endroit très précis de l'ADN à l'aide d'un ARN simple guide (sgRNA) (Jinek *et al.*, 2012). Ceci favorise l'évènement de la recombinaison homologue pour modifier de manière efficace et spécifique l'ADN. Cette technologie d'ingénierie de génome a notamment permis à Emmanuelle Charpentier et Jennifer Doudna d'obtenir le Prix Nobel de Chimie en 2020. C'est à partir de ces premières démonstrations, et voyant le potentiel de cette technologie, que les études ont explosées. En 10 ans, c'est plus de 27 000 articles publiés, qui sont répertoriés sur Pubmed en utilisant le terme « CRISPR OR Cas9 ». C'est le point de départ d'une expansion importante

de la boîte à outils pour l'ingénierie des génomes, quels qu'ils soient, ainsi que pour des applications dérivées.

2.3.1. Premières applications d'un outil universel

Par leur simplicité et leur découverte préalable aux types V et VI, les systèmes de types II se sont imposés comme l'outil principal pour de nombreuses applications avec notamment le système de *S. pyogenes*. Comme précédemment décrits, ces systèmes nécessitent seulement la présence de la protéine Cas9 et du sgRNA pour cliver de manière très précise l'ADN. En 2012, c'est par une expérience *in vitro* de clivage de plasmide que cela a tout d'abord été démontrée avec l'utilisation d'un tracrRNA couplé au crRNA (Jinek *et al.*, 2012). C'est encore *in vitro* que de l'ADN génomique d'*E. coli* et humain ont été clivés de manière précise par le système de *S. thermophilus* (Karvelis, Gasiunas and Siksnys, 2013). Mais rapidement, la technologie s'est répandue et les preuves de concept dans de nombreux organismes *in vivo* se sont multipliées (Sander and Joung, 2014). Des simples bactéries aux organismes eucaryotes complexes, ce sont de multiples modèles qui ont montré que le système CRISPR est un outil universel et qu'il peut fonctionner chez les bactéries (Jiang, Bikard, *et al.*, 2013), chez les levures (DiCarlo *et al.*, 2013), chez les cellules humaines (lignées cancéreuses et cellules souches pluripotentes) (Cho *et al.*, 2013; Cong *et al.*, 2013; Jinek *et al.*, 2013; Mali *et al.*, 2013) mais aussi directement chez les plantes (J. . Li *et al.*, 2013; Nekrasov *et al.*, 2013; Xie and Yang, 2013), le zebrafish (Hwang *et al.*, 2013), la souris (Shen *et al.*, 2013; Wang *et al.*, 2013), le lapin (Yang *et al.*, 2014) et bien d'autres organismes. Cette utilisation du système CRISPR dans ces nombreux modèles a mis en exergue la précision de cet outil, son efficacité et son faible coût.

2.3.2. Outil génétique pour les bactéries

L'ingénierie de génome par le système CRISPR-Cas9 est à l'origine basée sur la coupure précise d'un endroit du génome, suivie par la réparation de l'ADN par la voie NHEJ (Non-homologous end joining) ou HDR (homologous direct repair) permettant ainsi l'édition du génome. Chez plusieurs espèces bactériennes, le système CRISPR-Cas9 a permis des modifications génétiques, mais étant donné que chez de nombreuses espèces ou souches, les voies de réparations de l'ADN ne sont pas suffisamment efficaces, le système a rapidement servi à une utilisation comme système de contre sélection (Cho, Shin and Cho, 2018; Strich and Chertow, 2018).

Chez les bactéries du genre *Bacillus*, le système a pu être utilisé chez le modèle *B. subtilis* (Altenbuchner, 2016) via l'utilisation de la recombinaison homologue, mais pas en utilisant la voie de réparation NHEJ (Bernheim *et al.*, 2017), ce qui implique une efficacité peu élevée de l'outil d'ingénierie. Chez *B. smithii*, le système a également été utilisé en système de contre sélection via l'induction de la recombinaison entre 45-55 °C chez cette bactérie thermophile, puis l'induction de la contre sélection par Cas9 à 37 °C (Mougiakos *et al.*, 2017). Chez les bactéries pathogènes du genre *Clostridium*, ici aussi malgré une faible efficacité de recombinaison homologue, le système CRISPR-Cas9 a permis d'obtenir des délétions de gènes chez *C. beijerinckii* ainsi que des intégrations et des SNV (Y. Wang *et al.*, 2015; Wang *et al.*, 2016). Il en est de même pour certaines bactéries du genre *Corynebacterium* (Cho, Shin and Cho, 2018), *Streptomyces* (Cobb, Wang and Zhao, 2015) ou sur le modèle bactérien le plus utilisé, *E. coli* (Jiang, Bikard, *et al.*, 2013; Li *et al.*, 2015).

Les diverses opportunités fournies par l'outil CRISPR-Cas9 chez les bactéries ont permis des avancées importantes dans le domaine de l'ingénierie de voies métaboliques comme l'augmentation de la production de butanol par les clostridia (S. Wang *et al.*, 2017), la production d'acide aminobutyric chez *C. glutamicum* (Becker and Wittmann, 2015) ou de Biofuel (n-Butanol) dans *E. coli* (Kim *et al.*, 2017). Mais cela étend également les possibilités dans l'étude fondamentale des bactéries ainsi que la compréhension des interactions hôtes-pathogènes comme pour *Clostridium septicum* (Koike-Yusa *et al.*, 2014).

Comme énoncé précédemment, chez les bactéries, peu d'espèces bactériennes possèdent une efficacité de recombinaison suffisante pour utiliser facilement le système CRISPR-Cas9 à des fins d'ingénierie directe de génome. La contre sélection médiée par le système CRISPR-Cas9 chez les bactéries s'est développée, comme par exemple dans *Staphylococcus aureus* où les souches pathogènes et celles possédant une séquence de résistance aux antibiotiques, ont été éliminées par un ciblage précis de séquences les discriminants des autres souches (Bikard *et al.*, 2014).

Les limites rencontrées dans les différents systèmes biologiques lors des essais d'utilisation de l'outil original CRISPR-Cas9 ont mené à des innovations technologiques importantes que ce soit chez les eucaryotes ou chez les procaryotes. Chaque limite est constamment repoussée au fur et à mesure des recherches et des innovations dans le

domaine de l'ingénierie de génome. Et comme l'a dit Albert Einstein, « A problem without a solution is a poorly stated problem ». Les prochains points vont traiter des limites du système original et des nombreuses améliorations qui y ont été apportées.

2.4. Limites et expansion de la boîte à outils

L'utilisation d'un tel système pour faire de l'ingénierie de génome dans une cellule n'est pas sans inconvénient. Différents obstacles sont répertoriés, ils concernent les différentes grandes étapes pour l'action du système CRISPR, à savoir : son entrée dans la cellule, le clivage de l'ADN et la survie de la cellule.

2.4.1. Entrée du système dans la cellule

L'entrée dans la cellule du système CRISPR est une limite importante particulièrement pour les systèmes eucaryotes complexes, mais également pour les procaryotes. Deux stratégies distinctes sont généralement utilisées : l'apport direct d'un complexe ribonucléique (protéine effectrice et sgRNA) sous forme « prêt-à-l'emploi » ou l'insertion d'ADN codant le système. Chaque méthode possède ses avantages et inconvénients et nécessite des technologies différentes.

Chez les bactéries, il n'existe pas de façon très différente d'insérer le système dans la cellule. En général, le système CRISPR est apporté sous forme de plasmide répliquatif ou intégratif qui permet l'expression des différentes composantes. Pour les bactéries, l'électroporation, le choc thermique et la transformation au Poly Ethylene Glycol sont les méthodes plus courantes de transformation (Jiang, Bikard, *et al.*, 2013), mais l'utilisation de la conjugaison naturelle peut également être envisagée comme c'est le cas entre *E. coli* et *Salmonella enterica* (Hamilton *et al.*, 2019). La principale limite pour cette étape chez les bactéries est la possibilité d'avoir un protocole de transformation efficace ainsi que des vecteurs disponibles pour l'espèce ciblée.

Pour les organismes eucaryotes, les méthodes sont beaucoup plus variées. Cela peut être aussi réalisé sous forme nucléique par transformation (nucléofection, électroporation, transfection) dans des organismes unicellulaires simples comme *S. cerevisiae* ou en culture cellulaire (DiCarlo *et al.*, 2013; Mali *et al.*, 2013; Tsarmopoulos *et al.*, 2016). Il est également possible de délivrer la nucléoparticule Cas9-sgRNA par injection directe dans le zebrafish, *Caenorhabditis elegans* ou dans des cellules embryonnaires (Sander and Joung, 2014). Une limite à l'utilisation de l'ADN codant le système CRISPR chez les

eucaryotes est lié à l'introduction d'ADN exogène dans la cellule qui pourrait entraîner une recombinaison dans le génome, la délétion potentielle de gènes et une exposition longue au système de clivage, multipliant les risques d'événements « Off-target » (Zuris *et al.*, 2015). Il est également difficile de prévoir quelle quantité de protéine Cas9 sera produite avec l'introduction d'ADN (Lino *et al.*, 2018). L'utilisation d'ARNm posait encore récemment de gros problèmes de stabilité et d'immunogénicité et n'est pas encore utilisé. La base uridine a été identifiée comme étant le principale problème entraînant une immunogénicité et une simple modification chimique de la base a permis notamment les récents exploits liés aux recherches de Katalin Kariko sur les vaccins à ARNm (Karikó, 2021). Ceci pourrait dorénavant être à l'origine d'un nouveau mode de cargo *in vivo* d'ARNm codant pour le système CRISPR.

La production *in vitro* de complexes ribonucléoprotéiques pour l'édition de génome par CRISPR a été particulièrement développée et est considérée comme l'option la plus sécurisée, la plus spécifique et la plus précise (Zuris *et al.*, 2015). Des méthodes comme l'utilisation de nanoparticules lipidiques et liposomes ou des billes d'or sont possibles pour délivrer ces complexes ribonucléiques *in cellulo*, mais leurs efficacités *in vivo* et l'inflammation que cela provoque sont des inconvénients importants. Pour des applications *in vitro* mais également *in vivo* dans la souris, une nouvelle technologie a fait son apparition avec l'utilisation de lipides cationiques comme moyen de transport, pour une efficacité d'édition très importante *in vitro* après un seul traitement (Zuris *et al.*, 2015).

Mais de nos jours, c'est toutefois le mode de livraison par un vecteur viral de l'ADN codant le système qui est particulièrement utilisé chez les animaux. Il est actuellement le meilleur moyen pour des applications *in vivo*, de par son efficacité (Lino *et al.*, 2018). L'inconvénient est la limite de taille de 4,7 kb qui peut être empaquetée par un vecteur viral AAV (Virus Adéno-Associé) (Patrick D Hsu, Lander and Zhang, 2014; Lino *et al.*, 2018). Ceci empêche par exemple l'utilisation de la classique spCas9 par ce moyen de transport et c'est une des raisons qui ont poussé la recherche de plus petits orthologues de spCas9 tels que la Cas9 de *C. jejuni*, la Cas9 de *S. aureus* (Benjamin P. Kleinstiver *et al.*, 2015; Ibraheim *et al.*, 2018) ou encore l'ingénierie directe de Cas12f pour la création d'une version « mini » (Xu *et al.*, 2021).

Finalement, l'entrée dans la cellule va poser particulièrement des problèmes dans le cas des systèmes eucaryotes complexes et non des organismes unicellulaires. Des solutions sont développées continuellement avec l'identification de protéines Cas orthologues suffisamment petites pour être empaquetées et utilisées dans des vecteurs AAV par exemple.

2.4.2. Clivage de l'ADN

Une fois à l'intérieur de la cellule, le système CRISPR-Cas9 va agir et produire une cassure double brin de l'ADN sur une séquence précise de l'ADN. Plusieurs limites associées au clivage peuvent être répertoriées : (1) la cytotoxicité et les clivages off-target engendrés par le système CRISPR-Cas, (2) la limitation de choix de séquences cibles liée à la séquence PAM. Nous allons par la suite détailler ces problèmes et voir les solutions qui ont été apportées.

2.4.2.1. Toxicité cellulaire : Cytotoxicité et off-target

Chez les bactéries particulièrement, le problème de la cytotoxicité du système se pose. En effet, il a été montré que la surexpression de spCas9 chez certaines bactéries pouvait être cytotoxique et que cela pouvait être une barrière importante à son utilisation pour de l'édition de génome (Vento, Crook and Beisel, 2019). Cela pourrait être dû à des clivages non spécifiques de l'ADN, mais pas seulement puisque cette cytotoxicité peut également être observée sans sgRNA ou après inactivation des domaines catalytiques de l'endonucléase. De plus, chez *Corynebacterium glutamicum* par exemple, Cas9 seule est cytotoxique même en absence de gRNA (Jiang *et al.*, 2017). Les problèmes combinés de faibles efficacités de transformation et de cytotoxicité rendent ainsi le système SpCas9 inutilisable dans de nombreuses espèces. Utiliser une Cas9 dont l'utilisation des codons a été optimisée pour l'espèce bactérienne améliore l'efficacité du système, mais cela peut ne pas suffire (Li *et al.*, 2019). Comme démontré dans *Haloferax volcanii*, lorsque cela ne suffit toujours pas, les seules colonies qui sont obtenues hébergent un système CRISPR désactivé par mutation ponctuelle (Cas9) ou perte du tracrRNA ou dont la séquence protospacer cible a été éliminée ou modifiée (Vento, Crook and Beisel, 2019).

Un problème récurrent dans l'utilisation de l'outil CRISPR/Cas9 chez les procaryotes comme chez les eucaryotes est le problème des clivages off-target. Chez les bactéries ce problème de off-target, se traduit le plus souvent par la mort de la cellule. Cela pose tout

de même plus de problèmes chez les eucaryotes qui ont des génomes bien plus grands que celui des bactéries et qui ont des systèmes de réparations très efficaces. Cela met également en péril les nombreuses applications thérapeutiques *in vivo* qui sont envisagées par l'utilisation de la technologie CRISPR. Ces problèmes sont particulièrement dus à des similarités de séquence entre la cible et d'autres régions du chromosome. En effet, comme pour le système de défense naturelle, on sait que la séquence seed est essentielle à l'activation de l'interférence et que quelques mésappariements dans le reste du protospacer peuvent être tolérés par l'effecteur et entraîne tout de même le clivage (Szczelkun *et al.*, 2014). Tout comme pour le système de défense, la séquence seed semble être importante pour l'utilisation de CRISPR/Cas9 en termes d'outils dans d'autres organismes. Toutefois, la position et l'effet d'un mésappariement n'est pas toujours prédictible puisque certaines mutations en région distale du protospacer semblent parfois avoir des effets importants sur l'activité de Cas9 et a contrario certaines mutations en position proximale dans la séquence seed non (Fu *et al.*, 2013). Les cellules eucaryotes telles que les cellules humaines sont complexes et ont des génomes de l'ordre de 3,2 Milliards de pb. De nombreuses portions de séquences sont ressemblantes et il est décrit qu'avec la spCas9, 99,96 % des guides utilisables auraient de potentielles off-target avec une séquence PAM et seulement 0, 1 ou 2 nucléotides de différence avec la séquence guide (Fu *et al.*, 2013; Mali *et al.*, 2013). Certaines protéines Cas9 peuvent même tolérer jusqu'à 5 ou 6 différences entre la cible et son gRNA *in vivo* (Fu *et al.*, 2013; Tadic *et al.*, 2019), mais certaines substitutions à des positions précises ne vont pas avoir le même effet sur l'activité (Hsu *et al.*, 2013). A cela s'ajoute la possible reconnaissance de séquences PAM non canoniques pour cliver l'ADN, comme NAG pour spCas9 (Patrick D. Hsu, Lander and Zhang, 2014). Tout comme pour la cytotoxicité chez les bactéries, la concentration de Cas9 est un facteur très important dans les phénomènes de off-target (Patrick D. Hsu, Lander and Zhang, 2014). Les différences entre séquence cible et gRNA sont tolérées d'autant plus que la concentration de Cas9 est importante, ce qui entraîne de plus grands nombres de off-target (Hsu *et al.*, 2013; Pattanayak *et al.*, 2013).

Au vu du nombre de paramètres qui entrent en jeu, les clivages off-target peuvent être fréquents et très difficiles à prédire. Quelques outils bio-informatique tels que Cas-OFFinder (Bae, Park and Kim, 2014), sont développés afin de détecter des sites où le gRNA utilisé pourrait s'hybrider de manière plus ou moins spécifique. Mais ces outils ne

prennent en compte que les off-target médiées par une séquence spécifique du gRNA, et tout n'est pas encore décrypté concernant les systèmes CRISPR et leurs flexibilités. Afin de s'assurer du résultat, des solutions de contrôle des off-target ont été mises en place pour valider certaines expériences scientifiques. Par exemple, un système classique de complémentation d'un gène peut contribuer à la vérification du rôle d'un gène inactivé par CRISPR. L'utilisation systématique de différents gRNA qui auront logiquement différents Off-target peut également être envisagée. D'autres méthodes plus sensibles et basées sur le séquençage haut débit sont utilisées pour identifier des off-target à faible fréquence (Hsu *et al.*, 2013; Ran *et al.*, 2013; Y. Fu *et al.*, 2014).

Finalement pour pallier aux problèmes de cytotoxicité et de off-target chez les procaryotes comme chez les eucaryotes, différentes stratégies sont mises en œuvre. Cela va passer par le contrôle de la concentration du complexe sgRNA-Cas9, par exemple en privilégiant les approches d'entrée dans la cellule de la forme protéique et non nucléique du système, mais pas seulement. Le contrôle de l'expression du système dans le temps et l'espace s'est particulièrement développé.

2.4.2.2. Contrôle de l'expression dans le temps et dans l'espace

Le contrôle de l'expression de Cas9 peut se faire de manière classique par promoteur inductible à la tétracycline (Zhang *et al.*, 2019) ou à la doxycycline (Cao *et al.*, 2016), tout comme celle du sgRNA (Aubrey *et al.*, 2015). Mais les problèmes de fuite d'expression sont assez récurrents et ces méthodes de contrôle peuvent ne pas suffire. Le système Cre-lox a également été employé pour améliorer la répression, en induisant la recombinase Cre, le gRNA est par exemple activé (Chylinski *et al.*, 2019). Dans *B. subtilis*, l'utilisation d'un plasmide *OriC* sensible à la température a également permis le contrôle et l'utilisation de ce système de manière contrôlée (Altenbuchner, 2016), et dans *B. smithii*, la simple augmentation de la température à 42 °C permet d'inactiver spCas9 sans poser de problèmes de croissance à la bactérie (Mougiakos *et al.*, 2017). Malheureusement, de telles variations de température ne sont pas possibles pour de nombreux organismes. Une fois de plus, l'innovation technologique a pris le pas sur de nombreuses études et un certain nombre de systèmes sont maintenant disponibles pour le contrôle de Cas9 (Zhang *et al.*, 2019). Le domaine DHFR (Dihydrofolate reductase) est un domaine structural instable dérivé d'une protéine de *E. coli*. En ajoutant ce domaine à Cas9, la protéine est instable et cela empêche son action via l'élimination de la protéine par les processus

cellulaires naturels (Maji *et al.*, 2017). L'ajout d'une petite molécule, la TMP (Triméthoprime), stabilise le domaine et donc Cas9 et le système CRISPR est alors activé. Ce système a même été optimisé par une photo-activation de la TMP (Manna *et al.*, 2019). D'autres mécanismes plus complexes encore sont utilisés et permettent de réprimer le système au maximum. Une équipe a, par exemple, séparé la protéine Cas9 en deux sous unités qui vont se dimériser à l'aide de deux petites protéines en présence de rapamycine et une séquence NLS (pour l'adressage au noyau) a été ajoutée qu'à une des deux sous unités. Il existe bien d'autres systèmes qui sont répertoriés dans la littérature avec par exemple l'utilisation de peptides auto-excisables, ou le contrôle par de petits heptamères à ARN (Kundert *et al.*, 2019; Oakes *et al.*, 2019; Zhang *et al.*, 2019).

2.4.2.3. Limitation du nombre de cibles

La dernière limite importante liée à la partie clivage de l'ADN concerne la limitation du nombre de cibles par le motif PAM de spCas9 : NGG. Par exemple pour des bactéries comme les mycoplasmes qui possèdent des génomes riches en AT, un motif NGG est beaucoup moins fréquent que chez d'autres organismes. En ce qui concerne les applications thérapeutiques, le ciblage ultra spécifique est nécessaire dans l'optique de modifier certaines mutations dans l'ADN associées à des maladies génétiques.

En réponse à ce problème majeur, mais également pour répondre aux problèmes de taille de la spCas9 mais aussi de toxicité cellulaire et de off-target, un très grand nombre de recherches se sont développées autour de la production de variants de la spCas9 et la caractérisation de nouvelles protéines Cas9 et Cas12.

2.4.2.4. Utilisation de systèmes hétérologues.

Des équipes travaillant principalement sur les bactéries ont voulu utiliser des systèmes CRISPR endogènes à l'organisme sur lequel ils travaillaient pour limiter les problèmes de toxicité (Vento, Crook and Beisel, 2019). C'est par exemple le cas chez *Clostridium pasteurianum* avec l'emploi de la Cas9 endogène pour de l'ingénierie de génome (Pyne *et al.*, 2016) mais également chez d'autres organismes avec des systèmes à multiples effecteurs comme celui de type IE chez *E. coli* (Luo *et al.*, 2015). Outre l'emploi de système endogène, des différences selon le système hétérologue utilisé ont été observées, par exemple chez *Corynebacterium glutamicum*. Chez cet organisme, la Cas9 seule est cytotoxique alors que Cas12a a permis d'éditer le génome (Jiang *et al.*, 2017). Même si

l'essentiel des outils développés reposent sur les systèmes de Classe 2, on trouve aussi des exemples d'exploitation de systèmes endogènes de classe 1 comme chez *Pseudomonas aeruginosa* (Csörgő *et al.*, 2020).

Concernant les systèmes eucaryotes complexes, la découverte de spCas9 a poussé de nombreux chercheurs à se mettre en quête d'autres protéines Cas9 hétérologues, mais aussi de type Cas12, avec des atouts intéressants en termes de possibilités d'importation, de séquence PAM, de spécificité et d'efficacité. En effet, un an après la publication de la preuve de concept de spCas9, quatre Cas9 d'espèces différentes ont été testées et utilisées simultanément dans *E. coli*. En revanche, toutes, n'ont pas montré d'activité chez les cellules humaines (Esvelt *et al.*, 2013). Ceci a permis de démontrer que l'efficacité, mais également la spécificité, serait différente selon le système que l'on utilise. Cela a motivé la caractérisation fonctionnelle de nombreux systèmes provenant de nombreuses espèces de bactéries et archées différentes. La protéine Cas9 provenant de *S. aureus* (SaCas9) est rapidement devenu un candidat très intéressant, notamment de par sa plus petite taille que celle de spCas9 (1053 contre 1368 acides aminés) qui favorise son maniement *in vivo* à l'aide de vecteurs viraux de type AAV (Ran *et al.*, 2015). Mais elle a aussi montré une activité intéressante puisque contrairement à spCas9, SaCas9 est une enzyme à multiple turnover et qui clive *in vitro* beaucoup plus de substrat (Yourik *et al.*, 2019). C'est également le cas de celle de *Campylobacter jejuni* (CjeCas9), la plus petite Cas9 caractérisée, qui a été exploitée chez la souris à l'aide d'un vecteur AAV (Jo *et al.*, 2019). Ces deux Cas9 montraient également des séquences PAM bien distinctes de spCas9 NNGRRT pour SaCas9 et NNNRYAC pour CjeCas9. Avoir un grand nombre de systèmes fonctionnels implique d'avoir un plus grand choix pour l'utilisation de celui qui est le plus adapté.

Pour caractériser ces nouveaux systèmes, il a fallu en premier lieu décrypter la séquence PAM qui lui est nécessaire. Il existe différentes techniques qui sont couramment utilisées. La première est une méthode bio-informatique, avec l'extraction des spacers associés au système CRISPR, le BLAST des protospacers correspondants et l'analyse et l'alignement des séquences à proximité de ces protospacers. Cette méthode est limitée par le nombre de protospacers qui sont obtenus par BLAST ainsi que par la présence de séquence « faux positif » (Leenay *et al.*, 2017). Une méthode *in vitro* est particulièrement utilisée, elle consiste à la production de l'effecteur (Cas9 ou Cas12) et du gRNA qui vont servir *in vitro*

au clivage d'une banque de plasmides contenant la séquence cible suivie de 5 ou 7 nucléotides dégénérés. Après séquençage d'amplicons, cela permet d'identifier l'ensemble des 1024 ou 16 384 possibles PAM qui sont reconnues par le complexe ribonucléoprotéique (Karvelis *et al.*, 2015; Karvelis, Young and Siksnys, 2019). Cette méthode est simple d'utilisation, et elle a notamment permis à l'équipe du Professeur Syknys, de caractériser les séquences PAM de 79 orthologues Cas9 (Gasiunas *et al.*, 2020). Cette méthode nécessite une couverture de séquençage importante et n'est pas utilisable pour des systèmes complexes comme ceux de type I. Une méthode *in vivo* est également très employée, la méthode SCANR. Elle évalue la capacité d'un système CRISPR inactivé à réprimer ou non la GFP, le tri cellulaire par FACS sert ensuite à isoler les cellules ayant des séquences PAM fonctionnelles (Leenay *et al.*, 2017).

De très nombreux exemples d'utilisation d'autres systèmes Cas chez les procaryotes comme chez les eucaryotes sont disponibles. Chaque système a ses avantages et inconvénients, et le choix d'emploi de l'un d'eux va dépendre de ce que l'on veut faire avec ce système et dans quel organisme on va l'utiliser.

2.4.2.5. Evolution dirigée de l'effecteur

L'évolution en laboratoire de protéine Cas9 déjà caractérisée est une autre stratégie qui a été mise en place pour répondre rapidement aux problèmes de toxicité, de limitation de PAM, de spécificité et d'efficacité.

Chez les bactéries, par exemple chez *Bacillus licheniformis*, l'utilisation d'une Cas9 avec un des domaines catalytiques inactivé, appelé Cas9 nickase (Cas9n), a permis d'éliminer une partie de la toxicité de la Cas9 et d'augmenter l'efficacité d'édition de génome de manière importante (Li *et al.*, 2018). Il en est de même dans plusieurs espèces de *Lactobacillus* où la toxicité induite par Cas9 a été corrigée par l'utilisation d'une nickase pour de l'ingénierie de génome (Song *et al.*, 2017; Goh and Barrangou, 2021).

Pour modifier la spécificité de séquence PAM, certains groupes ont utilisé une méthode de greffe de domaine PI de la Cas9 en remplaçant celui de la spCas9 par celui de la Cas9 de *Streptococcus macacae* par exemple afin de donner un motif de reconnaissance de type NAAN (Ma *et al.*, 2019; Chatterjee *et al.*, 2020). D'autres systèmes encore plus élaborés ont été mis en place, via des systèmes de banque de Cas9 mutagénisés au niveau du domaine PI, suivi d'un système de sélection positive dans *E. coli*. Cela a permis de produire

de nombreux variants de spCas9 avec des PAM différentes telles que le variant VQR avec une séquence PAM de type NGA (Benjamin P. Kleinstiver *et al.*, 2015), mais également des variants de SaCas9 avec comme séquence PAM, NNNRRT vs NNGRRT préalablement, tout en conservant un nombre de off-target similaire chez les cellules humaines (Benjamin P Kleinstiver *et al.*, 2015).

Pour améliorer la spécificité en diminuant le nombre de off-target, la même stratégie a été employée avec des mutagenèses sur l'ensemble de la Cas9 afin d'identifier des variants intéressants. Un large panel de mutations intéressantes a été identifié, telles que les variants Cas9_R63A/Q768A, qui montrent une meilleure spécificité dans les cellules humaines via une augmentation de la sensibilité au mésappariement et une réduction de la sensibilité de la Cas9 à la PAM (Bratovič *et al.*, 2020). C'est alors que de nombreuses études ont tenté d'améliorer la spécificité du motif PAM, l'efficacité de clivage et la spécificité du système via la diminution des off-target (Hu *et al.*, 2018; Lee *et al.*, 2018; Babu *et al.*, 2019).

Pour résumer, l'entrée dans la cellule, la toxicité intrinsèque du système, les off-target qu'il génère, la spécificité de PAM et l'efficacité de clivage dans l'organisme ciblé, sont les principales limites du système. Le développement de nouvelles technologies de transformation, de transport, de contrôle du système et le développement d'un très grand panel de variants de Cas9 et d'orthologues, sont des réponses plus ou moins efficaces qui sont apportées par la communauté scientifique pour pallier à toutes ces limites. Avoir le choix simplifie grandement le travail des recherches à venir et cela ouvrira le champ des possibilités afin de sélectionner un système qui prenne en compte différents paramètres comme l'organisme auquel on s'intéresse et le but de l'ingénierie.

2.4.3. Réparation de l'ADN : voies et alternatives

L'utilisation de systèmes CRISPR permet d'avoir une coupure très précise de l'ADN, mais afin de faire de l'édition de génome, un système de réparation de l'ADN est nécessaire. Comme évoqué précédemment, le clivage de l'ADN double brin entraîne de la toxicité cellulaire si cette coupure n'est pas réparée. La plupart des systèmes eucaryotes ainsi que quelques bactéries vont alors s'appuyer sur leur système de réparation qui est suffisamment efficace pour survivre. Mais ce n'est pas le cas de tous les systèmes biologiques ou la réparation naturelle ne suffira pas. Nous arrivons sur la limite la plus importante qui est la réparation de l'ADN. Pour répondre à ce problème, différentes voies

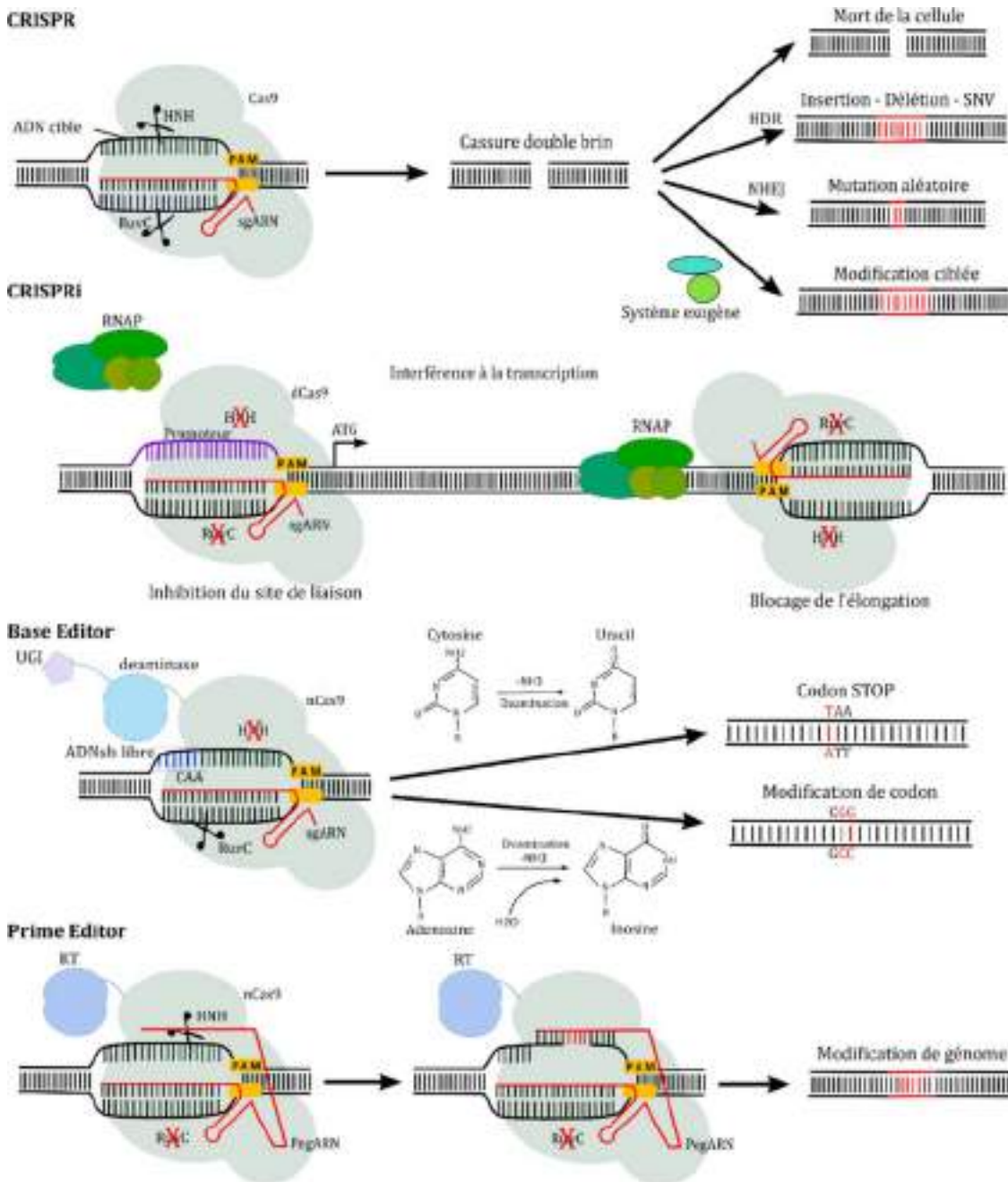


Figure 16 : Les outils CRISPR et leurs dérivés. Représentation schématique des principaux outils utilisés par la communauté scientifique : le CRISPR classique, le CRISPRi, le CRISPR Base Editor, le CRISPR Prime Editor. Les domaines catalytiques de la Cas9 HNH et RuvC peuvent être représentés barrés d'une croix rouge afin de décrire l'inactivation du domaine. Les mutations de l'ADN engendrées par le système sont représentées en rouge. Le promoteur du gène théorique est représenté en violet, HDR = Homologous Direct Repair, représente la voie de réparation homologue de l'ADN, NHEJ= Non-Homologous End Joining, représente la voie de réparation non homologue de l'ADN. RNAP= ARN polymérase. UGI= Uracil Glycosyltransferase Inhibitor. RT= reverse transcriptase.

ont été explorées et les différentes alternatives qui sont utilisées sont présentées ici (Figure 16).

2.4.3.1. Voies endogènes de réparation

Dans la nature, il existe plusieurs voies de réparation de l'ADN mais deux voies sont principalement utilisées (Figure 16). La réparation homologue de l'ADN (HDR : homologous directed repair), en présence d'une matrice de recombinaison (ADN modèle), la machinerie de réparation va s'inspirer de celui-ci pour restaurer de manière homologue au modèle. Si aucun modèle n'est disponible, c'est la réparation NHEJ (Non-homologous end joining) qui va permettre une maintenance rapide mais moins précise de l'ADN (M. Liu *et al.*, 2019). Cette dernière va parfois entraîner des mutations au niveau de la cassure.

Chez la plupart des eucaryotes, les deux voies sont très efficaces et l'utilisation de l'une ou l'autre de ces voies après clivage par un système CRISPR va dépendre du but de l'expérience. Si une modification précise est désirée, telle qu'une insertion, délétion ou substitution, la réparation HDR va être induite en fournissant un modèle d'ADN avec des bras homologue au niveau de la coupure et l'efficacité peut être améliorée en inhibant la réparation NHEJ (M. Liu *et al.*, 2019). Pour des applications d'inactivation de gènes, il est également possible d'utiliser la voie NHEJ qui va alors ajouter ou éliminer des nucléotides de manière aléatoire et possiblement inactiver le gène par décalage du cadre de lecture. L'avantage à avoir ces deux voies de réparation est la facilité de pouvoir réaliser des modifications précises de l'ADN, mais c'est aussi un inconvénient puisque contrairement aux bactéries qui vont mourir le plus souvent (Vento, Crook and Beisel, 2019), une coupure de type off-target, sera également le plus souvent réparée et possiblement entraîner des mutations non souhaitées à un autre endroit du génome.

Chez les bactéries, les voies de réparations de l'ADN ne sont pas aussi nombreuses que chez les eucaryotes, mais il existe tout de même la recombinaison homologue médiée principalement par la protéine RecA et la machinerie RecBCD à partir d'un ADN modèle, (Del Val *et al.*, 2019). La liaison des extrémités de l'ADN ciblé de manière aléatoire et non homologue est également disponible chez certaines espèces et se fait par la voie de réparation NHEJ ou MMEJ (Microhomology-mediated end joining) (Finger-Bou *et al.*, 2020). Les protéines principalement responsables de la voie NHEJ sont les protéines Ku et LigD, et, pour la voie MMEJ, le complexe RecBCD (tout comme la recombinaison homologue, mais associé à LigA et non RecA). L'utilisation de ces voies endogènes de

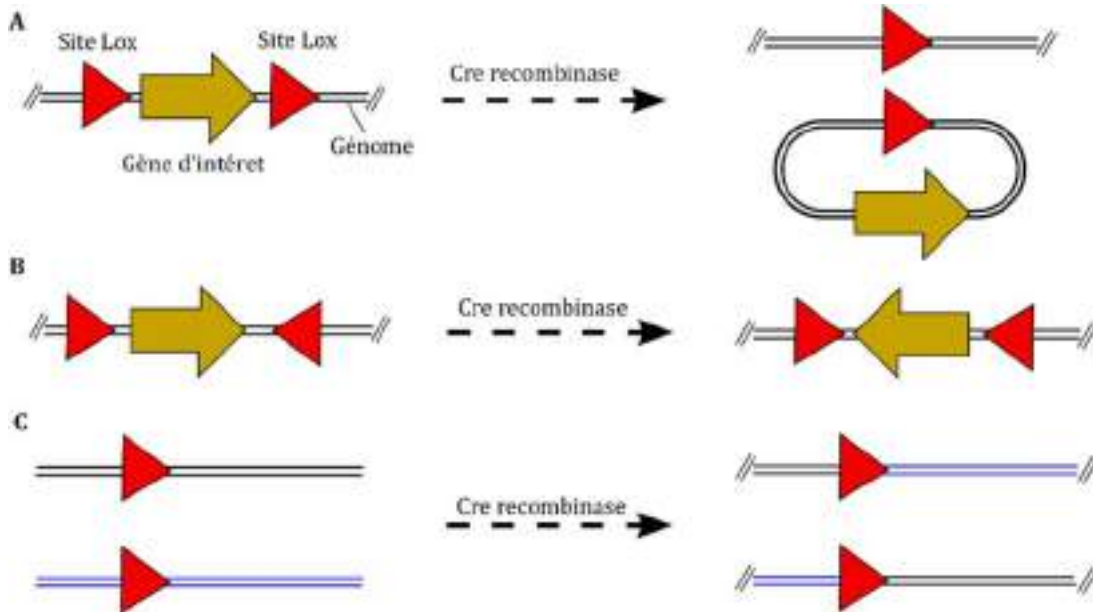


Figure 17 : Représentation des opportunités d'ingénierie par le système Cre-lox. La disposition des sites lox utilisés va permettre d'arriver à différents résultats qui sont représentés ici.

réparations pour de l'ingénierie de génome en combinaison avec l'utilisation du système CRISPR ont plusieurs fois été mis en œuvre (Finger-Bou *et al.*, 2020), mais leur présence ou leurs efficacités ne suffisent parfois pas pour l'obtention de mutants en fonction des espèces, voire même des souches (Vento, Crook and Beisel, 2019). L'importation de systèmes hétérologues bactériens a également été testée et peut parfois fonctionner via l'importation de gènes codant une RecA efficace ou des gènes impliqués dans la voie NHEJ. Ceci a particulièrement fonctionné pour *E. coli* où de nombreuses données et connaissances sont disponibles (Wright *et al.*, 2017; Zheng *et al.*, 2017; Finger-Bou *et al.*, 2020).

Finalement, les voies endogènes de réparation de l'ADN sont couramment utilisées chez les organismes où ces voies fonctionnent et sont efficaces. Mais chez de nombreuses bactéries, ce n'est pas le cas. Les protéines virales comme le système Cre-lox, sont connues pour être fonctionnelles dans de nombreux hôtes, et sont en général très efficaces. De nombreuses équipes se sont ainsi intéressées à l'importation de systèmes de recombinaisons exogènes pour réaliser de l'ingénierie de génome dans des bactéries récalcitrantes.

2.4.3.2. Systèmes exogènes phagiques

Depuis quelques années, de nombreux systèmes exogènes d'origine phagique le plus souvent, sont utilisés dans les bactéries ne possédant pas une efficacité de réparation suffisante pour faire de l'ingénierie de génome. Ils peuvent être couplés ou non au système CRISPR et ce sont de puissants outils pour l'ingénierie de génome (Figure 17 et 18).

2.4.3.2.1. Systèmes SSR

Les recombinaisons à site spécifique (SSR), sont une classe de protéines qui ont pour fonction la recombinaison entre deux sites spécifiques. Le système le plus connu est celui de Cre-lox qui appartient à la classe des tyrosines recombinase, mais il en existe d'autres tels que les sérines recombinases (Xu and Brown, 2016). Ces systèmes proviennent de phages et sont très présents dans la nature, le système Cre-lox, par exemple, provenant du bactériophage P1 est capable d'infecter différentes espèces bactériennes. Ce système repose sur la protéine Cre, qui possède une fonction recombinase, et sur des paires de sites lox qui constituent les points de recombinaison (Figure 17). Ces sites sont composés de 34 paires de bases dont 8 paires de bases centrales homologues, flanquées par

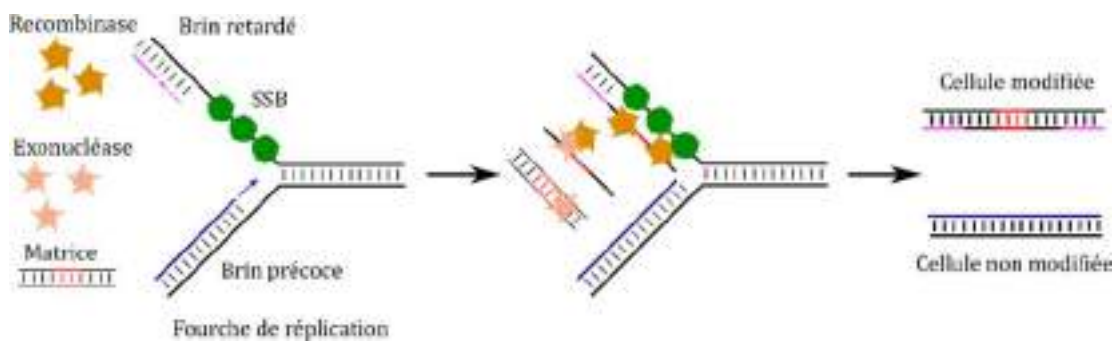


Figure 18 : Représentation d'un modèle d'ingénierie génétique médié par le système par RecET.

13 paires de bases inversées entre les deux sites (Shaw, Serrano and Lluch-Senar, 2021). Dans un contexte naturel, le système Cre-lox permet aux phages de s'insérer et de s'exciser du génome bactérien. Comme outils d'ingénierie chez les bactéries, ils sont particulièrement utilisés comme système d'élimination de marqueurs de résistance aux antibiotiques. En effet, en disposant des sites lox autour du marqueur de résistance à l'antibiotique, après ingénierie de génome, il est possible d'exciser et éliminer la séquence codante du marqueur. En combinaison avec le système CRISPR, ces systèmes peuvent être utilisés pour la modification d'une bactérie suivie de la contre sélection par le système CRISPR, mais cela implique la possibilité préalable de pouvoir modifier le génome pour y ajouter des sites lox. Il est possible d'utiliser ces systèmes pour la capture précise de grands fragments d'ADN bactériens dans *E. coli* (Enghiad *et al.*, 2021).

Finalement, les recombinases à site spécifique sont un point de départ des recherches de systèmes exogènes, mais ce sont d'autres systèmes qui ont plus largement intéressé les microbiologistes, les systèmes de recombinaison phagiques tels que le système RecET.

2.4.3.2.2. Systèmes RecET et lambda Red

Les « Phage-encoded ssDNA annealing proteins » (SSAPs) sont une nouvelle génération d'outils très puissants pour l'ingénierie de génome de bactéries manquant d'une voie de réparation de l'ADN efficace (Figure 18). Parmi les plus populaires et les plus anciens connus (années 1990), le système lambda red et le système RecT proviennent respectivement du phage lambda et du prophage Rac (Hall and Kolodner, 1994; Thresher *et al.*, 1995). Ces systèmes sont liés évolutivement et définissent une large famille de protéines pour la recombinaison de l'ADN (Iyer, Koonin and Aravind, 2002). Ces systèmes sont composés de différents partenaires protéiques plus ou moins essentiels. La protéine recombinase (Rec-T ou Red- β pour les deux systèmes cités), est la protéine la plus importante du système. Seule, elle a pour fonction l'intégration d'un ADN simple brin (ADNsb) dans le génome bactérien via trois mécanismes indépendants de RecA qui ont été plus ou moins précisément décrits : (1) l'incorporation de l'ADNsb comme fragment d'Okazaki pendant la réplication, (2) par liaison de brin ou (3) par invasion de brins (Hall and Kolodner, 1994; Poteete, 2008; Thomason, Costantino and Court, 2016). La protéine recombinase utilisée semble être hôte spécifique et l'utilisation d'un système endogène ou proche phylogénétiquement semble montrer une plus grande efficacité (Datta *et al.*, 2008; Sun *et al.*, 2015). Une des explications, mais pas la seule, serait en lien avec une

protéine qui lie l'ADN simple brin (SSB : single strand DNA binding) et qui est souvent associée au système. Celle-ci serait reconnue par RecT en C-terminale ce qui faciliterait l'interaction entre l'ADNsb et la recombinaise. L'ajout d'une SSB dans certains cas améliorerait de 1000 fois l'efficacité de recombinaison (Filsinger *et al.*, 2021). Outre la recombinaise seule, le système peut être accompagné d'un partenaire de type exonucléase (RecE ou Exo) digérant les extrémités de l'ADN à partir d'une matrice d'ADNdb afin de générer des fragments simple brin pour l'action de la recombinaise (Murphy, 2016; Thomason, Costantino and Court, 2016).

Depuis quelques années, les systèmes SSAPs proches de RecT et Lambda Red sont largement étudiés et utilisés pour de l'ingénierie de génomes bactériens comme chez *Bacillus subtilis* (Sun *et al.*, 2015), *Lactobacillus* (Yang, Wang and Qi, 2015; Huang, Song and Yang, 2019), *Acinetobacter* (Tucker *et al.*, 2019), *Shewanella* (Corts *et al.*, 2019; Z. Y. Wu *et al.*, 2019), *Pseudomonas* (Swingle *et al.*, 2010; Yin *et al.*, 2019), *Photorhabdus* (Yin *et al.*, 2015) ou *Corynebacterium* (Huang *et al.*, 2017).

Ces systèmes peuvent également être utilisés comme alternative à la recombinaison homologue endogène pour réparer une coupure effectuée avec le système CRISPR. Cela augmente l'efficacité de la recombinaison dans certains procaryotes ou sert à la contre sélection des événements de recombinaison (Huang, Song and Yang, 2019; Piñero-Lambea *et al.*, 2020; Zhao *et al.*, 2020).

2.4.3.3. Outils dérivés : de puissantes alternatives

Les systèmes d'édition CRISPR classiques sont tous basés sur le clivage de l'ADN et la réparation. Précédemment, les principales barrières à l'utilisation de cette technologie ont été énumérées et des solutions diverses ont été apportées dans de nombreux organismes. Malheureusement, pour certains organismes récalcitrants ou pour des applications nécessitant une extrême précision (telles que la modification *in vivo* à des fins thérapeutiques chez l'homme), cela ne suffit pas. Des outils dérivés des systèmes CRISPR sont apparus et la technologie ne cesse de progresser (Figure 16).

2.4.3.3.1. Modulation de la transcription

La première innovation est l'utilisation d'une Cas9 inactivée, ou « dead » (dCas9), par la mutation de ses deux domaines catalytiques HNH et RuvC (et non plus un seul comme pour la Cas9n) (Jinek *et al.*, 2012). Cette dCas9 possède toujours ses capacités de ciblage

précis de l'ADN et va pouvoir reconnaître l'ADN, se lier et produire une R-loop classique. En revanche, elle ne va plus pouvoir couper l'ADN et va rester fixée à cet ADN. C'est notamment grâce à l'inactivation de ces domaines que toutes les structures cristallines de la R-loop de l'ADN avec les différents complexes Cas9-sgRNA, ont pu être résolues (Jiang *et al.*, 2016). L'activité intéressante que va produire ce complexe stable est une activité d'inhibition de la transcription, comme cela a été rapidement montré dans *E. coli* (Bikard *et al.*, 2013; Qi *et al.*, 2013). Cet outil est parfaitement complémentaire au système classique puisqu'il résout à lui seul le problème de réparation de l'ADN et une partie du problème de toxicité cellulaire. Des analyses phénotypiques sont alors possibles dans des organismes où le système classique ne fonctionne pas et ceci particulièrement chez les procaryotes. De plus, l'interférence à la transcription permet de moduler des gènes essentiels tandis qu'une coupure va simplement éliminer la cellule (Peters *et al.*, 2016). Ce nouvel outil a été mis en place chez *Mycobacterium segmentis* (Choudhary *et al.*, 2015), *Staphylococcus aureus* (Dong *et al.*, 2017; Zhao, Shu and Sun, 2017), *Bacillus subtilis* (Peters *et al.*, 2016) et de nombreux autres systèmes, y compris eucaryotes chez *S. cerevisiae* ou dans des cellules humaines (Gilbert *et al.*, 2013; Qi *et al.*, 2013). L'activité d'interférence à la transcription est médiée par des fixations sur deux types de sites (Figure 16) : (1) le promoteur d'un gène où la formation d'une R-loop stable induite par dCas9 empêche le complexe RNAP de se lier ; (2) une séquence interne au CDS, ce qui va bloquer le complexe et l'élongation de la transcription (Bikard *et al.*, 2013). Basée sur l'utilisation de la dCas9 de *S. pyogenes* et d'un seul guide, la technique a profité de l'essor des caractérisations de diverses Cas9, en combinant par exemple plusieurs systèmes différents pour faire de la modulation de voie métabolique dans *E. coli* (Sung *et al.*, 2019). Des systèmes à partir de Cas12 (Zhang *et al.*, 2017) ou même à partir des protéines provenant du système CRISPR de type I se sont également développés (Luo *et al.*, 2015). Pour ces systèmes, quelques limites sont identifiées, que ce soit les mêmes que pour le système classique avec des problèmes de off-target ou de nombre de séquence cible limité par la PAM, ou d'autres. Il a été montré que la présence de mésappariements (jusqu'à 11) n'est pas un frein à la répression chez *E. coli* (Cui *et al.*, 2018) ; cela peut entraîner des problèmes de off-target évidents. De plus, une nouvelle difficulté a été mise en évidence, nommée « bad seed effect ». En utilisant une librairie de guides et une approche d'analyses informatiques poussée, il a été montré que certains guides possédant communément une signature particulière de 5 nucléotides en séquence seed peuvent entraîner des défauts

de croissance importants (Cui *et al.*, 2018). L'origine de cette toxicité n'est pour le moment pas décryptée et il ne semble pas qu'elle soit forcément liée à des interactions off-target sur des gènes essentiels (Vigouroux and Bikard, 2020). Pour les systèmes de types VI ciblant l'ARN, il est également possible d'utiliser Cas13 pour cibler et cliver un ARN cible et réaliser une interférence (Omar O. Abudayyeh *et al.*, 2016; Abudayyeh *et al.*, 2017). Une revue décrivant comment réaliser de l'interférence à CRISPR chez les bactéries a récemment été écrite ; elle liste un certain nombre de clés pour optimiser la stratégie (Depardieu and Bikard, 2019).

De la même manière, principalement chez les eucaryotes dans lesquels les protéines activatrices de la transcription sont connues, des systèmes dCas9 de régulation positive de la transcription ont été produits. Ceci combine l'utilisation d'une dCas9 qui est fusionnée à des protéines ou des domaines activateurs de la transcription. En ciblant ainsi la région en amont du promoteur, il a été démontré la possibilité de réguler de manière positive la transcription (Cheng *et al.*, 2013; Gilbert *et al.*, 2013; Perez-Pinera *et al.*, 2013). La protéine VP16 du virus de l'herpès possède un domaine activateur de la transcription d'un gène (Wysocka and Herr, 2003). L'utilisation de 3 copies fusionnées à la dCas9 a permis la construction de la dCas9-VP48 et de celle de 4 copies, la dCas9-VP64, pour l'activation de la transcription dans *S. cerevisiae* et dans les cellules humaines (Cheng *et al.*, 2013; Gilbert *et al.*, 2013; Perez-Pinera *et al.*, 2013). Chez les bactéries, peu de domaines de régulation sont connus mais un domaine synthétique SoxS, a récemment été mis en valeur pour cette utilisation dans *E. coli* (Dong *et al.*, 2018). Ce domaine a montré une activité particulièrement intéressante sur les promoteurs de la famille σ 70 et l'utilisation d'une Cas9 avec une PAM modifiée (dxCas9), a permis de développer sensiblement ces possibilités d'applications (Fontana *et al.*, 2020). D'autres domaines d'activation ont finalement été décrits comme fonctionnels et ont permis d'accroître les possibilités comme la sous-unité p65 du facteur de transcription NF-kappaB, Rta domaine, etc (Tadic *et al.*, 2019).

2.4.3.3.2. Systèmes Base Editor (BE) : mutagénèse dirigée

Les systèmes dérivés de CRISPR parmi les plus utilisés de nos jours sont les systèmes d'édition de base (BEs : Base Editor systems) (Figure 16). À l'origine, ils ont été mis en valeur en 2016 par deux équipes distinctes, celle de David Liu (USA) et celle de Akihiko Kondo (Japon) (Komor *et al.*, 2016; Nishida *et al.*, 2016). Les BEs combinent la précision

des systèmes CRISPR-Cas9 avec la fonction d'une protéine désaminase capable d'agir sur l'ADN simple brin. Ils sont indépendants de tout système de réparation de l'ADN puisqu'ils n'utilisent pas la fonction endonucléase de la Cas9. En effet, une dCas9 est utilisée pour cibler grâce au sgRNA, un endroit précis de l'ADN, créant une R-loop sur une vingtaine de paires de bases. Cette conformation de l'ADN est favorable à l'action de la désaminase fusionnée à la dCas9 qui va agir sur l'ADN simple brin libre pour effectuer l'édition de base. Les premières désaminases utilisées ont été de type cytosine désaminase (rAPOBEC1, hAID, pmcDA1) et entraîne le changement d'une base cytosine en uracile qui, lors de la réplication, est reconnue en thymine. En ciblant le brin sens et antisens, il est alors possible de manière programmée, de modifier les cytosines en thymines et les guanines en adénines. Les systèmes d'édition de base de type cytosine (CBEs) se sont développés puisqu'ils se sont montrés moins toxiques que le CRISPR dit classique, avec moins de problèmes de off-target, d'insertions et de délétions tout en étant très efficaces que ce soit chez les eucaryotes ou chez les procaryotes (Komor *et al.*, 2016; Anzalone, Koblan and Liu, 2020). Outre ces trois points positifs, et comme expliqué précédemment, peu de bactéries possèdent des systèmes de réparation efficaces de l'ADN. Le développement d'un tel outil a donc ouvert des portes pour la génomique fonctionnelle de nombreuses espèces bactériennes comme *Klebsiella* (Y. Wang *et al.*, 2018), *Pseudomonas* (Sun *et al.*, 2020), *Clostridium* (Li *et al.*, 2019), *Streptomyces* (Tong *et al.*, 2019) ou *Agrobacterium* (Rodrigues *et al.*, 2021).

Outre les CBE, d'autres systèmes d'édition de base ont ensuite vu le jour avec notamment l'utilisation d'une adénosine désaminase (système ABE) développée dans le groupe de David Liu ouvrant la possibilité de changer cette fois-ci, une adénine en guanine et une thymine en cytosine sur le brin inverse (Gaudelli *et al.*, 2017). Cette fois-ci contrairement aux CBEs où la cytidine désaminase utilisée modifie l'ADN naturellement, il a fallu passer par de l'évolution forcée sur une adénosine désaminase provenant de *E. coli* (TadA). La fonction naturelle de TadA de modification de l'ARN a évolué vers la modification de l'ADN simple brin pour en faire un BEs intéressant.

Les ABEs et CBEs ont rapidement soulevé un fort intérêt dans la communauté scientifique et de nombreuses améliorations et dérivés ont été obtenus. Une limite de ces systèmes se situe au niveau du choix des cibles et de la fenêtre d'action (ou fenêtre d'édition). En effet, dans les premiers ABEs et CBEs, la possibilité de modifier des bases est limitée à quelques

nucléotides sur l'ADN simple brin dans la R-loop qui ne sont pas encombrés de manière stérique par la dCas9. En ajoutant à cela la présence d'une séquence PAM qui doit être coordonnée à ces quelques nucléotides libres, il devient parfois compliqué d'avoir toutes les conditions nécessaires pour modifier à souhait une cible précise. Ainsi, à l'aide de variants ou homologues de Cas9, la spécificité de Cas9 pour la liaison sur des sites d'off-targets, la spécificité de la séquence PAM et la position de la fenêtre d'action ont pu être modifiés selon la Cas9 utilisée. Par exemple, pour des génomes riches en AT, le variant VQR de la SpCas9 pourrait être une option très intéressante avec une séquence PAM de type NGA versus NGG pour l'originale. Plusieurs dizaines de désaminases et de variants ou homologues de Cas9 ont été décrits dans la littérature ce qui laisse maintenant un choix important aux utilisateurs. (Evanoff and Komor, 2019; Huang, Newby and Liu, 2021)).

Pour aller toujours plus loin, un double système ABE-CBE appelé SPACE a également été produit pouvant générer les deux mutations et augmenter les possibilités de modifications (Grünewald *et al.*, 2020).

Les BEs étant limités aux simples modifications pyrimidines-pyrimidines ou purines-purines, d'autres systèmes ont vu le jour toujours dans l'optique de décupler les possibilités tout en conservant l'avantage d'une toxicité réduite, d'une spécificité augmentée et d'une grande efficacité. Le « prime editing » a été développé, pour le moment uniquement chez les eucaryotes (Figure 16). Ce n'est plus alors une protéine désaminase, mais une transcriptase reverse qui est fusionnée à une Cas9 nickase. L'ARN guide va non seulement cibler un endroit du génome, mais il va également posséder une extension artificielle pour devenir un ARN prime editing guide (pegRNA). Cette extension est composée de quelques nucléotides d'ARN, complémentaires au brin ciblé par la Cas9 nickase, suivi d'un fragment possédant la modification désirée qui peut être de tous types (insertion, délétion, transversion) et de la fin de la séquence cible. Ainsi la reverse transcriptase va synthétiser, à partir du pegRNA, la modification directement dans le brin d'ADN de la cellule (Anzalone *et al.*, 2019).

2.4.3.3.3. Autres outils d'application

D'autres applications plus complexes utilisant les propriétés de ciblage de Cas9 ont été développées, comme la modulation de la chromatine par la fusion d'un domaine PYL à la dCas9 (Chen *et al.*, 2017) ou d'autres modifications épigénétiques (Tadic *et al.*, 2019) ; l'immunoprécipitation de la chromatine (enChIP) utilisant différents orthologues de

Cas9 afin de comprendre les interactions moléculaires à certains endroits de l'ADN (Fujita, Yuno and Fujii, 2018) ; l'imagerie en direct du génome avec un Tag fluorescent fusionné à la dCas9 (Chen *et al.*, 2013) ; l'utilisation de banques de gRNA pour le criblage de gènes impliqués dans une très grande variété de processus biologiques (Koike-Yusa *et al.*, 2014; Shalem *et al.*, 2014) comme par exemple pour l'identification du récepteur d'entrée ACE2 au SARS-Cov-2 dans les cellules alvéolaires humaines (Daniloski *et al.*, 2021).

2.5. Applications et éthiques : quelles limites ?

À partir de toutes les connaissances fondamentales sur les systèmes CRISPR, des innovations biotechnologiques et médicales très concrètes ont vu le jour notamment avec des systèmes CRISPR inactivés (Vigouroux and Bikard, 2020). Des outils très utiles de diagnostic ont, par exemple, été développés, comme pour le diagnostic du virus Zika (Pardee *et al.*, 2016), la détection de bactéries résistantes aux antibiotiques telles que *Staphylococcus aureus* avec une technique de fluorescence *in situ* de la dCas9 (Guk *et al.*, 2017), l'utilisation de Cas13a pour détecter le virus Zika, génotyper l'ADN humain ou identifier des mutations liées aux tumeurs (Gootenberg *et al.*, 2017). Plus récemment et en lien avec l'actualité, la détection en 20 minutes du virus SARS-CoV-2, en utilisant Cas13 et Csm6 comme senseurs, a été développée de manière très rapide par une collaboration de nombreux chercheurs (Liu *et al.*, 2021). De nombreuses applications sont nées par la découverte de ces systèmes de défense bactériens, la liste des applications peut être très longue et de nombreuses autres feront leurs apparitions dans les prochaines années.

Des utilisations encore plus ambitieuses sont également en cours de développement et font l'objet de nombreuses recherches industrielles. A l'image de Félix Hubert D'Hérelle fondateur de l'idée de la phagothérapie dans le milieu des années 1900, il est imaginé qu'avec les systèmes CRISPR, des applications thérapeutiques pourraient voir le jour. Par exemple, pour lutter contre les maladies infectieuses et la résistance aux antibiotiques voire même contre les espèces vectrices de maladies. D'autres avancées médicales sont également à l'œuvre avec les nombreuses démonstrations dans des cellules humaines qui ont montré des preuves de concept fortes telles que l'élimination d'un chromosome 21 surnuméraire dans des cellules pluripotentes (Zuo *et al.*, 2017) mais également *in vivo* chez des embryons de souris. Le système de clivage a également été utilisé pour l'élimination de provirus persistants tels que le HIV-1 dans une souris humanisée

souffrant d'une infection chronique (Yin *et al.*, 2017). Une prouesse similaire a été réalisée pour supprimer une pathologie génétique de la rétine dans un modèle de souris, avec aucun indel (insertions-délétions) détecté dans les sites potentiels de off-target après 14 mois (Jo *et al.*, 2019). Plus récemment encore, il a été montré que par l'utilisation d'un système base editor, il a été possible de corriger la mutation qui est à l'origine du Syndrome de Hutchinson-Guilford (ou Progeria) dans un modèle de souris. Une simple injection a permis d'avoir des tissus sains avec 90 % de correction de la mutation dans les fibroblastes et une durée de vie 2,5 fois supérieure à la normale (Koblan *et al.*, 2021).

Tous ces succès donnent de l'espoir pour l'apparition de « traitements » novateurs pour de nombreuses maladies génétiques. Mais comme dirait Winston Churchill (Et Oncle Ben de spider man), « Where there is great power, there is great responsibility ». Nous arrivons aujourd'hui à la croisée des chemins de l'éthique scientifique et sociétale pour répondre ou non aux enjeux majeurs de ce monde via l'utilisation de la modification génétique. La génétique de l'homme et les possibilités de connaître voire de modifier son génome sont apparues depuis bien longtemps comme des questions cruciales de société, dans des œuvres littéraires et cinématographiques (« Le meilleur des mondes » de Aldous Huxley, « Bienvenue à Gattaca » de Andrew Niccol), mais aussi dans l'histoire de l'humanité, avec l'idéologie de l'Eugénisme de Francis Galton (1822-1911) et les événements dramatiques qui en découlent jusqu'à la Seconde Guerre mondiale, ou avec l'émergence de pouvoir modifier ou même « designer » un bébé (Agar, 1995). Mais l'apparition des outils tels que CRISPR, Base-Editor, Prime-Editor, a permis d'obtenir, en quelques années seulement, la technologie nécessaire pour imaginer de nombreuses applications. Avec les très nombreuses connaissances accumulées sur les maladies génétiques, beaucoup de scientifiques et personnes atteintes de ces maladies peuvent espérer. Mais il peut alors en ressortir des essais cliniques prématurés, éthiquement non recevables et illégaux tels qu'on l'a vu avec la naissance de jumelles en 2018, dont les embryons ont été édités par CRISPR. Actuellement les législations diffèrent selon les pays, mais de manière globale, l'édition de cellules somatiques humaines est autorisée, car ces mutations ne pourront être transmises à la descendance. Cela a permis notamment par des méthodes *ex vivo*, de modifier une cellule hématopoïétique de patients atteints de drépanocytose (Sickle Cell Disease), puis de la réimplanter après ablation de la moelle osseuse. Plus d'un an après, les deux patients présentent toujours un niveau important d'allèles édités (Frangoul *et al.*, 2020). En revanche, la modification génétique sur des

cellules germinales est plus sujette à débat, car l'on ne veut pas répandre des fonctions non naturelles chez l'homme sans savoir ce qu'il pourrait en découdre. Ceci est autorisé par certains pays pour des modifications sur les plantes et animaux, mais pour ce qui est de l'embryon humain, cela peut être fait seulement sur des embryons non viables ou de moins de 14 jours, sans autorisation possible de réimplantation de ces embryons pour faire naître un embryon modifié (Doudna, 2020). De nombreuses publications font l'objet des questions éthiques soulevées par l'utilisation des outils CRISPR (Doudna, 2020; Collins *et al.*, 2021; Piergentili *et al.*, 2021) et des groupes de discussion, tentent de mettre au clair des règles qui vont harmoniser les besoins scientifiques, les enjeux médicaux ou sociétaux et l'éthique associée à tout cela comme par exemple le « International Summit on Human Gene Editing » en 2015, ou le comité multidisciplinaire formé par l'organisation « World Health Organization » pour l'édition de génome Humain.

Le débat entre ce que l'on sait faire et ce que l'on va faire, prend une place de plus en plus importante depuis l'apparition de ces nouveaux systèmes d'édition de génomes et les avancées importantes qui en ressortent. Les applications les plus basiques que mon avis jugerait essentielles pour faire évoluer ce monde sont encore loin d'être acceptés par une majorité de la population qui n'est pas suffisamment sensibilisé à ce domaine et qui malgré tout possède une partie du pouvoir décisionnaire. Quand l'on voit encore actuellement la médiatisation de personnes jugeant sans comprendre les technologies les plus basiques tels que les vaccins, l'utilisation d'outils impliquant la modification génétique à des fins thérapeutiques devra forcément passer par une sensibilisation du public non scientifique. Cela concerne particulièrement la santé humaine précédemment développée, mais aussi l'écologie animale, végétale et microbienne qui sont des enjeux importants dans le futur avec les thématiques de conservation de la biodiversité ou de production de nourriture suffisante pour la population mondiale en expansion constante.

3. OUTILS GENETIQUES CHEZ LES MOLLICUTES

L'étude des bactéries en général et donc des mycoplasmes dans son ensemble, de la compréhension fondamentale à l'application, nécessite la possibilité de réaliser de l'ingénierie génétique que ce soit pour de l'expression de protéines hétérologues, la complémentation fonctionnelle ou pour de la modification génétique (insertion, délétion et substitution). La mise en place de telles approches est nécessaire pour approfondir de

Espèce	Branche phyle	Transformation	Antibio-résistance	Type de plasmide	Recombinaison	Applications	Références	
<i>M. sp.</i>	S	PEG	<i>erm</i> <i>tet(M)</i>	Naturels p <i>oriC</i> Tn4001 et dérivés	R1	Démonstration Expression Mutagenèse	13-14-20	
<i>M. sp.</i>		PEG	<i>tet(M)</i>	p <i>oriC</i> Suicide Tn4001 et dérivés	R1 RecA E. coli	Démonstration Mutagenèse	13-21	
<i>M. sp.</i>		PEG	<i>aacA-aphD</i> <i>erm</i> <i>tet(M)</i>	Tn916 et dérivés Tn916 et dérivés Naturels p <i>oriC</i>	Son	Élimination marqueur Expression Mutagenèse	19-22-23-24-25	
<i>M. fortuitum</i>		Electroporation	<i>aacA-aphD</i> <i>cat</i>	Tn4001 et dérivés	Son	Mutagenèse	28	
<i>M. goodii</i>		Electroporation	<i>tet(M)</i>	Naturels	Son	Démonstration	48	
<i>M. goodii</i>		Copagation Electroporation PEG	<i>pac</i> <i>specR</i> <i>cat(M)</i>	p <i>oriC</i> Tn5	Son	Mutagenèse	17-18	
<i>Spiroplasma citri</i>		Electroporation PEG	<i>aacA-aphD</i> <i>tet(M)</i>	Tn4001 et dérivés Tn5 Naturels p <i>oriC</i>	R2	Complémentation Élimination marqueur Mutagenèse	10-11-12-13-14-15-74	
<i>Spiroplasma ericksonii</i>		Electroporation	<i>tet(M)</i>	p <i>oriC</i>	Son	Expression	14	
<i>Spiroplasma kunkinii</i>		Electroporation	<i>tet(M)</i>	Naturels	Son	Démonstration	75	
<i>Spiroplasma rhodesianum</i>		Electroporation	<i>tet(M)</i>	Naturels	Son	Démonstration	75	
<i>M. hominis</i>		H	Copagation PEG	<i>tet(M)</i>	Tn916 et dérivés Tn4001 et dérivés	Son	Mutagenèse	29-30
<i>M. arifloridis</i>			Copagation PEG	<i>tet(M)</i>	Tn916 et dérivés Tn4001 et dérivés	Son	Expression Mutagenèse	31-32-33
<i>M. pulchra</i>			PEG	<i>aacA-aphD</i> <i>cat</i> <i>tet(M)</i>	Tn916 et dérivés Tn4001 et dérivés p <i>oriC</i>	R1	Expression Mutagenèse	1-32-34-37-38-39
<i>M. bovis</i>			Electroporation PEG	<i>aacA-aphD</i> <i>tet(M)</i>	Tn4001 et dérivés p <i>oriC</i>	Son	Complémentation Expression Mutagenèse	25-44-46-72
<i>M. agalactiae</i>			Electroporation	<i>aacA-aphD</i> <i>cat</i> <i>tet(M)</i>	Tn4001 et dérivés p <i>oriC</i>	R1	Complémentation Mutagenèse Promoteur inducible	14-20-44-45-46-47
<i>M. ginseng</i>	Electroporation		<i>tet(M)</i>	p <i>oriC</i>	Son	Démonstration	49	
<i>M. hyopneumoniae</i>	Electroporation		<i>pac</i> <i>cat(M)</i>	Tn4001 et dérivés p <i>oriC</i> Suicide	R1 RecA E. coli	Complémentation Expression Mutagenèse	40-41-42-43	
<i>M. hyoviscera</i>	Electroporation PEG		<i>tet(M)</i>	Tn916 p <i>oriC</i> Suicide	R1 RecA E. coli	Mutagenèse	34-35-36	
<i>M. indicus</i>	Electroporation		<i>tet(M)</i>	p <i>oriC</i>	Son	Démonstration	58	
<i>M. gallinigenum</i>	Copagation Electroporation PEG		<i>aacA-aphD</i> <i>cat</i> <i>tet(M)</i>	Tn4001 et dérivés Tn916 et dérivés p <i>oriC</i>	R1	Complémentation Expression Mutagenèse	5-6-28-50-51-52-53 54-55-56-57-58	
<i>M. pneumoniae</i>	P	Electroporation	<i>aacA-aphD</i> <i>cat</i> <i>tet(M)</i>	Tn4001 et dérivés p <i>oriC</i> Suicide	R2 Cre-lox SSAP BAGE	Expression Mutagenèse	26-28-39-64-61-64-65 66-67-68-69-70	
<i>M. genitalium</i>		Electroporation	<i>aacA-aphD</i> <i>cat</i> <i>pac</i> <i>tet(M)</i>	Tn4001 et dérivés Suicide	R1 R2 Cre-lox	Mutagenèse Promoteur inducible	62-63-70-71-73	
<i>G. aralyticum</i>		Natural ecology	<i>tet(M)</i>	Tn1545	Son	Démonstration	8	
<i>G. parvum</i>	PEG	<i>aacA-aphD</i>	Tn4001 et dérivés	Son	Démonstration	9		
<i>Acholeplasma laidlawii</i>	A	PEG	<i>aacA-aphD</i> <i>erm</i> <i>neoR</i> <i>tet(M)</i>	Tn4001 et dérivés Tn916 et dérivés Naturels	R1	Démonstration Expression	1-2-3-4-5	
<i>Acholeplasma oculi</i>		Lipofection PEG	<i>aacA-aphD</i>	Tn4001 et dérivés	Son	Démonstration	6-7	

Table 4 : Ingénierie génétique publiées chez les mollicutes. Table représentant l'ensemble des espèces de mollicutes pour lesquelles au moins un outil génétique est disponible. La branche phylogénétique à laquelle appartient l'espèce est représentée dans la seconde colonne (S : Spiroplasma ; H : Hominis ; P : Pneumoniae ; A : Acholeplasmae. Les protocoles de transformation répertoriés sont présentés dans la troisième. Les marqueurs antibiotiques utilisés sont présents dans la quatrième colonne et comprennent le marqueur de résistance à la tétracycline : *tet(M)*, au chloramphénicol : *cat*, à la gentamicine : *aacA-aphD*, à l'érythromycine : *erm*, à la néomycine : *neoR* et à la puromycine : *pac*. L'origine des plasmides qui sont utilisés sont précisés, et peuvent être de type suicide, dérivés du transposon Tn4001, dérivés du transposon Tn916 ou de type plasmide *OriC*. La présence d'une publication relatant un évènement de recombinaison est présentée : R1 signifie une intégration naturelle par simple crossing-over, R2 signifie un évènement naturel de recombinaison par double crossing-over et l'utilisation d'un système exogène est précisé.

nombreux axes intéressants étudiés chez les mycoplasmes tels que la notion de cellule minimale (Glass *et al.*, 2006), la compréhension de l'émergence de ces espèces dans de si nombreux hôtes avec l'interaction hôtes pathogènes (Citti and Blanchard, 2013), la compréhension des mécanismes de résistance aux antibiotiques dans les cas cliniques (Chernova *et al.*, 2016) et la recherche de solutions pour lutter contre ces pathogènes (Gómez-Martín *et al.*, 2013; Ishfaq *et al.*, 2020). De nombreux outils de mutagenèse sont disponibles pour bon nombre d'espèces bactériennes, notamment la recombinaison homologue. Malheureusement, chez les mycoplasmes une seule espèce a été décrite comme ayant une recombinaison homologue assez efficace pour l'utiliser à des fins d'ingénierie de génome, il s'agit de *M. genitalium* (Dhandayuthapani, Rasmussen and Baseman, 1999). Les outils génétiques chez les mycoplasmes sont restés pendant longtemps peu nombreux et dépendants des espèces, ce manque constituant une limitation importante pour leur étude. Dans ce chapitre va être répertorié ce que l'on sait actuellement faire en termes d'ingénierie génétique chez les mollicutes (Table 4), les limites des outils actuels et les enjeux à en développer de plus performants.

3.1. Transformation et vecteurs

La première étape majeure à l'ingénierie génétique est de faire entrer de l'ADN dans la cellule bactérienne, il s'agit de la transformation. La compétence naturelle qui permet l'acquisition d'ADN à partir du milieu, comme par exemple chez *Bacillus subtilis*, n'a pas été observée chez les mycoplasmes. Les seuls exemples d'entrée naturelle de l'ADN connus sont actuellement les transferts d'ICE par des mécanismes apparentés à la conjugaison (Citti *et al.*, 2018; Sagné, Citti and Dordet-Frisoni, 2021) et l'infection par des phages que l'on retrouve intégrés sous forme de prophages complets ou dégradés dans certaines espèces comme MAV1 dans *M. arthritidis* (Washburn *et al.*, 2004). Pour faire rentrer de l'ADN dans les cellules de mycoplasmes, la conjugaison bactérienne d'*Enterococcus faecalis* vers une espèce de mycoplasme *M. hominis* (Roberts and Kenny, 1987), *M. arthritidis* (Voelker and Dybvig, 1996; Washburn, Bird and Dybvig, 2003), *Mgal* (Ruffin *et al.*, 2000)) a été utilisée, mais cette méthode a rapidement laissé place à des protocoles de transformations artificielles plus robustes qui ont été développées en laboratoire pour de nombreuses espèces de mycoplasmes. Deux méthodes majeures sont utilisées : l'électroporation et la transformation par traitement au poly-éthylène glycol (PEG). L'électroporation consiste à soumettre des cellules à un choc électrique de quelques ms à

Naturels	Tn4001 et dérivé	OriC
pIK - pKMK1 (<i>Mmc</i>)	pMT85	pMC03 (<i>Mcap</i>)
pMyBK1 (<i>M. yeatsii</i>)	pMT85_2RES	pMYCO1(<i>Mmc</i>)
pSci – spV1 (<i>S. citri</i>)	pIVT	pMYSO1 (<i>Mmm</i>)
pVA868 - VA920 - pNZ18 (<i>Streptomyces</i>)	pIVC	pMfIT (<i>M. florum</i>)
	pISM31	pBOT - pOT -pSD4 (<i>S. citri</i>)
	pIV	pMPD (<i>M. pulmonis</i>)
	pISM2062	p20-1miniO/T - pMM21-7 - pIRR5632 (<i>M. agalactiae</i>)
	pMT	pIRR45v (<i>M. bovis</i>)
	pKV	pMGAL (<i>M. gallisepticum</i>)
		pMPNOriC2 (<i>M. pneumoniae</i>)
		pMAS-LoriC (<i>M. synoviae</i>)
		pMHO (<i>M. hyopneumoniae</i>)
		pGEMT-LoriC - pGEMT-MoriC (<i>M. hyarhinis</i>)

Table 5 : Récapitulatif des différents noms de plasmides utilisés dans la littérature.

haut voltage ce qui va induire la déstabilisation de la membrane cellulaire et l'entrée de l'ADN par diffusion dans la cellule. Les protocoles de transformation au PEG tirent profit de l'absence naturelle de paroi des mycoplasmes. Le PEG est un polymère qui interagit avec la membrane et la déstabilise, ce qui facilite l'entrée de molécules d'ADN dans la cellule ou la fusion entre cellules. Chaque protocole de transformation est dépendant de l'espèce bactérienne et les deux protocoles sont parfois utilisés. L'efficacité varie selon de nombreux paramètres tels que le plasmide utilisé, le marqueur de sélection et les promoteurs utilisés, le contenu du plasmide et la présence de systèmes de défense contre l'ADN exogène du mycoplasme.

3.1.1. Marqueurs de sélection et promoteurs

L'absence de paroi chez les mollicutes limite l'utilisation d'une classe d'antibiotique, les β -lactamines ainsi que d'autres molécules qui inhibent la synthèse de la paroi bactérienne. Les mycoplasmes sont également naturellement résistants à la rifampicine, les polymyxines et les sulfonamides. La mutation naturelle de RpoB, une sous unité de la RNA polymérase, empêche la liaison de la rifampicine. Les polymyxines et les sulfonamides ciblent la synthèse de lipopolysaccharides et de l'acide folique or, chez les mycoplasmes, ces voies métaboliques ne sont pas présentes. De plus, les mycoplasmes sont capables d'adaptation rapide et, depuis quelques années, l'augmentation des résistances aux antibiotiques a été observée pour de nombreuses espèces d'intérêt médical ou vétérinaire (Pereyre and Tardy, 2021). Ceci va d'autant plus limiter l'utilisation de certains antibiotiques en laboratoire pour le travail sur des souches isolées récemment (Taylor-Robinson and Bébéar, 1997; Chernova *et al.*, 2016; Le Roy *et al.*, 2016; Gautier-Bouchardon, 2018; Valentine-King *et al.*, 2020). Quelques marqueurs de résistance sont malgré tout disponibles et couramment utilisés pour des applications de transformation de mycoplasmes, notamment des marqueurs conférant des résistances à la tétracycline, la gentamicine, le chloramphénicol, l'érythromycine et la puromycine (Dybvig and Cassell, 1987; Dybvig, French and Voelker, 2000; Algire *et al.*, 2009). Le gène de résistance à la tétracycline, *tet(M)*, est couramment utilisé chez de nombreuses espèces de mollicutes. Il confère un haut niveau de résistance à la tétracycline en liant le ribosome et empêchant son action bactériostatique. La gentamicine est également un antibiotique bactériostatique de la famille des aminoglycosides, cette molécule se lie au ribosome bactérien et augmente fortement le taux d'erreurs de lecture lors de la synthèse

protéique. Le gène de résistance à la gentamycine *aacA-aphD*, (Rouch *et al.*, 1987) est utilisé également chez les mollicutes et modifie directement l'antibiotique empêchant ainsi sa liaison au ribosome. La résistance au chloramphénicol est induite par le gène de résistance chloramphénicol acetyltransferase (*cat*) qui catalyse l'acétylation du chloramphénicol, altérant sa capacité de fixation au ribosome bactérien. Le gène *erm* est responsable de la résistance à l'antibiotique érythromycine, mais comme pour la gentamicine, l'apparition de résistances spontanées est plus fréquente en présence de ces molécules et ce marqueur n'est donc pas fréquemment utilisé (Renaudin, Breton and Citti, 2014). Plus récemment, le gène *puromycine N-acetyltransferase* a été évalué et utilisé avec succès dans plusieurs espèces de mycoplasmes (Algire *et al.*, 2009).

Afin de contrôler l'expression du marqueur de résistance, mais également de gènes d'intérêt, il est nécessaire d'avoir des promoteurs plus ou moins forts selon l'application envisagée, et qui fonctionne de préférence chez plusieurs espèces. Chez les mollicutes, relativement peu d'études se sont focalisées sur les promoteurs de mycoplasmes. Il est courant d'utiliser les promoteurs naturels des gènes de résistance comme pour le gène *aacA-aphD*. Mais de plus en plus, on se rend compte que le choix du promoteur contrôlant l'expression du gène de résistance à l'antibiotique est un facteur déterminant de l'efficacité de transformation (données non publiées ; Montero-Blay *et al.*, 2019). Des promoteurs comme celui de la spiraline pS provenant de chez *Spiroplasma citri* (Lartigue *et al.*, 2002) ou p438 de chez *M. genitalium* (Pich *et al.*, 2006) sont couramment utilisés pour conduire l'expression de certains gènes dans des espèces très variées de mollicutes comme *M. pneumoniae*, *M. genitalium*, *Mgal*, *M. agalactiae* ou les espèces du cluster mycoïdes (Janis *et al.*, 2005; Pich *et al.*, 2006; Algire *et al.*, 2009; Breton *et al.*, 2010). Mais pour certaines espèces chez lesquelles ils ne sont pas reconnus ou dans le but d'obtenir une expression plus importante, des promoteurs endogènes tels que celui du gène *tuf*, du gène *vsa* ou du gène *P97* sont utilisés (Dybvig, French and Voelker, 2000; Zimmerman and Herrmann, 2005; Ishag *et al.*, 2016). Récemment, un promoteur synthétique consensus basé sur celui du gène *tuf* de dix espèces de mycoplasmes a été construit et montré comme étant fonctionnel chez quatre espèces de mycoplasmes très éloignées (Montero-Blay *et al.*, 2019). Le contrôle de l'expression offre de nombreuses possibilités intéressantes mais très peu de promoteurs inductibles sont actuellement disponibles pour les mycoplasmes. Un promoteur inductible à la tétracycline a été construit et a été testé dans plusieurs

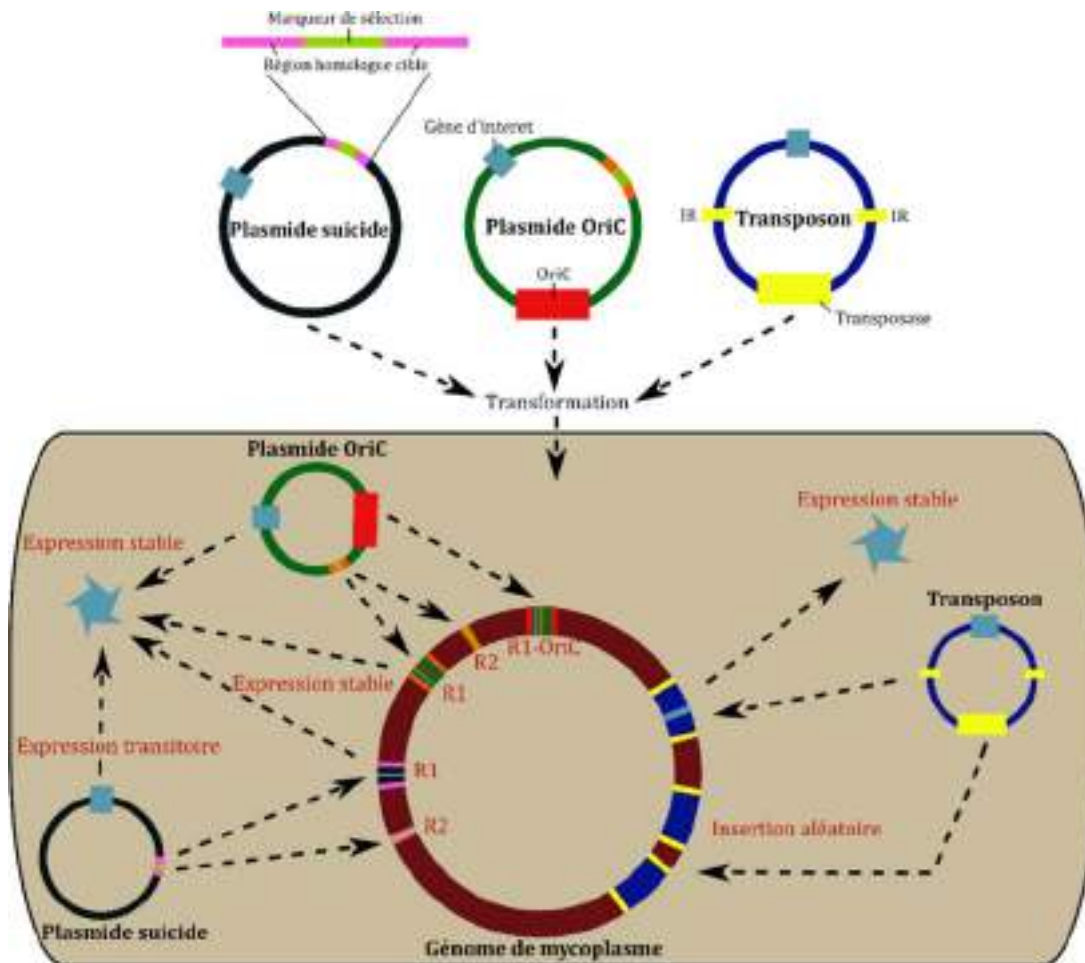


Figure 19 : Représentation schématique des principaux outils disponibles chez les mycoplasmes. Les trois types de plasmides sont représentés en noir, vert et bleu. Les différents éléments composant ces plasmides sont représentés avec des boîtes de couleur. Les IR sont les séquences répétées inversées et sont nécessaires à l'intégration de la séquence plasmique par la transposase associée. La recombinaison peut être liée à un événement simple (R1) qui intègre l'ensemble du plasmide dans la cible, ou à deux événements (R2) permettant l'échange précis de fragments d'ADN.

espèces de mollicutes, notamment *S. citri*, *M. agalactiae* et *M. genitalium* (Breton *et al.*, 2010; Mariscal *et al.*, 2016).

Trois types de vecteurs sont actuellement disponibles : les plasmides suicides, les transposons et les plasmides réplicatifs *oriC*.

3.1.2. Plasmide suicide

Le plasmide suicide est un plasmide qui, une fois transformé dans la bactérie, est rapidement éliminé sauf s'il s'intègre dans le génome. Chez les mycoplasmes, il est utilisé principalement pour deux applications : l'inactivation de gènes par recombinaison homologue et l'expression transitoire d'un gène particulier (Figure 19).

La première consiste à insérer dans un plasmide suicide deux séquences homologues ciblant un locus particulier sur le chromosome et encadrées par un marqueur de résistance à l'antibiotique. Cette méthode visant l'insertion, la délétion ou le remplacement d'une partie du génome, repose sur la recombinaison naturelle de la bactérie et la sélection des recombinants par l'antibiotique correspondant. Cette méthodologie a permis par exemple d'obtenir des mutants chez *M. genitalium* (Dhandayuthapani, Rasmussen and Baseman, 1999; Dhandayuthapani *et al.*, 2001) qui est la seule espèce de mycoplasme actuellement décrite capable d'un double crossing-over naturellement (Table 4, Figure 19 : R2). Un simple crossing-over (R1) conduisant à l'insertion du plasmide suicide au niveau de la région complémentaire peut être suffisant pour inactiver un gène. Peu de rapports en font état dans la littérature hormis chez *M. genitalium* ou *M. pneumoniae* (Krishnakumar *et al.*, 2010) (Table 4) et plus largement pour les mollicutes, chez *Acholeplasma laidlawii* (Dybvig and Woodard, 1992). Le faible succès obtenu lors de ces tentatives est à mettre en lien avec les faibles fréquences de recombinaison homologue chez les mycoplasmes, aux efficacités de transformation qui restent souvent relativement faibles et au fait qu'un simple plasmide suicide est le plus souvent éliminé avant que l'évènement de recombinaison n'ait pu avoir lieu.

La seconde application consiste à l'introduction d'une séquence codant une protéine pour une expression transitoire dans la cellule. Cela permet par exemple d'exprimer de manière transitoire la protéine RecA de *E. coli*, pour induire la recombinaison et produire ainsi des mutants. Trois rapports en font état dans la littérature, produisant, par un évènement de simple crossing-over, l'inactivation d'un gène par intégration chez *Mmc* et

de *M. hyopneumoniae* (Allam *et al.*, 2010; Clampitt, Madsen and Minion, 2021) et la délétion d'un gène par double crossing-over chez *Mycoplasma hyorhinis* (Ishag *et al.*, 2017b).

Ces plasmides sont également utilisés pour l'expression transitoire de recombinaisons spécifiques comme celle codée par le gène *tnpR* (Janis *et al.*, 2008) ou la recombinaison Cre (Mariscal *et al.*, 2016; Garcia-Morales *et al.*, 2020), utilisées après ingénierie du génome pour l'élimination du marqueur de résistance par exemple. En introduisant le gène codant de telles recombinaisons dans un plasmide suicide, il est alors possible de garder suffisamment longtemps le système dans la cellule pour permettre l'expression et l'action de ces enzymes conduisant à l'élimination du marqueur intégré au génome. Ce plasmide est ensuite naturellement éliminé par la cellule sans pression de sélection.

3.1.3. Plasmides *oriC*

Les plasmides *oriC* sont des plasmides répliatifs contenant une origine de réplication chromosomique compatible avec l'espèce à transformer. Le gène *dnaA* et les régions à proximité permettent l'initiation de la réplication et la propagation du plasmide au fil des générations, la présence de celui-ci étant sélectionnée par une pression de sélection apportée par un marqueur de résistance à un antibiotique. L'utilisation de plasmide répliatif a été déclencheur d'un important développement chez les mollicutes puisque la conservation d'une molécule d'ADN circulaire dans le temps au sein d'une cellule a des avantages par rapport au transposon. Aucun gène ou région chromosomique n'est impacté par la présence de cette molécule, il est possible de l'éliminer sans pression de sélection, il est possible de le maintenir dans une population de cellules en croissance afin de favoriser l'observation d'évènements rares tels que les évènements de recombinaison homologue (Figure 19).

Ce type de plasmide a d'abord été développé chez *Spiroplasma citri* pour de l'expression de protéines hétérologues (Renaudin *et al.*, 1995). Rapidement il a été utilisé chez cette même espèce comme outil de mutagenèse dirigée, le maintien de la matrice de recombinaison clonée dans le plasmide *oriC* pendant plusieurs générations favorisant des évènements de recombinaisons (Duret *et al.*, 1999). Ces plasmides ont alors été développés pour de nombreuses espèces : *M. pulmonis* (Cordova *et al.*, 2002), *Mmm*, *Mmc* et *Mcap* (Lartigue *et al.*, 2003), *M. agalactiae* (Chopra-Dewasthaly *et al.*, 2005), *M. imitans* et *Mgal* (Lee, Browning and Markham, 2008), *M. hyopneumoniae* (Maglennon *et al.*, 2013),

Mycoplasma synoviae (Shahid *et al.*, 2014), *Mbov* (Sharma, Citti, *et al.*, 2015), *Spiroplasma eriocheiris* (Terahara, Tulum and Miyata, 2017), *Mesoplasma florum* (Matteau *et al.*, 2017) et *M. pneumoniae* (Blötz *et al.*, 2018). Dans ces espèces, quelques rares évènements de recombinaison ont parfois pu alors être observés. Ainsi, des articles relatent la construction de mutants par simple crossing-over chez *Mcap* (Janis *et al.*, 2005; Lartigue *et al.*, 2009), *Mgal* (Lee, Browning and Markham, 2008) et *M. hyorhinae* (Ishag *et al.*, 2017a).

Comme pour tous les outils, il existe quelques inconvénients. Le premier est l'intégration parasite du plasmide au niveau de l'*oriC* qui est observée chez la plupart des espèces. Cela peut engendrer des problèmes de faux positifs et l'impossibilité d'éliminer un élément du plasmide. L'utilisation d'une région réduite de l'*oriC* (Lartigue *et al.*, 2002) ou comme chez les espèces du groupe mycoïdes, chez *Mgal* et *M. synoviae*, l'utilisation de l'*oriC* d'une espèce proche (Lartigue *et al.*, 2003, 2009; Shahid *et al.*, 2014), permet de limiter ce phénomène.

Le second inconvénient est la nécessité d'utiliser une séquence *oriC* provenant ou très proche de l'espèce que l'on veut transformer. Cela implique le développement d'un plasmide *oriC* pour chaque espèce ou groupe d'espèces et non un outil universel à tous les mollicutes. Le manque de contrôle du nombre de copies présentes dans une cellule peut également être problématique pour certaines applications.

En tous les cas, le faible nombre de publications rapportant la production de mutant montre bien que ces technologies sont limitées et ne sont pas des outils suffisamment efficaces et pratiques pour faire de l'ingénierie complexe de génome.

3.1.4. Transposons

Les transposons sont à l'origine des éléments mobiles naturels présents dans la plupart des phyla bactériens, comme j'ai pu le décrire dans la partie précédente (CF CRISPR). Dès 1978, ils ont été utilisés comme outils génétiques pour des applications d'inactivations aléatoires de gènes non essentiels chez les bactéries : la mutagenèse à transposon (Hernalsteens *et al.*, 1978, no date). Les outils utilisés pour ce type de mutagenèse sont en fait des plasmides non réplicatifs dans l'hôte d'intérêt et portant un transposon plus ou moins modifié. Ils sont composés, a minima, d'une transposase et de séquences IR (répétées inversées) qui vont permettre l'insertion de la séquence présente entre les IR

dans le génome, de manière aléatoire. Cette insertion aléatoire va avoir comme conséquence, la mort de la cellule si un gène essentiel est inactivé, la survie si l'insertion s'est faite dans une région intergénique ou dans un gène non essentiel. La caractérisation de la localisation du transposon permet d'une part d'identifier les gènes responsables d'un phénotype particulier comme par exemple les gènes responsables de la formation des biofilms chez *Streptococcus gordonii* (Loo, Corliss and Ganeshkumar, 2000) et d'autre part, d'évaluer l'essentialité des gènes dans une condition définie, par mutagenèse à saturation combinée à du séquençage haut débit. Cela a permis, par exemple, de déterminer quels gènes pouvaient être éliminés pour construire la cellule minimale JCIV-Syn3.0 (Clyde A. Hutchison *et al.*, 1999).

Chez les mycoplasmes, deux transposons ont été particulièrement utilisés : le Tn916 (Clewell and Gawron-burke, 1986) provenant de *Enterococcus faecalis*, le premier utilisé, avait l'avantage de posséder un marqueur *tet(M)* qui a été rapidement décrit comme fonctionnel chez *Acholeplasma laidlawii* et *M. pulmonis* (Dybvig and Cassell, 1987). Le Tn916 a d'abord été utilisé par conjugaison bactérienne qui s'est révélée possible avec certaines espèces de mycoplasmes (Roberts and Kenny, 1987) (Table 4). En revanche, c'est un transposon très grand (18 kb) pouvant changer de positions plutôt fréquemment (King and Dybvig, 1991). Le Tn4001 (Rouch *et al.*, 1987) est arrivé rapidement après le Tn916, lui aussi ayant été décrit comme fonctionnel chez *Acholeplasma laidlawii* et *M. pulmonis* (Mahairas and Minion, 1989; Mahairas, Jian and Minion, 1990). Celui-ci possède le gène *aacA-aphD* de résistance à la gentamicine et est plus petit (4,7 kb). Il a été utilisé dans de très nombreuses espèces de mollicutes, et des modifications y ont été apportées comme le changement du marqueur de résistance, des promoteurs et l'inversion des séquences IR afin d'exclure le gène de la transposase après insertion. Ce transposon est le plus utilisé et de nombreux dérivés en découlent portant différents noms comme pISM, pIV, pMT, Mini. Ces deux transposons ont permis l'étude fonctionnelle de nombreux gènes en permettant de montrer leur lien avec des processus particuliers tels que la cytoadhérence chez *M. pneumoniae* (Hedreyda and Krause, 1995) ou l'activité ATPase de la F1-likeX0 chez *Mmm* (Béven *et al.*, 2012). Ces transposons permettent également d'insérer facilement dans le génome des gènes hétérologues pour des applications telles que l'utilisation du système de colonie blanc-bleu par insertion du gène *lacZ* chez *Mgal* et *Acholeplasma laidlawii* (Knudtson and Minion, 1993). Dernièrement, un autre transposon a servi d'outil pour l'étude de plusieurs mycoplasmes, il s'agit du Tn5, dérivé du phage λ

de *E. coli* (Berg *et al.*, 1975). Ce transposon a fait l'objet de nombreuses études pour l'optimiser et le commercialiser. Il est maintenant l'outil idéal pour la production d'une banque de mutants conséquente, cela a notamment été réalisé pour *Spiroplasma citri* (Mutaqin *et al.*, 2011), *Mesoplasma florum* (Baby, Lachance, *et al.*, 2018) et JCVI-Syn2.0 (Hutchison *et al.*, 2019).

Les transposons ont tout de même des inconvénients, notamment de par leurs insertions aléatoires, et ils ne peuvent donc pas être utilisés pour une ingénierie précise des génomes. Afin d'obtenir l'inactivation d'un gène particulier, il faut une librairie conséquente couplée à un crible important si aucun test phénotypique n'est disponible. De plus, des biais d'intégration sont fréquemment observés pour certains transposons comme le Tn916 (Nelson, Richardson and Dougherty, 1997). Finalement, seuls les gènes non essentiels peuvent être ainsi étudiés et des effets locaux indésirables sont également possibles, comme l'interruption d'un opéron ou la modulation de la transcription d'un gène par insertion dans une séquence régulatrice (Hutchison *et al.*, 2019).

La possibilité d'utiliser les transposons et les plasmides *oriC* comme vecteurs de transports d'autres systèmes a permis tout de même de développer de manière modeste les outils disponibles chez certaines espèces de mycoplasmes. Par exemple, des techniques de fluorescence et de microscopie sont disponibles chez *Mbov* et *Mmm* avec l'expression de protéines comme MneonGreen et MkO2 (Bonnefois *et al.*, 2016) ou chez *Spiroplasma eriocheiris* et l'expression d'une protéine EYFP (Terahara, Tulum and Miyata, 2017). Cela peut permettre d'étudier certaines interactions et relations entre l'hôte et le pathogène. De plus, actuellement au laboratoire, la microscopie à Super Résolution (Betzig *et al.*, 2006) est en cours de développement et semble montrer qu'il est possible de marquer, localiser et caractériser une protéine d'intérêt au sein d'une cellule de mycoplasme (Rideau *et al.* In preparation). Outre les aspects de microscopie, certains systèmes comme le système dCas9 de *S. pyogenes*, ont été utilisés chez *M. pneumoniae* et JCVI3.0 pour réaliser de l'interférence à la transcription (Mariscal *et al.*, 2018).

3.1.5. Systèmes de recombinaison importés

Le manque d'efficacité de recombinaison naturelle pour la plupart des espèces de mycoplasmes a poussé certains groupes à s'intéresser à l'importation de systèmes

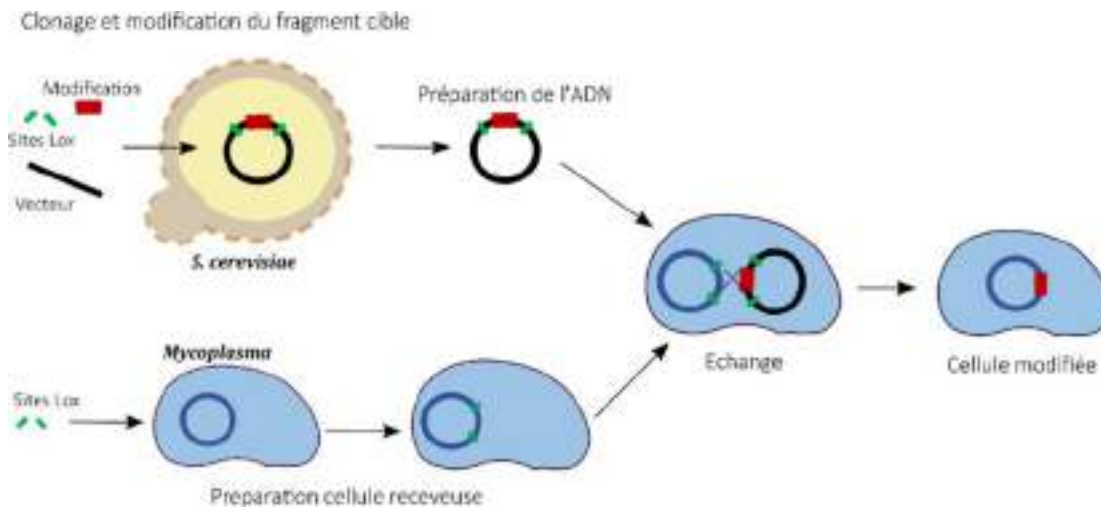


Figure 20 : Représentation de la méthode RAGE. D'une part, le vecteur contenant la séquence codante pour la Cre recombinase et la modification du génome de mycoplasme encadré par deux sites lox est assemblé dans la levure *S. cerevisiae*. D'autre part, le cellule receveuse de mycoplasme est modifié à l'aide du système Gp35 pour ajouter au niveau de la région de génome à modifier, deux sites lox. Finalement l'ADN provenant de la levure est transformé dans la cellule receveuse de mycoplasme et l'échange peut alors être stimulé par le système Cre-lox. Ceci permet d'obtenir une cellule modifiée d'intérêt.

hétérologues de recombinaison comme avec le système Cre-lox précédemment décrit. Une nouvelle application a également vu le jour chez *M. pneumoniae* avec l'utilisation d'une recombinase du phage SPP1 de *Bacillus subtilis*: Gp35 pour de l'ingénierie génétique à partir d'oligonucléotides pour de petites insertions et délétions (Piñero-Lambea *et al.*, 2020). Par la suite, la méthode RAGE a été développée pour des modifications génomiques de plus grande ampleur chez *M. pneumoniae* pour échanger des parties du génome de cette bactérie en passant par la capture et la modification de fragments dans la levure (Garcia-Morales *et al.*, 2020). Cette dernière combine la technologie Gp35 pour modifier une cellule receveuse de Mycoplasme en lui insérant des sites lox autour d'une région d'intérêt, à la technologie Cre-lox pour transformer un fragment d'ADN contenant les sites lox complémentaire et la Cre, ce qui va permettre d'échanger un fragment d'ADN dans le génome ciblé (Figure 20).

Face au manque d'outils d'ingénierie et surtout face à l'intérêt croissant de la communauté scientifique pour ces bactéries minimales, le domaine de la biologie de synthèse s'est particulièrement intéressé aux mycoplasmes.

3.2. Approches de biologie de synthèse chez les mycoplasmes

La biologie de synthèse est un domaine scientifique plutôt récent et en pleine expansion depuis deux décennies. Deux types d'approches sont utilisés dans ce domaine : la minimisation de systèmes et mécanismes biologiques complexes en systèmes simples et maîtrisés (approche « top-down »), et la construction de systèmes biologiques (nouveaux ou pas) à partir de briques élémentaires, des données OMICS disponibles et de leur modélisation (approche « bottom-up »).

Les mycoplasmes ayant les plus petits génomes connus capables de vie autonome, il a été imaginé au John Craig Venter Institut (JCVI), qu'il était possible de designer, reconstruire et transplanter un de ces génomes dans une cellule. Ceci permettrait de construire littéralement de toute pièce un génome de mycoplasme et de produire une souche synthétique. Une première innovation a tout d'abord consisté à montrer qu'il était effectivement possible de transplanter un génome entier de *Mycoplasma mycoides* subsp. *capri* (*Mmc*) dans une cellule de *Mcap* changeant ainsi une espèce en une autre (Lartigue *et al.*, 2007). Par la suite, une plateforme pour modifier ou créer un génome de mycoplasme a été utilisée : la levure. En effet, le clonage de génome entier de *Mmc* dans la

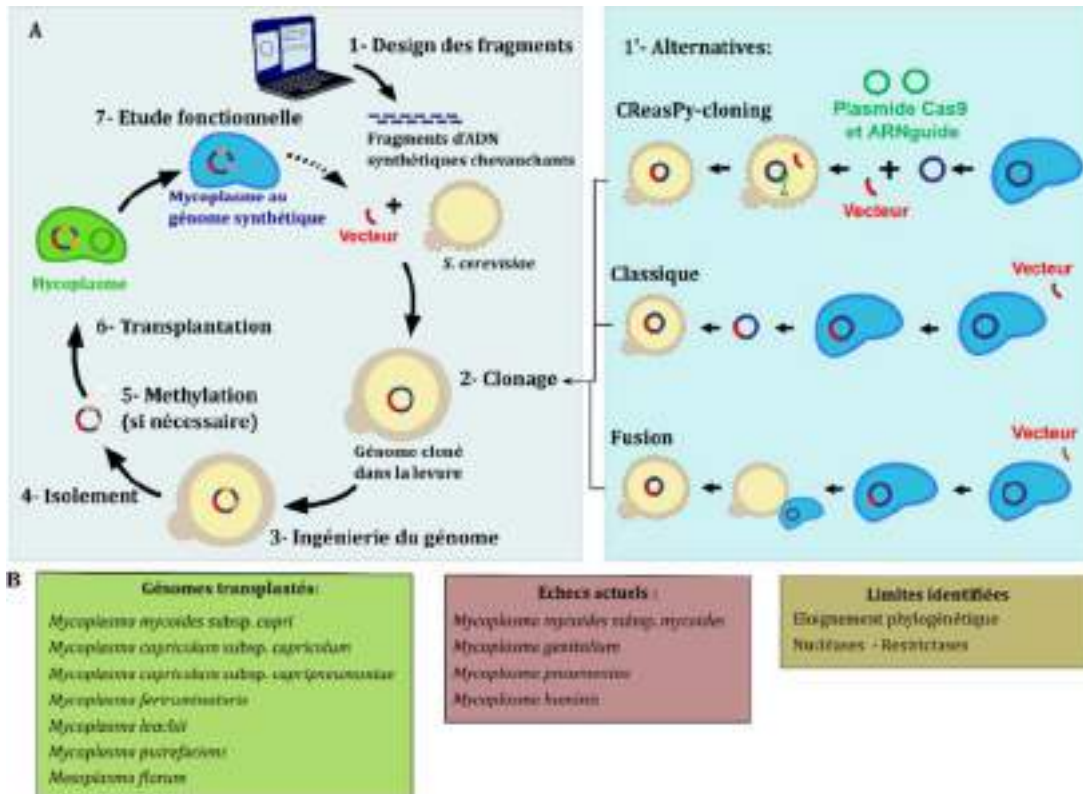


Figure 21 : Cycle d'ingénierie de génome de mycoplasme dans la levure. A- (1) Le génome est construit *in silico* et des fragments d'ADN chevauchants sont synthétisés. (1') Le génome peut également être d'origine naturelle et trois techniques sont utilisées pour arriver à l'étape de clonage. (2) Le génome est ensuite assemblé par recombinaison homologue dans la levure *S. cerevisiae* avec un vecteur contenant les éléments. (3) Des outils génétiques sont disponibles pour modifier le génome ultérieurement. (4) Le génome est isolé et (5) méthylyé afin de le protéger des systèmes de restriction présents dans la cellule receveuse. (6) Ce génome est transplanté dans une cellule receveuse adéquate. (7) Une bactérie à génome synthétique est obtenue et peut être caractérisée au niveau génotypique et phénotypique. B- Récapitulatif chez les mollicutes des espèces chez qui le cycle est fonctionnel, pour lesquels il a échoué et les limites identifiées.

levure a permis la possibilité de le modifier très facilement. Après modification, la transplantation dans *Mcap* permet alors d'obtenir une souche modifiée de *Mmc* (Lartigue *et al.*, 2009). Pour aller plus loin, le génome de *Mmc* a été totalement synthétisé et assemblé dans la levure, permettant la construction de la première souche bactérienne au génome synthétique : JCVI1.0 (Gibson *et al.*, 2010) (Figure 21).

Le cycle clonage/manipulation/transplantation de génome est un outil extrêmement puissant, qui permet de nombreuses applications. Dans la recherche pour la biologie de synthèse tout d'abord, une souche minimale de JCVI3.0 a été développée avec l'élimination du génome de la plupart des gènes non essentiels à la bactérie (Hutchison *et al.*, 2016). La souche JCVI3.0 ne possède plus que 531 kb contre 1085 kb pour la bactérie originale. Ceci permet des avancées importantes dans la compréhension des mécanismes nécessaires à la vie d'une bactérie par exemple.

De nombreux projets se sont construits autour du clonage de génomes bactériens dans la levure (de mycoplasmes, mais également d'autres) dans l'espoir de créer des souches synthétiques et/ou minimales. Les génomes de nombreuses espèces de mycoplasmes ont été clonés, mais également le génome de *Haemophilus influenza* ou même le génome minimal de *E. coli* dernièrement en 3 fragments (Karas *et al.*, 2013; Vashee, Arfi and Lartigue, 2020; Yoneji *et al.*, 2021). Cette méthode a également pu être appliquée pour la construction de génomes viraux suivi de la production de la particule virale comme celui du Cytomégalovirus (Vashee *et al.*, 2017) ou plus récemment, en moins d'un mois, celui du Sars-CoV2 (Thi Nhu Thao *et al.*, 2020).

Au laboratoire cette approche de biologie de synthèse est parfaitement maîtrisée grâce à l'arrivée du Docteur Carole Lartigue ayant mis au point la technique au JCVI. C'est une technique qui est utilisée couramment au laboratoire puisque, outre l'aspect de biologie de synthèse, elle permet de modifier facilement un génome de mollicutes dans la levure, que ce soit pour une simple délétion de gène non-essentiel ou pour la modification de gènes essentiels. Plusieurs techniques de clonage de génomes de mycoplasmes dans la levure ont été développées au laboratoire ou par d'autres équipes, permettant par exemple le clonage et la modification du génome de mycoplasme par le système CRISPR en une seule étape (Ruiz *et al.*, 2019). Cette technique de modification de génome et de transplantation s'est étendue à d'autres espèces: *Mycoplasma leachii* et *Mycoplasma putrefasciens* (Labroussaa *et al.*, 2016), *Mycoplasma feriruminatoris* (Vashee, Arfi and

Lartigue, 2020) et même *Mesoplasma florum* (Baby, Labroussaa, *et al.*, 2018). D'autres applications ont vu le jour, par exemple la production de châssis bactérien de mycoplasmes comme souche vaccinale en supprimant des facteurs de virulence (Jores, Ma, *et al.*, 2019) ou en ajoutant des antigènes de l'espèce d'intérêt (Données non publiées).

La transplantation de génome ne semble pour le moment fonctionner que sur des espèces très proches phylogénétiquement, et nous sommes encore loin de comprendre l'ensemble des facteurs de compatibilité entre génome donneur et cellule receveuse. Il faut énormément de travail pour mettre au point la technique sur une nouvelle espèce et une fois la technique en place, quelques semaines pour obtenir un mutant d'intérêt.

Malgré le potentiel considérable que présente cette méthodologie, un outil plus simple, rapide et facile à adapter est encore nécessaire pour la majorité des espèces de mycoplasmes.

3.3. Outils génétiques pour *Mycoplasma gallisepticum*

Malgré son importance dans l'industrie de la volaille ainsi que l'impact de son émergence chez les rosélins aux USA, comme pour de nombreuses espèces de mycoplasmes, les outils génétiques disponibles actuellement chez *Mgal* sont peu nombreux (Table 4). Il est possible de cultiver *Mgal* en laboratoire dans un milieu riche complété de sérum de cheval. Un plasmide *oriC* (Lee, Browning and Markham, 2008), des transposons (Cao, Kapke and Minton, 1994) additionnés de 4 marqueurs de sélection et des promoteurs fonctionnels sont disponibles pour la transformation de *Mgal*. Des études sur l'activation ou l'inactivation des promoteurs permettant l'expression des protéines de surfaces variables de *Mgal* ont pu être menées à l'aide de l'expression du gène *lacZ* et de la sélection blanc-bleu (Liu *et al.*, 2000). Nombreux facteurs de virulence ont été également caractérisés suite à des mutagenèses par transposon (Tseng *et al.*, 2017). En revanche, la recombinaison homologe n'a pu permettre l'obtention que d'un unique mutant, sur une protéine de surface variable (Lee, Browning and Markham, 2008). Des tests de complémentation de souches par l'expression de protéines hétérologues peuvent également être utilisés. Par exemple la souche R_{high} qui, après de nombreux passages est devenue avirulente, et dans laquelle la réinsertion de trois gènes lui a permis de retrouver sa virulence (Papazisi *et al.*, 2002). Mais même si elle a permis des avancées scientifiques, cette méthodologie donne peu de flexibilité sur les mutants produits et reste limitée.

3.4. Outils génétiques pour *Mycoplasma mycoides subsp. mycoides* et *Mycoplasma bovis*

Mmm et *Mbov* sont, comme *Mgal*, parmi les espèces de mycoplasmes les plus importantes sur le plan vétérinaire. Ces espèces causent encore de nos jours d'importantes épidémies dans le monde que ce soit en Afrique pour la première ou dans des pays développés comme la Nouvelle-Zélande ou les USA pour la seconde. Les travaux de génomique fonctionnelle sur *Mmm* et *Mbov* restent pour le moment, limités, seuls les plasmides *oriC*, les transposons et la transformation sont disponibles (Table 4). Aucun exemple de mutants obtenus par recombinaison homologue n'a pour le moment été répertorié et cela rend la recherche difficile (Janis *et al.*, 2008; Thiaucourt *et al.*, 2011). De nombreuses souches sont disponibles en séquençage ce qui facilite les observations basées sur de la génomique comparée. Les expériences de transplantation de génome n'ont pour le moment pas permis de compléter le cycle clonage/ingénierie/transplantation pour ces deux espèces. Des outils de fluorescence (Bonnefois *et al.*, 2016) et d'infections cellulaires *in vitro* sont disponibles et fréquemment utilisés pour étudier ces pathogènes et leurs interactions avec les cellules eucaryotes (Thomas *et al.*, 2003; Dedieu, Chapey and Balcer-Rodrigues, 2005). Pour *Mmm*, malgré une distance phylogénétique très proche de *Mcap*, et une transplantation validée chez tous les autres membres du cluster mycoides, il n'a pour le moment pas été possible d'obtenir de transplants (Labroussaa *et al.*, 2016). La recherche du ou des facteurs bloquants est actuellement en cours au laboratoire et cela souligne la diversité de ces bactéries minimales qui ne vont pas répondre de la même manière aux différents outils génétiques.

Malgré le peu d'outils disponibles, ces bactéries font tout de même l'objet de nombreuses recherches et caractérisations. L'utilisation des transposons pour de la mutagenèse à saturation chez *Mbov* (Chopra-Dewasthaly *et al.*, 2005; Sharma, Markham and Browning, 2014) et *Mmm* (Janis *et al.*, 2008) a par exemple permis de caractériser des facteurs génétiques associés à différents phénotypes (Sharma, Tivendale, *et al.*, 2015). *Mbov* est par ailleurs étudié depuis de nombreuses années avec *M. agalactiae*, comme modèles pour l'étude des transferts d'ICE par conjugaison entre mycoplasmes (Citti *et al.*, 2020).

3.5. Conclusion

Dans l'ensemble, les mollicutes sont des bactéries diverses et variées parmi lesquelles, certaines espèces ont des impacts très importants dans le monde. Le peu d'outils disponibles limite grandement l'étude de ces bactéries même si dernièrement, les

ÉTUDE BIBLIOGRAPHIQUE : Outils génétiques chez les mollicutes

nouvelles technologies telles que la biologie synthétique, les systèmes SSAPs et les systèmes CRISPR, tendent à être développés chez certaines espèces. Pourtant aujourd'hui, aucun outil d'ingénierie globale des différentes espèces de mycoplasme, ne permet la construction précise et rapide de mutants.

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

L'évolution rapide des mollicutes et leur caractère minimal, leur capacité à s'adapter à de très nombreux hôtes et leur code génétique atypique en font des modèles d'études très intéressants où de nombreux mécanismes sont encore à découvrir. Le nombre et l'efficacité des outils génétiques disponibles pour étudier les mycoplasmes n'est pas suffisant face à l'importance de ces bactéries et le développement d'outils précis et efficaces de modification génétique est un enjeu majeur. L'équipe Mollicutes de l'UMR BFP 1332 travaille principalement sur les mollicutes, notamment les mycoplasmes, les phytoplasmes et les spiroplasmales. Depuis une vingtaine d'années, de nombreux outils y ont été développés comme des plasmides *oriC*, des transposons et, plus récemment, des approches de biologie de synthèse initialement mises en place au JCVI. L'équipe Mollicutes fait notamment partie des 3 équipes au monde qui maîtrise le clonage et la transplantation de génomes. Malheureusement, cette approche n'est actuellement disponible que pour certaines espèces de mollicutes et aucun outil universel et simple d'utilisation n'est disponible pour la construction de mutants chez de nombreuses espèces d'intérêt médical, vétérinaire ou agronomique.

C'est dans ce contexte que, depuis 2014, l'équipe s'est intéressée aux systèmes CRISPR. Cela s'est tout d'abord traduit par l'utilisation de ce système pour cloner et modifier les génomes de mycoplasmes dans la levure, mais l'idée d'une utilisation plus directe de ces systèmes dans les mycoplasmes a également fleuri et un travail en ce sens a été initié par le Dr. Pascal Sirand-Pugnet et son ancien étudiant en thèse, le Dr. Iason Tsarmpopoulos. Les systèmes CRISPR étant parfois toxiques pour certaines bactéries, nous avons tout d'abord envisagé l'utilisation d'un système endogène chez les mollicutes. Dans un premier temps, nous avons donc entrepris l'étude des systèmes CRISPR chez l'ensemble des mollicutes par une série d'analyses de génomique comparée.

1. Principaux résultats

Des systèmes complets et dégradés de type II-A et moins fréquemment de type II-C, ont été trouvés dans 21 des 52 espèces représentatives des mollicutes que nous avons étudié. Une reconstruction phylogénétique a permis la mise en évidence d'une origine commune pour tous les systèmes CRISPR/Cas9 des mycoplasmes. En revanche, les résultats suggèrent deux origines pour les systèmes de spiroplasmales. L'étude de la protéine Cas9 des mollicutes indique une forte proximité structurale avec la protéine Cas9 de *Staphylococcus aureus* à l'exception du domaine PI qui intervient dans l'interaction avec

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

la séquence PAM. Ceci suggère que les séquences PAM des systèmes de mollicutes sont sans doute très différentes. Malgré leurs très petits génomes, les mollicutes ont maintenu plusieurs mécanismes de protection contre l'ADN exogène, principalement les systèmes de Restriction-Modification et CRISPR/Cas9, suggérant l'importance de ces systèmes dans leur contexte écologique naturel.

2. Article publié dans *Frontier in Microbiology*
(21/11/2019 ; *Front Microbiol.* 2019 ; 10: 2701, PMID: 31824468)



Multiple Origins and Specific Evolution of CRISPR/Cas9 Systems in Minimal Bacteria (*Mollicutes*)

Thomas Iouricha^{1,2}, Jason Tsampopoulos^{1,2}, Vincent Talenti^{1,2}, Christine Gaspin³, Annick Molari⁴, Caray A. Walker⁵, Joe Brownie⁶, Alain Blanchard^{1,2}, Patricia Thebaud¹ and Pascal Strand-Pugnet^{1,2*}

¹UMR, UMRI 1330 de Biologie du Fruit et Pathologie, Université d'Orléans, France, ²Université de Bordeaux, UMR 1330 de Biologie du Fruit et Pathologie, Université d'Orléans, France, ³UMR, Méthodes et Informatique Appliquées de Toulouse, Université de Toulouse, Toulouse, France, ⁴School of Life Sciences, Anglia Ruskin University, Cambridge, United Kingdom, ⁵Department of Pathobiology and Population Sciences, Royal Veterinary College, University of London, London, United Kingdom, ⁶UMR 5003, LaBRI, CNRS, University of Bordeaux, Bordeaux, France

OPEN ACCESS

Edited by:

Chih-Hong Ahn,
Academia Sinica, Taiwan

Reviewed by:

Shigeyuki Nakajima,
National Institute of Advanced
Industrial Science and Technology
(AIST), Japan
Yuhua Gu,
Shanghai Normal University, China

*Correspondence:

Pascal Strand-Pugnet
pascal.strand-pugnet@univ.fr

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 26 September 2019

Accepted: 07 November 2019

Published: 21 November 2019

Citation:

Iouricha T, Tsampopoulos J,
Talenti V, Gaspin C, Molari A,
Walker CA, Brownie J, Blanchard A,
Thebaud P and Strand-Pugnet P
(2019) Multiple Origins and Specific
Evolution of CRISPR/Cas9 Systems
in Minimal Bacteria (*Mollicutes*).
Front. Microbiol. 10:2701.
doi: 10.3389/fmicb.2019.02701

CRISPR/Cas systems provide adaptive defense mechanisms against invading nucleic acids in prokaryotes. Because of its interest as a genetic tool, the Type II CRISPR/Cas9 system from *Streptococcus pyogenes* has been extensively studied. It includes the Cas9 endonuclease that is dependent on a dual-guide RNA made of a tracrRNA and a crRNA. Target recognition relies on crRNA annealing and the presence of a protospacer adjacent motif (PAM). *Mollicutes* are currently the bacteria with the smallest genome in which CRISPR/Cas systems have been reported. Many of them are pathogenic to humans and animals (mycoplasmas and ureaplasmas) or plants (phytoplasmas and some spiroplasmas). A global survey was conducted to identify and compare CRISPR/Cas systems found in the genome of these minimal bacteria. Complete or degraded systems classified as Type II-A and less frequently as Type II-C were found in the genome of 21 out of 52 representative *Mollicutes* species. Phylogenetic reconstructions predicted a common origin of all CRISPR/Cas systems of mycoplasmas and at least two origins were suggested for spiroplasmas systems. Cas9 in *Mollicutes* were structurally related to the *S. aureus* Cas9 except the P1 domain involved in the interaction with the PAM, suggesting various PAM might be recognized by Cas9 of different *Mollicutes*. Structure of the predicted crRNA/tracrRNA hybrids was conserved and showed typical stem-loop structures pairing the Direct Repeat part of crRNAs with the 5' region of tracrRNAs. Most *Mollicutes* crRNA/tracrRNAs showed G + C% significantly higher than the genome, suggesting a selective pressure for maintaining stability of these secondary structures. Examples of CRISPR spacers matching with *Mollicutes* phages were found, including the textbook case of *Mycoplasma cynos* strain C142 having no prophage sequence but a CRISPR/Cas system with spacers targeting prophage sequences that were found in the genome of another *M. cynos* strain

that is devoid of a CRISPR system. Despite their small genome size, mollicutes have maintained protective means against invading DNAs, including restriction/modification and CRISPR/Cas systems. The apparent lack of CRISPR/Cas systems in several groups of species including main pathogens of humans, ruminants, and plants suggests different evolutionary routes or a lower risk of phage infection in specific ecological niches.

Keywords: CRISPR/Cas9, mollicutes, mycoplasmas, spiroplasmas, phage, mobile genetic elements, horizontal gene transfer, evolution

INTRODUCTION

CRISPR systems are natural prokaryotic adaptive immune systems involved in protection against invading DNAs and especially viruses (Díello et al., 2018). Memory of the systems is stored in genomes as typical arrays of spacers interspaced with direct repeats. Effector proteins are encoded by *cas* genes involved in adaptation (acquisition of new spacers), pre-crRNA maturation, and target recognition and cleavage. Since their discovery, and even more, their successful adaptation as molecular scissors for genome engineering in nearly all organisms (Doudna and Charpentier, 2014; Knott and Doudna, 2018), CRISPR systems have been described in about 30–40% bacteria and in most archaea (Koonin et al., 2017). Based on specific signature *cas* genes, CRISPR systems are currently distributed into two classes and 6 types, each of which is sub-divided into up to 7 subtypes. CRISPR systems belonging to Class 2 Type II were the first to be widely used in a large variety of organisms. The iconic *spCas9* endonuclease from *S. pyogenes* was used to target DNA by a single guide hybrid RNA designed from the natural structure, which was formed by annealing of matured crRNA and highly structured tracrRNA (Jinek et al., 2012). All Type II systems rely on a Cas9 effector module and a set of other *cas* genes involved in acquisition namely, *cas1*, *cas2* and, in some systems, *cas3* or *cas2*. Based on *cas* genes organization and Cas proteins phylogenies, three Type II systems were identified. Type II-A, which includes the system of *S. pyogenes*, has a typical operon formed by *cas9*, *cas1*, *cas2*, and *cas2* genes. In Type II-B, the *cas2* gene is replaced by *cas4*. In Type II-C, neither *cas2* nor *cas4* are present. Expression of Type II systems involves transcription of a precursor crRNA (pre-crRNA) that is subsequently matured by endogenous non-specific RNase III after pairing of pre-crRNA direct repeats with the structured tracrRNA in presence of Cas9. Matured individual crRNA/tracrRNA duplexes, still bound to Cas9, constitute the active complexes involved in targeted DNA cleavage. Target specificity of Type II systems involves recognition of short protospacer adjacent motif (PAM) located downstream of the protospacer on the non-target DNA strand. The PAM sequence is required both in the new spacer acquisition process and for target recognition and cleavage. Because of their relative simplicity and high potential as programmable DNA cleavage tools, a huge amount of data has been obtained on Type II CRISPR systems, including high resolution structures of *S. pyogenes*, *Staphylococcus aureus* and *Campylobacter*

jejuni Cas9 in an interaction with gRNA and target DNA (Nishimura et al., 2014, 2015; Yamaoka et al., 2017). Despite their popularity, Type II systems are among the rarest found in bacteria with a detection in only 5–7% of bacterial genomes (Chylinski et al., 2014; Bernheim et al., 2019). This type has been detected neither in archaea nor in several bacterial phyla including some photosynthetic bacteria as Cyanobacteria, Chlorobi, and Chloroflexi, thermophilic bacteria as Thermotogae, Aquificae, Deinococcus-Thermus and intracellular pathogens Chlamydia.

Mollicutes are small bacteria, most of which have a parasitic lifestyle, infecting a wide diversity of eukaryotic hosts including humans, many other mammals, birds, reptiles, fishes, arthropods and plants (Gaspardich, 2014; Marini et al., 2014; May et al., 2014; Repass, 2014). In accordance with their parasitic life, these bacteria have lost many metabolic pathways including the ability to synthesise cell-wall components, nucleic acids and protein precursors. Consequently, these minimal bacteria require complex culture media and some of them (phytoplasmas and hemotropic mycoplasmas) remain uncultured. Evolution of mollicutes has been described as mostly driven by genome reduction, resulting in minute-sized genomes ranging from 0.58 to 2.2 Mbp with an average size around 1 Mbp. Most genera have a specific usage of the UGA codon to encode tryptophan instead of being a stop codon, which might have accentuated a genetic isolation. However, the growing number of genomes available has revealed that mobile elements including plasmids, phages, transposons, and ICE have circulated among mollicutes (Citiu et al., 2013). Moreover, comparative genomics and wet-lab experiments have shown large genomic exchanges could occur among mollicutes by horizontal gene transfer (HGT), bringing a new light on the evolution of these minimal bacteria (Dordet-Frisoni et al., 2014; Lo et al., 2015; Bazzanioli et al., 2018; Tsai et al., 2018; Music et al., 2019). In many bacterial groups, invasion by mobile elements is thwarted by strain-specific systems including restriction-modification (RM), bacteriophage abortive infection mechanisms (Abi), and CRISPR systems. In mollicutes, RM and Abi systems have been described in many species, with various repertoires depending on the strains. More recently, some CRISPR systems were identified in several species. To the best of our knowledge, mollicutes are the bacteria with the smallest genomes where CRISPR systems have been observed (Toucheff et al., 2016). In order to get an overview on CRISPR/Cas systems in mollicutes, a global survey was performed using a representative set of 52 mollicutes species.

MATERIALS AND METHODS

Comparative Genomics and Analysis of CRISPR Loci

CRISPR systems were searched in a selection of 52 mollicutes genomes (Supplementary Table S1) available in the MolBGen database (Barré et al., 2004) or in GenBank (Benson et al., 2017) using (i) Blastp (Johnson et al., 2008) search of Cas proteins, (ii) analysis of direct repeats at the CRISPR database (Grissa et al., 2007) with the included CRISPR finder tool (Courvin et al., 2018). A manual analysis of all candidates was then achieved to get a precise annotation of the loci. In order to predict DR/tracrRNA hybrid secondary structure, sequences of DR and tracrRNA were concatenated and the secondary structures of the hybrids were simulated using the mfold software (Zuker, 2003) available at <http://unafold.rna.albany.edu/>. Cas9 proteins from mollicutes were compared to structurally characterized homologs using HHpred (Soding et al., 2005) and manual inspection of conserved positions. Logos were constructed from direct repeats sequences using weblogo¹. Consensus direct repeats sequences were submitted to an automated classification process using CRISPRMap tool (Lange et al., 2013).

Phylogenetic Tree Reconstructions

The phylogenetic tree of the 52 selected mollicutes was generated from the concatenated multiple sequence alignments of 62 selected orthologous proteins involved in translation (Supplementary Table S2), as described in Gimpstein et al. (2014). Multiple alignments were generated with MAFFT (Katoh et al., 2007), curated from unreliable positions with GBLOCKS (Talavera and Castresana, 2007), and concatenated with Seaview (Gouy et al., 2010). The final concatenated alignment contained 9,990 sites. The phylogenetic tree was constructed by the Maximum Likelihood method using PhyML (Guindon et al., 2010) (substitution model: WAG, number of categories: 4) available on the web server Phylogeny.fr (Dereeper et al., 2008). For phylogenetic analyses of Cas9 (Supplementary Table S1), multiple alignments were generated using MSAProbs (Gorntler-Domínguez et al., 2016) at <https://toolkit.tuebingen.mpg.de> (FASTA format available in Supplementary Figure S1). Phylogenetic trees were inferred using a Maximum Likelihood method with PhyML (substitution model: WAG, number of categories: 4) available on the web server Phylogeny.fr.

Spacer Analysis

Several custom databases were built to perform spacer sequences analyses. First, a custom database was assembled from the complete genomes of the NCBI referenced with the Taxonomy identifier as tenericutes (taxid = 544448). 415 complete genomes were retrieved and used to build two custom databases composed of a complete set of CDSs and a collection of intergenic regions, respectively. A prophage database was also built by retrieving sequences from the European Nucleotide Archive (contains 3180 phage sequences) and the prophage and virus database

¹<https://weblogo.berkeley.edu/>

(Last update: 2017) provided by Zhou et al. (2011). Using these three databases, blastn (nbi-blast 2.7.1 + version) analyses were carried out to search sequences similar to the spacers sequences (562 unique sequences) retrieved from mollicutes CRISPR arrays. An identity of 95% and a coverage of 99% were used as parameters. A total of 12, 569 and 2 hits were retrieved with the CDS, Intergenic and prophage databases, respectively.

Search for Mobile Genetic Elements in Mollicutes Genomes

GenBank genome files of selected mollicutes were downloaded from NCBI Website. (1) Integrative and conjugative elements (ICEs) were identified following the method described in Tardy et al. (2015). Briefly, tblastn searches were conducted using four genes as queries (CDS 5, 17, 19, 22 from *Mycoplasma agalactiae* PG2 ICEs and *Mycoplasma capricolum* subsp. *capricolum* ATCC 27345 ICEs). Hits were manually analyzed and located on the related chromosome. Co-location of the 4 CDS and consistent blast results were interpreted as signs for the presence of complete ICE. Co-location of 2 or 3 CDS and consistent blast results were considered as revealing degraded or atypical ICEs. (2) Insertion sequence were identified using IS finder (Siguier, 2006) at <https://finder.biotoul.fr/>. (3) Restriction systems were predicted using the REbase database (Roberts et al., 2015) at <http://rebase.neb.com/rebase/rebase.html>. When several genes were missing, manual check of the region was performed with a Blastp to identify if the system was complete or not. (4) Phage detection was performed using PHAST (Zhou et al., 2011) at <http://phast.wishartlab.com/>. Additional manual comparisons between mollicutes bacteriophages described in the literature and mollicutes genome sequences were performed using tblastn.

RESULTS

CRISPR/Cas Systems Are Present in Most but Not All Mollicutes Groups

CRISPR/Cas systems were first searched in a representative set of 52 mollicutes genomes composed of species from the phylogenetic AAP branch (*Acholeplasma*-*Asteroleplasma*-*Phytopleasma*) and from the three main groups of the SEM branch (*Spiroplasma*-*Entomopleasma*-*Mycoplasma*), namely the *Hominiis*, *Puccinonniae* and *Spiroplasma* phyla (Supplementary Table S1). Complete or degraded Type II CRISPR/Cas systems were detected in 21 out of 52 complete or draft genomes of mollicutes (Figure 1). CRISPR/Cas systems including, in the following order, *cas9*, tracrRNA, *cas1*, *cas2*, and a CRISPR array, were found in most species. Some inversions were observed for several strains such as *Mycoplasma dispar* ATCC27140, *Mycoplasma argenteum* NM2010, *Mycoplasma hyoscyamine* 232, *Mycoplasma arginini* HAZ145.1, and *Mycoplasma arthritidis* 158L3.1. The overall picture is that CRISPR/Cas systems are detected in most of the main phylogenetic phyla with a more frequent occurrence in the *Hominiis* group. No other type of CRISPR/Cas system other than Type II was predicted.

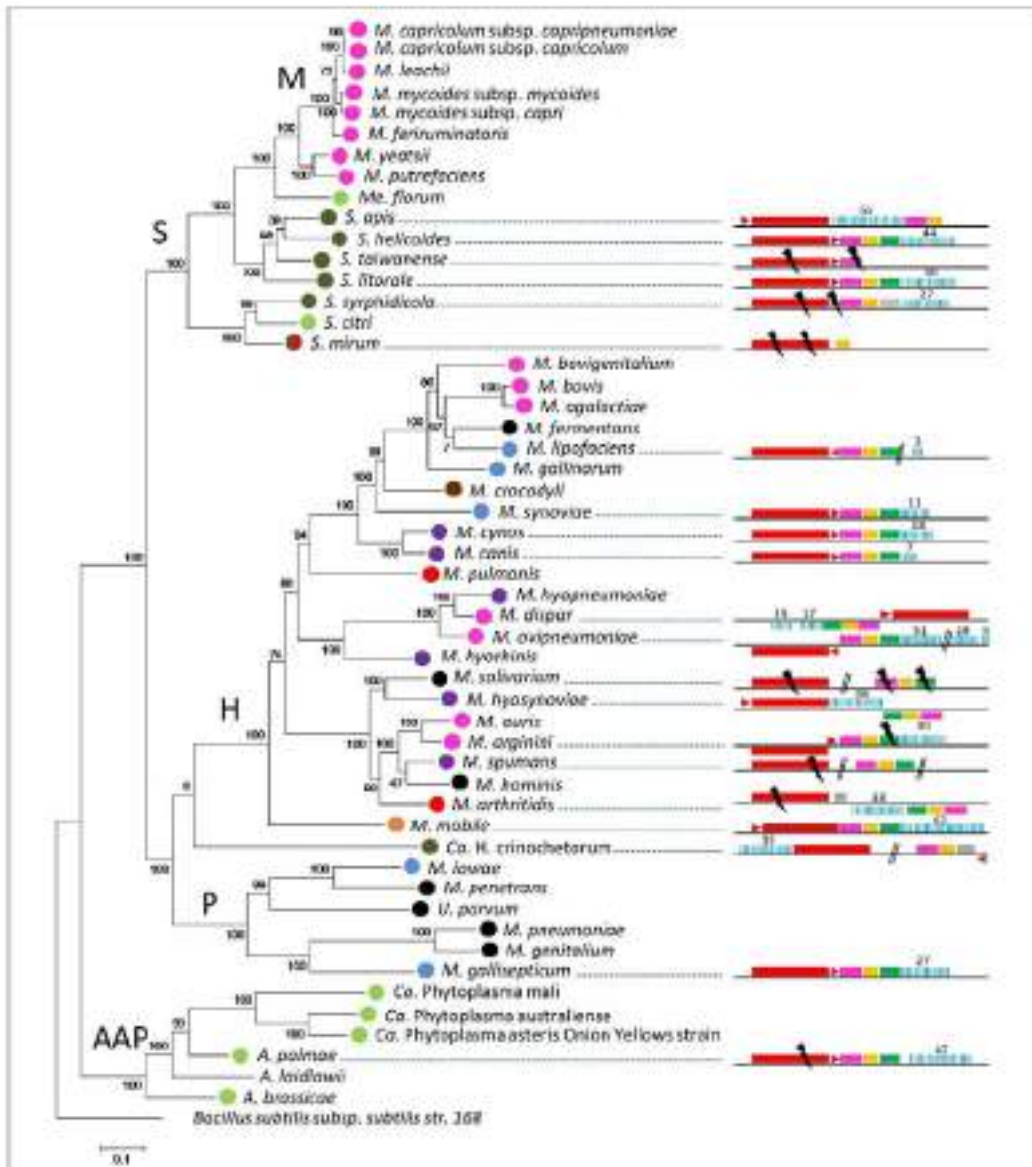


FIGURE 1 | Distribution of CRISPR systems in mollicutes. Organization of the CRISPR systems predicted in mollicutes genomes are presented on the right part. Red rectangle, casR; red triangle, IntraRNA; pink rectangle, cas1; orange rectangle, cas2; green rectangle, cas3; grey rectangle, Cas5 not related to CRISPR; blue bars, CRISPR array. Above numbers indicate the number of spacers. Double bars, genome insertion; black fish, disrupted gene. The phylogenetic tree was generated using the maximum likelihood method from the concatenated multiple sequence alignments of selected 52 orthologous proteins involved in invasion. Main phylogenetic groups are indicated: S, Spirochaetales; H, hemolysales; P, Pasteurellales; AAP, actinobacteria/phytoplasma; M, mycoidae cluster; D, subtilis was used as an outgroup. Statistical values from an Approximate Likelihood Ratio Test are indicated on branches. Hosts are indicated by colored circles: arthropod, dark green; ruminant, pink; rodent, red; bird, blue; others remain in colorless (grey), purple, human, black, plant/arthropod, light green, reptile, bat and fish, orange.

CRISPR/Cas systems were found to be widespread throughout the Hominis phylum, in the genomes of species infecting a variety of animal hosts, with complete structures predicted in *Mycoplasma synoviae* 53, *Mycoplasma cynos* C142, *Mycoplasma canis* PG14, *M. dispar* ATCC27140, *M. ovipneumoniae* NM2010, *Mycoplasma hyopneumoniae* NPL1, *Mycoplasma mobile* 163K and possibly *Mycoplasma lipofaciens* ATCC35015, and incomplete structures in *Mycoplasma salinarium* ATCC23064, *Mycoplasma spumans* ATCC19526, *Mycoplasma urefratilis* 158L3_1 as well as the outgroup *Ca. Hepatoplasma crinocetorum* Ar.

In the Pneumoniae phylum, a CRISPR/Cas system was found only in the bird pathogen *Mycoplasma gallisepticum* 56. For this species, complete or incomplete forms of the CRISPR locus have been characterized in the 12 genomes available (Supplementary Figure S2). Depending on the strain, CRISPR arrays included 23 to 105 spacers and their evolution has been previously associated with a recent host shift from poultry to American house finches (*Haemaphysalis mexicanus*) (Delaney et al., 2012; Tulman et al., 2012).

In the Spiroplasma phylum, positively complete CRISPR/Cas systems were found in *Spiroplasma helicoideis* TABS-2 and *Spiroplasma litore* TN-1 whereas more or less degraded forms were found in *Spiroplasma timentense* CT1 and *Spiroplasma nitrum* ATCC29335. In *Spiroplasma syphilicola* EA1 and *Spiroplasma apis* B31, all the above mentioned genetic elements were predicted (complete or disrupted), except *cas2* that remained undetected. Remarkably, a second CRISPR locus including an array of 9 direct repeats and 8 spacers was detected in *S. syphilicola* EA1 genome but *cas* genes could not be identified, whereas a putative *tracrNA* was predicted (not shown). No CRISPR/Cas system could be detected in *Spiroplasma citri* GI-3 and other closely related species (i.e., *S. kankohi* and *S. mullifurum*, not shown). In the AAP branch, a degraded CRISPR/Cas system was found in *Acholeplasma palmate* 123 but not in the 11 other *acholeplasma* genomes available (extended search on all genomes available in June 2019).

Remarkably, some groups of species belonging to the same phylogenetic branch or associated with the same host appear to lack CRISPR/Cas systems, at least from the currently available genomes. Most noteworthy examples are human pathogenic mollicutes infecting pulmonary and urogenital tracts. Despite a growing number of genome sequences (extended search on all genomes available in June 2019), no CRISPR/Cas were found in *M. pneumoniae* (90 genomes available), *M. genitalium* (6), *M. hominis* (23), *M. penetrans* (2), *U. urealyticum* (18), and *U. parvum* (14). CRISPR/Cas systems also remained undetected in the ruminant pathogens belonging to the Mycooides cluster (*Spiroplasma* phylum and 35 genomes available) as well as the *M. agalutiae*/*M. horvi* group (Hominis phylum and 39 genomes available), they were, however, detected in other ruminant mycoplasmas, namely *M. ovipneumoniae*, *M. dispar* and *M. arginini*. Finally, no CRISPR/Cas system was predicted in the uncultivated and plant pathogenic phytoplasma (AAP phylum, 26 genomes available).

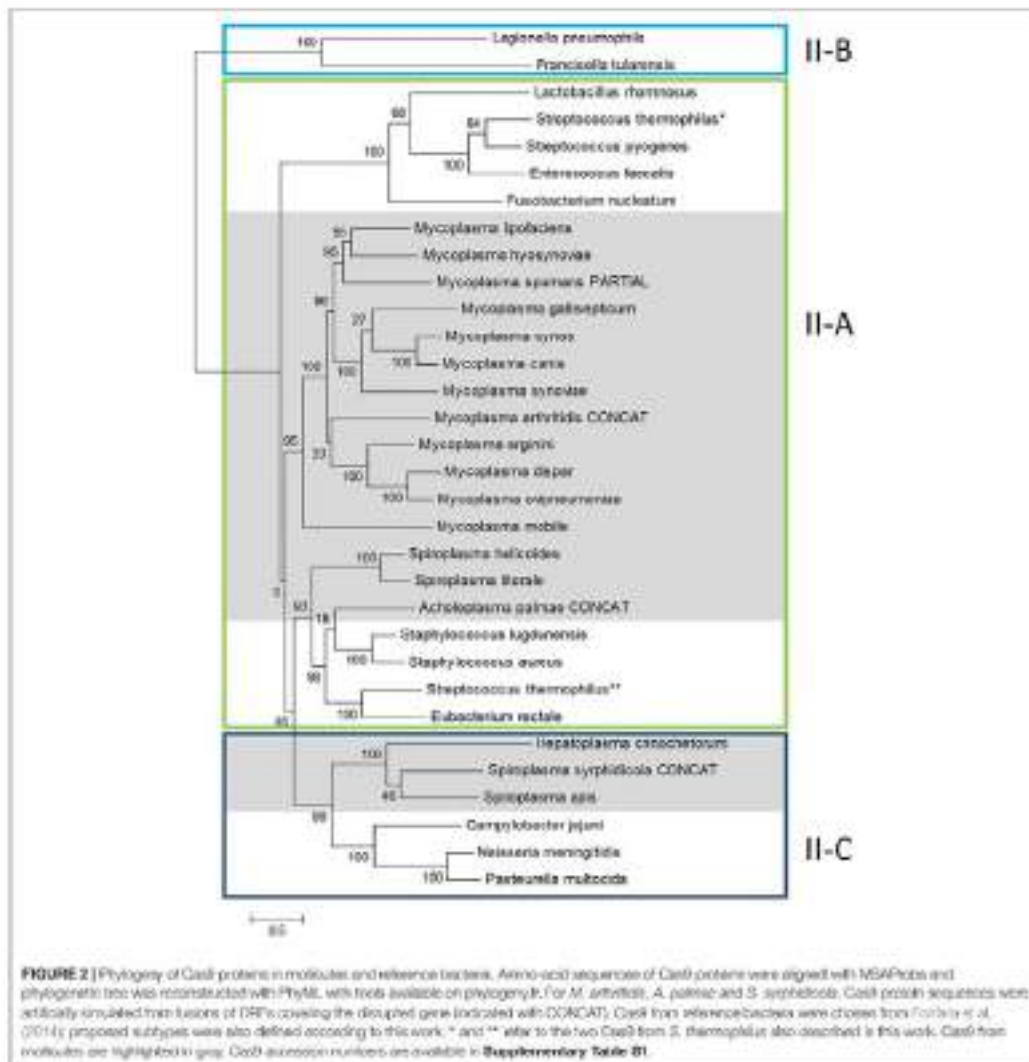
Therefore, Type II CRISPR/Cas systems appear to be widespread among mollicutes with some particular

groups currently lacking any trace of these microbial defense mechanisms.

CRISPR/Cas Systems From Mycoplasmas Have a Common Origin but Not Those From Spiroplasmas

Previous phylogenomic studies on the diversity of bacterial CRISPR/Cas systems including some mycoplasma systems have classified them as Type II CRISPRs, with all representatives studied gathering in a specific branch of subtype II-A (Porfano et al., 2014). In order to get a more complete picture, we performed similar phylogenomics focusing on mollicutes CRISPR/Cas systems. Amino-acid sequence of Cas9 from mollicutes were aligned together with a set of reference Cas9 proteins from subtypes II-A, II-B and II-C, as defined in Fontana et al. (2014) (Supplementary Table S1). A phylogenetic tree was then inferred from the multiple alignment (Figure 2).

General topology of the Cas9 tree only partially reflects species phylogeny with proteins of several mollicutes species clustering with non-mollicutes homologs. However, protein Cas9 from all mycoplasmas clustered in a statistically highly supported branch (aLRT value, 95%). This suggested a common origin of all CRISPR/Cas systems currently identified in mycoplasma species. Intra-branches only partially reflected species phylogeny. For example, Cas9 from *M. gallisepticum* 56 was found in a 100% supported subgroup including *M. synoviae* 53, *M. cynos* C142 and *M. canis* PG14. While those last three species are phylogenetically closely related in the Hominis group (see Figure 1), *M. gallisepticum* is a remote species from the Pneumoniae group. Similarly, Cas9 from *M. lipofaciens* ATCC35015 clustered with homologs found in relatively remote species *M. hyopneumoniae* NPL1 and *M. spumans* ATCC 19526. Cas9 from other mollicutes were found more widely distributed in the phylogenetic tree. Cas9 from *A. palmate* 123 was found in a branch with homologs from several reference Gram positive bacteria (staphylococci, *Streptococcus thermophilus*, *Enterococcus faecalis*) which is consistent with their common ancestral origin. Cas9 orthologs from *Spiroplasma* and *Ca. H. crinocetorum* Ar were distributed in two remote subgroups with no correlation to their relative phylogenetic position. Indeed, Cas9 from *S. apis* B31 was found closely related to that of *S. syphilicola* EA1 whereas those two *spiroplasmas* belong to two clearly distinct phylogenetic subgroups. By contrast, Cas9 from *S. helicoideis* TABS-2 appeared remote from that of *S. apis* B31 while those two species are very closely related. These observations suggested different origins for the CRISPR/Cas systems found among *spiroplasmas*. In addition, we noticed that no trace of *cas2* gene could be predicted in the genomes of *S. apis* B31, *S. syphilicola* EA1 and *Ca. H. crinocetorum* Ar. Interestingly, Fontana et al. have proposed a fine classification of Type II CRISPR/Cas systems with a subtype II-C characterized by the lack of *cas2* or *cas1* that are found in subtypes II-A and II-B, respectively (Fontana et al., 2014). This suggested that the CRISPR/Cas systems of these three mollicutes might be evolutionarily related to subtype II-C systems by contrast to all



other CRISPR/Cas systems from mollicutes for which a *cas2* gene has been predicted.

Cas9 Proteins of Mollicutes and *S. aureus* Have Similar Domains Except the PAM-Interacting Domains

We compared predicted Cas9 proteins from mollicutes with structurally characterized Cas9 using HHpred. In all cases, the best hit was SaCas9 from *S. aureus* (PDB ID 3CZZ), with 100%

probabilities and *e*-values ranging from 4.9E-146 to 2.6E-103. HHpred alignments were extended over all the proteins except for the Type II-C Cas9 of *S. apis* (B31), *S. syphilicola* EA1 and *C. H. cinnamomum* Ar for which the highly significant alignments stopped after the RuvC-III domains. Still with SaCas9, similarity ranged from 31% to 38%, with the best similarities observed for the Cas9 of *A. palmis* (concatenation of the two protein parts predicted from the ORF interrupted by a mismatch), *S. helioides* and *S. litoreale*. This was in accordance with their phylogenetic related positions in the Cas9-inferred phylogenetic

tree (Figure 2). Structural domains of each of the Cas9 of mollicutes were inferred from the pairwise HHpred alignments with SaCas9 (Supplementary Figures S1, S3). Cas9 proteins from mollicutes are larger than SaCas9 (1059 aa) but smaller than SpCas9 (1368 aa), with sizes ranging from 1069 aa to 1272 aa. Compared to SaCas9, conserved domains with larger sizes are the REC (recognition) lobe and to a lesser extent the L1, RuvC-III and WED domains, all involved in the NUC (nuclease) lobe.

The most conserved regions among Cas9 of mollicutes are also similar to SaCas9, suggesting key positions involved in the protein structure, furthermore the interactions with gRNA and target DNA might be globally conserved. A more detailed analysis was undertaken (Supplementary Material SM1), based on the key amino acids identified from the crystal structure of SaCas9 (Nishimatsu et al., 2015). Overall, many key residues involved in the interaction between Cas9 and the gRNA/target DNA hybrid are conserved among mollicutes Cas9. By contrast, the amino acids from the WED and PI domains involved in PAM duplex recognition are not conserved within mollicutes Cas9 and nor when compared to SaCas9, with the exception of Tyr739 that is found strictly conserved among Type II-A Cas9 but not Type II-C (Supplementary Figure S4). Globally, the protein regions involved in PAM duplex recognition in SaCas9 were found very variable among mollicutes with many indels and non-similar amino acid changes.

Direct Repeats/tracrRNA Hybrids of Mollicutes CRISPR/Cas Systems Form Typical Structures but With Highly Variable Sequences

Typical CRISPR arrays with direct repeats (DR) interspaced with unique spacers were predicted in all mollicutes with a CRISPR/Cas system, with the exception of *S. mitium* ATCC29335, *M. salivarium* ATCC23664 (draft genome) and *M. apium* ATCC19526 (draft genome). Depending on the species and the strain, the number of spacers was found highly variable, from 3 spacers in *M. lipofaciens* ATCC35015 (draft genome) up to 105 in *M. gallisepticum* (strain R_{Agg}). Consensus sequences of DR were determined for 17 mollicutes CRISPRs showing an identical length of 36 bp (Supplementary Table S3). A logo plot was designed (Figure 3) showing some conserved positions on both sides of the motif as well as a few positions inside the motif. The logo plot designed from the 11 DR from mycoplasmas evidenced an internal motif GTACAAT conserved at position 12–18 in all species except *M. mobile* 163K. All 17 consensus DR sequences from mollicutes were submitted to an automated classification process using CRISPRMap tool. By comparing mollicutes DR with a database of 4719 consensus repeats covering 24 families and 18 structural motifs, CRISPRMap assigned all DR to superclass P except DR from *Ca. H. crinocholorum* which was not assigned. Superclass F gather DR from various bacteria with a high level of sequence diversity. Notably, this superclass includes Family F13 to which DR from the typical *S. pyogenes* CRISPR/Cas Type II system has been assigned.

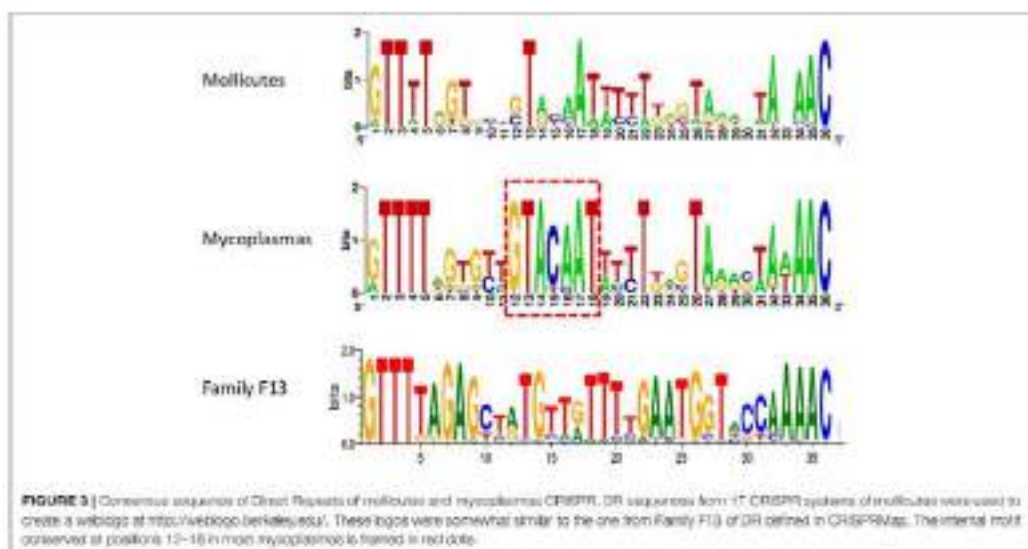
In Type II CRISPR, pairing between the typical antirepeat region of the tracrRNA and the DR from the pre-crRNA is

essential for the RNase III processing of the hybrid structure into a mature dual-guide RNA (Delicbeva et al., 2011). In order to predict tracrRNA/DR interactions in mollicutes CRISPRs, secondary structures of virtual RNA molecules consisting in the 36bp-long DR sequence concatenated with predicted tracrRNA were simulated using the mfold program. As a control, the same process was applied to concatenated DR and tracrRNA from the typical *S. pyogenes* CRISPR01 system. Secondary structure predicted for this was in accordance with previous work (Delicbeva et al., 2011), showing a nearly perfect pairing of DR with the 5' region of tracrRNA (Figure 4). Similarly, long stem-loops involving both RNA molecules were predicted for the 15 mollicutes studied, as exemplified for *M. gallisepticum* 56, *S. apis* B31, *A. pulvis* J283, *M. lipofaciens* ATCC35015 and *M. synoviae* 53 (Figure 4 and Supplementary Figure S5). As shown for other Type II-A systems, direct repeat/antirepeat stem can be divided in a lower stem, a bulge and an upper stem. The predicted hybrid structure for *S. apis* B31 is highly similar to that of other mollicutes despite the relatively remote position of this Type II-C CRISPR/Cas system. Remarkably, no bulge or mispairing was observed in the structure predicted in *M. synoviae* 53 and *Ca. H. crinocholorum* Av (not shown).

Regions of the Type II tracrRNA downstream of the hybrid include the nexus stem-loop and a Rho-independent terminator ending with a poly-U track. Those typical features were predicted for all mollicutes but with a remarkable diversity of sequences and stem lengths (Supplementary Table S4).

Multiple alignment of mollicutes tracrRNAs showed very low conservation except a few positions in the AT-rich base of the lower stem and the GC-rich base of the nexus. The relative conservation of these particular regions of mollicutes tracrRNAs is consistent with general pattern observed in Type II tracrRNAs (Foure et al., 2019). In particular, when focusing on the mycoplasma genus, the upper stem part adjacent to the bulge (when existing) is highly conserved with the GUACAAU motif from the DR paired with a AUUGUAC motif in the antirepeat part of tracrRNAs in most species. For other mollicutes as spiroplasmas and schloplasmas, pairing of this region has been maintained through evolution, despite significant sequence divergence. The conservation of this part of the hybrid suggests a selective pressure for the co-evolution of DR and tracrRNA, as it has been noticed in other groups of bacteria.

The lower stem was found to start with a G:U wobble base pair in only 7 mollicutes dual-gRNAs out of 15, whereas this was shown to be the case in the vast majority of Type II-A dual-gRNAs (Foure et al., 2019). By contrast, a G:C and a A:U pair were predicted for *A. pulvis* J283 (Supplementary Figure S5) and *M. hyosynoviae* NPL1, respectively. Moreover, the one or two first bases starting the DR immediately downstream to the spacer were predicted unpaired for *S. helicoidei* B31, *S. itorale* TN-1, *M. cynos* C142, *M. canis* PG14, *M. arginini* HAZ145_1, and *M. mobile* 163K. These uncommon features were conserved in closely related species (*S. helicoidei* TABS-2 and *S. itorale* TN-1, *M. cynos* C142 and *M. canis* PG14), suggesting no sequencing artifact or specific mutation had occurred. Therefore, unpaired bases might reflect secondary structure inaccurate predictions or a particular flexibility in the dual-gRNA structures of those mollicutes.



The junction between the hybrid region and the nexus is AA in 8 (possibly 11) species, as reported for most Type II-A. This observation is consistent with the previously mentioned conservation of Asn780 and Len783 that are known to interact the first A (A35) of this junction. However, other sequences (i.e. AAA, CAA, and AAU) were predicted for 4 species. The nexus stems include up to 8 GC-pairs, providing highly stable structures. In 9 species, the nexus sequence starts with a GGC motif, as observed for many Type II-A. However, the nexus stem exhibits a remarkable variability in length (5 to 18 bp) and sequence (Supplementary Table S4 and Supplementary Figure S6).

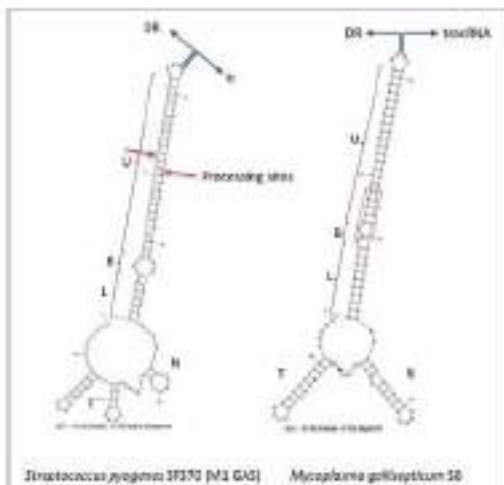
From a general perspective, the predicted DR/tracrRNA secondary structures from mollicutes show AG free energies ranging from -47.80 to -73.20 kcal/mol, which is comparable or even lower than the values predicted from bacteria with higher G + C genome contents (Supplementary Table S4). Genes encoding Cas9 have a slightly lower G + C% than genomes with average values of 24.2% and 27.1%, respectively. By contrast, most tracrRNAs from mollicutes show G + C% significantly higher (average value 33.5%) than corresponding genome (Wilcoxon rank sum test, p -value = 0.0004, Supplementary Table S4). This tendency is not observed for the non-mollicutes bacteria used in this study. This suggests that CRISPR/Cas systems from mollicutes evolved under a selective pressure to keep some G-C pairs in DR/tracrRNA for the stability of secondary structures.

Some Spacers From Mollicutes CRISPR Arrays Target Phage and Bacterial Genes

In order to get some insights on the potential origin of the spacers present in CRISPR loci of mollicutes, all spacers were retrieved from the 17 identified CRISPR arrays and

potential corresponding protospacers were searched in mollicutes (database of 224 genomes) and prokaryotic phages (database of 2480 phages) using a blastn approach. Out of the 595 spacers retrieved, 561 were unique (Supplementary Table S5). Repeated spacers were observed in CRISPR arrays of *M. apis* B31 (1 repeat), *M. pulmo* D33 (8 repeats), *M. gallisepticum* S6 (1 repeat), *M. disjunctum* ATCC27140 (1 repeat), *M. ovipneumoniae* NM2010 (5 repeats), *M. hyopneumoniae* NPL1 (1 repeat), *M. arthritis* 158L3_1 (2 repeats), *M. mobile* 163K (2 repeats), and *Ca. H. crinochetorum* Av (2 repeats). Most spacers did not have any hits in the databases used, however, for 15 of them, highly similar sequences were retrieved. Each potential protospacer was manually checked. One spacer from the CRISPR array of *S. apis* B31 (spacer 24) had a perfect hit in a CDS of the *Spiroplasma furvovola* 23-6 genome (SFLOV_RS04885). Interestingly, this CDS is part of a predicted ICE and is homologous to CD619 described in other ICE from various mollicutes species. A total of 5, 1 and 2 spacers targeting phage and prophage sequences were identified in the CRISPR arrays of *M. cynos* C142, *M. canis* PG14 and *M. arthritis* 158L3_1, respectively. A detailed analysis was conducted on these species.

Comparison of the genomes of *M. cynos* C142 and 210 strains showed that the complete CRISPR/Cas locus identified in C142 was replaced by genes encoding a Type III restriction/modification system in strain 210 (Figure 5). Moreover, 4 genome regions encompassing a total of nearly 50 kbp and showing high similarity with a MAV1 prophage locus of *M. arthritis* 158L3_1 genome (Supplementary Figure S7), were identified in *M. cynos* 210 but were missing in strain C142. Out of the 18 spacers predicted in the CRISPR array of C142, five of them (spacers 1, 2, 3, 9 and 18) presented 100% identical sequences with some genes encoding phage proteins in



Streptococcus pyogenes JF370 (M1 GAS) *Mycoplasma gallisepticum* 28

FIGURE 4 | Predicted DR/tracrRNA hybrid secondary structures. Sequences of DR and tracrRNA were concatenated and the secondary structures of the hybrids were simulated using the mfold software at <http://unimol.cba.hawaii.edu/>. Position of the DR/tracrRNA concatenation is indicated by divergent blue arrows. Predicted stems involving DR and tracrRNA are depicted in black (a lower stem L), a bridge B), and an upper stem U). For *S. pyogenes*, processing sites of the natural hybrid by RNase III are indicated by red arrows. For *M. gallisepticum* 28, DR and tracrRNA sequences were defined based on (Chylinski et al., 2010) and our own work. N, neck stem-loop; T, terminator. The base of the upper stem loop that is highly conserved in mollicutes is framed in red.

strain 210. A total of 10 CDS spread in the four phage regions of the strain 210 genome were targeted by the CRISPR/Cas system of strain C142 and seven of them included two different protospacers. In *M. cynos*, comparison between strains PG14 and IV showed an inactive CRISPR/Cas system (mutation in *cas9*

and IS integrated in CRISPR array) in IV strain. As observed in *M. cynos* example, two regions of *M. cynos* strain IV genome showed high similarity with phage MAV1 which were not found in strain G14. Blast analysis of spacers from *M. cynos* PG14 showed that spacer 6 was matching (96.7%) in the two regions of *M. cynos* IV corresponding to the putative prophage MAV1. All these results suggest that these CRISPR/Cas systems identified in two mycoplasmas infecting dogs are highly dynamic and probably fully active to prevent those minimal bacteria against phage infections.

In the CRISPR array identified in *M. arthritidis* 158L3_1 genome, 2 spacers (spacer 14 and 26) out of 33 matched with potential protospacers identified in the genome of the bacteriophage MAV1 but also in the own genome of the mycoplasma. Indeed a complete prophage MAV1 was found integrated in the bacterial genome (15 CDS from position 766929 to 721810) of strain 158L3-1. In addition, a protospacer with sequence identical to spacer 13 was identified in a *M. arthritidis* 158L3_1 gene encoding a phenylalanine-tRNA ligase subunit beta (MARTH_RS00670). Remarkably, the gene encoding *Cas9* in this mycoplasma was predicted to be disrupted by frameshift mutation generating an UAG STOP codon at the amino acid 394, providing an explanation for tolerating self targeting spacers.

Other cases of self targeting spacers were identified in other mycoplasmas. Two *M. dispar* ATCC27140 spacers (spacer 13 and 19) matched its own genome and the genome of other mollicutes species in the 23S ribosomal RNA gene. Three others spacers (*M. dispar* ATCC27140 spacer 7, *A. palmis* J233 spacer 13, and *M. mobile* 163K spacer 56) were shown to match on two protein encoding genes (respectively: MDS_RS03805: ABC-transporter ATP binding protein, BN854_RS00695: FAD-binding protein and MMOV_RS02220: glycosyltransferase family 2 protein). Surprisingly, while *cas9* gene was predicted disrupted in *A. palmis* J233, CRISPR/Cas systems in *M. dispar* ATCC27140 and *M. mobile* 163K were predicted to be complete and functional.

In order to identify PAM sequences recognized by Cas9 of mollicutes, upstream and downstream sequences of the

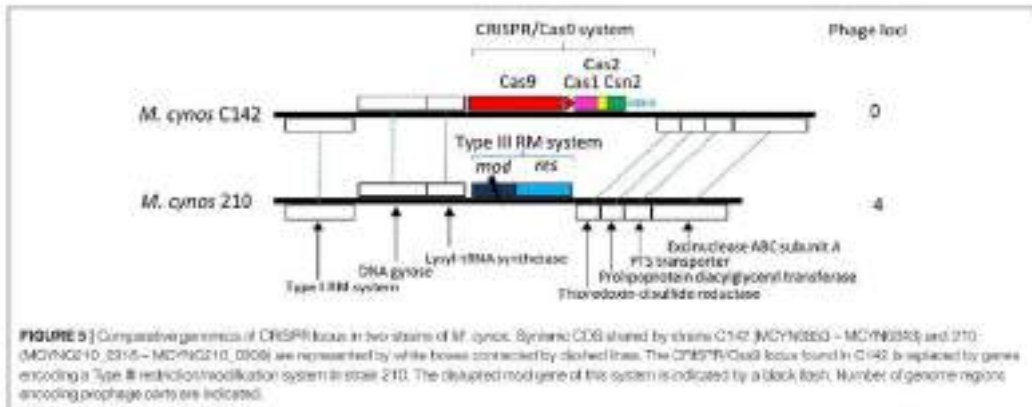


FIGURE 5 | Comparative genomics of CRISPR loci in two strains of *M. cynos*. Syntactic CDS shared by strains C142 (MCYN0263 - MCYN0263) and 210 (MCYN0210_0215 - MCYN0210_0208) are represented by white boxes connected by dashed lines. The CRISPR/Cas0 locus found in C142 is replaced by genes encoding a Type II restriction/modification system in strain 210. The disrupted *mod* gene of this system is indicated by a black box. Number of genome regions encoding prophage parts are indicated.

identified protospacers were retrieved and compared. We did not identify any clear consensus from this set of sequences. A more focused analysis was undertaken using the CRISPR array of *M. cynos* C142 for which five spacers with hits had been found. However, even in this specific case the protospacer adjacent sequences remained highly variable and no clear consensus could be defined.

DISCUSSION

Distribution and Origin of CRISPR/Cas Systems in Mollicutes

CRISPR/Cas systems were found in 21/52 mollicutes species (40%) included in this study, which is a significant proportion considering the fact that some systems might be detected in seemingly CRISPR-lacking species as long as more genomes became available. CRISPR/Cas systems were detected in the genomes of a variety of species belonging to most major phylogenetic branches and infecting various hosts including pigs, ruminants, dogs, birds, fish, and arthropods. However, CRISPR/Cas systems remain undetected in some groups of species, belonging to the same phylogenetic branch or associated with the same host. This is the case of human pathogenic mollicutes infecting pulmonary and urogenital tracts (153 genomes available). This suggests that despite being found in ecological niches shared with a wide diversity of bacteria having CRISPR/Cas systems (i.e., lactobacilli, streptococci, *Acinetobacter*, *Pseudomonas*, *E. coli* ...), these mollicutes have not acquired CRISPR/Cas systems from non-mollicutes bacteria. CRISPR/Cas systems also remained undetected in the ruminant pathogens belonging to the Mycoioides cluster and the *M. agalactiae*/*M. bovis* group (74 genomes available) they were, however, found in other ruminant mycoplasmas, namely *M. capreolensis*, *M. divae* and *M. arginini*. Interestingly, *M. arginini* is frequently co-isolated with other ruminant mycoplasmas and especially *M. bovis* (P. Tardy, personal communication). Therefore, it appears rather surprising that CRISPR/Cas systems do not seem to have spread among all ruminant mycoplasmas. The lack of CRISPR/Cas systems in some specific species might be explained by an ecological niche where viruses are poorly abundant. This might be the case for some intracellular haemotrophic mollicutes (i.e., *Mycoplasma haemovivax*, *Mycoplasma haemofelis*, *Mycoplasma suis*, not shown) as it has been observed in other obligate intracellular bacteria such as *Chlamydiae* (Borstein et al., 2016). However, this possible explanation cannot be invoked for all mollicutes groups lacking CRISPR/Cas systems, some of them showing many traces of phage attacks (see below).

All CRISPR/Cas systems identified in mollicutes were classified as Type II. This finding is in accordance with previous studies showing this type is mainly found in commensal and parasitic species (Borstein et al., 2019). Cas9 phylogeny suggested most CRISPR/Cas systems of mollicutes can be classified as Type II-A, a subtype mostly found in Firmicutes which is in accordance with the common origin of both bacterial phyla (Segata et al., 2013). In mycoplasmas, Cas9 phylogeny

suggests that all CRISPR/Cas systems derive from a common ancestor. However, as CRISPR/Cas systems are prone to circulate via HGT (Jiang et al., 2013), it remains highly speculative to predict such a system was present in mycoplasma last common ancestor. In accordance with the idea that CRISPR/Cas systems might have occasionally spread among mycoplasmas after speciation, Cas9 phylogeny within the mycoplasma branch only partially reflects species phylogeny. For instance, Cas9 from *M. lipofaciens* ATCC35015 clustered with Cas9 from remote species *M. hyosynoviae* NPL1 and *M. spumans* ATCC19256. Moreover, Cas9 from the bird pathogens *M. gallisepticum* S6 and *M. synoviae* S5 clustered in a highly supported branch despite these species being phylogenetically remote. This suggests possible exchange through HGT, as it has been shown for other parts of the genome (Vasconcelos et al., 2005; Strand-Pedersen et al., 2007; Arfi et al., 2016). While our data suggested possible HGT of CRISPR/Cas systems among mycoplasmas, no CRISPR/Cas system was found in associated plasmids, ICE or phages. Therefore, molecular and cellular mechanisms involved in spreading CRISPR/Cas systems among mollicutes remain elusive. We found no evidence of HGT with non-mollicutes prokaryotes. This might result from the particular genetic code used by most mollicutes. For example, with the exception case of *A. palmae* [233 that uses universal genetic code, genes encoding Cas9 in other mollicutes contain up to 18 UGA codons (average 9 UGA) that encode tryptophan but are interpreted as stop codon in other bacteria. Moreover, we observed that Cas9 encoding genes have evolved toward a low-GC content, in accordance with the general bias of mollicutes genomes. These genetic features of mollicutes CRISPR/Cas systems suggest they have been present in this class of bacteria for enough time to adopt the typical genetic particularities of their genomes. Moreover, presence of many UGA codons in cas genes most probably limits the possibility of transfer or at least persistence in other bacteria. In spiroplasma, a phylogenetic tree inferred from Cas9 proteins and comparisons of CRISPR locus organization and Cas9 nuclease domains showed that both Type II-A and Type II-C systems were present in this large group of mollicutes found in arthropods. Furthermore, clear inconsistencies between Cas9 and species phylogenies indicated a complex history of CRISPR/Cas systems in spiroplasma, with several origins and probable HGT. Complementary analyses using additional genome sequences of spiroplasma (extended search, August 2019, 28 genomes available) confirmed the presence of CRISPR/Cas systems from one of the two subtypes in 10 genomes, with Type II-A predicted in *S. litoreale*, *S. helioideae*, *Spiroplasma taraxacum*, *Spiroplasma gladstatoris*, *Spiroplasma taiwanense* and *Spiroplasma alghemense* and Type II-C in *S. apis*, *S. syriphalloicola*, *Spiroplasma clarifii* and *Spiroplasma eriocheiris* (this study and C.H. Kuo, personal communication).

Adaptation of Cas9 and tracrRNA in Mollicutes

CRISPR/Cas systems of mollicutes appear as typical Type II systems with the same set of cas genes including cas1, cas2, cas9 and cas2, the latest lacking in the Type II-C systems found

In *S. apis* B31, *S. syphilidicola* EAI and *Ca. H. crinodetronum* Av. Fine analysis of main players Cas9 and tracrRNA showed some particularities of mollicutes CRISPR/Cas systems. While mollicutes are known to be minute-sized bacteria with minimal and highly compact genomes, Cas9 proteins from mollicutes include all domains described in structurally and functionally characterized homologs with a global length even slightly larger than the closely related SaCas9 from *S. aureus*. Main variations within mollicutes Cas9 and compared to SaCas9 occurred in the WED and PE domains that are involved in PAM duplex recognition. This suggests that PAM specificity is probably different among mollicutes systems and different from the NNGRRR PAM sequence recognized by SaCas9 (Ravi et al., 2013). The PAM specificity of mollicutes Cas9 could not be predicted from the few protospacers identified. Further studies including *in vivo* or *in vitro* assays will be necessary to characterize the functionality of mollicutes Cas9 and their PAM specificities. The global divergence observed among tracrRNA groups and phylogeny of the corresponding bacteria has been interpreted as a resulting from a combination of HGT and convergent evolution (Faure et al., 2015). In the case of mollicutes, the remarkable sequence diversity of tracrRNA is consistent with the fast evolution of these bacteria that has been concluded from many studies and the particularly long branches observed in phylogenetic trees (Ciccotelli et al., 2006; Wu et al., 2009). Analyses of predicted tracrRNA/DR hybrid structures showed many typical features of Type-IIA RNA components in other bacteria. Notable is the wide diversity of length of different stem loops and the absence of bulge in some species whereas it has been shown to be required in other systems. By contrast with Cas9 encoding gene, tracrRNA from mollicutes have significantly higher GC% than the corresponding genome. This suggests that evolutionary constraints to keep RNA hybrid stability (i.e., dG that is not lower than for non-mollicutes systems) are strong enough to prevent decrease of GC%. This trait was also observed for some other highly structured cellular RNAs of mollicutes, including rRNAs (Sirand-Pugnet et al., 2007).

Impact of CRISPR/Cas Systems in the Control of Phages and Mobile Genetic Elements

One of the main questions about CRISPR/Cas systems in the bacterial world is their efficacy in controlling invading DNAs, and the evolutionary benefits and costs of such systems. This question is particularly interesting for mollicutes which evolution is mainly marked by significant genome down-sizing, a global streamlining of the basic functions for cell life and general adoption of a host-dependent lifestyle. To the best of our knowledge, mollicutes are the bacteria with the smallest genome where CRISPR/Cas systems have been found. Some endosymbiotic bacteria from the *Proteobacteria* (i.e., *Buchnera*, *Ca. Tremblaya*, *Ca. Hodgkinia*, and *Ca. Carsonella*...) and *Bacteroidetes* (i.e., *Ca. Sibiria* and *Ca. Uzinara*) phyla have genomes even smaller than mollicutes. Besides being even more dependent on their host metabolism, such bacteria seems to be more protected against potential invasive DNA, as suggested by the lack of prophages, mobile

elements and also by the absence of defense systems such as CRISPR/Cas and RM in most of them (Touchon et al., 2016). In mollicutes, most genomes have traces of mobile elements including prophages, IS, ICE, and plasmids, suggesting they are still subject to MGE-carried fluxes of genetic information (Supplementary Material SM2 and Supplementary Table S6). Some genomes as *M. mycoides* subsp. *mycoides* PG1 (IS represent 13% of the genome) (Westberg et al., 2004), *Ca. phytoplasma asteris* Onion Yellow strain (8.6% are potential mobile unit) (Bai et al., 2006), *M. agalactiae* 5632 (ICE and IS represents > 10% of the genome) (Nouvel et al., 2010) are particularly illustrative of the significant impact of these invading DNAs in reshaping mollicutes genomes. Substantial viral invasion in some specific phyla of spiroplasma was concluded from the genome analysis of *S. citri* GH-3 (> 20% genome are prophage remnants) (Carle et al., 2010) and related species *S. hunkahii* (Hal and Hegermann, 2002), *S. molijerum* (Lo et al., 2013), *X. peaufonii* (Dreves et al., 2015) while typical plectroviral sequences were not found in other spiroplasma phyla (Ku et al., 2013). Because of the specific genetic code used by most mollicutes, circulation of MGE is most probably favored within this group of bacteria. This has been suspected from phylogenomic studies indicating some MGE had been recently exchanged among species sharing the same host, including IS among human mollicutes from the urogenital tract (Pereyre et al., 2009), IS and ICE among ruminant mycoplasmas (Thomas et al., 2005; Cribi et al., 2018) and ICE-related elements in some spiroplasma (Lo et al., 2013). In parallel to this evolutionary trait, highly variable repertoires of genes encoding defense systems in mollicutes genomes suggest that most of these minimal bacteria undergo a selective pressure toward maintaining means to control invading DNAs. However, we did not find a clear correlation in the distribution of CRISPR/Cas systems and mobile elements.

A global survey indicated that nearly half of the bacterial genomes contained at least one prophage but that prophages are rare in small bacterial genomes (Touchon et al., 2016). Accordingly, our survey indicated traces of prophages in about 16% of mollicutes genomes. Surprisingly, lysogen bacteria are statistically slightly more likely to have CRISPR/Cas systems (52%) than non-lysogens (43%). More precisely, Type II CRISPR/Cas systems are as frequent in lysogen and non-lysogen bacteria, suggesting these systems might not be as efficient as previously thought in immune defense against phages. Despite the mollicutes genomes with prophages and CRISPR/Cas being currently too rare to get a real picture, this tendency might be also observed in mollicutes, with 4 out of 5 lysogen species having a complete or incomplete CRISPR/Cas system. Therefore, as more genomes become available, it will be interesting to re-evaluate the tendency of CRISPR/Cas systems to occur in mollicutes phyla where traces of phage attacks can be evidenced.

Persistence of Degraded CRISPR/Cas System in Mollicutes

Inactivation and more or less advanced degradation of CRISPR/Cas systems have been predicted for at least 8 out of 21 CRISPR/Cas systems of mollicutes (38%). Overall,

our data suggest that CRISPR/Cas systems were frequently acquired during the evolution of main mollicutes groups but also frequently inactivated and possibly completely lost in several phyla. This assumption is in accordance with other studies proposing that fast evolution of CRISPR loci including acquisition by HGT, expansion of CRISPR arrays, degradation and complete loss might be driven by opposing positive and negative selection for immunity keeping CRISPR/Cas systems in a continuous state of flux (Jiang et al., 2013). CRISPR/Cas systems would be acquired in environments where phages are a major source of mortality while being quickly inactivated and lost when the benefit of the protection system does not counterbalance the cost of expressing constantly the Cas proteins and reducing the possibility to acquire new genetic information by HGT. This would make sense in the context of minimal bacteria whose most cellular energy is likely devoted to translation, and where HGT has been shown to play a significant evolutionary role at least for some groups.

In addition, it has been proposed that accidental incorporation of self-targeting spacers might be as frequent as 18% in CRISPR-containing prokaryotes (Stern et al., 2010). This could incur an autoimmune fitness cost, and therefore help to explain the abundance of degraded CRISPR/Cas systems across prokaryotes. Self-targeting spacers were found in some CRISPR arrays in several mollicutes species (i.e., *M. dispar* ATCC27140, *M. mobile* 163K, *M. gallisepticum* 56 and *A. palmarum* J233) suggesting avoidance of self-targeting is not perfect in these species, as in many other bacteria. Non-functional CRISPR/Cas systems were predicted for *A. palmarum* J233 and *M. arbeitsidi* 158L3_1 where mutations in *Cas9* were detected, providing an explanation for the survival of the bacteria. However, for *M. dispar* ATCC27140, *M. gallisepticum* 56 and *M. mobile* 163K, the CRISPR/Cas systems were predicted to be functional. Therefore, our current hypothesis for these cases is that PAM sequences might be mutated avoiding target cleavage. In minimal bacteria as mollicutes, persistence of non-functional CRISPR/Cas systems remains somewhat puzzling as genome streamlining appears to be a strong evolutionary force. Interestingly, in the best documented example currently available for studying the evolution of CRISPR locus, *M. gallisepticum*, fast degradation of the locus was observed in some isolates after host shift from poultry to House Finches (DeLaney et al., 2012; Tulman et al., 2012). As a growing number of genomes becomes available for different species, we will probably soon learn more on the acquisition and persistence of CRISPR/Cas systems of mollicutes and their biological importance in the fast evolution of these minimal bacteria.

REFERENCES

Arif, Y., Minder, I., Pirota, C., Di Ray, A., Le Béd, C., Despat, L., et al. (2016). MR-REP is a mycoplasma system that captures and cleaves transposons. *G. Proc. Natl.Acad.Sci. U.S.A.* 113, 5436-5441. doi:10.1073/PNAS.1605460113

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. Genome data can be found at GenBank, under the following accession numbers: GCA_000731685.1, NC_007633.1, NC_014751.1, NC_005364.2, NC_015431.1, NZ_ANFU00000000.1, GCA_000380285.1, NC_021083.1, NC_006055.1, NC_021998, NZ_CP017015, NC_021846.1, NZ_CP012357.1, NC_021284.1, NZ_CP002082.1, NZ_ADRH00000000.1, NC_014760.1, NC_009497.1, NC_014552.1, NZ_JMKY0100006.1, NZ_HJEED0000000.1, NC_014014.1, NC_007294, NC_019949, NZ_CP014281.1, NC_002771.1, NC_006360.1, NZ_CP007129, NZ_JAKV91000001.1, NC_014448.1, GCF_000485555.1_ASM48555v1, JFKL01000035.1, GCA_000367765.1, NZ_AP014657_HVFB1000003, NC_013511.1, NC_011025, NC_006908, CP000932.1, GCA_000227555.2, NC_004432.1, NC_002162.1, NC_000912.1, NC_000908.2, NC_023030, NC_011047.1, NC_010544.1, NC_005303.2, NC_022538, NC_010163.1, NC_022549.1, NC_008532.1, NC_008533.1, NC_006368.1, NC_003737.2, NC_013198.1, NC_003454.1, AECED01000042.1, NC_012781.1, NC_002163.1, NC_003116.1, NC_002663.1, GCA_000185485.1, and NC_008901.1.

AUTHOR CONTRIBUTIONS

TI and PS-P designed the study. TI, IT, VT, CG, AM, CW, PT, and PS-P performed the analyses and analyzed the data. TI, IT, CW, JB, AB, PT, and PS-P wrote and improved the manuscript.

FUNDING

This research was supported by the National Science Foundation (Grant Number IOS-1110151). Additional support was received from the INRA and University of Bordeaux.

ACKNOWLEDGMENTS

The authors thank Chih-Hong Kuo and Florence Tardy for sharing knowledge and thoughts about this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.02701/full#supplementary-material>

Bai, A., and Hegdebar, S. A. (2006). A genome sequence survey of the mollicute *com. strain* *Spiroplasma furciferi*. *FEBS Microbiol. Lett.* 216, 7-12. doi:10.1111/j.1574-6968.2001.tb17151.z

Bai, X., Zhang, J., Feing, A., Miller, C. A., Janusz Babik, A., Shostakovskii, D. V., et al. (2006). Living with genome instability: the adaptation of *Spiroplasma*

- in diverse environments of their insect and plant hosts. *J. Bacteriol.* **188**, 3672–3686. doi: 10.1128/JB.118.10.3672-3686.2006
- Banasinski, E., Dworkin-Flesher, E., Saghai, E., Hygenberg, M.-C., Prete, G., Ghemal, S., et al. (2018). The integrative conjugative element (ICE) of *Mycoplasma agalactiae* key elements involved in horizontal dissemination and influence of coexistent ICEs. *MSystems* **9**, e0075–e018. doi: 10.1128/mSystems.00973-18
- Barré, A., de Dierckx, A., and Wainchund, A. (2010). *molSigs*, a database dedicated to the comparative genomics of mollicutes. *Nucleic Acids Res.* **38**, D387–D390. doi: 10.1093/nar/gph114
- Basson, D. A., Cavanagh, M., Clark, K., Karim-Mirzaei, I., Lyman, D. J., Oriel, J., et al. (2017). GenBank. *Nucleic Acids Res.* **45**, D37–D42. doi: 10.1093/nar/gkx097
- Berthelin, A., Nkanti, D., Youcher, M., and Rocha, E. P. (2018). Co-occurrence of multiple CRISPRs and *cas* genes suggest episodic insertions. *BioRxiv [Preprint]*. doi: 10.1101/292900
- Beskers, D., Sun, C. L., Brown, C. T., Stamm, L., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* **7**, 10613. doi: 10.1038/ncomms10613
- Carls, P., Sallsted, C., Corvino, M., Corvino, S., Duret, S., and Tarkenton, S., et al. (2013). Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral insertion and impact on gene decay. *Appl. Environ. Microbiol.* **79**, 3426. doi: 10.1128/AEM.01951-12
- Cajkovic, K., Makarova, K. S., Charpentier, E., and Koonin, E. V. (2016). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6101. doi: 10.1093/nar/gkv241
- Cavonius, J. D., Dworkin, T., von Merzig, C., Cronley, C. J., Seal, B., and Barré, A. (2016). Toward accurate reconstruction of a highly recombined tree of life. *Science* **351**, 1283–1287. doi: 10.1126/science.1273981
- Choi, C., Dworkin-Flesher, E., Norvell, L. X., Rao, C., and Banasowski, E. (2018). Horizontal gene transfer in *Mycoplasma* (Mollicutes). *Curr. Opin. Microbiol.* **39**, 3–11. doi: 10.1016/j.cmi.2018.07.003
- Coariza, D., Becardain, A., Tefarik-Nocka, C., Yeadon, M., Michálek, I., Némec, S., et al. (2018). CRISPR/CasVinder, an update of CRISPRfinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W216–W223. doi: 10.1093/nar/gky125
- Chakrav, N. P., Hedges, S., Hernandez, G., Mars, C. L., Hill, D. H., Ferguson-Nixon, N., et al. (2012). Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS Genet.* **8**(10):e1002511. doi: 10.1371/journal.pgen.1002511
- Delcher, E., Chylinski, K., Szymca, C. M., Grotzels, G., Cava, T., Fenala, Z. A., et al. (2011). CRISPR RNA maturation by trans-processed small RNA and host factor RNase III. *Nature* **471**, 607–607. doi: 10.1038/nature09886
- Devezer, A., Gagnon, V., Blac, G., Audic, S., Buffet, S., Charpentier, E., et al. (2008). Phylogeny-independent phylogenetic analysis for the non-specificist. *Nucleic Acids Res.* **36**, W460–W469. doi: 10.1093/nar/gln180
- Dworkin-Flesher, E., Saghai, T., Banasowski, E., Broten, M., Norvell, L. X., Blanchard, A., et al. (2014). Chromosomal transfer in *Mycoplasma* when minimal genomes go mobile. *MSystems* **5**, e00938. doi: 10.1128/mSystems.00938-14
- Dworkin, J. A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR/Cas9. *Science* **346**, 1258096. doi: 10.1126/science.1258096
- Eason, G., Banasinski, E. A., Makarova, K. S., Wolf, Y. I., Casadevall, A. B., Banasinski, E., et al. (2015). Comparative genomics and evolution of trans-activating CRNAs in Class 2 CRISPR-Cas systems. *RNA Biol.* **16**, 435–448. doi: 10.1080/15476281.2014.1003331
- Fredin, L., Le Rhun, A., Chylinski, K., Makarova, K. S., Leclercq, A. L., Brodeur, J., et al. (2014). Phylogeny of Cas9 determines functional interchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 2577–2596. doi: 10.1093/nar/gkt1874
- Gaspard, G. E. (2014). "The family *erythromycesaceae*," in *The Prokaryotes*, ed. E. F. DeLong, S. Lory, E. Stackman, and F. Thompson. (Heidelberg: Springer Berlin Heidelberg), 469–504
- Gerasimov-Dimitrova, E., Liu, Y., Youstov, I., and Schmidt, B. (2016). MSAProvB (MSP) parallel multiple sequence aligner for distributed-memory systems. *Bioinformatics* **32**, 2626–2628. doi: 10.1093/bioinformatics/btw038
- Geis, M., Grotzels, S., and Gascard, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224. doi: 10.1093/molbev/mtp259
- Gilua, L., Vergnaud, G., and Paunou, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**(17). doi: 10.1186/1471-2105-8-17
- Groves, B., Dworkin, M., Vissel-Bagnat, P., Tsody, E., Thiesenauer, F., Choi, C., et al. (2014). Predicting the minimal translation apparatus lessons from the reductive evolution of mollicutes. *PLoS Genet.* **10**(04):e1004161. doi: 10.1371/journal.pgen.1004161
- Gunders, S., Dalgaard, L.-F., Leber, V., Anthonisen, M., Hordijk, W., and Gascard, O. (2018). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321. doi: 10.1093/sysbio/syx010
- Hill, P., Richter, H., Wang, S. P., Brinson, M., Bassel, S., and Charpentier, E. (2018). The biology of CRISPR-Cas: hallowed and honored. *Cell* **171**, 1239–1259. doi: 10.1016/j.cell.2017.11.002
- Hing, W., Martin, L., Anelli, F., Wang, Y., Levin, B. R., and Marzocchi, L. A. (2013). Dealing with the evolutionary diversity of CRISPR immunity in bacteria and beneficial plasmids. *PLoS Genet.* **9**(10):e1003441. doi: 10.1371/journal.pgen.1003441
- Ivick, M., Chylinski, K., Fredin, L., Bass, M., Dworkin, J. A., and Charpentier, E. (2013). A programmable dual-ENA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821. doi: 10.1126/science.1229229
- Johansen, M., Zaslavskii, L., Rappelli, T., Miroshnik, V., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9. doi: 10.1093/nar/gkn121
- Katoh, K., Rozewicz, J., and Tamada, E. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1168. doi: 10.1093/bib/bbx010
- Kim, G. I., and Doudna, J. A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 896–899. doi: 10.1126/science.1310111
- Koonin, E. V., Makarova, K. S., and Zhang, S. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–76. doi: 10.1016/j.cmi.2017.03.008
- Ku, C., Lu, W.-S., Chen, L.-L., and Kao, C.-H. (2013). Complete genomes of two *Spiroplasma* associated bacteriophages provided insight into the origin, dynamics, and impact of viral insertion in *Spiroplasma*. *Genome Biol. Evol.* **5**, 1151–1161. doi: 10.1093/gbe/abt084
- Lango, S. J., Alkharbush, O. S., Rose, D., Will, S., and Buchstein, B. (2013). CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.* **41**, 8014–8044. doi: 10.1093/nar/gkt590
- Lu, W.-S., Chen, L.-L., Cheng, W.-C., Gagnon, G. U., and Kao, C.-H. (2013). Comparative genome analysis of *Spiroplasma volutum* DSM 9886, a bacteriophage-associated bacterium. *BMC Genomics* **14**(22). doi: 10.1186/1471-2105-14-22
- Lu, W.-S., Gagnon, G. U., and Kao, C.-H. (2015). Inward and Inset: the fate of horizontally acquired genes in *erythromyces* *Spiroplasma*. *Genome Biol. Evol.* **7**, 2438–2472. doi: 10.1093/gbe/evb160
- Marini, M., Mincione, C., Liu, L.-M., and Ferris, G. (2014). "The family *erythromycesaceae* (including *phytoplasmaceae*)" in *The Prokaryotes*, ed. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackman, and F. Thompson. (Heidelberg: Springer Berlin Heidelberg), 469–504
- May, M., DeGisi, M. P., and Blaserud, A. (1974). "The order *erythromycesales*," in *The Prokaryotes*, ed. E. F. DeLong, S. Lory, E. Stackman, and F. Thompson. (Heidelberg: Springer Berlin Heidelberg), 515–530
- Miao, M. S., Samayra, T., Hagenbush, S. A., Harrison, M., Cho, S.-T., and Kao, C.-H. (2019). The genome of *Yersinia* *phytoplasmata* strain SA-1 is highly dynamic and prone to adapting foreign sequences. *Appl. Appl. Microbiol.* **41**, 117–127. doi: 10.1002/oaapm.2018.01008
- Nishimura, H., Gong, L., Yan, W. X., Bao, F. A., Zeng, B., Li, Y., et al. (2015). Crystal Structure of *Spiroplasma* *myces* Cas9. *Cell* **162**, 1113–1126. doi: 10.1016/j.cell.2015.04.007
- Nishimura, H., Bai, F. A., Wu, P. D., Kurokawa, S., Shibata, S. Y., Dohma, N., et al. (2014). Crystal structure of *cas9* in complex with guide rna and target DNA. *Cell* **156**, 935–940. doi: 10.1016/j.cell.2014.02.001
- Norvell, L. X., Stamm-Pagnat, P., Maranda, H. S., Saghai, E., Karim, V., Margoset, E., et al. (2010). Comparative genomics and proteomic analyses of two *Mycoplasma agalactiae* strains: clues to the macro- and micro-evolution that are shaping *erythromyces* diversity. *BMC Genomics* **11**(66). doi: 10.1186/1471-2105-11-66

Dandekar, J. C., Herrero, J. L., Schröpfer, F., Muto, K., Claverol, S., Esco, C. H., et al. (2015). Genome sequence of the *Streptococcus melanosphaera* milk-killing *Streptococcus* strain MS803 endosymbiont. *MBio* 6:e02437. doi: 10.1128/mBio.02407-1434

Freyer, S., Strand-Pedersen, P., Jensen, L., Charrier, A., Renaudin, H., Baril, A., et al. (2006). Life on arginine for *Mycoplasma bovis*: clues from its minimal genome and comparison with other known minimal mycoplasmas. *PLoS Genet* 2:e1000477. doi: 10.1371/journal.pgen.1000477

Hsu, T. A., Cong, L., Yin, W. X., Scott, D. A., Greenberg, J. S., Kriz, A. J., et al. (2015). In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 526, 188–191. doi: 10.1038/nature14299

Segawa, L. B. (2014). "The family Mycoplasmatinae," in *The Prokaryotes*, eds E. E. Neilson, S. Levy, I. Stachurski, and J. Thompson, (Springer: Berlin Heidelberg), 561–567.

Robert, B. J., Vacim, T., Postle, J., and Maclell, D. (2015). EEBASE—4 database for DNA restriction and modification enzymes, genes and genomes. *Nucleic Acids Res.* 43, D286–D289. doi: 10.1093/nar/gku336

Segata, N., Bertone, D., Margat, E. C., and Clemente, C. (2013). PhyloPhlAn: a new method for improved phylogenetic and taxonomic placement of microbes. *Nu. Genome* 4:254. doi: 10.1093/ng/gkt034

Spacie, P. (2006). 158bp: the reference code for bacterial insertion sequences. *Nucleic Acids Res.* 34, D61–D66. doi: 10.1093/nar/gkj014

Strand-Pedersen, P., Olsen, C., Benn, A., and Bendtsen, A. (2007). Evolution of modification: down a bumpy road with *in vivo* and *in vitro*. *Res. Microbiol.* 158, 754–766. doi: 10.1016/j.resmic.2007.09.007

Ising, J., Rogge, A., and Lipus, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi: 10.1093/nar/gki408

Sherr, A., Keres, I., Ferrer, O., Amadi, G., and Serik, E. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26, 389–396. doi: 10.1016/j.tig.2010.05.009

Salavara, C., and Castresna, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *BMC Evol. Biol.* 7, 364–377. doi: 10.1186/1471-2147-7-364

Tardy, F., Mich, V., Darda-Prison, E., Marenca, M. S., Strand-Pedersen, P., Hauchard, A., et al. (2015). Integrative conjugative elements are widespread in field isolates of *Mycoplasma agalactiae* pathogens for ruminants. *Appl. Environ. Microbiol.* 81, 1634–1641. doi: 10.1128/AEM.01723-1434

Thomas, A., Linden, A., Mairal, J., Bickel, D. F., Frey, J., and Vlot, H. M. (2001). *Mycoplasma bovis* strain insertion sequences with *Mycoplasma agalactiae* and *Mycoplasma mycoides subsp. mycoides SC* evolutionary and developmental signals. *FEBS Microbiol. Lett.* 240, 249–253. doi: 10.1016/S0168-2267(01)03013

Toranzo, M., Berthelin, A., and Rocha, E. P. (2016). Genetic and Bin-Library traits associated with the distribution of prophages in bacteria. *ISME J.* 10, 2746–2754. doi: 10.1038/ismej.2016.47

Tsai, Y.-M., Chang, A., and Ross, C.-H. (2010). Horizontal gene acquisition contributed to genome expansion in insect-symbiotic *Mycoplasma chikil*. *Genome Biol. Evol.* 10, 1526–1532. doi: 10.1093/gbe/evy113

Tufman, D. R., Liu, X., Sacrepanek, S. M., Lee, D. H., Kirsch, G. E., and Geary, S. J. (2012). Extensive variation in surface lipoprotein gene content and genomic changes associated with virulence during evolution of a novel north american house finch epidemic strain of *Mycoplasma gallisepticum*. *Microbiology* 158, 2073–2084. doi: 10.1099/mic/0/000560-0960

Vasconcelos, A. T. R., Ferreira, H. B., Diniz, C. V., Batista, S. L., Carvalho, M. G., Pinto, P. M., et al. (2015). *Yersinia* and poultry pathogenic: the complete genome sequence of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma mycoides*. *J. Bacteriol.* 187, 5085–5077. doi: 10.1128/JB.187.16.5084-5077.2015

Wastberg, J., Penase, A., Holmberg, A., Gostmann, A., Lundberg, J., Johansson, K. E., et al. (2004). The genome sequence of *Mycoplasma mycoides subsp. mycoides SC* type strain PG11, the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Res.* 14, 221–227. doi: 10.1101/gr.247384

Wu, D., Hagerlitsch, P., Mironov, K., Pikaál, R., Dalia, G., Ivanova, N. N., et al. (2009). A phylogeny-driven genomic encyclopedia of Bacteria and Archaea. *Nature* 462, 1094–1098. doi: 10.1038/nature08696

Yanaki, M., Wakabayashi, Y., Gostenberg, I. S., Hirano, H., Fan, F. A., Nakano, T., et al. (2017). Crystal structure of the minimal Cas9 from campylobacter jejuni reveals the molecular diversity in the CRISPR-Cas9 systems. *Mol. Cell* 65:1199–1211.e5. doi: 10.1016/j.molcel.2017.02.007

Zhou, Y., Jiang, Y., Lynch, K. H., Dennis, J. L., and Waters, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W357–W362. doi: 10.1093/nar/gkr443

Zúñiga, M. (2005). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 33, 1426–1435. doi: 10.1093/nar/gki494

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Josticia, Euzepopoulos, Tselentis, Gajjar, Mavrou, Walker, Kocovic, Bouchard, Thebaud and Strand-Pedersen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

3. Données supplémentaires

(21/11/2019 ; Front Microbiol. 2019 ; 10: 2701, PMID: 31824468)

Supplementary Material

SM1. Key amino acids involved in gRNA and target recognition are conserved in Cas9 proteins of mollicutes and *S. aureus*

Comparison of various Cas9 proteins and crystal structures of Cas 9 of *S. pyogenes* and *S. aureus* have evidenced a global organisation with two endonuclease domains. The HNH domain cleaves the DNA strand complementary to the guide RNA whereas the RuvC domain cleaves the non-complementary strand (Nishimasu *et al.*, 2015). Structural studies have shown that SpCas9 and saCas9 bound to single guide RNA adopt a flexible bilobed structure consisting in a recognition lobe (REC) and a nuclease lobe (NUC) with a channel wrapping the RNA-DNA duplex (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014, 2015). The C-terminal part of the Cas9 proteins includes the PAM-interacting domain (PI) which interaction with the PAM sequence located on the non-target strand of DNA is required for the initiation of the cleavage process. While they both exhibit the same general domain organisation, SpCas9 and SaCas9 share only 17% similarity and have significantly different sizes, 1368 aa and 1053 aa, respectively. They also have different PAM specificity, SpCas9 recognizing NGG and SaCas9, NNGRRT (R represents A or G). Finally, fine structure comparison revealed some noticeable differences including interacting regions of REC lobe with sgRNA, of PI domain with the PAM sequence and a highly divergent WEDGE (WED domain).

Based on multiple alignment of the Cas9 proteins and SaCas9, conservation of key amino acids was manually inspected (Figure S4).

In the following description, numbering of the amino acids refers to SaCas9 (Nishimasu *et al.*, 2015). Among the SaCas9 amino acids bound with gRNA and target DNA, most conserved ones (identical or similar) in Cas9 from mollicutes were involved at three levels:

Recognition of the heteroduplex formed by pairing of the gRNA with the target DNA. Most conserved residues include Tyr651 (RuvC domain), Trp229, Gly235, Thr392 and Arg245 (REC lobe) that were shown to interact with the target DNA strand. Many amino acids known to be involved in recognition of the paired gRNA were also well conserved. Those are part of the Bridge helix (Arg48, Arg51, Arg55, Arg59, Arg60) and the REC lobe (Arg209, Arg116, Gly117, Arg165). Some SaCas9 residues from the REC lobe interacting with the gRNA nucleotides paired with the non-PAM extremity of the target DNA strand were replaced by similar residues in most mollicutes, including Thr238 (Ser in most mollicutes) and Tyr256 (Phe in most mollicutes).

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

Interaction with the nexus stem-loop formed by self folding of the tracrRNA part of the gRNA. Conserved amino acids includes Arg47, Arg55, Arg54, Arg58 (Bridge helix) and Ser219, Gly216 and Arg209 (REC lobe). The Phosphate lock loop amino acids Asn780 and Leu783 that are involved in the recognition of the first Adenine nucleotide (A55) downstream of the lower stem of the repeat-antirepeat duplex are highly conserved within Cas9 of mollicutes.

Binding to the Repeat-antirepeat duplex. Some of the SaCas9 amino acids involved were also conserved in mollicutes Cas9, including Lys114, Gly162, Arg165 (REC lobe) and Lys878, Lys881 (WED domain). However, some other residues involved in this interaction were only poorly conserved, especially those bound to the gRNA at the bottom of the lower stem-loop.

SM2. Distribution of CRISPR systems in mollicutes is not correlated with presence of MGE and other defense systems

In order to investigate the global impact of CRISPR systems against phages and other mobile genetic elements (MGE) during the evolution of mollicute genomes, repertoires of MGE including prophages, IS, ICE, plasmids were predicted in the 52 genomes set. In addition, non-CRISPR defense systems against invading DNAs were also listed, namely, RM and Abi systems (Table S6).

Among MGE, IS were the most frequent, with 86% of the selected 52 genomes harboring elements of one or several IS families. IS were predicted in all groups of mollicutes and in most genomes, with the exception of *Me. florum* L1, *S. taiwanense* CT-1, *M. crocodyli* MP145, *M. mobile* 163K, *M. genitalium* G37 and *M. pneumoniae* M129. Complete and potentially degraded ICEs were found in 35 % of the selected genomes, widespread in the Spiroplasma and Hominis phylogenetic groups while not predicted in the mollicutes from the Pneumoniae and Acholeplasma/Phytoplasma groups. Large Potential Mobile Units (PMU) have only been described in phytoplasmas. Plasmids and prophages were more rarely detected, in 11 % and 13% of the selected genomes, respectively. With the exception of a putative prophage remnant in *U. parvum* serovar 3 ATCC 700970, these MGE were not observed in the genomes belonging to the Pneumoniae group. Regarding defense systems against invading DNAs, the most massively prevalent ones are the RM systems. They were predicted in 96% of the selected genomes, with the notable exception of *S. taiwense* and *S. mirum*. Type I and II RM systems were more frequently identified than Type III and Type IV which were only found in 4 genomes. Besides CRISPR systems, abi systems were also detected in mollicutes, but only in 5 genomes distributed in the Hominis group and in the mycoides cluster.

Statistical analyses conducted to evaluate whether some correlation or anti-correlation could be found between the distribution of MGE and defense systems only indicated correlations for the

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

occurrences of IS and RM systems (kendall's $\rho = 0.258$) and IS and prophages (kendall's $\rho = 0.212$). No statistically significant correlation could be obtained between CRISPR systems and other MGE or defense systems. However, out of five mollicute genomes where potentially complete prophages were detected, three had uncomplete CRISPR systems (*M. arginini* HAZ145_1, *M. spumans* ATCC19526, *M. arthritidis* 158L3-1) and one had a potentially complete CRISPR (*M. lipofaciens* ATCC 35015). For this last case, examination of the spacers present in the CRISPR array showed no similarity with the prophage sequence.

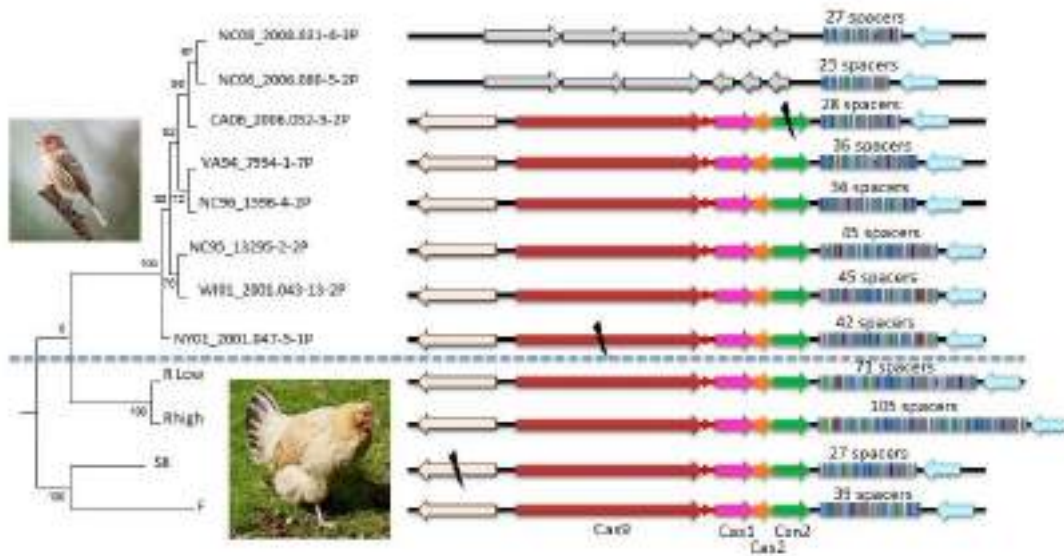


Figure S2. Comparative genomics of CRISPR locus in 12 strains of *Akkermansia mucosiphila*. Phylogenetic tree of the 12 strains was downloaded from MGD database (<http://mbdb.asm.org/>). The eight strains isolated from American House Finches are clustered in a statistically supported branch (upper part of the tree), whereas the four strains isolated from adults are distributed in two branches (lower part of the tree). A complete CRISPR locus was found in all strains except MCD6_2008 and MCD6_2006 where all core genes have been lost. Depending on the strain, CRISPR arrays included 27 to 105 spacers. Orthologous genes are represented as same colored arrows. Disrupted genes are indicated with black slashes. This figure is based on the work of Tullner et al., 2012.

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

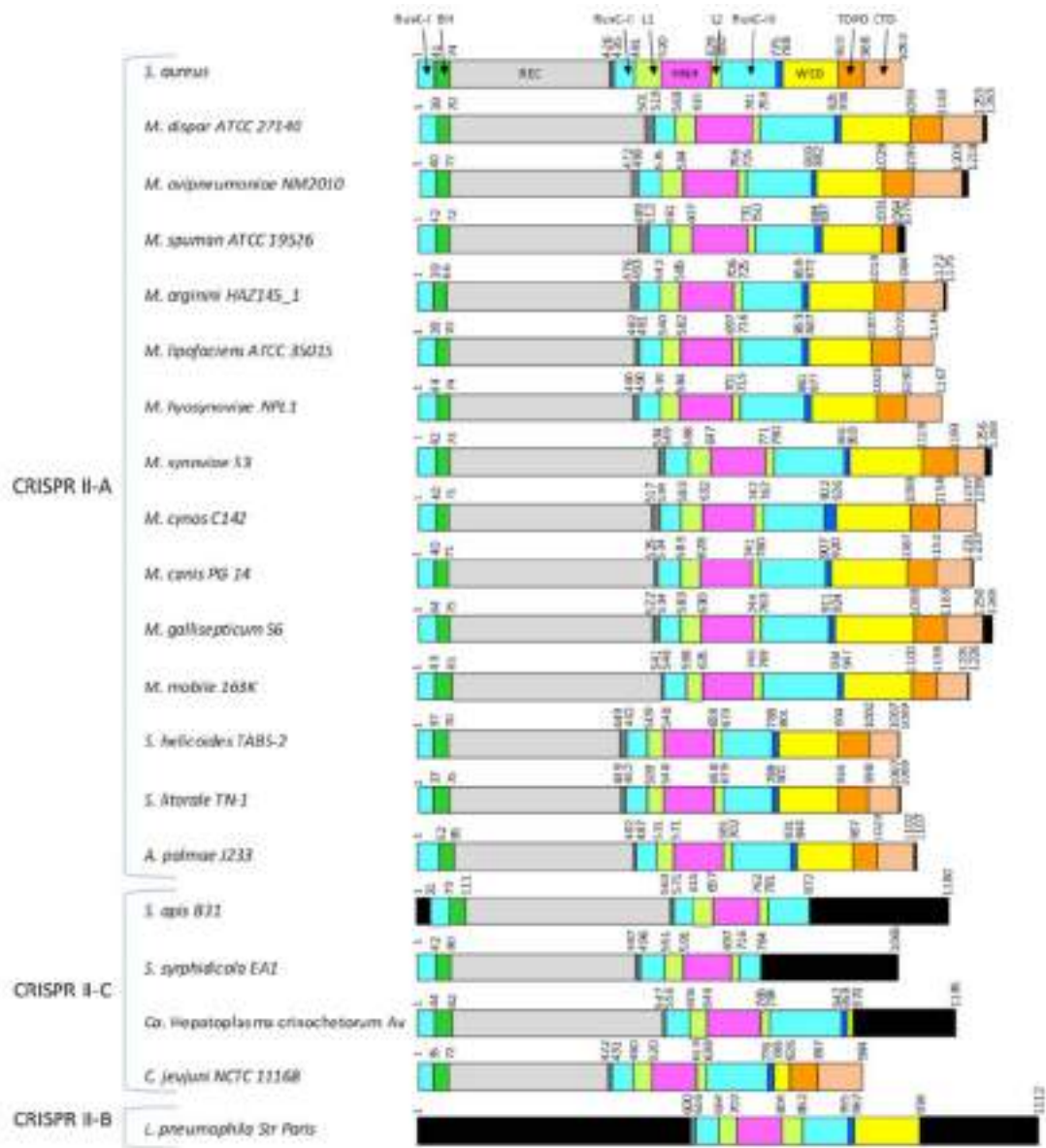
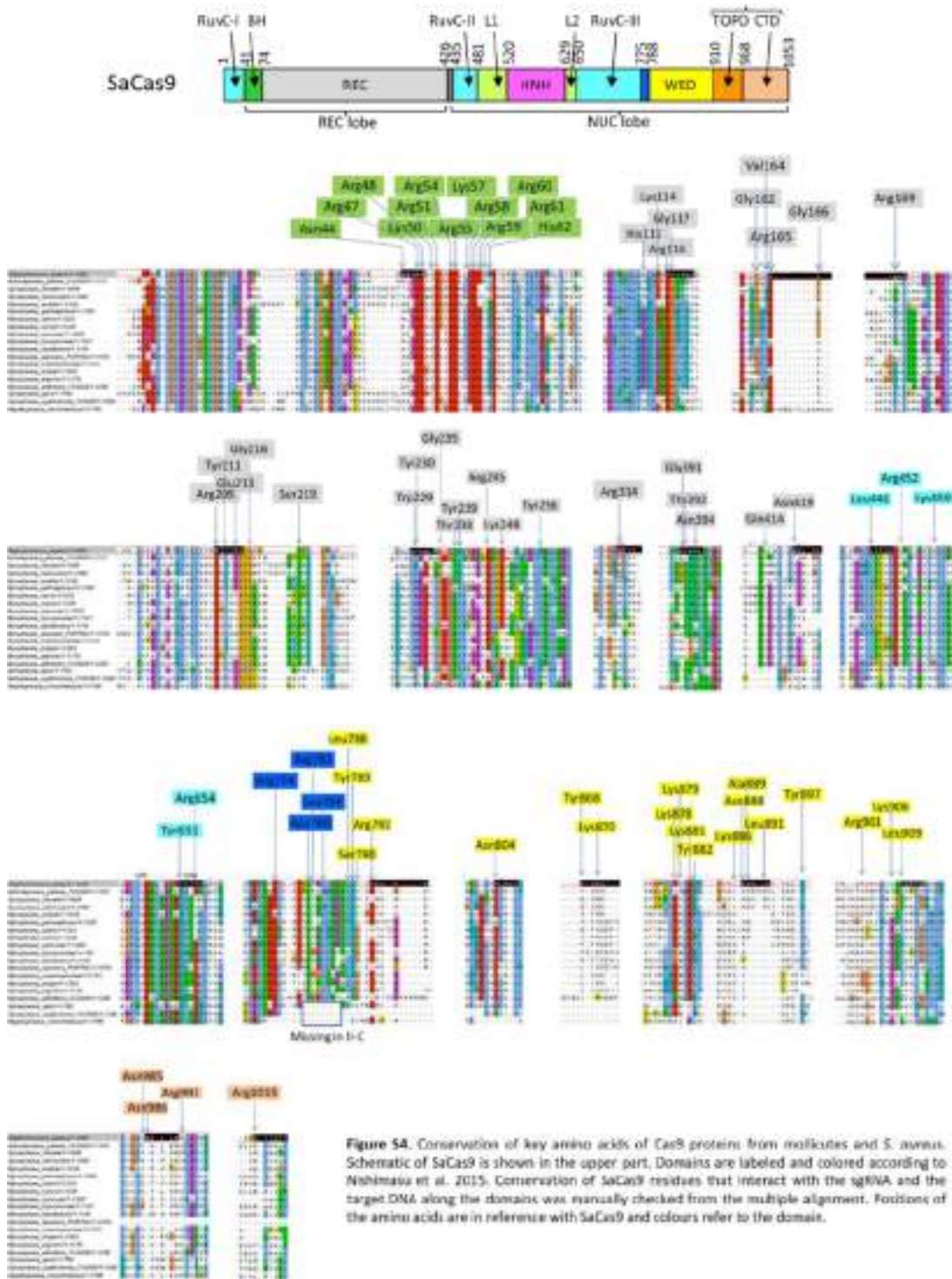


Figure S3: Domain structure of Cas9 proteins from mollicutes and reference bacteria. Domains were inferred from HHpred comparisons and positions are numbered according to SaCas9 positions described in Nashimasu 2015. Black, unaligned region using HHpred.

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes



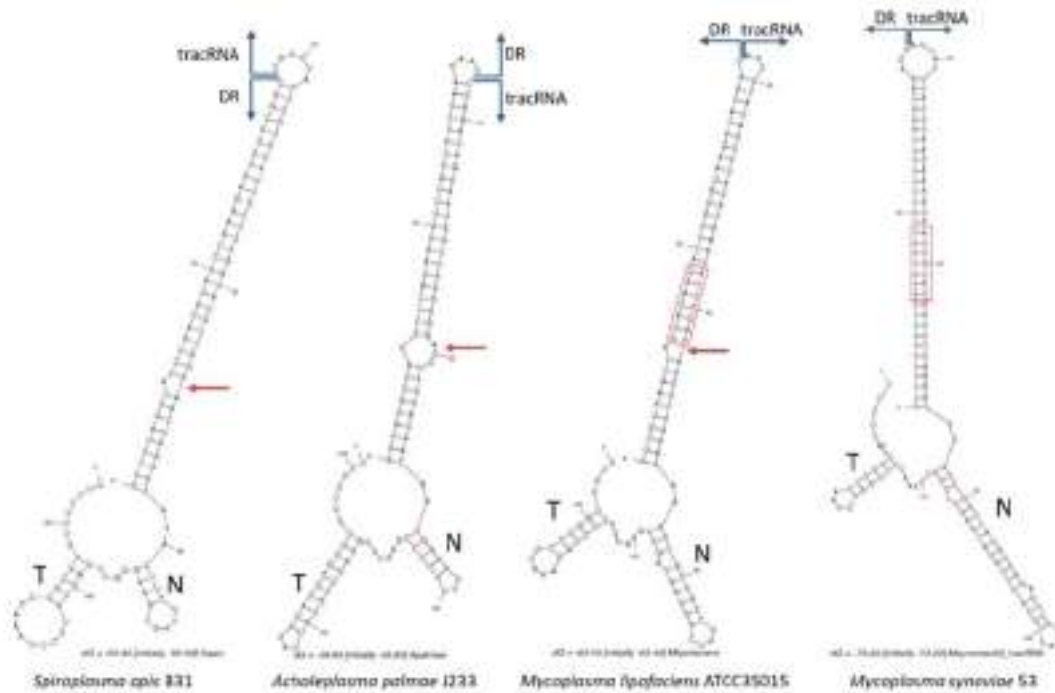


Figure 55. Predicted DR/tracrRNA hybrid secondary structure. Sequences of DR and tracrRNA were concatenated and the secondary structures of the hybrids were simulated using mfold software at <http://unafold.rna.albany.edu/>. Position of the DR/tracrRNA concatenation is indicated by divergent arrows. N, nexus ; T, terminator stem-loops. Bulges connecting upper and lower stems are indicated by red arrows. DR and tracrRNA sequences were defined based on Chylinski 2014 and our own work. The base of the upper stem loop that is highly conserved in mycoplasmas is framed in red.

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

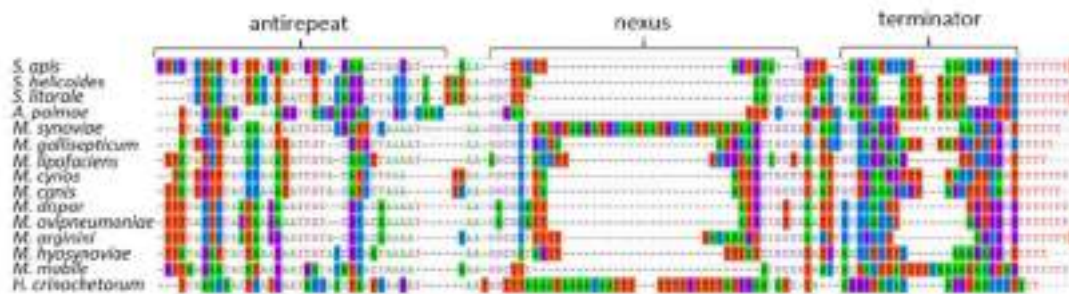


Figure S6. Conservation of tracrRNA in mollicutes. Multiple alignment of tracrRNA sequences predicted from mollicutes genomes. Alignment was manually optimized in Mega6 according to the predicted secondary structures of nexus and terminators and antirepeats pairing with direct repeat regions. Positions conserved at >60% were not highlighted.

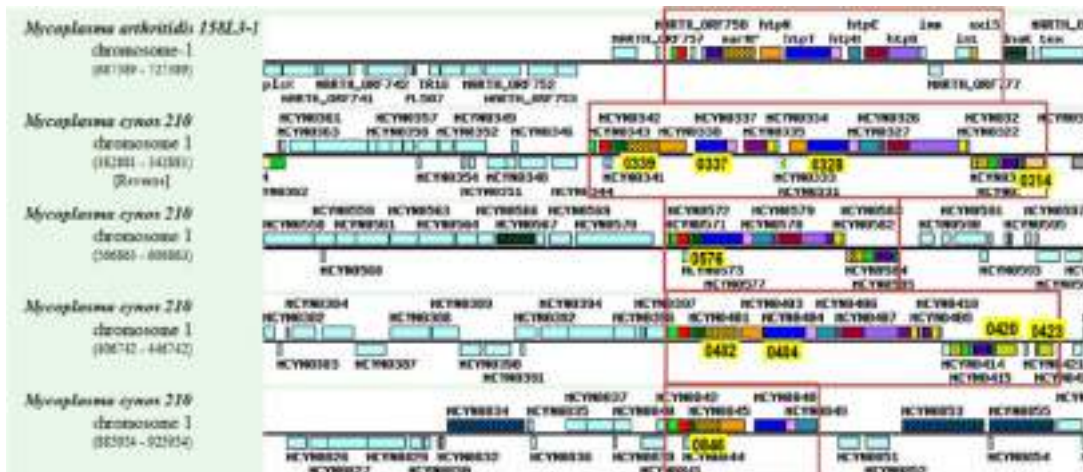


Figure S7. Genomic regions encoding phage proteins in *M. cynos* 210 and *M. arthritidis* 158L3-1. This figure is adapted from a screenshot generated with the MGD database (<http://mgd.genome.afip.ca>). CDS are represented as boxes. Their orientation is given by position from the vertical black line: above (forward orientation) and under (reverse orientation). Homologous genes are indicated by color code. Genomic regions encoding phage proteins are framed in red. CDS including proteoglycans matching spaces 1, 10, 16, 17, 18 from *M. cynos* C142 CRISPR array are highlighted in yellow.

CHAPITRE 1 : Etude des systèmes CRISPR chez les mollicutes

Table S3. Consensus sequence of Direct Repeats in CRISPR arrays of mollicutes.

Species	Consensus Direct Repeat
<i>Spiroplasma syrrhoidicola</i> EA1 (locus 1)	5' TTTT G T G G L I G T C A T T T T A T T G T A G A A T A T A A C
<i>Spiroplasma syrrhoidicola</i> EA1 (locus 2)	5' T A T T A G T C T T C T C C A A T T T T C A G G A A A T A A C A G C
<i>Spiroplasma apis</i> B31	5' T T T T G T T A T C T G A C A T A T C T A A G G A A T A G A C G A C
<i>Spiroplasma helicoides</i> TABS-2	5' T T A T G G T A C C C T G T A A A A T T A T G T A G T A G T A G A A C
<i>Spiroplasma litoreale</i> strain TN-1	5' T T T A T G G T A C C C T G T A A A A T T A T G T A G T A G T A G A A C
<i>Acholeplasma palmae</i> J233	5' T T T G T G T A C C C I C G T A A T T T T T G C T A T T T A A C A A C
<i>Mycoplasma synoviae</i> 53	5' T T T T G G G G I I G T A C A A T T A T T T T G T T A A G T A A A A C
<i>Mycoplasma gallisepticum</i> 56	5' T T T T A G L A C T G T A C A A T A C T T G T G T A A G C A A T A A C
<i>Mycoplasma lipofaciens</i> ATCC 35015	5' T T T T A A G T T A G I A C A A T A T T T G I G T A A G A T A T A A C
<i>Mycoplasma cynos</i> C142	5' T T T T A G T G I T G T A C A A T A T T T G G G T A A A C A A T A A C
<i>Mycoplasma canis</i> PG 14	5' T T T T A G T G T T G T A C A A T A T T T G G G T A A A C A A T A A C
<i>Mycoplasma dispar</i> ATCC 27140	5' T T T T T G T G C T G T A C A A T T T C T T A C T A G A G T A A A A C
<i>Mycoplasma ovipneumoniae</i> NM2010	5' T T T T T G T G C I G T A C A A T T T C T T A C T A G A G T A A A A C
<i>Mycoplasma arginini</i> HAZ145_1	5' T T T T T G T G L I G T A C A A T T T C T T A T T A G A G T A A A A C
<i>Mycoplasma hyosynoviae</i> strain NPL1	5' T T T T A T T G T C G T A C A A T T T T T A G T A A T C T A T A A C
<i>Mycoplasma arthritis</i> 158L3-1	5' C T T T T G G A C I G T A C A A T T T T A T A T A G A G T A A A G T
<i>Mycoplasma mobile</i> I63K	5' T T T T T G G T A G T A T C A T T C T T A T G T A T T C T T A A A C
<i>Ca. Hepatoplasma crinochetorum</i> Av	5' T T T T T G G T T A G T T G G T A T T C A T G A T G G T G T T T A A C C G

Table S4. Genome, Cas9 and DR/tracrRNA hybrid features in Type-II CRISPR systems of mollicutes and non-mollicutes bacteria.

Species	genome GC (%)	tracrRNA GC (%)	Cas9 GC (%)	Cas9 length (AA)	ΔG tracr	repeat-antirepeat stem (bp)			nexus stem (bp)	terminator stem (bp)
						upper stem	bulge	lower stem		
MOLLICUTES										
<i>Spiroplasma apis</i> B31	28.3	36.0	25.7	1180	-55.80	26*	yes	9	4	5
<i>Spiroplasma helicoides</i> TABS-2	26.8	30.9	23.5	1009	-54.90	24	yes	7	3	6
<i>Spiroplasma litoreale</i> strain TN-1	24.9	30.5	22.1	1009	-54.90	24	yes	7	3	6
<i>Acholeplasma palmae</i> J233	29.0	38.8	28.0	1107	-59.80	22	yes	9	5	10
<i>Mycoplasma synoviae</i> 53	28.3	33.6	22.6	1269	-73.20	33	No	-	18++	6
<i>Mycoplasma gallisepticum</i> 56	31.5	43.3	29.7	1269	-67.50	21	yes	10	8+	9
<i>Mycoplasma lipofaciens</i> ATCC 35015	25.3	37.1	22.3	1146	-63.40	24	yes	10	9+	6
<i>Mycoplasma cynos</i> C142	25.7	30.8	22.7	1239	-54.10	24	yes	5	5+	12**
<i>Mycoplasma canis</i> PG 14	27.2	33.3	26.0	1233	-54.10	25	yes	5	5+	13**
<i>Mycoplasma dispar</i> ATCC 27140	29.0	32.5	26.6	1263	-47.80	22**	yes	10	5+	6
<i>Mycoplasma ovipneumoniae</i> NM2010	29.0	35.8	26.6	1272	-53.10	22	yes	10	5+	6
<i>Mycoplasma arginini</i> HAZ145_1	26.4	37.8	22.5	1175	-54.40	24	yes	9	8+	5
<i>Mycoplasma hyosynoviae</i> strain NPL1	27.1	32.6	24.0	1167	-54.00	24	yes	10	8+	7
<i>Mycoplasma mobile</i> I63K	25.0	24.4	22.0	1226	-51.30	24	yes	8	3	11
<i>Ca. Hepatoplasma crinochetorum</i> Av	22.5	25.5	19.3	1195	-60.80	33	No	-	12+	7+
NON-MOLLICUTES										
<i>Streptococcus pyogenes</i> SF370 (M1 GAS)	38.5	39.3	35.1	1368	-51.30	25	yes	7	2	6
<i>Streptococcus thermophilus</i> LMD-9 *	38.9	35.6	31.2	1388	-51.00	25+	yes	6	2	6
<i>Enterococcus faecalis</i> TX0012	37.4	36.9	36.0	1337	-75.44	25	yes	9	3	10
<i>Lactobacillus rhamnosus</i> GG	46.7	43.1	42.3	1363	-90.90	33+	yes	7	9**	9
<i>Streptococcus thermophilus</i> LMD-9 **	38.9	40.4	31.5	1120	-63.20	24+	yes	9	3	13
<i>Eubacterium rectale</i> ATCC 32656	41.5	36.8	33.3	1114	-54.80	25*	yes	8	3	9
<i>Staphylococcus lugdunensis</i> M23590 *	33.9	29.8	26.4	1054	-53.10	22	yes	9	3	10
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168	30.5	39.7	29.0	984	-44.30	27+	No	-	6	9
<i>Neisseria meningitidis</i> Z2491	51.8	46.8	48.9	1082	-82.20	23*	yes	8	8+	17+
<i>Pasteurella multocida</i> str. <i>Pm70</i>	40.4	43.6	37.6	1056	-72.60	24*	yes	8	6+	18**+
<i>Staphylococcus aureus</i> subsp. <i>aureus</i>	32.7	34.17	30.04	1056	-53.60	21	yes	9	3	10

*Cas9 GC% from strain VCU150

* =1MistMatch

+ = additional bulge

CHAPITRE 2 : Le système CRISPR-Cas9 de
Mycoplasma gallisepticum : de la caractérisation du
système naturel à l'outil d'ingénierie des génomes

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

En parallèle des analyses de génomique comparée des systèmes CRISPR de mollicutes, nous avons décidé de nous concentrer sur celui de *Mgal*. Les raisons sont multiples et ce choix avait été fait en amont de mon arrivée au laboratoire. Tout d'abord, le système CRISPR de *Mgal*, tout en étant typique des systèmes de mycoplasmes, présente une évolution particulière, associée à un récent changement d'hôte de *Mgal*, de la volaille vers une espèce d'oiseaux migrateurs aux USA : les roselins (House finch) (Fischer *et al.*, 1997; Delaney *et al.*, 2012). Une perte soudaine du locus cas chez certaines souches de roselin ainsi que l'étude des spacers présents dans différentes souches (de poulet ou de roselins), laissent penser que le système aurait perdu de son utilité chez le nouvel hôte alors qu'il était apparemment fonctionnel dans les souches de volaille. Par ailleurs, des études transcriptomiques ont montrés que le locus CRISPR est bien transcrit dans une souche de poulet S6 (Mazin *et al.*, 2014). Enfin, 12 génomes de *Mgal* étaient disponibles, ce qui, à ce moment-là, n'était pas courant pour les mollicutes. Sur ces 12 génomes, le locus CRISPR était prédit fonctionnel pour 8 souches. Toutes ces indications laissaient penser que le système CRISPR de *Mgal* était bien fonctionnel. De plus, un plasmide *oriC* ainsi qu'un protocole de transformation étant déjà mis en place au laboratoire, tout cela a fait de *Mgal* le meilleur candidat pour aller plus loin dans l'analyse du système CRISPR.

L'objectif était alors de caractériser fonctionnellement le système CRISPR de *Mgal* pour comprendre son mode de fonctionnement et le transformer si possible en outil d'ingénierie de génome pour les mycoplasmes. Pour cela plusieurs questions se posaient : le système de défense est-il effectivement fonctionnel ? Quelle est la séquence PAM qui détermine la spécificité du système ? La transformation de ce système en outil est-elle possible ? Quelles possibilités cela va-t-il ouvrir pour l'ingénierie de génome des mycoplasmes ?

Le précédent doctorant, Iason Tsarmopoulos, avait déjà pu montrer que le système de défense de la bactérie était bien toujours actif chez la souche de poulet S6 par des essais de transformations avec des plasmides contenant des spacers suivis de séquences PAM putatives déterminés bio-informatiquement. Mon travail s'est appuyé sur ces résultats préliminaires pour continuer l'exploration du système CRISPR de *Mgal*.

Ce travail nous a emmené bien plus loin que nous ne l'avions imaginé au départ, notamment grâce aux discussions que nous avons eu avec le Dr. Camille Bonneaud sur l'évolution de *Mgal* suite à son émergence comme pathogène d'un nouvel hôte. Notre

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

étude a alors pris une nouvelle orientation, en développant particulièrement le lien entre les systèmes de défense de *Mgal* contre les phages et l'évolution spécifique de cette bactérie liée au changement d'hôte.

1. Principaux résultats

Cette partie de ma thèse a conduit à l'étude fonctionnelle du système CRISPR de *Mgal* et sa déclinaison sous forme d'outil génétique. La diversité de séquence du locus CRISPR et particulièrement au niveau du domaine PI de MgCas9 nous a amené à étudier fonctionnellement cette protéine clé du système CRISPR chez deux souches de *Mgal* provenant l'une de poulet (S6) et l'autre de roselin (CA6). La liste exhaustive des séquences PAM reconnues par les deux MgCas9 a été établie par la combinaison d'une technique de clivage *in vitro* d'une banque de plasmide portant une séquence PAM dégénérée sur 7 positions et de séquençage haut débit d'amplicons. Une différence de spécificité a été mise en évidence par cette approche et elle a été confirmée *in vivo* lors d'une expérience d'interférence, par transformation de la souche S6 et d'une souche de S6 exprimant une MgCas9 dont le domaine PI a été échangé avec celui de la souche CA06 (Construite à l'aide des outils développés dans le chapitre 3). D'un point de vue évolutif et en retraçant le contexte écologique de ces souches, nous avons pu y découvrir des systèmes de défenses qui semblent génétiquement très dynamiques et pouvant être activés et désactivés au cours de l'évolution du pathogène. Parmi ces systèmes de défense de la bactérie contre des éléments génétiques invasifs, le système CRISPR présente une évolution majeure en termes de spécificité, apparemment corrélée à l'émergence de la bactérie dans un nouvel hôte. L'inactivation de MgCas9 pour toutes les souches de roselin postérieures à 2007 semble également être en lien avec l'apparition d'une résistance au pathogène de ce nouvel hôte. De nouvelles études pourraient permettre de comprendre le rôle exact que le système CRISPR a pu avoir lors de l'émergence de 1994. Finalement, la transformation du système de la souche S6 en système minimal a permis de développer une série de plasmides « all-in-one » permettant l'expression d'un sgRNA et de MgCas9. Ces plasmides ont été utilisés comme outil de contre sélection chez plusieurs espèces de mycoplasmes éloignés phylogénétiquement. Nous avons pu ainsi sélectionner une souche de *M. capricolum* subsp. *capricolum* curée de son élément mobile ICE.

2. Article en finalisation pour soumission à CRISPR journal

Intraspecific evolution and engineering of a CRISPR/Cas9 defense system associated with the host shift of a minimal pathogen

Authors: Thomas IPOUTCHA¹, Iason TSARMOPOULOS¹, Géraldine GOURGUES¹, Fabien RIDEAU¹, Vincent BABY¹, Paul DUBOS¹, Geoffrey E. HILL², Andrea DOWLING³, Alain BLANCHARD¹, Yonathan ARFI¹, Carole LARTIGUE¹, Patricia THEBAULT⁴, Camille BONNEAUD³, Pascal SIRAND-PUGNET¹ *

¹ Univ. Bordeaux, INRAE, UMR BFP, F-33882, Villenave d'Ornon, France,

² Department of Biological Sciences, Auburn University, Auburn, Alabama, 36849-5414, United States of America

³ Biosciences, College of Life and Environmental Science, Penryn Campus, University of Exeter, Cornwall, TR10 9FE, UK

⁴ LaBRI, Univ. Bordeaux, F-33405 Talence,

*Corresponding author: Pascal Sirand-Pugnet, pascal.sirand-pugnet@inrae.fr

Abstract:

Successful emergence of a bacterial pathogen into a new host depends on its ability to colonize this host and to cope with its defense systems. It also potentially depends on the presence of compatible phages and other invading DNAs. In this study, we used the recently reported host jump of *Mycoplasma gallisepticum* from poultry to American House Finch as a case study to investigate this aspect of emergence that has been greatly overlooked. Mycoplasmas are minimal bacteria that have no cell wall and use UGA stop codon to encode Trp. They are considered as the smallest bacteria able to grow in axenic media and most of them have a parasitic lifestyle. Contrasting with an evolution mostly marked by genome reduction, fast evolution of gene sequences and horizontal gene transfers have been described as an underestimated source of genetic diversity that might contribute to the wide distribution of hosts these minimal bacteria can colonize, including human and many other animals. Recent host jump from poultry to American House Finch of *M. gallisepticum* was associated with genome evolution of potential virulence factors but also of a predicted CRISPR system. Here we functionally characterize the Type II CRISPR system of *M. gallisepticum*. We demonstrate that the evolution of MgCas9 PI domain drove a complete reset of the defense system, with a change of the PAM motif that is involved in self/non-self recognition specificity and of the set of spacers found in the CRISPR array. We revisited the current model of evolution of the bird pathogen, integrating the observed modifications of the CRISPR and other bacterial defense systems encoded within a previously undescribed defense genomic island. Finally, we demonstrate the possibility to use this endogenous CRISPR system of mycoplasma as an effective tool for counter selection applications in a diversity of mycoplasmas.

Introduction

CRISPR systems are used by prokaryotes to adapt their defense against phages and other sources of invading DNAs. While a growing diversity of systems have been described, their flexibility and efficacy to protect populations of bacteria from phage attacks commonly relies on the ability to acquire new spacers in the CRISPR array, that stands as the physical memory of the system. Whereas their occurrence in most phyla suggest the central role they play in the evolution of prokaryotes, maintaining active CRISPR systems has a cost that has been invoked to explain their frequent inactivation and disappearance from genomes (Koonin 2019; Makarova et al. 2020). Therefore, fast adaptation of spacer arrays and gain/loss events of active CRISPR systems over the evolutionary trajectories of prokaryotes may result from complex ecological constraints combined with cost/benefit balance for selecting the most adapted set of defense systems in a specific context. For pathogenic bacteria, some main ecological constraints are driven by the interaction with the host, and especially by the ability to cope with the host immune system. The constant armed race between bacteria and

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

their hosts shape the evolution of their genomes with small- and large-scale changes associated with remarkable events like host jumps. Currently, very little is known about the evolution of bacterial defense systems against phages and other parasitic DNAs in the context of emergence and adaptation to a new host. The bird pathogen *Mycoplasma gallisepticum* (*Mgal*) represents an unique model for studying such evolution.

Mycoplasmas are minimal pathogens belonging to the class *Mollicutes* (May, Balish, and Blanchard 2014) that are characterized by a streamlining genome evolution from Gram-positive ancestor with a drastic genome reduction partially counter balanced by horizontal gene transfers (HGT) and exposure to numerous mobile genetic elements as potential sources of genetic diversity and genome reorganization (Citti et al. 2018; Sirand-Pugnet, Citti, et al. 2007). Despite an evolutionary tendency to genome minimization, most mycoplasmas have retained various defense systems against phages and invading DNAs. Indeed, mycoplasmas are the second group of bacteria with the greatest number of Restriction-Modification systems (Oliveira, Touchon, and Rocha 2014). They are also the bacteria with the smallest genome where CRISPR systems have been described. Phylogenetic studies suggest that Type II-A CRISPR system was already present in mycoplasma LCA and current CRISPR systems are found in ~50% of the species, with some diversity observed in the protein sequences of Cas9, Cas1, Cas2 and Csn2, and in the spacers sets and direct repeats of CRISPR arrays (Ipoutcha et al. 2019).

Among mycoplasma CRISPR systems, the *Mgal* CRISPR locus has undertaken a remarkable evolution correlated with a recent host shift that occurred in 1994 in the US from poultry hosts (chicken, turkey) to House Finch (HF) and caused the death of about 225 million birds in the three first years of the outbreak (Delaney et al. 2012; Dhondt, Tessaglia, and Slothower 1998). Comparisons of the unique CRISPR loci in the genomes of House Finch isolates and ancestral poultry strains revealed a complete modification of the spacer set, with not a single spacer of the 147 detected across the 4 poultry strains studied also found in the HF strains (Delaney et al. 2012). By comparison with poultry strains, the reduced diversity of spacers found in HF strains (50 unique spacers) suggests that the host shift was associated with a sudden slow-down in the recruitment of new spacers. Loss of spacers and degradation of the CRISPR locus was also observed in some HF strains, suggesting this defense was evolving towards becoming non-functional in the new host. **This fast evolution of CRISPR system at the intraspecific level and in association with a host-shift is unique and raises the question of the role played by prokaryotic defense mechanisms in the successful colonization of new hosts by pathogenic bacteria.**

Beside the first observation of the remarkable change of spacers set, no functional characterization has been made of the CRISPR system, especially for the main actor Cas9. In this paper, we functionally characterized the natural CRISPR system of the minimal bacteria *Mgal*. Using *in vitro* and *in vivo* approaches, we determined the PAM specificities of the MgCas9 and showed that the evolution of spacer repertoire during the host shift was correlated with a clear modification of the PAM signature which was associated with the evolution of the PI domain of MgCas9. We revisited the current model of evolution of the bird pathogen, considering the observed modifications of the CRISPR and other predicted bacterial defense systems. From the *Mgal* natural CRISPR system, we finally derived a minimal genetic tool made of MgCas9 and a compatible single-guide RNA which functionality was demonstrated in four unrelated mycoplasma species.

Results

Genomic comparison of CRISPR locus in *M. gallisepticum* strains.

To extend the 2012 study, a new set of *Mgal* strains isolated from HF was genome sequenced, including 78 US strains sampled between 1994 to 2015 in Alabama, Ohio, Missouri, Virginia, South Carolina, Texas, Georgia, Maryland and Pennsylvania (Table S1). Draft genomes and 19 other *Mgal* genomes available in public databases were analyzed and compared. First, a phylogenetic analysis was

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

performed including the 97 *Mgal* genomes. A set of 466 protein families predicted from all genomes was selected to build a high-resolution tree (Figure S1).

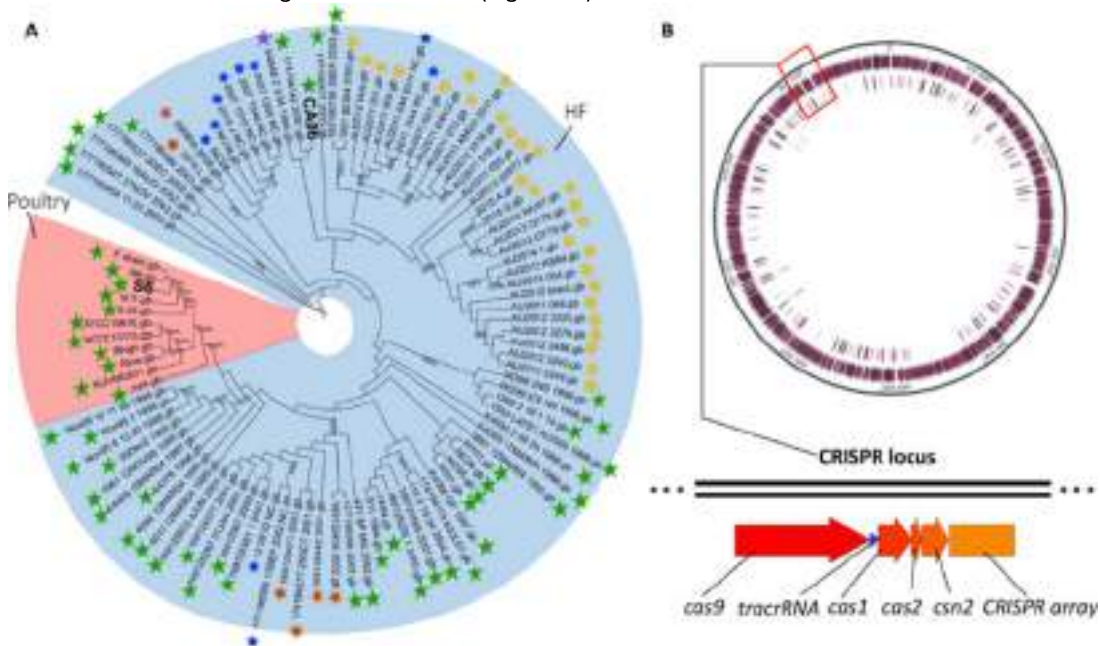


Figure S1: Presence and integrity of the CRISPR defense system in *M. gallisepticum* strains. A- Phylogenetic tree of *Mgal* strains based on 466 conserved proteins. Presence and integrity of the CRISPR defense system is indicated for all strains. Green star, complete CRISPR locus; blue star, strains with a CRISPR locus deleted from the tracrRNA and Cas genes; others signs, inactivated CRISPR with different truncated forms of MgCas9 sequence. B- *Mgal* genome is represented here and CRISPR locus is pointed and represented.

As expected, phylogenetic tree of all those strains indicated that *Mgal* poultry strains are well separated than HF ones and oldest HF strains are closer to poultry strains (Figure S1). A single CRISPR locus was found in all *Mgal* strains but for 10 of them (2 HF strains isolated in 2002, 5 in 2007, 2 in 2011 and 1 in 2015), part of the locus harboring the tracrRNA and the cas genes (i. e. *cas9*, *cas1*, *cas2* and *csn2*) has been lost, leaving only the CRISPR array (Figure S1A). Further *in silico* analyses were undertaken to document the intraspecific diversity of *Mgal* CRISPR system among the strains that contained an entire locus. Protein sequences of the *Mgal* Cas9 (MgCas9) were aligned and a phylogenetic tree was constructed (Figure 1A). A clear distribution into two main branches was observed with a clustering of MgCas9 from poultry strains within one branch and all HF MgCas9 clustering in the other branch and showing more than 99% identity in sequence. Three truncated versions of MgCas9 were predicted among HF strains. MgCas9 truncated form V1 is truncated by a frameshift mutation at position 3445 of the gene, leading to a protein truncated of the last 120 amino acids within the PAM Interacting domain (PI). These truncated forms were clustered in the MgCas9 phylogenetic tree and were identified in some HF strains isolated in 2001, 2002, 2009 and 2015. The truncated form V2 was found in strains collected in years 2011 to 2015 and is truncated at position 1018 within the REC lobe of the protein by a single nucleotide variation changing a GAA Glu codon into a TAA STOP codon. The V3 form was only observed in a 1998 strain and is truncated at position 1498 in the REC domain by a single base change that generates a TAG STOP codon instead of CAG Gln codon. The occurrence of various truncated forms of MgCas9 in different branches of the *Mgal* tree suggest the inactivation has occurred several times during the evolution of *Mgal*. All HF strains with a predicted complete MgCas9 were isolated during years 1995-2004, suggesting co-occurrence of strains with active and inactive MgCas9 has been limited to years 2001 to 2004. By contrast, all poultry strains

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

collected in chicken and turkeys over a large period of time (~1958-2019) were predicted to harbor an active CRISPR system with no inactivating mutations observed in MgCas9.

Comparing one representative copy of HF MgCas9 (from strain CA06) with the typical MgCas9 from a poultry strain (S6), we observed a concentration of amino acid variations within the PI domain that is involved in the specific interaction between Cas9 and the PAM sequence located downstream of the protospacer of the target DNA (Figure 1B). At the protein level, a sequence variation of 8.7% was calculated for the PI domain for 3.1% for the remaining domains of MgCas9. At the nucleotide level, percentages of variations were 4.2% and 2.2% for the PI and remaining domains, respectively. Regarding other Cas proteins encoded within the CRISPR locus, Csn2 was shown more variable in sequence (5.1%) compared to the Cas1 (0.3%) and Cas2 (0.9%). Therefore, we concluded that the host jump of *Mgal* from poultry to HF had been correlated with a marked evolution of MgCas9 PI domain, before a global inactivation of the defense system. **From these observations, we hypothesized that the PAM recognition specificity of MgCas9 might be different between typical poultry strains and HF strains isolated during the 10 first years after the host jump.**

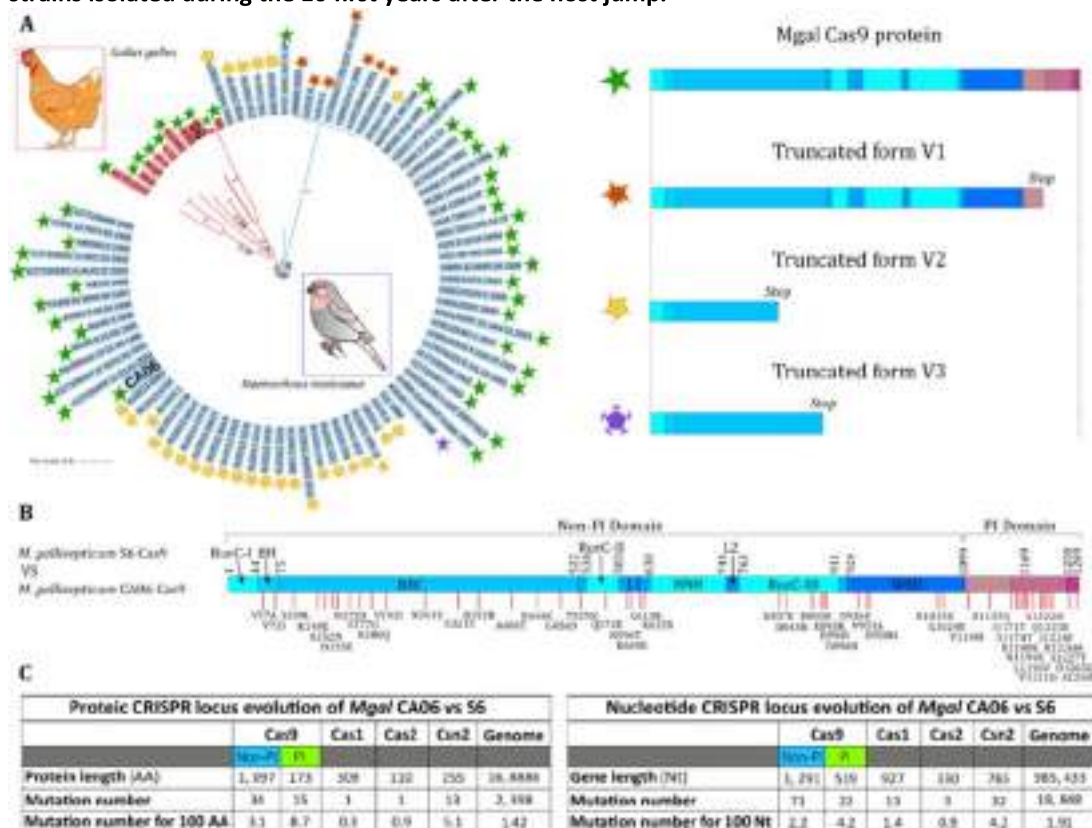


Figure 1: Evolution of the *M. gallisepticum* CRISPR locus. A- Phylogenetic tree of MgCas9. Strains isolated from poultry hosts and HF are highlighted in red and blue, respectively. MgCas9 forms are indicated for all strains. Green star represents complete MgCas9 sequence. Others signs represent three MgCas9 truncated forms. B- Representation of MgCas9 protein domains and mutations between strains S6 (poultry) and CA06 (HF). V57A indicate the 57th amino acid in S6 MgCas9 sequence is Valine and is replaced by Alanine in CA06 MgCas9. C- Nucleic and proteic sequence comparisons between CRISPR effectors of S6 and CA06 strains. Genome mutation number was calculated comparing 466 core genome proteic sequences or using all genome nucleic sequences.

***In vitro* PAM specificity is different between MgCas9 from poultry (S6) and HF (CA06) strains**

In order to compare the PAM recognition preference of MgCas9 from a poultry strain (S6) and a HF strain (CA06), we used an *in vitro* assay previously developed by Siksnys 's group (Gasiunas et al. 2020;

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

Karvelis, Young, and Siksnyš 2019). MgCas9 proteins and sgRNAs were produced using a cell-free in vitro translation (IVT) method. Afterwards, IVT mixtures were used as sources of MgCas9/sgRNA ribonucleoparticles for the cleavage of a library of plasmids containing a target sequence and a PAM sequence randomized on 7 positions (Figure 2A). The randomness of the plasmid library was first evaluated by amplicon sequencing on three replicates, using an Illumina MiSeq sequencer. Out of the 16,384 possible PAM sequences, 16,306 were detected in one or several replicates with a number of reads varying from 1 to 420 (Figure S2A). Base frequency at each of the 7 positions was calculated showing nearly perfect reproducibility among the replicates and observed average frequencies of A, C, G and T ranging from 0.16 to 0.35 over the 7 positions (Figure S2B). Average frequencies of all detected PAM sequences were calculated (Figure S2C) and a normalization factor was determined for each PAM sequence as the ratio between observed and theoretical frequencies (Table S2).

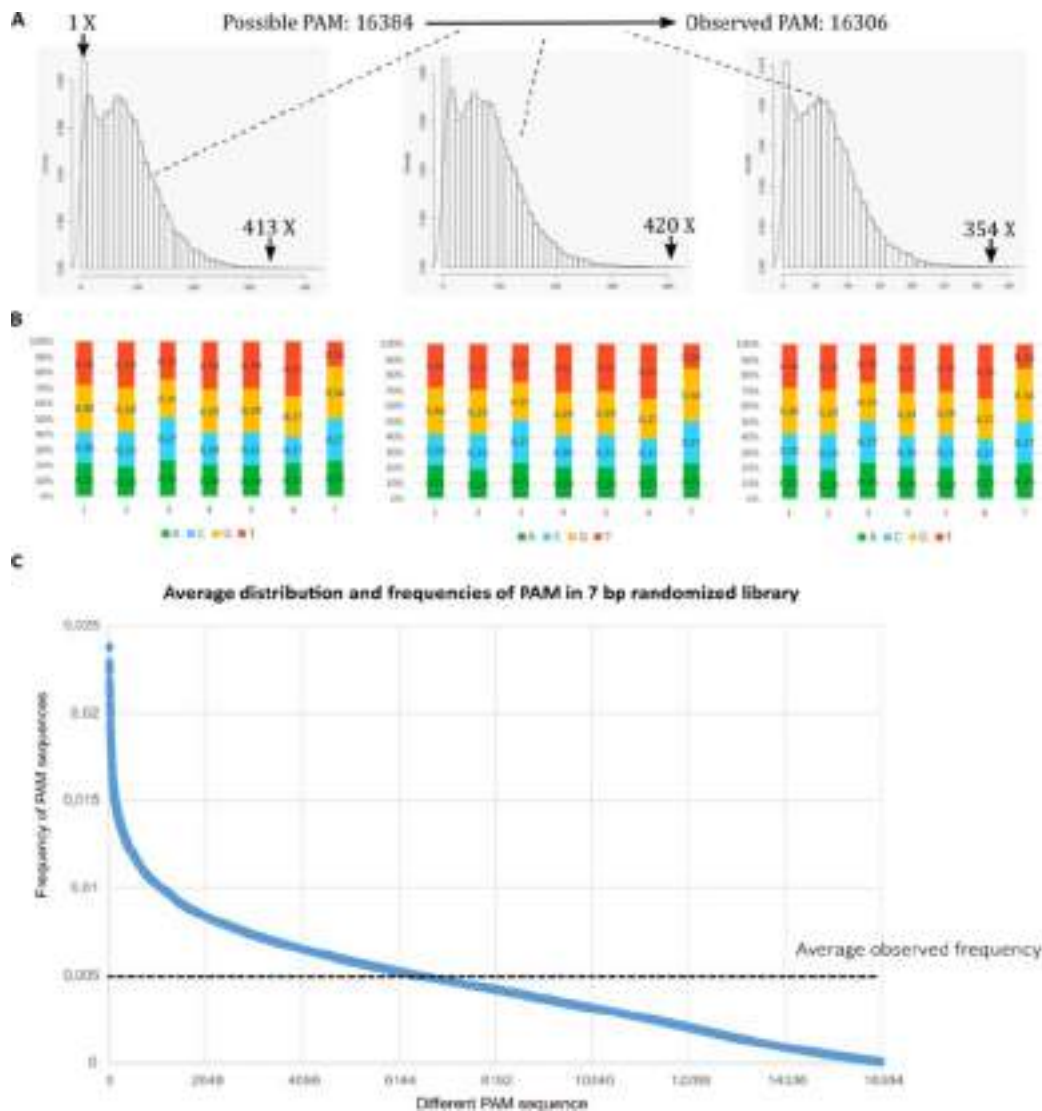


Figure S2 : Plasmid library validation. **A-** Histograms of PAM motif distributions for each replicates are represented with minimum and maximum coverage. **B-** Position frequency matrix of the 7 bp randomized for each replicates. **C-** Average frequency of each PAM sequence observed in at least one of the three replicates was calculated. **C-** Average frequencies of PAM sequences in the 7 bp randomized library were ranked and represented as a graph. Average observed was indicated.

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

After cleavage of the plasmid library by MgCas9 S6 or MgCas9 CA06, molecules containing a recognized PAM sequence were captured, enriched by PCR amplification and submitted to deep sequencing. Three independent *in vitro* cleavage and further processing were performed in each case. Cleavage was observed three bases upstream from the PAM sequence and results obtained for the two series of replicates were nearly identical (Figures S3A and S3B). Out of the 16,306 PAM sequences observed in the library, 10928 and 9709 PAM sequences were detected at least once after cleavage by MgCas9 S6 and MgCas9 CA06, respectively. In order to validate the specificity of the cleavage induced by the IVT, a PAM sequence corresponding to the 7 first positions of Direct Repeat sequence present in the CRISPR arrays of *M. gallisepticum* strains CA06 and S6 (GTTTTAGCACTGTACAATACTTGTGTAAGCAATAAC) was searched for in the initial and cleaved libraries. Indeed, this sequence should not be found in the list of PAM sequences recognized by MgCas9, as it would mean suicide for the bacteria. While the GTTTTAG sequence was the 497th most represented motif in the initial PAM library, it was not detected in any of the replicates after cleavage with MgCas9 CA06 or MgCas9 S6. This result suggested the high specificity of the *in vitro* cleavage with IVT mixtures. However, distribution study showed that ~90% of the detected PAM sequences had nearly null frequencies and were therefore considered as poorly significant (Figures S3C and S3D). The 1000 most frequent PAM sequences were selected and normalized frequencies were calculated, taking the bias of the original plasmid library into account (Table S3). Frequency matrices and sequence logos were produced (Figures 2B and 2C). PAM sequences recognized by MgCas9 from S6 and CA06 strains are both 5 letters motifs with only positions 3, 4 and 5 showing marked bias. Remarkably, sequence logo comparison showed that PAM preferences of MgCas9 proteins were clearly different in the two strains of *M. gallisepticum*. Main difference was observed at position 3 with 99% being A in S6 and G in CA06. Position 4 was preferentially A or G in both cases, with a bias towards A more marked in S6. In position 5, T was present at 73% in S6 compared to 30% in CA06, suggesting a difference of importance of this position in the two strains. This experiment demonstrates that MgCas9 recognize 5-positions PAM sequences and that recognition preferences *in vitro* are clearly different in the poultry strain S6 and the HF strain CA06. This was confirmed when comparing the Top 10 PAM sequences most frequently recognized by S6 and CA06 MgCas9 (Figure 2D).

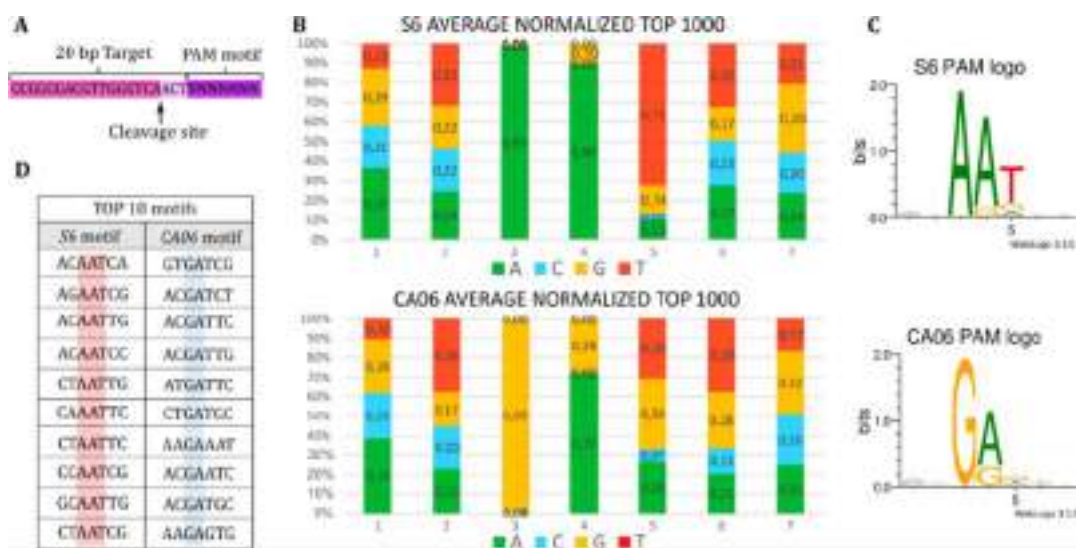


Figure 2 : *In vitro* characterization of PAM sequences recognized by MgCas9 from poultry (S6) and HF(CA06) strains. **A-** Position of the observed cleavage site for the two MgCas9. **B-** Position Frequency Matrix (PFM) of observed after cleavage of a 7-positions PAM degenerated library by MgCas9 from S6 and CA06 *Mgal* strains. Frequencies were averaged on three replicates and corrected considering bias in the distribution of PAM sequences in the initial plasmid library. Adenine are represented in orange, Cytosine in grey, Guanine in yellow and Thymine in blue. Frequencies of each base at each position are indicated on the graph. **C-** Web logos of consensus PAM sequences recognized by MgCas9 S6 or CA06. **D-** Top 10 of the best motifs recognized *in vitro* by MgCas9 S6 and CA06.

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

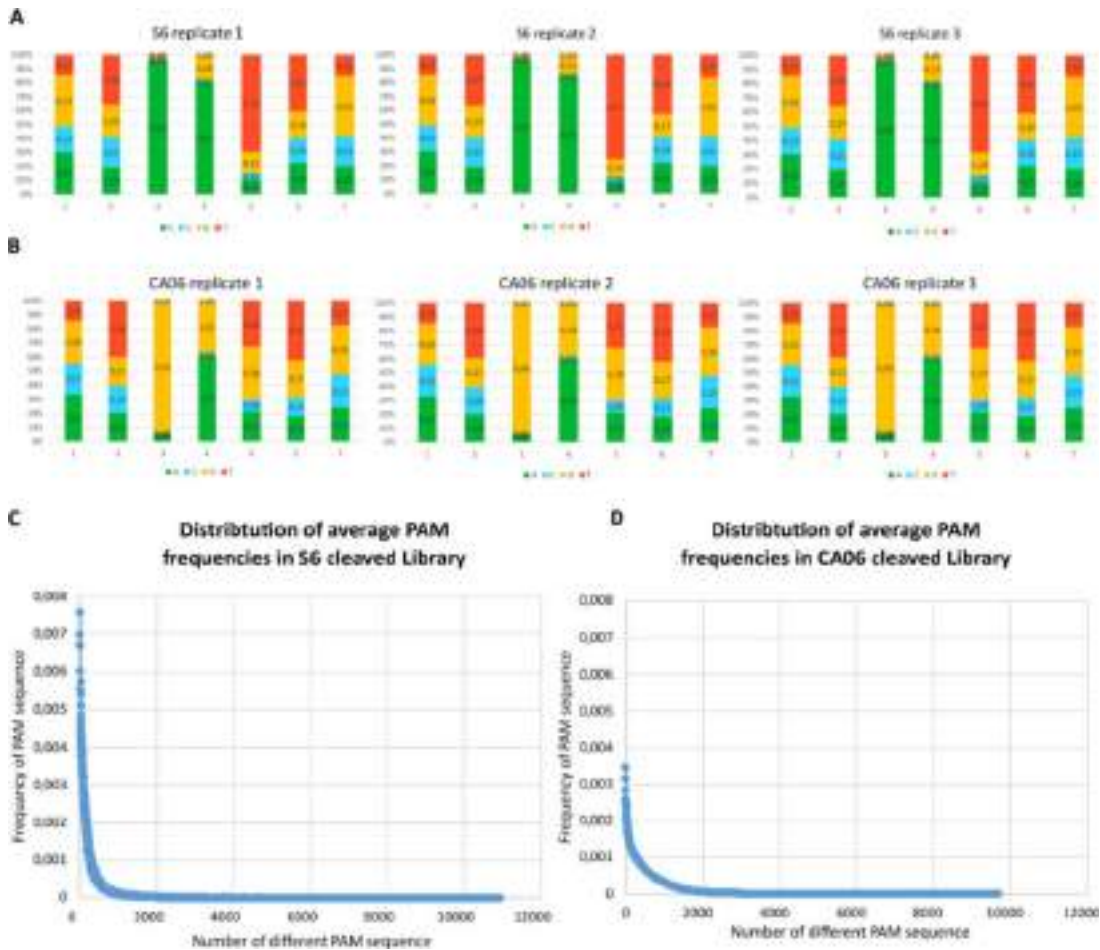


Figure S3 : PAM analyses after cleavage with MgCas9 from S6 (A) and CA06 (B) strains. A- Position frequency Matrix of PAM cleaved sequence by S6 Cas9 for each replicate. **B-** Position frequency Matrix of PAM cleaved sequence by CA06 Cas9 for each replicate. **C-** Average frequency of each PAM sequence observed in at least one of the three replicates was calculated. Average frequencies were ranked and represented as a graph.

To confirm the PAM recognition specificity observed by *in vitro* cleavage of a degenerated plasmid library, we selected two PAM sequences, specifically recognized by MgCas9 S6 (PAM p1) and MgCas9 CA06 (PAM p2), respectively. Three plasmids were constructed containing the target sequence followed by one of the two selected PAM sequences (pTZ57-Motif1 and pTZ57-Motif2) or by the 7 first nucleotides of Direct Repeat sequence (pTZ57-Motif3)(Figure S4A). After *in vitro* cleavage of these plasmids by IVT-produced MgalCas9/sgRNA complexes, we observed that MgCas9 S6 efficiently cleaved PAM p1 but not PAM p2 whereas MgalCas9 CA06 preferentially cleaved the PAM p2 (Figure S6B). By contrast, the control PAM p3 remained uncleaved after incubation with one MgalCas9 or the other. We then confirmed the difference of PAM specificity of MgCas9 originating from poultry and HF strains.

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

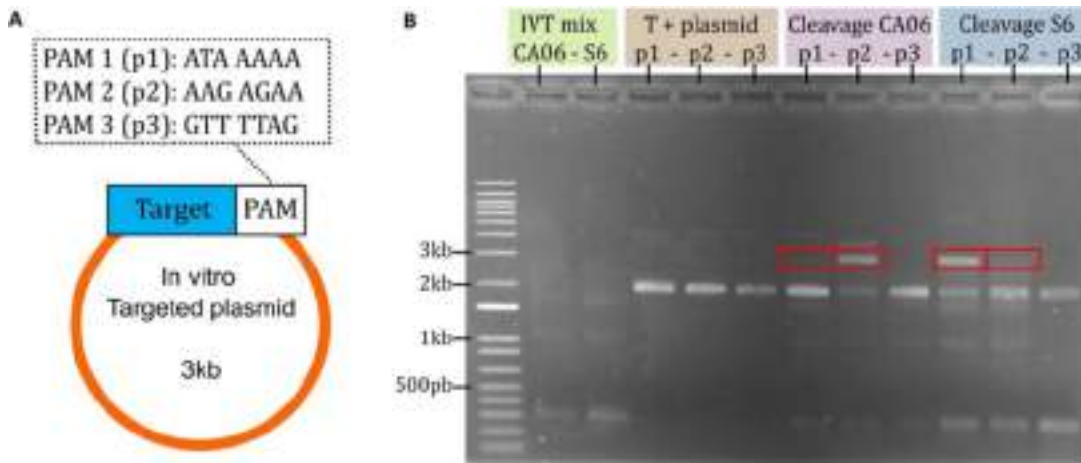


Figure S4 : *In vitro* cleavage of selected PAM sequences. **A-** Design of the plasmids used for *in vitro* cleavage assay. PAM sequences m1, m2 and m3 were inserted downstream of a spacer targeted by the sgRNA to generate plasmids pTZ57-Motif1, pTZ57-Motif2 and pTZ57-Motif3, respectively. **B-** Result of the *in vitro* cleavage after incubation of the plasmids with MgCas9/sgRNA complexes produced by IVT. Uncleaved plasmids were used as control (T+). Reaction products of IVT were also loaded on gel as a control of background bands.

***In vivo* interference experiment confirmed differences of PAM recognition specificity of MgCas9 depending on the strain origin.**

To go further and evaluate if the difference of PAM specificity observed *in vitro* was consistently correlated with a difference of specificity of the bacterial defense system *in vivo*, we set a cleavage assay based on the transformation of *Mgal* with a replicative plasmid (pMGAL-Spacer and derivatives) containing the third spacer naturally found in the CRISPR array of strain S6, followed by a selected PAM candidate (Figure 3A, and 3B). Four PAM motifs were chosen based on *in vitro* results. Motifs S6-1 and S6-2 were well recognized by MgCas9 S6 but poorly by MgCas9 CA06, *in vitro*. CA06-3 motif was not found cleaved *in vitro* by MgCas9 S6 but very well by MgCas9 CA06. On the contrary, CA06-4 motif was efficiently recognized by MgCas9 CA06 but weakly by MgCas9 S6. Interference assays were performed using these plasmids in *Mgal* S6 strain and a genetically engineered *Mgal* S6 (thereafter named *Mgal* S6 mod PI CA06) which MgCas9 PI domain had been replaced by the homologous domain of the MgCas9 CA06. A plasmid that contained the Direct Repeat of the CRISPR array as a PAM sequence was used as a negative control. After transformation, highly different numbers of colonies were observed on selective plates, depending on the PAM sequence and the PI domain. By comparison with the negative control plasmid, transformations with plasmids containing PAM sequences expected to be cut by MgCas9 S6 (S6-1 and S6-2) lead to 100-1000 times less transformants (Figure 3C). This confirmed that the CRISPR/Cas9 system is functional as a defense system in *Mgal* S6 and that PAM sequences efficiently recognized *in vitro* by MgCas9 S6 are efficiently recognized *in vivo*. Similarly, plasmids harboring PAMs frequently cleaved *in vitro* by MgCas9 CA06 were actively cleaved *in vivo* by MgCas9 mod PI CA06, demonstrating that the hybrid MgCas9 was fully active. Concerning the difference of specificity, we observed, as expected from the *in vitro* results, that the PAM CA06-3 was not recognized by MgCas9 S6 but efficiently drives cleavage by MgCas9 with a CA06 PI domain. This result confirms that the specificity of MgCas9 is mediated by the PI domain and that an important difference of PAM recognition specificity can be observed depending on the origin of the PI domain. Interestingly, motifs that were considered as background on the *in vitro* assay because they were found at very low level (Ca06-4 for MgCas9 S6, S6-1 and S6-2 for MgCas9 CA06), were efficiently cut *in vivo*.

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

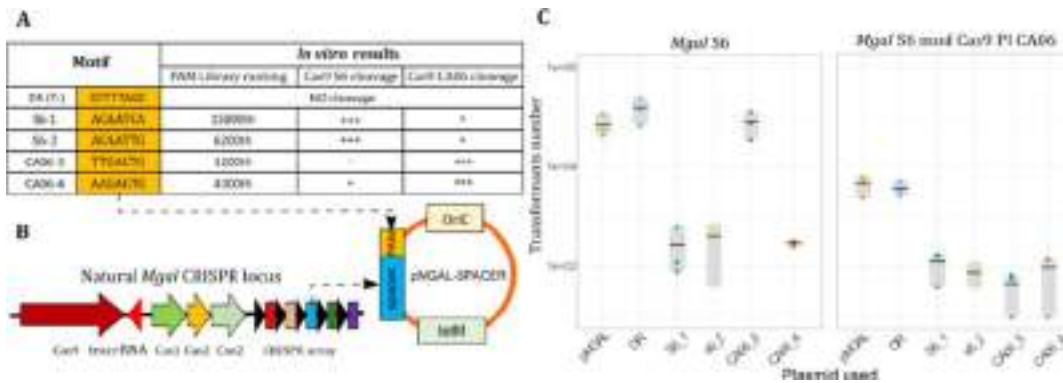


Figure 3 : Evaluation of PAM specificity *in vivo* using plasmid interference assays. A- PAM sequences for *in vivo* interference assays were selected from *in vitro* results. DR is the direct repeat sequence naturally present in the *Mgal* CRISPR array; this motif cannot be recognized by *Mgal*Cas9 and is used as a positive control of transformation and negative control for cleavage. Four others PAM sequences were chosen, that are unevenly cut *in vitro* by *Mg*Cas9 S6 and CA06. PAM sequences frequently retrieved after cleavage are noted (+++) whereas lowly and not recognized sequences are indicated by (+) and (-), respectively. B- Plasmid design for interference assays in *Mgal* S6 and *Mgal* S6 mod PI CA06. Plasmid pMGAL contains *Mgal* oriC, a *tet*(M) resistance marker and the third spacer which is present in the S6 CRISPR array, followed by selected PAM sequences. C- *In vivo* result of transformation assays for the two *Mgal* strains. The numbers of transformants obtained after three transformation replicates are represented as a box plot.

Understanding the evolution of *Mg*Cas9 in its genomic context

In order to better understand the link between the *Mgal* adaptation to HF and the evolution of *Mg*Cas9, an extended analysis of the CRISPR locus and other predicted bacterial defense systems was undertaken. First, the set of spacers inferred from *Mgal* CRISPR locus was compared in the different *Mgal* strains. We identified 458 unique spacers from 19 poultry strains and 76 from 85 HF strains. From all poultry strains, only one spacer with 1 mutation was shared between the vaccinal strain ts-11 (and derivatives K6372, K6356, K6212B, K6216D, K6208B, K5322C, K2966 and K6222B) and HF strains. Among *Mgal* poultry strains, a wide diversity of spacers was observed and no spacer was conserved in all of the strains. In comparison, among CRISPR arrays of HF strains, a low diversity of spacers was found and most spacers were highly conserved among strains (Table S4). Most spacers have been integrated before 2007 and this is in concordance with the predicted inactivation of *Mg*Cas9 since 2007. However, even if *Mg*Cas9 seems to be inactivated, some recent spacer acquisitions were predicted from the CRISPR arrays of strains from 2011, 2013 or 2015. We are currently looking for hits in various databases including phages and metagenomic databases to investigate the origin of the spacers.

Others defense systems against phages and invading DNAs were predicted in *Mgal* genomes (Table S5). Several of them were identified in the vicinity of the CRISPR locus, suggesting this system might be part of a more complex defense island containing also a type I R-M system (hereafter named RM 1) and *letA-letS* system recently characterized as bacterial defense systems against phage (Gao et al. 2020) (Figure S5). The last one, is composed by an ATPase (*letA*) and a S8 protease (*letS*) that has been functionally characterized as Toxine-Antitoxine system in *M. mycoides* subsp. *capricolum* and other mycoplasmas (Hill et al. 2021). The RM 1 that is found in some poultry strains close to the CRISPR locus, is lacking in the genomes of all HF strains. In addition, a transposase gene of an insertion sequence element (IS256) located immediately downstream of the defense island was found conserved in all strains. Two genes encoding a predicted S8 serine peptidase and a hypothetical protein, were also located upstream from the island. S8 serine peptidases are usually found in defense systems against phages (Gao et al. 2020) or sometimes against host immune system (McKenna et al. 2021). The putative S8 peptidase gene was predicted inactivated in several poultry strains but functional in all HF

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

strains except strains where the set of *cas* genes had been lost, suggesting both loci were lost during a single deletion event (Figure S5). In two strains, another type I R-M (hereafter named RM 2) was found close to another dynamic genome locus containing genes encoding variable surface proteins (VlhA) (Liu et al. 2000; Pflaum et al. 2015) (Figure S6, Table S5). **Finally, we show that all defense systems against phages showed a remarkably dynamic evolution among *Mgal* strains.**

	Poultry						Non-poultry					
	Strains	S8 family	CRISPR	letA/S	RM 1	RM 2	Strains	S8 family	CRISPR	letA/S	RM 1	RM 2
	S6	*	1	1	1	1	CA06	1	1	1	0	0
	Rlow	*	1	1	1	*	NC06_2006.081	0	0	1	0	0
	F	*	1	1	1	0	NC08_2008.031	0	0	1	0	0
	NTCC10105	1	1	1	1	0	NC95_13295-2	1	1	1	0	*
	Mx4	1	1	1	0	1i	NY01_2001.047	1	1	1	0	0
	f99	*	1	1	1	0	NC96_1596-4	1	1	1	0	0
	685	0	1	1	0	0	WI01_2001.043	1	1	1	0	0
	TS-11	*	1	1	0	0	VA94_7994	1	1	1	0	0
	K6356	0,5	1	1	0	ND	New 69 strains with Cas genes	1	1 or 0,5	1	0	0
	KUVMG	0	1	1	0	0	Strains without Cas genes	0	0	1	0	0

1 = Present

0 = Absent

* = Inactivated by DNA mutation

i = DNA region is invert

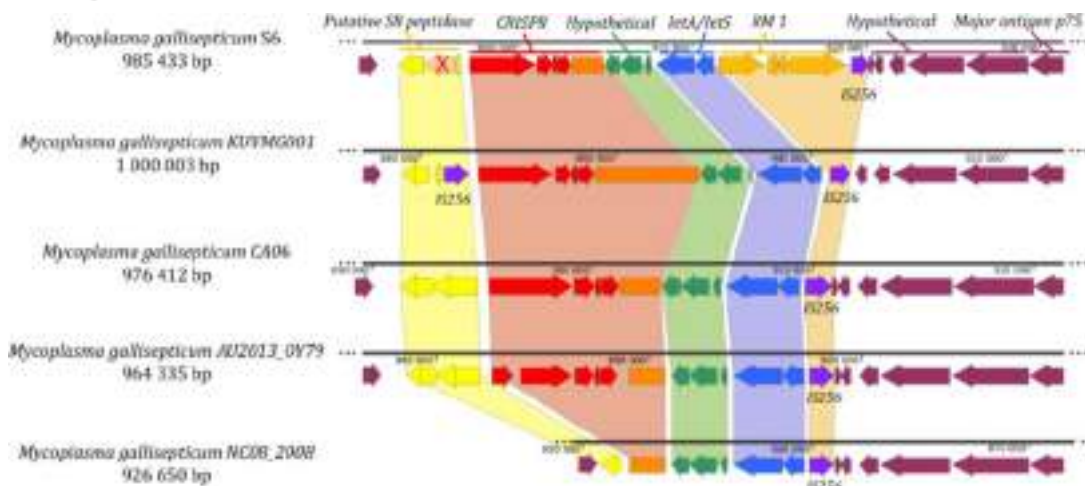


Figure S5 : Putative defense island in *Mgal* genome. Genes present in the putative defense island was shown for five representative *Mgal* strains (Table S5). All defense systems were associated with color: CRISPR (red), letA/letS (Blue), RM 1 (Orange). The two others putative systems are also associated with color: system containing S8 peptidase (Yellow) and hypothetical system (green).

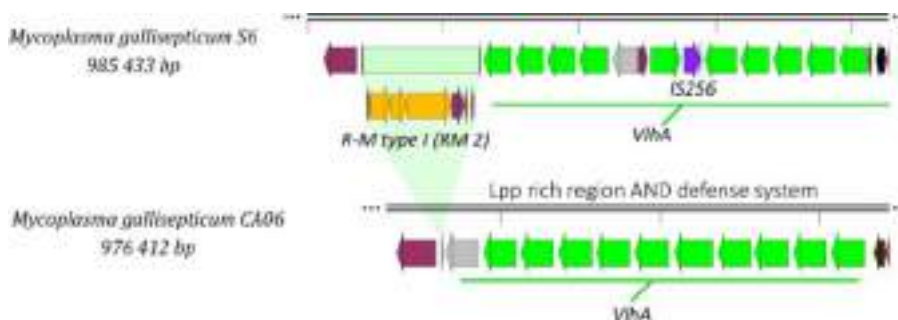


Figure S6 : Highly dynamic genome region with a RM system in *Mgal*. The region is composed by multiples copies of genes encoding for VlhA proteins (green) and by the RM 2 system in *Mgal* S6 genome. A transposase associated with IS256 is also found in the genomic locus of *Mgal* S6.

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

Building a genetic tool for mycoplasmas from *Mgal* natural CRISPR system

To evaluate the *Mgal* CRISPR/Cas9 system as a potential genetic tool for mollicutes, a minimal system was designed and introduced in an all-in-one plasmid (pTI2.0), including (1) the MgCas9 gene and (2) a sgRNA, both under the control of the constitutive pSpi promoter. Other genetic elements were (3) an antibiotic resistance cassette suitable for selection in most mollicutes (*tet(M)*) and, (4) an *oriC* from *Mgal* chromosome to ensure replication. A simplified version without MgCas9 gene was also constructed (pTI1.0) for use in *Mgal* that naturally expresses MgCas9. Double BseRI restriction sites were added in the sgRNA cassette to facilitate the introduction of a target sequence with annealed oligonucleotides. Unique restriction sites were inserted all around the plasmid to easily change the *oriC*, the antibiotic resistance marker and others genetic components. A set of plasmids was developed for four mycoplasma species distributed in the three main phylogenetic groups (*Mycoplasma capricolum* subsp. *capricolum* (*Mcap*), *Mycoplasma mycoides* subsp. *mycoides* (*Mmm*), *Mycoplasma gallisepticum* (*Mgal*) and *Mycoplasma pulmonis* (*Mpulm*) (Figure 4A). Transformation assays were performed in the different mycoplasmas to demonstrate the functionality of these all-in-one plasmids. In all cases, the sgRNA was designed to target the mycoplasma chromosome. Comparing transformation efficiencies of targeting and non-targeting plasmids, we observed a counter-selection effect close to 100% when targeting the chromosome (Figure 4B). Five surviving colonies of *Mgal* were analyzed and showed mutations on MgCas9 encoding gene (Gly₂₂₅ to Asp₂₂₅, Ala₁₁₁ to Asp₁₁₁, Ser₂₂₉ to Arg₂₂₉, Gln₅₈₂ to Lys₅₈₂ and a frameshift mutation introduced in Leu₁₇ codon). We also tried the system in *Mycoplasma pneumoniae* but any transformant was observed even without sgRNA, suggesting an intrinsic toxicity of MgCas9 in this species. Transformation assays of all-in-one plasmids in presence of recombination templates were attempted, but any recombinant was observed (data not shown).

As an example of application of *Mgal* CRISPR system as a counter-selection tool, we evaluated its ability to cure a mycoplasma genome from an integrated mobile element (Figure 4C). The selected target was an Integrative Conjugative Element (ICE) present in the *Mcap* chromosome that is known to self-excise at low frequency (Tardy et al. 2015). The gene encoding MgCas9 was first introduced in the *Mcap* genome and a replicative *oriC* plasmid harboring a sgRNA targeting a gene of the ICE, MCAPO559 (pMYCO1_sgRNA0559) was constructed. Transformation with a control plasmid without target (pMYCO1_sgRNA) generated 136,000, 475,000 and 20,000 transformants in three assays. In comparison, only 9, 7 and 0 transformants were retrieved when transforming the pMYCO1_sgRNA0559 (Figure 4D). A PCR screening and further sequencing were used to analyze the ICE locus of all these clones. One clone where the ICE had been properly cured from the genome was isolated, demonstrating the usability of *Mgal* CRISPR-based tools for selection of rare events in heterologous mycoplasma species. Three transformants that still possess the integrated ICE were further investigated by sequencing the *Mgal* Cas9 gene. Mutations were evidenced in all three cases (Glu₁₇₉ to stop, Gly₁₈ to Cys₁₈ and Phe₁₁₇₉ to Asp₁₁₇₉). **In conclusion, a functional minimal system was successfully derived from the natural *Mgal* CRISPR/Cas9 system and can be used as a powerful counter-selection tool in various mycoplasmas.**

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

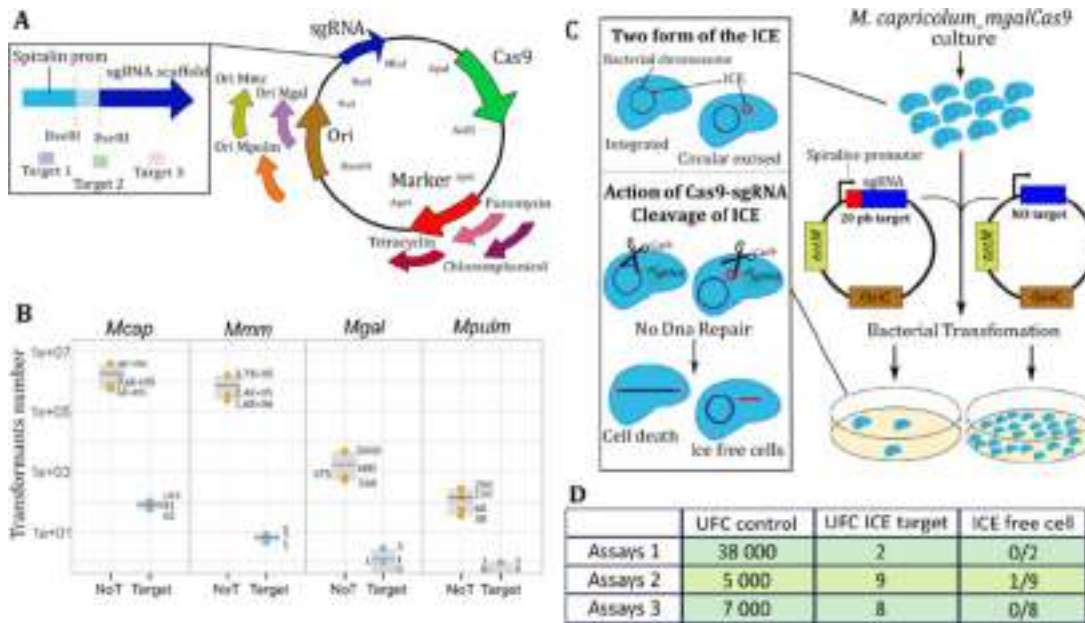


Figure 4 : Development of a counter-selection tool from *Mgal* CRISPR system. **A-** All-in-one plasmid backbone with main genetic elements including *MgCas9* gene (green arrow), sgRNA (blue), mycoplasma *oriC* (brown) and resistance marker (red). Restrictions sites introduced for easy change of main components are represented. **B-** Result obtained for *in vivo* transformation assays in four species of mycoplasma: *M. capricolum* subsp. *capricolum* (*Mcap*), *M. mycoides* subsp. *mycoides* (*Mmm*), *M. gallisepticum* (*Mgal*) and *M. pulmonis* (*Mpulm*). Numbers of transformants obtained using plasmid without target (NoT) or targeting the genome (Target) are indicated. **C-** Schematic of the experiment designed to cure *Mcap* genome from an Integrated Conjugative Element (ICE) using a minimal CRISPR/Cas9 system. **D-** Number of transformants obtained in three replicates of the ICE targeting experiment. Plasmids used contained a sgRNA with a spacer targeting the ICE (ICE target) or no (control).

Discussion

Emergence of a pathogen into a new host is an upheaval for both partners that can be considered as an inflection point in their evolutionary trajectories. New host pathogen interactions impose new constraints that will drive their genome evolution towards an ecologically sustainable equilibrium or an extinction of one partner. For a bacterial pathogen, changing of host means adapting virulence factors and metabolism to create a place for itself, potentially among other competitive species. In this perspective, the successful emergence of a bacterial pathogen into a new host would also depend on its ability to cope with a new environment of phages and other invading DNAs. In this study, we used the recently reported host jump of *Mgal* from poultry to HF as a case study to investigate this aspect of emergence that has been greatly overlooked.

The wide diversity of hosts colonized by mycoplasmas and the general absence of consistency between the phylogenetic positions of these minimal bacteria and their animal hosts suggest a number of host jumps have occurred along their evolution. In a number of cases, colonization of a new host has most probably been fostered by HGT, as it is now well established among ruminant mycoplasmas (Sirand-Pugnet, Lartigue, et al. 2007) or between distantly related mycoplasmas infecting birds. For instance, *M. gallisepticum* and *M. synoviae* belong to the Pneumoniae and Hominis phylogenetic groups, respectively, and the HGT of some genes encoding variable surface proteins Vh1a, neuraminidase or immunoglobulin cysteine protease CysP may have played a significant role in their ability to infect birds (Cizelj et al. 2011; Sirand-Pugnet, Lartigue, et al. 2007; Vasconcelos et al. 2005).

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

Surprisingly, the recent host shift of *Mgal* from poultry to HF does not seem to be associated with large-scale modifications of the genome as observed for other bacteria such as *S. aureus* subsp. *anaerobius* (Yebra et al. 2021). No gene acquisition was evidenced but rather an accumulation of thousands of SNPs and small-scale modifications affecting host pathogen interaction (Delaney et al. 2012; Tulman et al. 2012). In this context, the evolution of *Mgal* CRISPR locus is remarkable as it is marked by both SNP accumulation in *cas* genes and a complete change in the repertoire of spacers. CRISPR systems are defense systems against phages and other invading DNAs and they have been envisioned as barriers to HGT, which is one of the main driving forces of evolution in prokaryotes (Koonin 2019). In the case of the recent *Mgal* host shift, the lack of evidence for significant HGT is in accordance with a CRISPR system remaining active while evolving. Our data show that MgCas9 evolution during the host shift kept it functional but with noticeable changes in the AA sequence, especially in the PI domain which evolution modified its PAM recognition specificity.

Differences in the repertoires of PAM sequences recognized by MgCas9 originating from poultry and HF strains were particularly highlighted when using an *in vitro* assay and a library of plasmids with all possible combinations of PAM sequences. When comparing these results with those of an *in vivo* assay based plasmid interference, we confirmed that a PAM not recognized *in vitro* did not lead to the cleavage and loss of the plasmid, and that recognition specificity was dependent on the PI domain of MgCas9. We also noticed that PAM sequences only poorly recognized *in vitro* could lead to enough plasmid cleavage for stopping colonies from growing. This difference can be explained because the highly sensitive *in vitro* assay is designed to quantify cleavage efficacy of PAM sequences after one hour of incubation with the MgCas9/sgRNA complex. This assay revealed, in a quantitative manner, PAM sequences that are thermodynamically more susceptible to interact with MgCas9 PI domain. On the contrary, the *in vivo* assay is based on the cleavage of a low-copy number *oriC* plasmid by the endogenous CRISPR system of *Mgal*. Even if not frequent, any cleavage of the plasmid occurring before division of the bacteria (generation time for *Mgal* is 1.2 hour (Beaman and Pollack 1983) will hamper the propagation of the plasmid in the population and stop colony development on selective medium.

The complete reset of *Mgal* CRISPR system concomitantly with the 1994 host jump of the pathogen suggests this main event selected *Mgal* mutants not only with virulence factors evolved for an improved colonization of a new host, but also with defense systems adapted to a new phageome. Among poultry strains, the diversity of spacers was high, suggesting a fast turnover fostered by regular phage attacks. This hypothesis is in accordance with the demonstration that the CRISPR system is fully active in the S6 strain and the prediction that an active system has been preserved in all strains isolated from chicken and turkeys. On the contrary, in HF strains, spacers are very conserved among strains, suggesting the bacteria has to face a low diversity of phages. HF Strains isolated during years 1994 to 2007 kept the CRISPR system functional with a MgCas9 adapted to a new set of PAM sequences. This suggests a benefit-cost balance in favor of maintaining active the CRISPR defense system (Figure 5). As such, the CRISPR system of *Mgal* might have played an important role in *Mgal* emergence in HF. This further suggests that phage diversity may be an important barrier to bacterial establishment in a new host. A role of the phage community in facilitating or hindering emerging bacterial establishment may explain why *Mgal* is able to spillover, but not persist, in a large number of bird species (Staley et al. 2018).

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

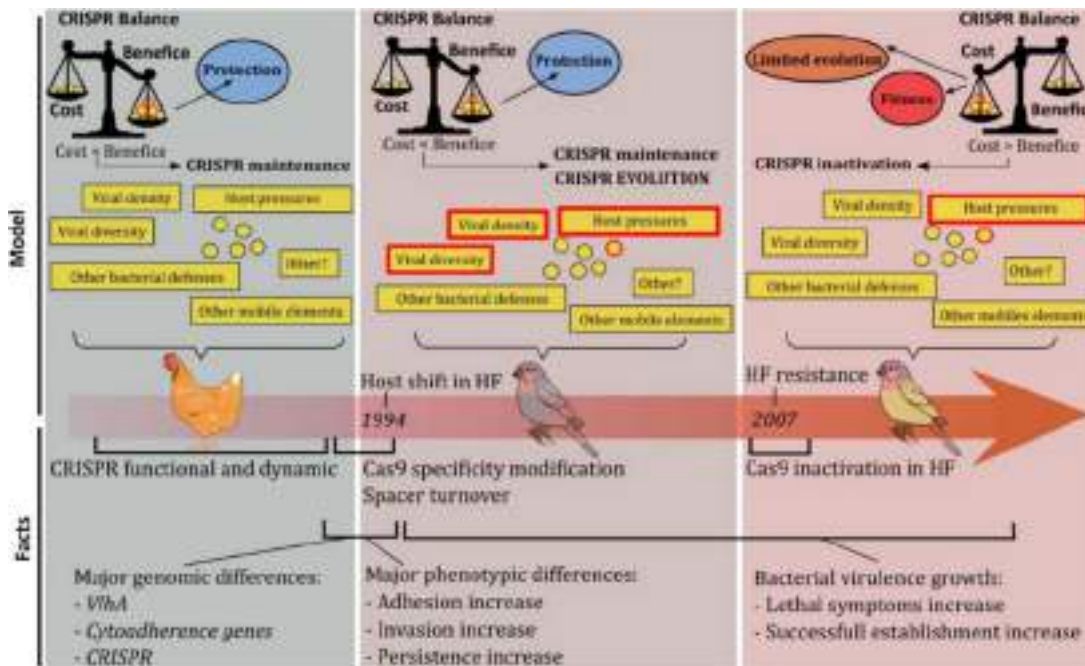


Figure 5 : Model of evolution of the CRISPR system in the context of *Mgal* emergence and adaptation to HF. The main steps of emergence of *Mgal* in HF are indicated. Data documented in literature and in this study are gathered as facts in the bottom of the figure. The model proposed integrating the benefit-cost balance and various constraints and parameters that may have driven the evolution of the CRISPR system is presented at the top of the figure.

Some models propose that the cost–benefit balance is the main driver for persistence of an active adaptive immunity system like CRISPR (Koonin 2019). At low diversity of phage, the balance is not in favor of maintaining an active CRISPR system, and less costly systems like Restrictions-Modifications systems are sufficient for resistance. Conversely, with a very high phage diversity, many spacers would be needed to ensure phage immunity, which might be more costly than other phage defense systems with fixed cost, like modification of phage surface receptor (Broniewski et al. 2020). In bacteria, different ways to switch the expression of CRISPR systems have been described (Patterson, Yevstigneyeva, and Fineran 2017) like quorum sensing that is responsible of CRISPR inactivation in *Serratia sp. ATCC39006* (Høyland-Kroghsbo et al. 2017; Patterson et al. 2016). This loss of Cas9 might also be the result of an evolutionary trade-off between defense and evolution as suggested in *E. coli* (García-Martínez et al. 2018).

In the case of *Mgal*, the observed inactivation and, in some strains, degradation of the CRISPR system in HF strains isolated after 2007 suggest the benefit-cost balance for maintaining an active CRISPR-Cas system has changed, leading to the observed mutations in MgCas9. Interestingly, we did not find bacteriophages in the genomes of the strains included in our study, suggesting the bacteria might still be protected. To investigate further this hypothesis, we extended *Mgal* genome analyses to other defense systems. Several systems including the recent Toxin-Antitoxin system characterized in *Mycoplasma mycoides* subsp. *capri* (Hill et al. 2021) and a RM systems were identified in the vicinity of the CRISPR locus, suggesting those were gathered in a defense island, as described in other bacteria (Bernheim and Sorek 2020). Only the TA system was found conserved among all poultry and HF strains. By contrast, the two RM systems located in the defense island and at another locus were systematically predicted inactivated in HF strains. Therefore, it does not seem that the inactivation of the CRISPR system in post-2007 HF strains had been compensated by a reinforcement of the RM systems.

Evaluating the current selective pressure imposed by phages on contemporary HF strains will require further investigations, but our results also raised the question of other defense mechanisms that might

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

bring enough protection to avoid maintaining an active CRISPR system. The conservation of the TA system suggests it might be involved. Alternatively, modifications of the phage receptors may also help the bacteria to resist to phage attacks. Currently, any phage receptor has been characterized in *Mgal*. However, some surface proteins associated with adhesion of the bacteria to the host cells have encountered remarkable changes during host shift, especially the V1ha variable surface proteins. (Delaney et al. 2012; Tulman et al. 2012). Possibly, such surface modifications of *Mgal* could have played a role in phage defense during host shift and subsequent evolution in HF. In this context, isolating *Mgal* phages would open interesting opportunities to investigate these hypotheses.

Other constraints related to the armed race between *Mgal* and its new host might have impacted the benefit-cost balance shaping the evolution of *Mgal* defense systems. Recent studies have shown that the virulence of HF strains has adaptively increased (Staley et al. 2018) while a widespread resistance in the host population of HF since 2007 (Bonneaud et al. 2011) (Figure 5). This host population resistance and the induced selective pressure imposed on *Mgal* to adapt its fitness possibly changed the benefit-cost balance of CRISPR and other defense systems. In this perspective, losing a functional CRISPR by mutation might be regarded as a way to switch-off a costly defense mechanism in a context of low phage pressure, while opening new evolution routes by relaxing barriers to HGT.

In conclusion, we have shown here that the emergence of a minimal pathogenic bacteria in a new host has been associated with a complete reset of the CRISPR defense system with several evolution steps including (1) a modification of Cas9 PAM recognition specificity, (2) a drastic change and a reduction of the diversity of the repertoire of spacers and, (3) multiple inactivation events after the early years of the emergence. A global study of the bacterial defense systems against phages and invading DNAs did not reveal new systems that may have compensated the loss of active CRISPR. On the contrary, some of the other defense systems (i.e. RM systems) were also systematically lost in recent strains, suggesting a general tendency to lower the guard against phages. This study is an invitation to consider the emergence of a pathogen into a new host in an enlarged ecological context, including not only the the pathogen and the host, but also the pathogens of the pathogen. Taking these multi-partner interactions into account will help understanding why some host-pathogen interactions are evolutionary sustainable and other not.

Material & Methods

Oligonucleotides and plasmids:

All the oligonucleotides used for this study were supplied by Eurogentec and are described in Table S1. All the plasmids constructed and used in this study are listed in Table S2.

Comparative genomics and *in silico* comparison of MgCas9 strains S6 vs CA06

Phylogenetic tree was made using Hiphygeny pipeline.

Phylogenetic analysis of MgCas9 was performed using proteic sequences and the phylogeny pipeline available at <http://www.phylogeny.fr/>. This pipeline includes the softwares Muscle, Gblocks, PhyML and TreeDyn. Representation of the tree was designed on iTOL (<https://itol.embl.de/>). Protein alignment was analyzed on MEGA-X using Muscle alignment software (Kumar et al. 2018). Mutations rates between S6 or CA06 were calculated from SNP detected using whole genome nucleotide alignment in MAUVE (Darling et al. 2004). For mutation rates of proteins, concatenated protein sequences of whole genome were compare using Blast-p.

Construction of a minimal *Mgal* sgRNA

Sequences of DR and tracrRNA were concatenated and the secondary structures of the hybrids were simulated using the mfold software at <http://unafold.rna.albany.edu/>. The sgRNA was designed according to (Jinek et al. 2012). The chimeric molecule was synthesized by IDT DNA company and further cloned in the pPS3.1 plasmid.

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

Plasmids Constructions

Plasmids pTZ57-Motif1, pTZ57-Motif2 and pTZ57-Motif3. The pTZ57-7nlibrary vector provided by CasZyme (Gasiunas et al. 2020) was used as PCR template. Primers D17/D20, D18/D20 and D19/D20 and the Q5 High-Fidelity DNA Polymerase kit (NEB, M0491) were respectively used to amplify linear DNA fragment for pTZ57-Motif1, pTZ57-Motif2 and pTZ57-Motif3. PCR products were submitted to restriction with the DpnI enzyme (NEB, R0176S) following the manufacturer's protocol. Both products were purified using GFX™ PCR DNA or Gel Band Purification Kit (GE-healthcare). The NEBuilder HiFi DNA Assembly Cloning Kit (NEB, E5520S) was used to assemble the fragments into the pTZ57-Motif1, pTZ57-Motif2 and pTZ57-Motif3 plasmids. Two µL of the assembly was transformed into *E. coli* NEB 5-alpha (NEB, C2987H). Transformants were screened after DNA minipreps using NucleoSpin Plasmid kit (Macherey-nagel, 740588.50) and enzymatic digestion. Sanger sequencing was performed for final verification.

Plasmids pET28-MgalS6Cas9 and pET28-sgRNA. The pET28-MgalCA06-Cas9 vector derived provided by CasZyme (Gasiunas et al. 2020) was used as PCR template. Primers B7/B8, A3/A4, A1/A2 B5/B6 the Q5 High-Fidelity DNA Polymerase kit (NEB, M0491) were respectively used to amplify MgalS6Cas9, sgRNA and backbone fragments. Cloning was performed using the same procedure as for pTZ57-Motif1 (see above) and resulted in the plasmids pET28-MgalS6Cas9 and pET28-sgRNA.

Plasmids pMGAL, pMGAL-Spacer-DR, pMGAL-Spacer-S6-1, pMGAL-Spacer-S6-2, pMGAL-Spacer-CA06-3, pMGAL-Spacer-CA06-4. The backbone shuttle vector pSRT2 was used. The origin of replication of *M. gallisepticum* has been identified by (Lee, Browning, and Markham 2008) as the area between MGA_0618 and MGA_0619. Primers E21/E22 were used and the Q5® High-Fidelity DNA Polymerase (M0491) kit to amplify the desired region. PCR product and pSRT2 vector were digested with the BamHI enzyme, following the suggested protocol. Both products were ligated following the T4 Ligase of Promega's protocol thus creating the pMGAL plasmid. The construction was completed with an introduction through an XmaI digestion and Ligation of annealed Primers E23/E24 carrying the third spacer sequence from the *Mgal* S6 CRISPR locus. Primers E25/E26, E27/E28, E29/E30, E31/E32, E33/E34 and the Q5 High-Fidelity DNA Polymerase kit (NEB, M0491) were respectively used to change nucleotide sequence after target sequence and insert desired PAM sequence to construct pMGAL-Spacer-DR, pMGAL-Spacer-S6-1, pMGAL-Spacer-S6-2, pMGAL-Spacer-CA06-3, pMGAL-Spacer-CA06-4. PCR products were submitted to restriction with the DpnI enzyme (NEB, R0176S) following the manufacturer's protocol. Both products were purified using GFX™ PCR DNA or Gel Band Purification Kit (GE-healthcare). The NEBuilder HiFi DNA Assembly Cloning Kit (NEB, E5520S) was used to assemble the fragments into pMGAL-Spacer-DR, pMGAL-Spacer-S6-1, pMGAL-Spacer-S6-2, pMGAL-Spacer-CA06-3, pMGAL-Spacer-CA06-4. Two µL of the assembly was transformed into *E. coli* NEB 5-alpha (NEB, C2987H). Transformants were screened after DNA minipreps using NucleoSpin Plasmid kit (Macherey-nagel, 740588.50) and enzymatic digestion. Sanger sequencing was performed for final verification.

Plasmid pTi1.0. The pMGAL vector was used as vector for pTi1.0. pSpi and sgRNA sequence were respectively amplified from pMYCO1 and XXXx by PCR using the Q5® High-Fidelity DNA Polymerase (M0491) kit and primers G41/G42, G43/G44. pMGAL and sgRNA cassette were digested with the EcoRI-HF enzyme (NEB, R3101S), following the suggested protocol. Digested pMGAL was dephosphorylated by Antarctic Phosphatase kit (NEB M0289S) 2 h at 37°C. After purification using GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, 28-9034-70), ligation was performed over night at 4°C using T4 DNA ligase (Promega M1801). Two µL of the assembly mixture were transformed into *E. coli* NEB 5-alpha (NEB, C2987H). Transformants were screened after DNA minipreps using NucleoSpin Plasmid kit (Macherey-nagel, 740588.50) and enzymatic digestion. Sanger sequencing was performed for final verification of pTi1.0 plasmid.

Plasmid pMYCO1_sgRNA. pSpi and sgRNA sequence were respectively amplified from pMYCO1 and XXXx by PCR using the Q5® High-Fidelity DNA Polymerase (M0491) kit and primers F35/F36, F37/F38. Overlap PCR were achieved between the two fragments using primers F35/F38. pMYCO1 and sgRNA cassette were digested with the EcoRI-HF enzyme (NEB, R3101S), following the suggested protocol.

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

Cloning was performed using the same procedure as for pTI1.0 (see above) and resulted in pMYCO1_sgRNA.

Plasmid pTi2.0. The pTI1.0 vector was used as vector for pTI2.0. MgCas9 sequence were amplified using Q5[®] High-Fidelity DNA Polymerase (M0491) kit with primers H45/H46. MgCas9 sequence and pTI1.0 were digested with the DraIII-HF (NEB), following the suggested protocol. Digested pTI1.0 was dephosphorylated by Antarctic Phosphatase kit (NEB M0289S) 2h at 37°C. Cloning was performed using the same procedure as for pTI1.0 (see above) and resulted in pTI2.0 plasmid.

Plasmids pTi2.1 and pTi2.3. The pTI2.0 vector was used as vector for pTI2.1 and pTI2.3. The *oriC* from the genomes of *M. mycoides* subsp. *capricolum* and *M. pulmonis* were amplified respectively from pMYCO1 and pMPO1 plasmids (Lartigue et al. 2003) using primers I46/I47, I48/I49 and Q5[®] High-Fidelity DNA Polymerase (M0491) kit. OriC fragments and pTI2.0 were digested with the BamHI and NsiI enzyme (NEB R3136S and R3127S), following the suggested protocol. Digested pTI2.0 was dephosphorylated by Antarctic Phosphatase kit (NEB M0289S) 2h at 37°C. Cloning was performed using the same procedure as for pTI1.0 (see above) and resulted in pTI2.1 and pTI2.3.

In all constructed that needed target sequence, 20 nucleotides targets were added using the same process. Targets were design as oligonucleotides R and F and are described in Table S1. 10 µL of each primer (100 µM) were heated with 2 µL of Adv2 polymerase buffer (TAKARA, 639232) at 95°C during 5 min and cooled down slowly (-0,1°C/sec) to room temperature. Plasmids were digested with BseRI (NEB, R0581S) 2 h at 37°C and after purification, targets sequences and plasmids were ligated with T4 DNA ligase (Promega) over night at 4°C. Transformation of 2 µL of ligation mix was performed in NEB 5α competent *E. coli* (NEB, C2987H). Colonies were screened by miniprep, restriction enzyme digestion and target sequence verified by sanger sequencing (Genewiz).

PAM library validation

The 7N degenerated PAM plasmid library was provided by Dr. Gasiunas from CasZyme (Gasiunas et al. 2020). Randomness of the 7 positions of the randomized PAM was validated as described in (Karvelis et al. 2019). Briefly, triplicate of short PCR of 15 cycles were performed on 2.5 ng of pTZ57-7Nlibrary using Phusion High-Fidelity DNA Polymerase kit (ThermoFisher, D-530XL) and primers C9/C10. After purification of triplicate pooled product using GeneJET PCR Purification Kit (ThermoFisher, K0701) new step of short PCR of 10 cycles was performed in triplicates using primers C11/C12 and pooled again. PCR products from 3 replicates were sequenced on an Illumina MiSeq V2 (2x150 bp). Datasets treatment was achieved using the tools available at <https://usegalaxy.eu/>. After trimming with Trimmomatic (Galaxy Version 0.36.5, average quality required: Q20) and quality check, the 3 datasets included 1.1 to 1.3 M of paired reads. For each dataset, reads R1 and R2 were merged using FLASH (Galaxy Version 1.2.11.4, Minimum overlap: 50, Maximum overlap: 65, Maximum mismatch density: 0.25) and filtered to perfectly match the expected sequence CGGCGACGTTGGGTCAACTNNNNNNNTGTCCTCTTCTTTAGCGTTA. Sequence logos were generated with the Sequence Logo software (Galaxy Version 3.5.0) and frequency matrices were calculated for each of the 7 randomized positions. The average frequency of each represented PAM was calculated (Figure S2) and a normalization factor N was calculated as the ratio between the average observed frequency and the theoretical frequency (1/16384).

In vitro cleavage of the PAM degenerated plasmid library

The assay was performed as described in (Karvelis et al. 2019). Cleavage assays using MgCas9 S6 and CA06 were performed on 1 µg of 7N Degenerated plasmid library. The complete digestion reactions were incubated with 2.5 U of DreamTaq DNA polymerase (5u/µL Thermo Fisher Scientific) and 0.5 µL of 10 mM dATP for an additional 30 min. at 72 °C. DNA were purified with GeneJET PCR Purification Kit. Two oligonucleotides adapters C13/C14 were annealed in primer annealing buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.5) during 95°C for 5min and cooling temperature until 25°C. 100 ng of the resulting adapter was ligated to an equal concentration of the purified 3' dA overhanging cleavage products for 1 h at 22°C in a 25 µL reaction volume in ligation buffer (40 mM

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

Tris-HCl pH 7.8 at 25°C, 10 mM MgCl₂, 10 mM DTT, 0.5 mM ATP, 5 % (w/v) PEG 4000, and 0.5 U T4 Ligase; Thermo Fisher Scientific). Using 10 µL of ligation reaction mixtures as a template and primers C13/C16, PCR were performed with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) for 15 cycles amplification in 100 µL total volume. Purification using GeneJET PCR Purification Kit and a second PCR using primer C11/C15 were performed.

For each MgCas9, 3 independent cleavage assays were performed. After capture of cleaved molecules by adapter ligation and PCR amplification, products were submitted to deep sequencing and datasets were analyzed as previously. Distributions of average PAM frequencies after cleavage by CA06 and S6 MgCas9 were calculated and the top 1000 more frequent PAM sequences (~ top 10%) were selected for further analyses to reduce the impact of background noise generated by rare or non-specific cleavage by the IVT mixture. Normalized frequencies were calculated as the ratio between the average observed frequencies after cleavage and the corresponding normalization factor N. PAM frequency matrices and sequence logos were then generated as previously.

In vitro cleavage of specific plasmids

Codon optimized sequence of MgCas9 strain CA06 under the T7 promoter and T7 terminator was provided by Dr. Gasiunas from CasZyme (Gasiunas et al. 2020). The pET28_MgalS6Cas9 plasmid with codon-optimized sequence of MgCas9 strain S6 was constructed in the lab. A PCR was performed using Advantage HF 2 PCR Kit (Takara, 639123) to produce linear matrices containing sgRNA from pET28-sgRNA plasmid. Cas9 was produced using PURExpress bacterial IVT kit (New England Biolabs). 0.5 µg of MgCas9 plasmid and a 100-fold molar excess of sgRNA linear template was added to the IVT reaction mix. 1 µL of RiboLock RNase Inhibitor (40 U; Thermo Fisher Scientific) were added in the IVT reaction mix. After 4 h of incubation, reaction was stopped by fast cooling on ice. 5 µL of the resulting Cas9-sgRNA ribonucleoprotein (RNP) complex were then combined with 1 µg of plasmid containing target in a 100 µL reaction buffer (10 mM Tris-HCl pH 7.5 at 37°C, 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT). After one hour of cleavage, DNA were purified using GFX™ PCR DNA or Gel Band Purification Kit (GE-healthcare). Results were observed using Gel electrophoresis.

Bacterial strains, culture conditions

Mgal S6 (Tax ID: 1006581) was cultured in Hayflick (Freundt 1983) modified medium under a 5 % CO₂ atmosphere. Tetracycline 5 µg/mL was used in selective conditions. *M. capricolum* subsp. *capricolum* (Tax ID: 40479) and *M. mycoides* subsp. *mycoides* (Tax ID: 2103) were cultured in SP5 medium (Labroussaa et al. 2016). Tetracycline 5 µg/mL was used in selective conditions. *M. pulmonis* UATP (Tax ID: 2107) was cultured in Hayflick medium (Freundt 1983), Tetracycline 5 µg/mL was used in selective conditions. *Mgal S6* strain with a modified PI domain of MgCas9 was obtained as described in (Ipoutcha 2021, submitted to ACS Synthetic Biology).

Transformation of *Mgal*

Mgal cells were grown for 36 h before transformation. At pH 6.2 to 6.5, aliquots of 10 mL were centrifuged during 15 min at 6,000 x g, 10°C. Cells were then resuspended with 5 mL of HBSS 1X wash buffer (ThermoFisher, 14065056) and centrifuged 15 min at 6000 x g, 10°C. The pellet was resuspended in 250 µL of CaCl₂ 0.1 M and incubated for 30 min on ice. Cold CaCl₂-incubated cells were gently mixed with plasmid DNA (10 µg) and 10 µg of yeast tRNA (ThermoFisher, AM7119). Then, 2 mL of 40 % polyethylene glycol 6000 (PEG) (Sigma, 11130) dissolved in HBSS 1X buffer were added to the cells. After 2 min of incubation at room temperature, contact with the PEG was stopped by addition of 20 mL of HBSS 1X wash buffer. Mixture was centrifuged 15 min at 6,000 x g, 10°C and the cells were resuspended in 1 mL of modified Hayflick medium pre-warmed at 37°C. After 2 hours of incubation at 37°C, cells were plated on Hayflick selective plates. After 10-15 days at 37°C with 5 % CO₂, colonies were counted, picked and re-suspended in 1 mL of Hayflick for three passages (one passage ~48 h).

Transformation of *M. capricolum* subsp. *capricolum*

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

Culture of *M. capricolum* subsp. *capricolum* were grown in SOB+ medium (pH=7,5) at 30°C. At pH 6,2 – 6,4, 6 mL of culture was centrifuged during 15 min at 5,800 x g, 10°C. Cells were then resuspended in 3 mL of wash buffer (Tris 10 mM - NaCl 250 mM, pH=6,5). After centrifugation 15min at 5,800 x g, 10°C, the pellet was resuspended in 200 µL of anhydrous CaCl₂ 0,1 M and incubated 30 min in ice. The mix was transferred in a tube containing 400 µL of SP5 without serum and 10 µg of plasmid. Then the same volume of 2X Fusion Buffer (Tris-HCL 20 mM, MgCl₂-6H₂O 20 mM; NaCl 500 mM; PEG 6000 (FLUKA) 10%) was added and incubated 90 min at 30°C. After incubation, 5mL of SP5 medium was added and the pellet was resuspended in 500 µL of SP5 medium after a centrifugation 15 min at 5,800 x g, 10°C. Finally, cells were plated on SP5 selective plates. After 3-10 days at 37°C, colonies were counted, picked and re-suspended in 1 mL of SP5 for three passages (one passage ~24 h).

Transformation of *Mycoplasma mycoides* subsp. *mycoides* T1/44

M. mycoides subsp. *mycoides* was grown in SP5 medium until pH=7. 4mL was centrifuged during 15 min at 5,800 x g, 10°C. Cells were then resuspended in 2mL of ST buffer (Tris 10 mM, Sucrose 0,5 M, pH=6,5) and centrifuged again 15 min at 5,800 x g, 10°C. Cells were resuspended in 250 µL of CaCl₂ 0,1 M and incubated 30 min in ice. 10 µg of yeast tRNA and 10 µg of plasmids were transferred in cells mixture and 2mL of PEG 8000 70 % dissolved in ST buffer was added. After 2 min of incubation, centrifugation was achieved 15 min at 12,000 x g, 10°C. Cell pellet was resuspended in 500 µL of warm SP5 and after 2 h of incubation at 37°C. Finally, cells were plated on SP5 selective plates. After 3-10 days at 37°C, colonies were counted, picked and re-suspended in 1 mL of SP5 for three passages (one passage ~24 h).

Transformation of *M. pulmonis*

M. pulmonis was grown in hayflick medium until pH= 6,80. 3 mL was centrifuged during 10 min at 10, 000 x g, 10°C. The pellet was resuspended in 2 mL of PBS 1X buffer (NaCl 13,7 mM, KCl 0,27 mM, KH₂PO₄ 0,15 mM, Na₂HPO₄ 12H₂O 8 mM, pH=7,3) and centrifuged again 10 min at 10, 000 x g, 10°C. Cells were resuspended in 250 µL of PBS 1X. 10 µg of yeast tRNA and 10 µg of plasmids were transferred in cells mixture and 2 mL of PEG 40 % dissolved in PBS 1X was added. After 2min of incubation, centrifugation was achieved 12 min at 12, 000 x g, 10°C. Cell pellet was resuspended in 300 µL of warm hayflick and after 3 h of incubation at 37 °C, cells were plated on hayflick selective plates. After 10-15 days at 37 °C with 5 % CO₂, colonies were counted, picked and re-suspended in 1 mL of Hayflick for three passages (one passage ~48 h).

Construction of *M. capricolum* subsp. *capricolum* strain expressing MgCas9

M. capricolum subsp. *capricolum* strain expressing MgCas9 from an expression cassette introduced in the genome was constructed using synthetic biologic approach (Lartigue et al. 2009). *M. capricolum* subsp. *capricolum* genome was modified in yeast to introduce the MgalCas9 expression cassette and back-transplanted in mycoplasma cells to obtained a mutant strain expressing MgCas9.

Colony analysis

PCR screening was performed on all transformants using Advantage HF 2 PCR Kit (Takara, 639123). List of all primers used are detailed in Table S1.

Acknowledgements

We are very thankful to Dr. Gasiunas from CasZyme for sharing the 7N PAM-degenerated pTZ57-7n library, the pET28ID116 and for appreciated advices on the protocol used for in vitro cleavage.

References

Beaman, K. D., and J. D. Pollack. 1983. "Synthesis of Adenylate Nucleotides by Mollicutes (Mycoplasmas)." *Journal of General*

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

- Microbiology* 129(10):3103–10. doi: 10.1099/00221287-129-10-3103.
- Bernheim, Aude, and Rotem Sorek. 2020. "The Pan-Immune System of Bacteria: Antiviral Defence as a Community Resource." *Nature Reviews Microbiology* 18(February):113–19. doi: 10.1038/s41579-019-0278-2.
- Bonneaud, Camille, Susan L. Balenger, Andrew F. Russell, Jiangwen Zhang, Geoffrey E. Hill, and Scott V. Edwards. 2011. "Rapid Evolution of Disease Resistance Is Accompanied by Functional Changes in Gene Expression in a Wild Bird." *Proceedings of the National Academy of Sciences of the United States of America* 108(19):7866–71. doi: 10.1073/pnas.1018580108.
- Broniewski, Jenny M., Sean Meaden, Steve Paterson, Angus Buckling, and Edze R. Westra. 2020. "The Effect of Phage Genetic Diversity on Bacterial Resistance Evolution." *The ISME Journal* 14(3):828–36. doi: 10.1038/s41396-019-0577-7.
- Citti, C., E. Dordet-Frisoni, L. X. Nouvel, C. H. Kuo, and E. Baranowski. 2018. "Horizontal Gene Transfers in Mycoplasmas (Mollicutes)." *Current Issues in Molecular Biology* 29:3–22. doi: 10.21775/cimb.029.003.
- Cizelj, Ivanka, Rebeka Lucijana Berčič, Daliborka Dušanić, Mojca Narat, Janko Kos, Peter Dovč, and Dušan Benčina. 2011. "*Mycoplasma Gallisepticum* and *Mycoplasma Synoviae* Express a Cysteine Protease CysP, Which Can Cleave Chicken IgG into Fab and Fc." *Microbiology* 157(2):362–72. doi: 10.1099/mic.0.045641-0.
- Darling, Aaron C. E., Bob Mau, Frederick R. Blattner, and Nicole T. Perna. 2004. "Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements." *Genome Research* 14(7):1394–1403. doi: 10.1101/gr.2289704.
- Delaney, Nigel F., Susan Balenger, Camille Bonneaud, Christopher J. Marx, Geoffrey E. Hill, Naola Ferguson-Noel, Peter Tsai, Allen Rodrigo, and Scott V. Edwards. 2012. "Ultrafast Evolution and Loss of CRISPRs Following a Host Shift in a Novel Wildlife Pathogen, *Mycoplasma Gallisepticum*." *PLoS Genetics* 8(2). doi: 10.1371/journal.pgen.1002511.
- Dhondt, Andre A., Diane L. Tessaglia, and Roger L. Slothower. 1998. "Epidemic From Conjunctivitis America in House Finches." *Journal of Wildlife Diseases* 34(2):265–80.
- Freundt, E. A. 1983. "CULTURE MEDIA FOR CLASSIC MYCOPLASMAS." in *Methods in Mycoplasmaology*.
- Gao, Linyi, Han Altae-Tran, Francisca Böhning, Kira S. Makarova, Michael Segel, Jonathan L. Schmid-Burgk, Jeremy Koob, Yuri I. Wolf, Eugene V. Koonin, and Feng Zhang. 2020. "Diverse Enzymatic Activities Mediate Antiviral Immunity in Prokaryotes." *Science*. doi: 10.1126/science.aba0372.Diverse.
- García-Martínez, Jesús, Rafael D. Maldonado, Noemí M. Guzmán, and Francisco J. M. Mojica. 2018. "The CRISPR Conundrum: Evolve and Maybe Die, or Survive and Risk Stagnation." *Microbial Cell* 5(6):262–68. doi: 10.15698/mic2018.06.634.
- Gasiunas, Giedrius, Joshua K. Young, Tautvydas Karvelis, Darius Kazlauskas, Tomas Urbaitis, Monika Jasnauskaitė, Mantvyda M. Grusyte, Sushmitha Paulraj, Po-hao Wang, Zhenglin Hou, Shane K. Dooley, Mark Cigan, Clara Alarcon, N. Doane Chilcoat, Greta Bigelyte, Jennifer L. Curcuro, Megumu Mabuchi, Zhiyi Sun, Ryan T. Fuchs, Ezra Schildkraut, Peter R. Weigele, William E. Jack, G. Brett Robb, Č. Venclovas, and Virginijus Siksnys. 2020. "A Catalogue of Biochemically Diverse CRISPR-Cas9 Orthologs." *Nature Communications*. doi: 10.1038/s41467-020-19344-1.
- Hill, Virginia, Hatice Akarsu, Rubén Sánchez Barbarroja, Valentina L. Cippà, Peter Kuhnert, Martin Heller, Laurent Falquet, Manfred Heller, Michael H. Stoffel, Fabien Labroussaa, and Joerg Jores. 2021. "Minimalistic Mycoplasmas Harbor Different Functional Toxin-Antitoxin Systems" edited by J. Stülke. *PLOS Genetics* 17(10):e1009365. doi: 10.1371/journal.pgen.1009365.
- Høyland-Kroghsbo, Nina M., Jon Paczkowski, Sampriti Mukherjee, Jenny Broniewski, Edze Westra, Joseph Bondy-Denomy, and Bonnie L. Bassler. 2017. "Quorum Sensing Controls the Pseudomonas Aeruginosa CRISPR-Cas Adaptive Immune System." *Proceedings of the National Academy of Sciences* 114(1):131–35. doi: 10.1073/pnas.1617415113.
- Ipoutcha, Thomas, Iason Tsampanopoulos, Vincent Talenton, Christine Gaspin, Annick Moisan, Caray A. Walker, Joe Brownlie, Alain Blanchard, Patricia Thebault, and Pascal Sirand-pugnet. 2019. "Multiple Origins and Specific Evolution of CRISPR / Cas9 Systems in Minimal Bacteria (Mollicutes)." *Frontiers in Microbiology* 10(November):1–14. doi: 10.3389/fmicb.2019.02701.
- Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337(6096):816–21. doi: 10.1126/science.1225829.
- Karvelis, Tautvydas, Joshua K. Young, and Virginijus Siksnys. 2019. *A Pipeline for Characterization of Novel Cas9 Orthologs*. Vol. 616. 1st ed. Elsevier Inc.
- Koonin, Eugene V. 2019. "CRISPR: A New Principle of Genome Engineering Linked to Conceptual Shifts in Evolutionary Biology." *Biology and Philosophy* 34(1). doi: 10.1007/s10539-018-9658-7.
- Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." *Molecular Biology and Evolution* 35(6):1547–49. doi: 10.1093/molbev/msy096.
- Labroussaa, Fabien, Anne Lebaudy, Vincent Baby, Géraldine Gourgues, Dominick Matteau, Sanjay Vashee, Pascal Sirand-Pugnet, Sébastien Rodrigue, and Carole Lartigue. 2016. "Impact of Donor-Recipient Phylogenetic Distance on Bacterial Genome Transplantation." *Nucleic Acids Research* 44(17):8501–11. doi: 10.1093/nar/gkw688.
- Lartigue, Carole, Alain Blanchard, Joël Renaudin, François Thiaucourt, and Pascal Sirand-Pugnet. 2003. "Host Specificity of Mollicutes OriC Plasmids: Functional Analysis of Replication Origin." *Nucleic Acids Research* 31(22):6610–18. doi: 10.1093/nar/gkg848.
- Lartigue, Carole, Sanjay Vashee, Mikkel A. Algire, Ray Yuan Chuang, Gwynedd A. Benders, Li Ma, Vladimir N. Noskov, Evgeniya A. Denisova, Daniel G. Gibson, Nacyra Assad-Garcia, Nina Alperovich, David W. Thomas, Chuck Merryman, Clyde A. Hutchison, Hamilton O. Smith, J. Craig Venter, and John I. Glass. 2009. "Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast." *Science* 325(5948):1693–96. doi: 10.1126/science.1173759.
- Lee, S. W., G. F. Browning, and P. F. Markham. 2008. "Development of a Replicable OriC Plasmid for *Mycoplasma Gallisepticum* and *Mycoplasma Imitans*, and Gene Disruption through Homologous Recombination in *M. Gallisepticum*." *Microbiology*. doi: 10.1099/mic.0.2008/019208-0.
- Liu, Li, Kevin Dybvig, Victor S. Panangala, Vicky L. Van Santen, and Christopher T. French. 2000. "GAA Trinucleotide Repeat Region Regulates M9/PMGA Gene Expression in *Mycoplasma Gallisepticum*." *Infection and Immunity* 68(2):871–76. doi: 10.1128/IAI.68.2.871-876.2000.
- Makarova, Kira S., Yuri I. Wolf, Jaime Iranzo, Sergey A. Shmakov, Omer S. Alkhnbashi, Stan J. J. Brouns, Emmanuelle Charpentier, David Cheng, Daniel H. Haft, Philippe Horvath, Sylvain Moineau, Francisco J. M. Mojica, David Scott, Shiraz A. Shah, Virginijus Siksnys, Michael P. Terns, Česlovas Venclovas, Malcolm F. White, Alexander F. Yakunin, Winston Yan, Feng Zhang, Roger A. Garrett, Rolf Backofen, John van der Oost, Rodolphe Barrangou, and Eugene V. Koonin. 2020. "Evolutionary Classification of CRISPR–Cas Systems: A Burst of Class 2 and Derived Variants." *Nature Reviews Microbiology* 18(2):67–83. doi: 10.1038/s41579-019-0299-x.
- May, Meghan, Mitchell F. Balish, and Alain Blanchard. 2014. *The Order Mycoplasmatales*.
- McKenna, Sophie, Kristin Krohn Huse, Sean Giblin, Max Pearson, Mohammed Said Majid Al Shibar, Shiranee Sriskandan, Stephen Matthews, and James Edward Pease. 2021. "The Role of Streptococcal Cell-Envelope Proteases in Bacterial

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

- Evasion of the Innate Immune System." *Journal of Innate Immunity* 1–20. doi: 10.1159/000516956.
- Oliveira, Pedro H., Marie Touchon, and Eduardo P. C. Rocha. 2014. "The Interplay of Restriction-Modification Systems with Mobile Genetic Elements and Their Prokaryotic Hosts." *Nucleic Acids Research* 42(16):10618–32. doi: 10.1093/nar/gku734.
- Patterson, Adrian G., Simon A. Jackson, Corinda Taylor, Gary B. Evans, George P. C. Salmund, Rita Przybilski, Raymond H. J. Staals, and Peter C. Fineran. 2016. "Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems." *Molecular Cell* 64(6):1102–8. doi: 10.1016/j.molcel.2016.11.012.
- Patterson, Adrian G., Mariya S. Yevstigneyeva, and Peter C. Fineran. 2017. "Regulation of CRISPR-Cas Adaptive Immune Systems." *Current Opinion in Microbiology* 37:1–7. doi: 10.1016/j.mib.2017.02.004.
- Pflaum, K., E. R. Tulman, J. Beaudet, X. Liao, and S. J. Geary. 2015. "Global Changes in *Mycoplasma Gallisepticum* Phase-Variable Lipoprotein Gene VlhA Expression during in Vivo Infection of the Natural Chicken Host." *Infection and Immunity* 84(1):351–55. doi: 10.1128/IAI.01092-15.
- Sirand-Pugnet, Pascal, Christine Citti, Aurélien Barré, and Alain Blanchard. 2007. "Evolution of Mollicutes: Down a Bumpy Road with Twists and Turns." *Research in Microbiology* 158(10):754–66. doi: 10.1016/j.resmic.2007.09.007.
- Sirand-Pugnet, Pascal, Carole Lartigou, Marc Marena, Daniel Jacob, Aurélien Barré, Valérie Barbe, Chantal Schenowitz, Sophie Mangenot, Arnaud Couloux, Beatrice Segurens, Antoine De Daruvar, Alain Blanchard, and Christine Citti. 2007. "Being Pathogenic, Plastic, and Sexual While Living with a Nearly Minimal Bacterial Genome." *PLoS Genetics* 3(5):744–58. doi: 10.1371/journal.pgen.0030075.
- Staley, Molly, Geoffrey E. Hill, Chloe C. Josefson, Jonathan W. Armbruster, and Camille Bonneaud. 2018. "Bacterial Pathogen Emergence Requires More than Direct Contact with a Novel Passerine Host." *Infection and Immunity* 86(3):1–9. doi: 10.1128/IAI.00863-17.
- Tardy, Florence, Virginie Mick, Emilie Dordet-Frisoni, Marc Serge Marena, Pascal Sirand-Pugnet, Alain Blanchard, and Christine Citti. 2015. "Integrative Conjugative Elements Are Widespread in Field Isolates of *Mycoplasma* Species Pathogenic for Ruminants." *Applied and Environmental Microbiology* 81(5):1634–43. doi: 10.1128/AEM.03723-14.
- Tulman, E. R., X. Liao, S. M. Szczepanek, D. H. Ley, G. F. Kutish, and S. J. Geary. 2012. "Extensive Variation in Surface Lipoprotein Gene Content and Genomic Changes Associated with Virulence during Evolution of a Novel North American House Finch Epizootic Strain of *Mycoplasma Gallisepticum*." *Microbiology (United Kingdom)* 158(8):2073–88. doi: 10.1099/mic.0.058560-0.
- Vasconcelos, Ana Tereza R., Henrique B. Ferreira, Cristiano V Bizarro, Sandro L. Bonatto, Marcos O. Carvalho, Paulo M. Pinto, Darcy F. Almeida, Luiz G. P. Almeida, and Rosana Almeida. 2005. "Swine and Poultry Pathogens: The Complete Genome Sequences of Two Strains Of." *Microbiology* 187(16):5568–77. doi: 10.1128/JB.187.16.5568.
- Yebrá, Gonzalo, Andreas F. Haag, Maan M. Neamah, Bryan A. Wee, Emily J. Richardson, Pilar Horcajo, Sander Granneman, María Ángeles Tormo-Más, Ricardo de la Fuente, J. Ross Fitzgerald, and José R. Penadés. 2021. "Radical Genome Remodelling Accompanied the Emergence of a Novel Host-Restricted Bacterial Pathogen" edited by A. Peschel. *PLoS Pathogens* 17(5):e1009606. doi: 10.1371/journal.ppat.1009606.

3. Article en finalisation pour soumission à CRISPR journal

Lors du développement des plasmides all-in-one, des tentatives d'utilisation de ce système CRISPR de mycoplasme pour de la mutagénèse directe par recombinaison homologue ont été réalisées en présence de matrice de recombinaison. Les premiers essais n'ayant pas été concluants, nous avons essayé de contrôler dans le temps l'expression de MgCas9 par le promoteur inductible à la tétracycline *pxyl-tetO2* précédemment développé au laboratoire chez *M. agalactiae* et *S. citri* (Breton *et al.*, 2010). L'idée était d'induire le clivage après quelques passages pour laisser plus de temps à la bactérie de réaliser la recombinaison homologue sur le site visé. Cette approche n'a pas permis d'obtenir de colonies vivantes avec un sgRNA ciblant le génome. Ceci est probablement dû à une importante fuite du promoteur, comme cela a été observé chez *Mmm* (chapitre 4), une espèce très proche de *Mcap*.

Des essais *in vitro* de production et purification de la protéine Cas9 de *Mgal S6* ont été réalisés avec l'aide de Laure Bataille, ancienne ingénieure au laboratoire. L'enjeu initial était la mise en place d'un test de clivage *in vitro* afin de tester la dégénérescence de la PAM par exemple. Il aurait également été envisageable par la suite de regarder la

CHAPITRE 2 : Le système CRISPR-Cas9 de *Mycoplasma gallisepticum* : de la caractérisation du système naturel à l'outil d'ingénierie des génomes

structure cristalline de cette dernière et possiblement de la comparer à celle de la souche CA06. Malheureusement, la protéine dotée d'une étiquette STREP-tag et n'a pu être correctement produite chez *E. coli*, les profils de chromatographie d'exclusion suggérant que la protéine soit agrégée. Une production directe dans un mycoplasme (*Mcap*) a été tentée, mais le même résultat a été obtenu. En augmentant la concentration d'agent réducteur, ces agrégats ont semblé moins importants, soulignant alors la possibilité que des ponts disulfures pourraient s'être formés entre protéines. La publication montrant l'utilisation des systèmes IVT (« *In vitro* translation ») avec un protocole optimisé pour la production d'une Cas9 de *Mgal* fonctionnelle a entraîné l'abandon de la stratégie de production d'une protéine recombinante (Gasiunas *et al.*, 2020).

L'utilisation directe de MgCas9 pour la mutagenèse chez les mycoplasmes se révélant impossible à mettre en place, nous avons imaginé utiliser une version inactivée de MgCas9 pour faire de l'interférence, comme cela avait été fait pour d'autres bactéries. Les sites catalytiques de SpyoCas9 et SauCas9 étant conservés chez MgCas9, nous avons inactivé les deux acides aminés responsables des activités nucléases, HNH et RuvC. Cette « dead MgCas9 » a ensuite été testée pour contrôler l'expression du gène *glpO* chez *Mcap*. Des colonies ont pu être obtenues après transformation avec un plasmide « all-in-one » exprimant la dead MgCas9 et un sgRNA ciblant le gène *glpO*, démontrant la non toxicité de la protéine mutée. Nous avons testé 7 sgRNA différents dans cette série d'expériences, afin de maximiser nos chances d'éteindre l'expression du gène *glpO*. Malheureusement, que ce soit par mesure du niveau d'expression de *glpO* par RT-qPCR, ou par un test quantitatif de la production d'H₂O₂ au cours du temps, aucune différence significative n'a pu être observée après quelques essais. Il faut noter que les sgRNA dessinés à l'époque correspondaient à des séquences PAM obtenues par analyses bio-informatique préalablement à mon arrivée au laboratoire (NNNAAAA). Même si certaines des séquences PAM utilisées pour ces 7 guides ont été retrouvées par la suite dans la liste exhaustive des PAM fonctionnelles *in vitro*, elles ne font pas partie des séquences PAM préférentiellement utilisées *in vitro* par MgCas9 de S6, et pourraient être qualifiées de PAM non canoniques (Yamano *et al.*, 2017). Ces PAM non canoniques ne seraient reconnues par le système que grâce à la « souplesse » de Cas9 et sa torsion. Il se peut que cela ne puisse fournir une stabilité suffisante pour bloquer le complexe RNAP de la transcription. Cet axe a également été abandonné, mais de nouveaux essais pourraient être envisagés avec des séquences PAM optimales.

CHAPITRE 3 : Utilisation d'un système de
recombinaison exogène chez *Mycoplasma*
gallisepticum comme outil d'ingénierie de génome

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

L'efficacité de la recombinaison homologue est faible chez les mycoplasmes et étant donné qu'elle ne semble pas stimulée efficacement par des cassures doubles brins de l'ADN, des solutions alternatives ont été recherchées. Nous nous sommes tout d'abord concentrés sur les possibilités d'ingénierie de *Mgal*. De nombreuses études chez les bactéries concernant l'importation de systèmes hétérologues de réparations de l'ADN couplé ou non à l'utilisation d'un système CRISPR ont été publiées début 2019 (Corts *et al.*, 2019; Tucker *et al.*, 2019; Y. Wu *et al.*, 2019; Yin *et al.*, 2019). Un approfondissement sur le sujet a permis d'imaginer l'utilisation de protéines SSAPs (Single Strand Annealing proteins) comme le système lambda Red ou le système RecET, pour apporter aux mycoplasmes un système de recombinaison homologue exogène utilisable comme outil d'ingénierie génomique. L'objectif était alors d'identifier un système hétérologue de recombinaison, fonctionnel dans *Mgal*, et permettant de réaliser des modifications génétiques, avec ou sans le système CRISPR. Mon travail a commencé par une analyse de la littérature ainsi que des analyses de génomique comparée pour identifier des systèmes de ce type afin de pouvoir ensuite les tester dans la bactérie.

1. Principaux résultats

Cette étude nous a permis de débloquent le verrou de la recombinaison homologue chez *Mgal*. À l'aide du système exogène RecET provenant du phage SPP1 de *Bacillus subtilis*, nous avons montré qu'en utilisant un promoteur adapté, il était possible d'obtenir des recombinants sur deux gènes différents de *Mgal*. Ce système, bien que limité à de petits événements de délétions, substitutions ou insertions, permet une avancée importante dans l'étude de ce pathogène pour lequel aucun outil de mutagenèse ciblée n'était disponible. Il ouvre aussi la voie au développement de stratégies similaires pour l'ingénierie de nombreuses espèces de mollicutes.

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

2. Article soumis à ACS Synthetic Biology (27/10/2021, en cours de révision)

Genome engineering in *Mycoplasma gallisepticum* using exogenous recombination systems

Authors: Thomas IPOUTCHA^{1,2}, Géraldine GOURGUES^{1,2}, Carole LARTIGUE^{1,2}, Alain BLANCHARD^{1,2, §}, Pascal SIRAND-PUGNET^{1,2, § *}

¹ INRAE, UMR 1332 de Biologie du Fruit et Pathologie, F-33140 Villenave d'Ornon, France

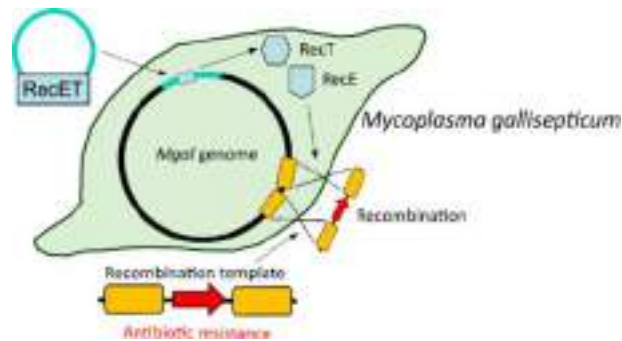
² Univ. Bordeaux, UMR 1332 de Biologie du Fruit et Pathologie, F-33140 Villenave d'Ornon,

*Corresponding author: Pascal Sirand-Pugnet, pascal.sirand-pugnet@inrae.fr

§ both authors contributed equally to the work

ABSTRACT:

Mycoplasma gallisepticum (*Mgal*) is an important pathogen affecting not only poultry but also wild birds (North American house finches) since a characterized jump of host that occurred in 1994. The molecular determinants of *Mgal* virulence and host specificity are still largely unknown, mostly due to the lack of efficient methods for functional genomics. After evaluating two exogenous recombination systems derived from phages found in the phylogenetically-related *Spiroplasma phoeniceum* and the more distant *Bacillus subtilis*, the RecET-like system from the latter was successfully used for gene inactivation and targeted replacement in *Mgal*. A limit of this system was reached when target sites increased. The additional Cre-lox recombination system was used for a second-step removal of the antibiotic resistance marker in recombinant mutants. This study therefore describes the first genetic tool for targeted genome engineering of *Mgal* and demonstrates the efficacy of heterologous recombination systems in minimal bacteria.



INTRODUCTION :

Mycoplasmas are minimal bacteria belonging to the class *Mollicutes*. They share a common ancestor with Gram-positive bacteria and their evolution was marked by a drastic reduction of genome size, but also horizontal gene transfers and exposure to various mobile genetic elements¹⁻³. Many mycoplasmas are pathogenic and responsible for diseases in humans as well as in wild and domesticated animals^{2,4,5}. *Mycoplasma gallisepticum* (*Mgal*) is a major pathogen in poultry with a wide variety of clinical manifestations and a worldwide distribution⁶. In the US, its economic annual impact on the meat and egg productions is estimated around \$780 million and \$150 million, respectively^{7,8}. *Mgal* also emerged in 1994 in a new host with an alarming jump that was traced back from domestic poultry into North American house finches (*Haemorrhous mexicanus*)^{9,10}. Over the years, the resulting and still ongoing epidemics massively impacted the new host population while spreading all over North America^{11,12}. The control of *Mgal* in the poultry industry includes methods of prophylaxis, antibiotics treatments and vaccines. In a general context of increasing antibiotic resistance¹³ and relative lack of efficacy of vaccines^{14,15}, there is a need for a better understanding of *Mgal* virulence basis and for developing improved vaccines¹⁴. The lack of genome engineering tools available for *Mgal* has largely hampered its study. Indeed, as for the vast majority of mycoplasma species, the extremely low efficacy of homologous recombination in *Mgal* prevents its use in experimental genetic approaches. Only few

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

studies confirm the rare occurrence of recombination events¹⁶ similarly to what was observed in other mycoplasmas¹⁷⁻¹⁹. Moreover, genome engineering methods relying on mycoplasma genome cloning and editing in yeast before back genome transplantation into a suitable recipient cell are only available for ruminant mycoplasmas from the mycoides cluster²⁰⁻²², and only a limited set of genetic tools is currently available for *Mgal*, including transposons (integrative)²³⁻²⁵ and *oriC* plasmids (replicative)¹⁶. With these tools, only random mutagenesis and complementation assays are currently possible, highlighting the need for more precise and efficient tools for studying *Mgal*. Phage-encoded ssDNA annealing proteins (SSAPs) constitute the basis for a new generation of powerful tools for genome engineering especially attractive for bacteria lacking an efficient endogenous homologous recombination machinery. Most well-characterized SSAPs have been known since the 1990' including the lambda red system (Gam, Exo et Beta) or the RecET system from the Rac prophage of *Escherichia coli*²⁶. RecT and Red β are phylogenetically related and belong to a large family of DNA recombination proteins²⁷. They have been used as single strand DNA (ssDNA) oligonucleotide recombineering system to modify bacterial genomes but also for double strand DNA (dsDNA) recombineering when combined with an exonuclease (RecE or Exo) that is usually encoded in close proximity in the phage genome²⁸⁻³⁵. Recombination based on phage-derived ssDNA-recombinase is independent of the host machinery but its efficacy depends on the phylogenetic distance between the native host of the phage and the targeted bacteria²⁸. The RecT-like system Gp35 recombinase from *Bacillus subtilis* (*Bsu*) phage SPP1 was used as an oligonucleotide recombineering system in *Bsu* and *E. coli* and more recently in the human pathogen *Mycoplasma pneumoniae*³⁶. This *Bsu* system is also functional for dsDNA recombination in *Bsu* using Gp34.1 as an exonuclease partner to Gp35^{28,37}.

In the present study, we have developed a toolbox for genome engineering in *Mgal*. The main tool is based on the Gp34.1-Gp35 system from the *Bsu* phage SPP1 (called *BsuRecET*-like hereafter) and was successfully used for targeted recombination in *Mgal*. We optimized the system and explore some of its limits. We also demonstrated the functionality of the Cre-lox system in *Mgal* to eliminate the residual antibiotic resistance marker.

RESULTS and DISCUSSION:

Selection of RecET-like systems suitable for genome engineering in *Mgal*

Two phage-derived ssDNA-recombinases were evaluated for *Mgal* genome engineering. The first one has already been described in *M. pneumoniae*³⁶, but using only the ssDNA recombinase (Gp35) from the *BsuRecET*-like system. Here, we used the two components of the *BsuRecET*-like system (Gp35 and Gp34.1) for our first attempts in *Mgal*. In addition, in order to evaluate a system that would be phylogenetically closer to *Mgal*, we extended the search for RecT-like systems in Mollicutes genomes, building upon previously published results³⁶. Using blastp search with previously described RecT protein sequences from four different bacteria^{28,30} used as queries, RecT-like proteins encoding genes were detected in the genome of some *Spiroplasma* species: *Spiroplasma phoeniceum* p40, *Spiroplasma kunkelii* CR2-3x, *Spiroplasma citri* R8A2 and *Spiroplasma poulsonii*. Multiple copies of two different *recT* genes were identified within two families of prophage-like elements of 22 kbp and 32 kbp found in the genome of *S. phoeniceum* p40 (*Spho*) (Figure S1A). Two copies of the 22 kbp and three copies of the 32 kbp element were observed in the *Spho* genome. As no RecE-like encoding gene was predicted within the 22 kbp region, further analyses were restricted to the 32 kbp element. Within this later, in the vicinity of the RecT-like encoding gene (SDAV_001194, Figure S1B), some other genes were predicted to encode an exonuclease (SDAV_001191), a single strand DNA binding protein (SDAV_001192), and a hypothetical protein of unknown function (SDAV_001193). Comparing this putative system with the *BsuRecET*-like system, amino acid identities of only 30% and 15% were calculated among recombinases and exonucleases, respectively, indicating the evolutionary distance

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

between the two systems (Figure 1A). The genes encoding the predicted RecET-like systems in *Bsu* and *Spho* were selected for cloning into a transposon-based vector (Figure 1B).

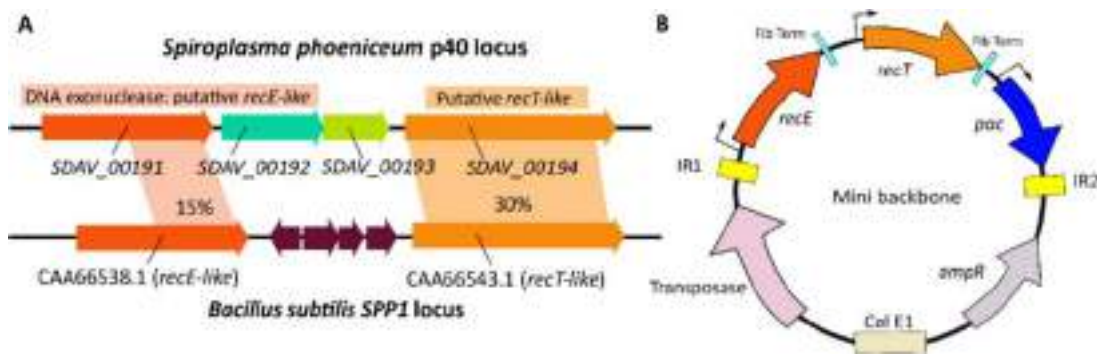


Figure 1. Predicted RecET-like locus in *Spho* genome and plasmid derived for genome engineering of *Mgal*. A- Alignment of the *Spho* and of the *Bsu* phage SPP1 loci encoding RecET-like systems (CAA66538.1; CAA66543.1). Significant local similarity percentages are indicated. B- Schematic of plasmid Mini backbone expressing RecET-like systems. Genes encoding RecT-like and RecE-like proteins from *Bsu* or *Spho* systems were each cloned between pSynMyco/p438 promoters (arrow) and the fibril terminator (Fib Term) from *S. citri*. Expression of the puromycin N-acetyltransferase (*pac*) confers puromycin resistance to *Mgal* transformants. All these components were cloned between two inverted repeats (IR) used by the transposase from Tn4001 transposon for integration into the genome of *Mgal*.

Production of *M. gallisepticum* strains expressing RecET-like systems from *S. phoeniceum* and *B. subtilis*.

Two sets of plasmids encoding the codon-optimized *Spho* and *Bsu* RecET-like systems were constructed (see Materials & Methods). The different plasmids differ in the use of promoters driving the expression of *recE*, *recT* and *puroR* genes. In the first set (Mini_SynMyco_RecET*Bsu* and Mini_SynMyco_RecET*Spho* plasmids), the *recET* and *puroR* genes are under the control of the strong pSynMyco promoter³⁸ and the weak p438 promoter from *Mycoplasma genitalium*³⁹ (unpublished work), respectively (SI-1, SI-2). In the second set of plasmids, the respective promoters for the *recE*, *recT* and *puroR* genes were exchanged compared to the first set (Figure 1B), resulting in the Mini_p438_RecET*Bsu* and Mini_p438_RecET*Spho* plasmids. Following first *Mgal* transformation assays, puromycin-resistant clones were only obtained with plasmids harboring the *recET-like* genes under the weak p438 promoter: 470 and 1,290 transformants were obtained with the plasmids encoding the *Bsu* and *Spho* RecET-like systems, respectively. This result suggested that the strong expression of RecET-like systems was toxic for *Mgal* cell. Alternatively, lack of transformants might result from a low expression level of the antibiotic resistance marker. However, given that the p438 promoter is functional in *Mgal* (unpublished data), this second hypothesis seems to us less plausible. Finally, five transformants were randomly selected for each plasmid and further characterized by single primer PCR and sequencing to determine the insertion sites of the transposon (Figure S2). Among these, two *Mgal* transformants were selected: clone 3 (with *Bsu* system) and clone 6 (with *Spho* system) (Figure S2B), with integration sites distant from mobile elements or repeated sequences to ensure the genomic stability of the mutant clones.

Evaluation of the *Bsu* and *Spho* RecET-like systems for producing small deletions in *Mgal* genome

We first evaluated the two selected *Mgal* clones harboring the *Bsu/Spho* RecET-like expression cassettes for producing targeted small deletions in the genome. The targeted site was located at the 3'-end of gene GCW_RS03665, that is part of a cluster of genes predicted to encode the mycoplasma-specific F1-likeX0 ATPase (Figure 2A). The target was selected as previous results from our laboratory

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

has indicated that this gene is not essential in mycoplasmas⁴⁰. Plasmid recombination templates containing the chloramphenicol acetyl transferase (*cat*) gene flanked by 1 kbp homologous arms from each side of the target site (pTempCAT) were produced (Figure 2A). Four different recombination template forms were evaluated as substrates for RecET-like mediated recombination: linear or circular double stranded DNAs, and linear or circular single stranded DNAs. Wild type and the two RecET-like recombinant *Mgal* strains were transformed using 20 µg of recombination template DNA, and transformants were selected on chloramphenicol plates. No colonies were observed with the control wild type strain and the strain expressing *Spho* system but 18 transformants were obtained when using *Mgal* strain expressing the *Bsu*RecET-like system. All transformants were picked (1, 4, 1, 12 clones using ds linear, ds circular, ss linear and ss circular templates, respectively) and grown in liquid media in the presence of chloramphenicol for further PCR screening of recombination events (Figure 2B). The expected recombination profile was observed in 4 clones resulting from transformation with ss linear (1 clone), ds circular (1 clone) and ss circular (2 clones) templates. Fourteen other transformants presented two-bands profiles typical of the recombinant and the wild type loci. Such profiles could result from a mixed population of mycoplasma cells within the same colony. However, the same profiles were observed after three filter-cloning steps, suggesting single recombination or non-homologous recombination events. From these first experiments, we concluded that only the *Bsu*RecET-like was efficient enough to promote a small deletion (107 bp) and the insertion of a selection marker (660 bp) at this *Mgal* genome locus. This imported recombination system was selected for further experimental studies.

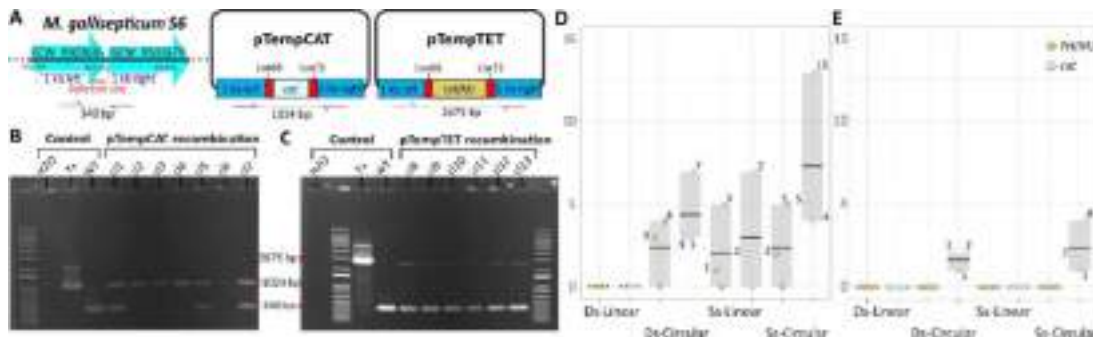


Figure 2. Targeted small deletion in *Mgal* genome using the *Bsu*RecET-like system. A- Schematic of the targeted region in *Mgal* genome and the pTempCAT and pTempTET suicide plasmids harboring recombination cassettes containing chloramphenicol acetyl transferase (CAT) or tetracycline resistance markers (*tet*(M)) flanked by 1 kbp homologous arms for recombination. Primers G55 and G56 were used to screen recombination events by PCR amplification with expected products that differed in size: WT DNA (340 bp) or recombinant DNA (1,024 or 2,675 bp). B- Example of colony screening using this scheme. Clones 2, 3, 4 and 6 showed recombinant profiles while mixed profiles were observed for other clones. C- Example of colony screening using Tet template. Only mixed profiles were observed. D and E- Results of three assays of recombination on locus GCW_RS03665 using CAT or Tet templates in blue and yellow, respectively. Four template forms (Double stranded (ds)-linear, ds-circular, single stranded (ss)-linear and ss-circular) were evaluated. Left graph (D) corresponds to all transformants obtained and the right one (E) corresponds to the number of isolated clones with a pure recombinant profile.

In order to compare the effectiveness of the *Bsu*RecET-like system using the four conformations of the template DNA, the experiment was repeated three times. In the same experiment, we also compared two resistance markers: *cat* (pTempCAT plasmid) and *tet*(M) (pTempTET plasmid) (Figure 2). Only 0 to 7 transformants were obtained for most recombination template forms, with the exception of the single stranded circular pTempCAT plasmid that led to 13 transformants in one assay (Figure 2D). After screening of recombinants, the expected profile was observed with both ds and ss circular pTempCAT as a template but not with pTempTET (Figure 2C). In addition to the impact of the form of DNA

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

template used for recombination, this experiment also shows the importance of choosing the most efficient selection marker (*cat* in this case). Finally, whole genome sequencing of a recombinant clone indicated that there was no major change in the genome structure, with only 4 small indels and no SNP variants (Table S3). We then concluded that the *Bsu*RecET-like system is functional in *Mgal* and can be used to generate precise deletions.

Determination of the optimal amount of template DNA for gene replacement in *Mgal* using the *Bsu*RecET-like system

Next, we optimized the amount of template DNA to be used for efficient targeted recombination. We selected another target and designed a recombination template for the precise replacement of a protein domain. More specifically, we chose to replace the PAM-Interacting (PI) domain of the Cas9 protein encoded in the CRISPR locus of the *Mgal* S6 strain by the PI domain of the Cas9 from *Mgal* strain CA06 (Figure 3A). As previously, a new recombination template pTempPI including the *cat* marker was built and transformation of *Mgal* cells was conducted using 20 µg of ss circular template. Out of 30 transformants, 20 clones presented a typical recombinant profile characterized by PCR products of 2,023 bp instead of 1,167 bp in the wild type (Figure 3B). One clone was verified by Sanger sequencing of the PCR product, confirming the precise exchange of PI domain in *cas9* gene. In order to determine the transformation efficiency, the assay was repeated three times with variable amounts of recombination template (5, 10, 20 and 50 µg). Analyses of 2 to 10 transformants per replicate per condition showed a total of 91 single-band profile recombinants out of 102 analyzed transformants. These results confirmed the efficacy of the *Bsu*RecET-like system for targeted genetic exchanges in *Mgal* and indicated better results with increasing amounts of template DNA (Figure 3C). We concluded that the *Bsu*RecET-like system is more efficient when using large amounts (20 to 50 µg) of recombination template.

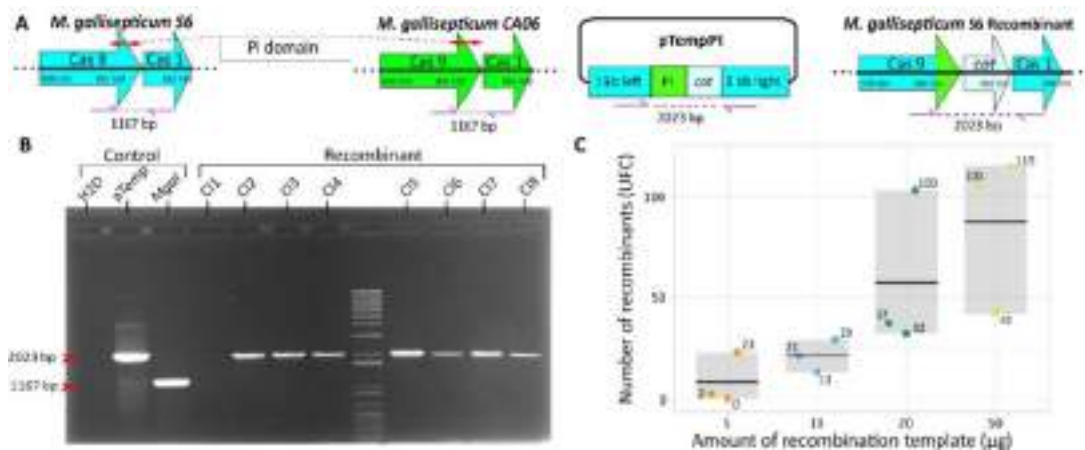


Figure 3. Accurate replacement of a protein domain and impact of recombination template amount. A- Schematic representation of the targeted gene region. The PI domain in Cas9-encoding gene is represented by a red arrow. The pTempPI is the template DNA that allows the PI domain replacement of S6 sequence by CA06 sequence and is composed by a 516 bp fragment encoding the PI domain of CA06 Cas9 and the CAT marker (755 bp) flanked by 1 kbp homologous arms from S6 strain genome. Primers H57 and H58 are represented in purple and allows to differentiate recombinants from wt strains by amplifying DNA fragments of 2,023 bp or 1,167 bp for strains S6 and CA06, respectively. B- Screening of recombinant clones using PCR amplification of the targeted locus and gel electrophoresis. In this example, all clones tested showed a recombinant profile. C- Evaluation of the effect of different amounts of recombination template on the efficiency of recombination in *Mgal*. The recombinant template was used as ss circular DNA, with increasing amounts of 5, 10, 20 and 50 µg. The experiment was conducted in triplicates for each template amount.

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

The RecET system efficacy decreases when targeting a larger genome region

We evaluated the limit of the *Bsu*RecET-like mediated recombination by trying to delete a larger fragment of the *Mgal* genome. The deletion of the entire F1-likeX0 ATPase locus (6.6 kbp), which contain five genes including the previously targeted GCW_RS03665 gene was chosen for this assay (Figure 4A-4B). A new recombination templates, pTempCAT-F1X0 was designed and built. After three transformation assays using 50 µg ss circular template, only three transformants were obtained (Figure 4C) and all presented mixed profiles as previously observed. The combination of several passages through a needle and filter cloning failed to isolate clones with pure recombinant profiles. The genome sequencing of a transformant indicated that a recombination event occurred at the targeted site but with a complete plasmid integration (Figure 4D). This result showed that the efficacy of the *Bsu*RecET-like system decreases when the size of the targeted genome region increases.

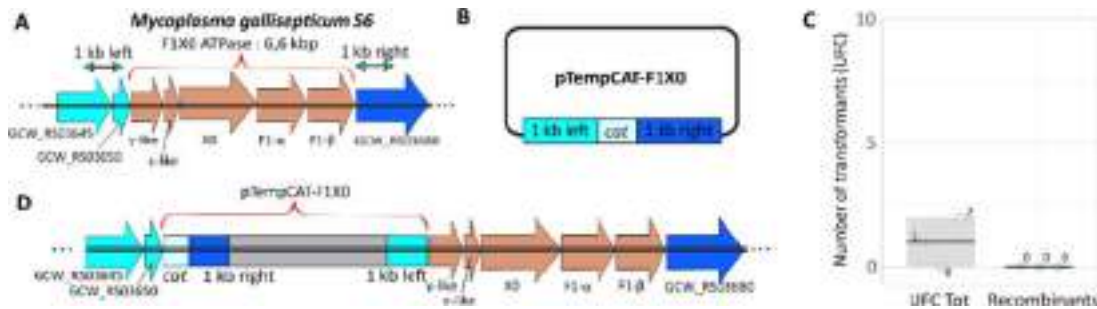


Figure 4. Attempt of targeted deletion of the F1-likeX0 ATPase locus (6.6 kbp). A- Schematic representation of the locus F1-likeX0 ATPase targeted for deletion. B- Recombination templates used for deletion containing the *cat* resistance marker flanked by 1 kbp homologous arms. C- Results of three transformation assays. Few colonies were obtained and showed mixed PCR profiles. D- Genome organization at the targeted locus in a clone with a mixed profile.

Cre recombinase is effective to remove resistance markers in Mgal genome.

In order to open the possibility of iterative genome editing steps, the Cre recombinase was evaluated to remove antibiotic resistance markers. During the design of the previously described 107 bp deletion in the GCW_RS03665 gene, two Lox sites (66 and 71) were included in the recombination template on each site of the marker. The Cre recombinase is expected to promote the excision of the antibiotic marker, leaving only a 34 bp scar consisting in the hybrid Lox66/71 site (Figures 2A and 5A). The suicide plasmid pCre_genta^{41,42} encoding the Cre recombinase was transformed into the previously obtained *Mgal* mutant clone 3 (Figure 2B) and cells were incubated in gentamycin containing liquid medium for 96 hours before plating on non-selective medium. Among thousands of colonies, 10 were randomly picked and grown in non-selective medium. One clone (clone 3.1) out of 10 tested showed a PCR profile indicating the loss of the antibiotic resistance marker (Figure 5B). Sanger sequencing of the locus indicated that the Lox71/66 site was the only scar left after this additional step of genome engineering (Figure 5B and 5C). We concluded that the Cre recombinase is effective in *Mgal* and can be used to eliminate a antibiotic resistance markers in recombinant strains.

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

MATERIAL AND METHODS

Bacterial strains and culture

M. gallisepticum S6 (Tax ID: 1006581) was cultivated at 37°C in Hayflick⁴⁵ modified medium, under a 5% CO₂ atmosphere. Tetracycline 5 µg.mL⁻¹, gentamycin 100 µg.mL⁻¹, chloramphenicol 15 µg.mL⁻¹ and puromycin 10 µg.mL⁻¹ were used in selective conditions. *E. coli* 5-alpha (NEB, C2987H) used for plasmid amplification were cultivated in LB Broth (ThermoFisher: 12795027). Ampicillin 100 µg.mL⁻¹ or kanamycin 50 µg.mL⁻¹ were added in selective conditions. Detailed protocols for *Mgal* transformation, transformants screening and characterization are available as SI.

DNA fragments design and synthesis

DNA sequence of *Bsu*RecE (*gp34.1*) (*recT*) and *Bsu*RecT (*gp35*) from *Bacillus* phage SPP1, SDAV_001191 (*Spho*RecE) and SDAV_001194 (*Spho*RecT) from *S. phoeniceum* putative system were codon optimized with the Optimizer online tool (<http://genomes.urv.es/OPTIMIZER/>) using the codon usage of *Mgal* from Kazusa database (<https://www.kazusa.or.jp/codon/>). Promoter pSynMyco, a synthetic promoter for mycoplasma³⁸ and terminator Fibril from *S. citri* were synthesized with the codon optimized genes. Manual modifications to avoid regions with a high AT content were performed to allow the synthesis of the DNA fragments by Twist Bioscience.

Oligonucleotides and plasmids

All the oligonucleotides used for this study were supplied by Eurogentec and are described in Table S1. All the plasmids constructed and used in this study are listed in Table S2. Detailed protocols for plasmid constructions, DNA sequences of *recE*, *recT* genes and plasmids are provided as Supplementary Information (SI-2).

Acknowledgements

Genome sequencing was performed by the Genome Transcriptome Facility of Bordeaux (<https://pgtb.cgfb.u-bordeaux.fr>, Grants from Investissements d'avenir, Convention attributive d'aide EquipEx Xyloforest ANR-10-EQPX-16-01). Plasmid pet28_MgalCasCA06 (pet28_ID116_His) was kindly provided by Dr Gasiunas from CasZyme.

References

1. Sirand-Pugnet P, Citti C, Barré A, Blanchard A. Evolution of mollicutes: down a bumpy road with twists and turns. *Res Microbiol.* 2007;158(10):754-766. doi:10.1016/j.resmic.2007.09.007
2. Razin S, Yorgev D, Naot Y. Molecular biology and pathogenicity of Mycoplasmas. *Microbiol Mol Biol Rev.* 1998;165(2):199-221.
3. Citti C, Dordet-Frisoni E, Nouvel LX, Kuo CH, Baranowski E. Horizontal gene transfers in mycoplasmas (Mollicutes). *Curr Issues Mol Biol.* 2018;29:3-22. doi:10.21775/cimb.029.003
4. Citti C, Blanchard A. Mycoplasmas and their host: Emerging and re-emerging minimal pathogens. *Trends Microbiol.* 2013;21(4):196-203. doi:10.1016/j.tim.2013.01.003
5. Rosengarten R, Citti C, Much P, Spersger J, Drosses M, Hewicker-Trautwein M. The changing image of mycoplasmas: from innocent bystanders to emerging and reemerging pathogens in human and animal diseases. *Contrib Microbiol.* 2001;8:166-185. doi:10.1159/000060409
6. Levisohn S, Kleven SH. Avian mycoplasmosis (*Mycoplasma gallisepticum*). *Sci Tech Rev Off Int des Epizoot.* 2000;19(2):425-442.
7. Hennigan SL, Driskell JD, Ferguson-Noel N, et al. Detection and differentiation of avian mycoplasmas by surface-enhanced raman spectroscopy based on a silver nanorod array. *Appl Environ Microbiol.* 2012;78(6):1930-1935. doi:10.1128/AEM.07419-11
8. Peebles ED, Branton SL. *Mycoplasma gallisepticum* in the commercial egg-laying hen: A historical perspective considering the effects of pathogen strain, age of the bird at inoculation, and diet on performance and physiology. *J Appl Poult Res.* 2012;21(4):897-914. doi:10.3382/japr.2012-00555
9. Delaney NF, Balenger S, Bonneaud C, et al. Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS Genet.* 2012;8(2). doi:10.1371/journal.pgen.1002511
10. Fischer JR, Stallknecht DE, Luttrell P, Dhondt AA, Converse KA. Mycoplasmal Conjunctivitis in Wild Songbirds : The Spread of a New Contagious Disease in a Mobile Host Population. *Emerg Infect Dis.* 1997;3(1):69-72. doi:10.3201/eid0301.970110
11. Dhondt AA, Altizer S, Cooch EG, et al. Dynamics of a novel pathogen in an avian host: Mycoplasmal conjunctivitis in house finches. *Acta Trop.* 2005;94(1):77-93. doi:10.1016/j.actatropica.2005.01.009

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

12. Tardy L, Giraudeau M, Hill GE, McGraw KJ, Bonneaud C. Contrasting evolution of virulence and replication rate in an emerging bacterial pathogen. *Proc Natl Acad Sci U S A*. 2019;116(34):16927-16932. doi:10.1073/pnas.1901556116
13. Gautier-Bouchardon A V. Antimicrobial Resistance in *Mycoplasma* spp. *Microbiol Spectr*. 2018;6(4). doi:10.1128/microbiolspec.arba-0030-2018
14. Ishfaq M, Hu W, Khan MZ, Ahmad I, Guo W, Li J. Current status of vaccine research, development, and challenges of vaccines for *Mycoplasma gallisepticum*. *Poult Sci*. 2020;99(9):4195-4202. doi:10.1016/j.psj.2020.06.014
15. Wijesurendra DS, Kanci A, Tivendale KA, et al. Development of a *Mycoplasma gallisepticum* infection model in turkeys. *Avian Pathol*. 2015;44(1):35-42. doi:10.1080/03079457.2014.992390
16. Lee SW, Browning GF, Markham PF. Development of a replicable oriC plasmid for *Mycoplasma gallisepticum* and *Mycoplasma imitans*, and gene disruption through homologous recombination in *M. gallisepticum*. *Microbiology*. Published online 2008. doi:10.1099/mic.0.2008/019208-0
17. Janis C, Lartigue C, Frey J, et al. Versatile Use of oriC Plasmids for Functional Genomics of *Mycoplasma capricolum* subsp. *capricolum*[†]. *Society*. 2005;71(6):2888-2893. doi:10.1128/AEM.71.6.2888
18. Clampitt JM, Madsen ML, Minion FC. Construction of *Mycoplasma hyopneumoniae* P97 Null Mutants. *Front Microbiol*. 2021;12(April):1-12. doi:10.3389/fmicb.2021.518791
19. Halbedel S, Stülke J. Tools for the genetic analysis of *Mycoplasma*. *Int J Med Microbiol*. 2007;297(1):37-44. doi:10.1016/j.ijmm.2006.11.001
20. Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (80-)*. 2010;329(5987):52-56. doi:10.1126/science.1190719
21. Labrousseau F, Lebaudy A, Baby V, et al. Impact of donor-recipient phylogenetic distance on bacterial genome transplantation. *Nucleic Acids Res*. 2016;44(17):8501-8511. doi:10.1093/nar/gkw688
22. Lartigue C, Glass JI, Alperovich N, et al. Genome transplantation in bacteria: Changing one species to another. *Science (80-)*. 2007;317(5838):632-638. doi:10.1126/science.1144622
23. Chen HJ, Zhao CM, Shen XY, Chen DQ, Yu SQ, Ding C. Construction of mini-Tn4001 transposon vector for *Mycoplasma gallisepticum*. *Sci China Life Sci*. 2010;53(11):1340-1345. doi:10.1007/s11427-010-4082-5
24. Whetzel PL, Hnatoow LL, Keeler CL, Dohms JE. Transposon mutagenesis of *Mycoplasma gallisepticum*. *Plasmid*. 2003;49(1):34-43. doi:10.1016/S0147-619X(02)00114-2
25. Ruffin DC, Van Santen VL, Zhang Y, Voelker LRL, Panangala VS, Dybvig K. Transposon mutagenesis of *Mycoplasma gallisepticum* by conjugation with *Enterococcus faecalis* and determination of insertion site by direct genomic sequencing. *Plasmid*. 2000;44(2):191-195. doi:10.1006/plas.2000.1485
26. Hall SD, Kolodner RD. Homologous pairing and strand exchange promoted by the *Escherichia coli* RecT protein. *Biochemistry*. 1994;91:3205-3209. doi:10.1073/pnas.91.8.3205
27. Iyer LM, Koonin E V., Aravind L. Classification and evolutionary history of the single-strand annealing proteins, RecT, Red β , ERF and RAD52. *BMC Genomics*. 2002;3:1-11. doi:10.1186/1471-2164-3-8
28. Sun Z, Deng A, Hu T, et al. A high-efficiency recombinering system with PCR-based ssDNA in *Bacillus subtilis* mediated by the native phage recombinase GP35. *Appl Microbiol Biotechnol*. 2015;99(12):5151-5162. doi:10.1007/s00253-015-6485-5
29. Yang P, Wang J, Qi Q. Prophage recombinases-mediated genome engineering in *Lactobacillus plantarum*. *Microb Cell Fact*. 2015;14(1):1-11. doi:10.1186/s12934-015-0344-z
30. Huang H, Song X, Yang S. Development of a RecE/T-Assisted CRISPR-Cas9 Toolbox for *Lactobacillus*. *Biotechnol J*. 2019;14(7):1-12. doi:10.1002/biot.201800690
31. Corts AD, Thomason LC, Gill RT, Gralnick JA. Efficient and Precise Genome Editing in *Shewanella* with Recombineering and CRISPR/Cas9-Mediated Counter-Selection. *ACS Synth Biol*. 2019;8(8):1877-1889. doi:10.1021/acssynbio.9b00188
32. Wu ZY, Huang YT, Chao WC, Ho SP, Cheng JF, Liu PY. Reversal of carbapenem-resistance in *Shewanella algae* by CRISPR/Cas9 genome editing. *J Adv Res*. 2019;18:61-69. doi:10.1016/j.jare.2019.01.011
33. Swingle B, Bao Z, Markel E, Chambers A, Cartinhour S. Recombineering using recTE from *Pseudomonas syringae*. *Appl Environ Microbiol*. 2010;76(15):4960-4968. doi:10.1128/AEM.00911-10
34. Yin J, Zhu H, Xia L, et al. A new recombinering system for *Photobacterium* and *Xenorhabdus*. *Nucleic Acids Res*. 2015;43(6):e36. doi:10.1093/nar/gku1336
35. Huang Y, Li L, Xie S, et al. Recombineering using RecET in *Corynebacterium glutamicum* ATCC14067 via a self-excisable cassette. *Sci Rep*. 2017;7(1):1-8. doi:10.1038/s41598-017-08352-9
36. Piñero-Lambeck C, Garcia-Ramallo E, Martínez S, Delgado J, Serrano L, Lluch-Senar M. *Mycoplasma pneumoniae* Genome Editing Based on Oligo Recombineering and Cas9-Mediated Counterselection. *ACS Synth Biol*. 2020;9(7):1693-1704. doi:10.1021/acssynbio.0c00022
37. Datta S, Costantino N, Zhou X, Court DL. Identification and analysis of recombinering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci U S A*. 2008;105(5):1626-1631. doi:10.1073/pnas.0709089105
38. Montero-Blay A, Miravet-Verde S, Lluch-Senar M, Piñero-Lambeck C, Serrano L. SynMyco transposon: engineering transposon vectors for efficient transformation of minimal genomes. *DNA Res*. 2019;26(4):327-339. doi:10.1093/dnares/dsz012
39. Pich OQ, Burgos R, Planell R, Querol E, Piñol J. Comparative analysis of antibiotic resistance gene markers in *Mycoplasma genitalium*: Application to studies of the minimal gene complement. *Microbiology*. 2006;152(2):519-527. doi:10.1099/mic.0.28287-0
40. Béven L, Charenton C, Dautant A, et al. Specific evolution of F 1-like ATPases in mycoplasmas. *PLoS One*. 2012;7(6).

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

doi:10.1371/journal.pone.0038793

41. Sternberg N, Hamilton D. Bacteriophage P1 site-specific recombination: I. Recombination between loxP sites. *J Mol Biol.* 1981;150(4):467-486. doi:10.1016/0022-2836(81)90375-2
42. Mariscal AM, González-González L, Querol E, Piñol J. All-in-one construct for genome engineering using Cre-lox technology. *DNA Res.* 2016;23(3):263-270. doi:10.1093/dnares/dsw015
43. Filsinger GT, Wannier TM, Pedersen FB, et al. Characterizing the portability of phage-encoded homologous recombination proteins. *Nat Chem Biol.* 2021;17(4):394-402. doi:10.1038/s41589-020-00710-5
44. Garcia-Morales L, Ruiz E, Gourgues G, et al. A RAGE Based Strategy for the Genome Engineering of the Human Respiratory Pathogen *Mycoplasma pneumoniae*. *ACS Synth Biol.* 2020;9(10):2737-2748. doi:10.1021/acssynbio.0c00263
45. Freundt EA. CULTURE MEDIA FOR CLASSIC MYCOPLASMAS. In: *Methods in Mycoplasmaology.* ; 1983. doi:10.1016/b978-0-12-583801-6.50029-9

3. Données supplémentaires de la publication

Supplementary Figures

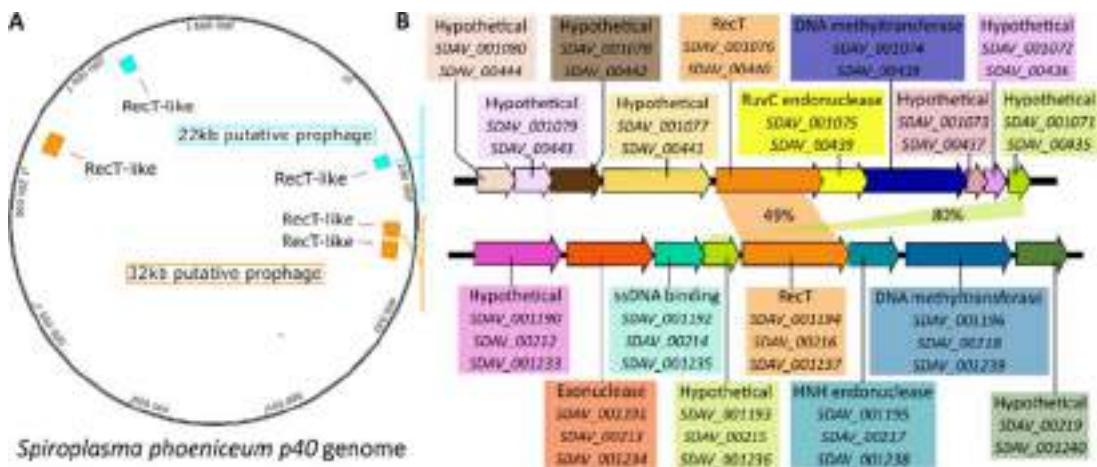


Figure S1. Putative prophage loci in *S. phoeniceum* p40 genome. A- Genome localization of putative prophages, as characterized by PHASTER analyses, in the *Spho* genome. Orange and blue boxes indicate the localization of the 32 kbp and 22 kbp predicted prophages. Genes encoding the RecT-like proteins are represented in red bar or purple bar. B- Comparison of loci around *recT*-like encoding genes predicted in 22 kbp (top) and 32 kbp (bottom) putative prophages families in *Spho* genome. Homologous genes between the two families are connected and protein similarity percentages are indicated. Predicted functions are indicated above the mnemonic list of *Spho* genes from the multiple prophage copies.

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome



Figure S2. Characterization of transposon integration site in *Mgal* genome. A- Example of PCR profiles obtained after single primer PCR on 5 clones of *Mgal* B- Transposon integration sites determined after Sanger sequencing and BLASTn against *Mgal* genome.

Supplementary Information

SI-1. Material and Methods

Identification of RecET-like systems in mollicutes genomes

mollicutes orthologues of RecT-like proteins from the *E. coli*-derived Rac system (NP_415865), the *Bacillus* phage SPP1 (CAA66543.1) and *Lactobacillus* (WP_015474118 and WP_011101092) were searched using BLASTp against Tenericutes database (taxid: 544448). The InterPro (<http://www.ebi.ac.uk/interpro/>) database was also used to identify functional domains. Hits were found in *S. phoeniceum* p40 (CP031088), *S. poulsonii*, *S. kunkelii* CR2-3x (CP010899.1), *S. citri* R8A2 (CP013197.1) and *S. melliferum* AS576 (CP029202). Comparison between proteins were made using Blastp and protein alignment in MEGA-X (Kumar *et al.*, 2018). Phages were further predicted in the *Spiroplasma* genome using PHASTER (<http://phaster.ca/>) (Arndt *et al.*, 2016) and a final prediction of putative phages and RecT-like proteins was achieved manually.

Plasmid construction

Construction of Mini_SynMyco_RecETBsu and Mini_SynMyco_RecETSpho. The Mini transposon vector derived from Tn4001 (Clyde A. Hutchison *et al.*, 1999) and the synthesized fragments provided by Twist biosciences (SI-2) were used as PCR template. Primers A1/A2, B4/B5 and B6/B7 and the Q5 High-Fidelity DNA Polymerase kit (NEB, M0491) were respectively used to amplify the plasmid backbone with puromycin resistance and the codon optimized sequences of *recT* and *recE* genes. PCR products were submitted to restriction with the DpnI enzyme (NEB, R0176S) following the manufacturer's protocol. Both products were purified using GFX™ PCR DNA or Gel Band Purification Kit (GE-healthcare). The NEBuilder HiFi DNA Assembly Cloning Kit (NEB, E5520S) was used to assemble the fragments into the Mini_SynMyco_RecETBsu plasmid. Primers A3/A2, B8/B9 and B10/B11 and the procedure used for the assembly of the Mini_SynMyco_RecETSpho plasmid are described in Table S1. Two μ L of the assembly was

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

transformed into *E. coli* NEB 5-alpha (NEB, C2987H). Transformants were screened after DNA minipreps using NucleoSpin Plasmid kit (Macherey-nagel, 740588.50) and enzymatic digestion. Sanger sequencing was performed for final verification.

Construction of Mini_p438_RecETBsu and Mini_p438_RecETSpho. In order to change the promoters driving the expression of the Rec proteins, PCR amplifications of the *recT* genes were performed using the Mini_SynMyco_RecETBsu and Mini_SynMyco_RecETSpho plasmids as templates and primer C14 or C18 integrating the p438 promoter in pair with primer C15 or C19. PCR amplifications of the *recE* genes were performed using the Mini_SynMyco_RecETBsu and Mini_SynMyco_RecETSpho plasmids as templates and primer C16 or C20 integrating the p438 promoter in pair with primer C17 or C21. The backbone was PCR amplified using the primers C12/C13 and the Mini-SynPuro plasmid as template. Cloning was performed using the aforementioned procedure to generate the plasmids Mini_p438_RecETBsu and Mini_p438_RecETSpho.

Construction of pTempCAT, pTempTET, pTempPI, pTempCAT-F1X0 and pTempTET-F1X0. Plasmids harboring the recombination templates were built using the pBlueScript II K- (Agilent, 212208) as backbone. The pBSK_loxCATlox or pBSK_loxTetlox plasmid and the *Mgal* gDNA were used as PCR templates. Primers E32/E33, E26/E27, E30/E31 and E28/E29, were respectively used to amplify the backbone vector, the 1 kbp left recombination arms, the 1 kbp right recombination arms and the antibiotic resistance marker. Cloning was performed using the same procedure as for Mini_SynMyco_RecETBsu (see above) and resulted in the pTempCAT, pTempTET, pTempPI, pTempCAT-F1X0 and pTempTET-F1X0 recombinant plasmids.

For recombination template production, primers E26/E31, were used to amplify linear template using Q5® High-Fidelity DNA Polymerase (M0491). PCR products were purified using GFX™ PCR DNA and Gel Band Purification Kit (GE-healthcare) and eluted in water. To prepare single strand templates, DNAs resuspended in water were heated at 95°C for 15 min, and then placed immediately on ice before transformation.

***Mycoplasma gallisepticum* transformation**

Mgal cells were grown for 36 h before transformation. At pH 6.2 to 6.5, aliquots of 10 mL were centrifuged during 15 min at 6,000 x *g*, 10°C. Cells were then resuspended with 5 mL of HBSS 1X wash buffer (ThermoFisher, 14065056) and centrifuged 15 min at 6000 x *g*, 10°C. The pellet was resuspended in 250 µL of CaCl₂ 0.1 M and incubated for 30 min on ice. Cold CaCl₂-incubated cells were gently mixed with plasmid DNA (10 µg) and 10 µg of yeast tRNA (ThermoFisher, AM7119). Then, 2 mL of 40 % polyethylene glycol 6000 (PEG) (Sigma, 11130) dissolved in HBSS 1X buffer were added to the cells. After 2 min of incubation at room temperature, contact with the PEG was

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

stopped by addition of 20 ml of HBSS 1X wash buffer. Mixture was centrifuged 15 min at 6,000 x *g*, 10°C and the cells were resuspended in 1 mL of modified Hayflick medium pre-warmed at 37°C. After 2 hours of incubation at 37°C, cells were plated on Hayflick selective plates. For recombination assays, the incubation at 37°C was extended for two additional hours. During the Cre-Lox experiment, the suicide plasmid pCre-genta containing the Cre recombinase encoding gene under the control of p438 promoter and a gentamycin resistance marker was used. After transformation and incubation at 37°C during 2h, gentamycin was added along with 3 mL of fresh Hayflick medium. After a 96h culture, cells were collected by centrifugation, resuspended in 500 µL of Hayflick medium and spread on Hayflick plates without antibiotic. After 10-15 days at 37°C with 5 % CO₂, colonies were counted, picked and re-suspended in 1 mL of Hayflick for three passages (one passage ~48h).

Localization of transposon by single primer PCR

The genome sites of inserted transposons were determined as described by Rideau et al (Rideau *et al.*, 2019). Briefly, DNA was extracted from 200 µL of transformant culture using genomic DNA extraction kit (Macherey-Nagel, 740952.50). Using primer D21, D22 or D23 and Taq DNA polymerase (NEB), single primer PCR reactions were performed (Buffer 1X, MgCl₂ 25 mM, primer D21, D22 or D23 100 µM, dNTPs 10 mM, Taq 0.5 U, gDNA 2.5 µL, H₂O for a total volume of 25 µL). The PCR program includes 20 cycles (95°C for 30 s, 60°C for 30 s and 68°C for 3 min), followed by 30 cycles (95°C for 30 s, 30°C for 30s and 68°C for 2 min), 30 cycles (95°C for 30 s, 60°C for 30 s, 68°C for 2 min) and final amplification at 68°C for 7 min. Amplification products were analyzed by electrophoresis on 1 % agarose gel and purified using GFX™ PCR DNA and Gel Band Purification Kit (GE-Healthcare). After Sanger sequencing (GENEWIZ), the localization of transposons was determined using BLASTn.

PCR screening of *Mgal* transformants

PCR screening was performed using the Advantage HF 2 PCR Kit (Takara, 639123). Primers used to screen recombinants on the different targets are listed in Table S1.

Whole genome sequencing of recombinant DNA

Genomic DNA of *Mgal* was extracted from a 10 mL culture using the Qiagen Genomic-Tips 100/G kit. Genome sequencing was performed by the Genome Transcriptome Facility of Bordeaux. Long reads and short reads were produced using a GridION (Oxford Nanopore) and a MiSeq (Illumina), respectively. For *Mgal* recombinant clone 3, ONT sequencing generated 31,164 reads and Illumina 1,637, 034 read pairs. Reads were submitted to SRA (accession number PRJNA769398). Analyses were performed on Galaxy (<https://usegalaxy.eu/>). Mutations were detected after mapping on the *Mgal* S6 genome (CP006916.3). Illumina reads were trimmed using Trimmomatic

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

(V 0.38.1; Sliding Window 10, 20 ; Drop read below Minimal length 250), mapped using BWA-MEM (V 0.7.17.1), Samtools sort (V 2.0.3), MPileup (V 2.1.1), and variants were detected using VarScan mpileup (V 2.4.3.1; Minimum coverage 30, Minimum supporting read 20, Minimum Base quality 30, Minimum variant allele frequency 0,8, Minimum homozygous variants 0,75). Mutations are showed in Table S3. Genome assembly was performed using the following steps: Nanopore reads were filtered using Filter FASTQ (V 1.1.5, Minimum size 45000 bp), assembled using Flye Assembly (V 2.6), polished using 4 rounds of Pilon (1.20.1) combined with Illumina short reads. Assembled genome was compared to the *Mgal* S6 (CP006916.3) reference genome using MAUVE software (Darling *et al.*, 2004). For the *Mgal* (Δ ATPase) with a mixed PCR profile, the same procedure was used with 24,443 ONT long-reads and 1,776,420 Illumina short reads pairs.

4. Données et discussions supplémentaires non publiées

Notre étude a montré tout l'intérêt du système exogène RecET pour la mutagenèse dirigée de *Mgal*. Il va permettre d'étudier les relations génotype-phénotype de manière ciblée chez ce pathogène majeur, comme nous avons pu le faire, par exemple, dans le chapitre 2 avec le remplacement du domaine PI de MgCas9 grâce à ce système. Il pourrait également être utilisé dans d'autres espèces de mycoplasmes, mais l'efficacité étant pour le moment limitée, il peut être difficile de le mettre en place. Un système endogène, tel que celui identifié chez les spiroplasmales cités dans l'étude, pourrait améliorer cette efficacité, mais une étude plus approfondie serait nécessaire pour aboutir à quelque chose d'intéressant. En effet, les essais préliminaires que nous avons réalisés avec le système endogène n'ont pas donné de résultats convaincants. Tout comme le système SPP, le système de spiroplasmale a été placé sous le contrôle du promoteur pSynMyco et décliné sous forme de différentes constructions incluant deux, quatre ou 5 des gènes présents dans le locus naturel. Malheureusement, en utilisant ce promoteur aucune colonie exprimant les constructions n'a pu être obtenue. Lors d'essai d'expression transitoire de ces mêmes systèmes par des plasmides suicide, aucun recombinant n'a pu être obtenu. Le temps consacré à ce projet étant limité, nous nous sommes finalement concentrés sur le système phagique SPP1 de *Bacillus subtilis* qui avait été décrit chez *M. pneumoniae* comme fonctionnel pour de la recombinaison d'oligonucléotide (Piñero-Lambea *et al.*, 2020) et le système minimal endogène de spiroplasmale avec les gènes putatifs RecE et RecT. Mais l'utilisation des autres gènes composant le locus de spiroplasmale pourrait permettre une nette augmentation de l'efficacité de ce système. L'un d'eux pourrait avoir un rôle dans la

CHAPITRE 3 : Utilisation d'un système de recombinaison exogène chez *Mycoplasma gallisepticum* comme outil d'ingénierie de génome

mécanistique du système RecET-like retrouvé chez les spiroplasmales. Des essais complémentaires chez *Mgal* ou directement chez des espèces de spiroplasmales pourraient être intéressants.

Des essais de co-expression du système CRISPR endogène de *Mgal* ont également été entrepris pour augmenter l'efficacité de recombinaison, mais aucun résultat intéressant n'a pu être obtenu. Ce système pourrait en revanche être utilisé en contre-sélection, pour éliminer les recombinants issus de recombinaisons impliquant un seul crossing-over. Il serait alors nécessaire d'avoir un système inductible sans fuite majeure, à contrario du promoteur *pxyl-tetO2* original utilisé dans notre étude. La version améliorée de ce promoteur, construite par l'équipe espagnole de Luis Serrano (Mariscal *et al.*, 2016), n'a pas encore été testée, mais elle pourrait être une solution intéressante.

Cet outil, bien que très utile pour l'étude de *Mgal*, ne répond pas encore aux critères d'un outil universel précis et facile d'utilisation pour de nombreux mycoplasmas. Un dernier axe de recherche a donc été suivi.

**CHAPITRE 4 : Construction d'un outil universel
pour des bactéries à faible efficacité de
recombinaison : les systèmes Base-Editor**

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

Comme expliqué dans l'étude bibliographie en introduction de cette thèse, les systèmes CRISPR ont abouti à de nombreux dérivés et certains d'entre eux ne sont plus dépendant de la réparation de l'ADN par la cellule. En 2016, deux laboratoires ont contribué à une avancée très importante dans ces développements, avec la création des systèmes de type Base-Editors (Komor *et al.*, 2016; Nishida *et al.*, 2016). Ces systèmes permettent la modification de l'ADN sur des bases spécifiques sans impliquer de cassure de l'ADN. Cet outil ayant été utilisé dans de nombreux modèles eucaryotes, mais aussi procaryotes, mon travail a consisté à l'identification et l'adaptation d'un système d'édition de base qui pourrait fonctionner chez les mycoplasmes.

1. Principaux résultats

Au cours de cette dernière partie de ma thèse, j'ai développé un système d'édition de base utilisable pour des espèces variées de mycoplasmes. Ce système est basé sur une protéine hybride fusionnant une Cas9 inactivée pour le clivage de l'ADN avec une cytosine désaminase. Un sgRNA guide l'enzyme vers sa cible et assure ainsi la spécificité d'action de la désaminase. La fonctionnalité du système a été démontrée par inactivation ciblée de gènes de virulence chez trois mycoplasmes pathogènes animaux pour lesquels aucun outil de mutagénèse ciblée n'était disponible : *M. gallisepticum*, *M. mycoides* subsp. *mycoides* et *M. bovis*. Un exemple particulièrement frappant a été l'inactivation de 54 IS1634 de *M. mycoides* subsp. *mycoides* avec un seul sgRNA.

2. Article en finalisation pour soumission à PNAS – mBio

Unblocking genome editing of major animal mycoplasmas using CRISPR-Cas9 Base Editor systems

Authors

Thomas IPOUTCHA¹, Fabien RIDEAU¹, Geraldine GOURGUES¹, Yonathan ARFI¹, Carole LARTIGUE¹, Alain BLANCHARD¹, Pascal SIRAND-PUGNET¹ *

¹ Univ. Bordeaux, INRAE, UMR BFP, F-33882, Villenave d'Ornon, France

*Corresponding author: Pascal Sirand-Pugnet, pascal.sirand-pugnet@inrae.fr

Abstract

Mycoplasmas are minimal bacteria that infect humans, wildlife, and all economically important livestock species. Mycoplasma infections cause a large range of chronic inflammatory diseases, eventually leading to death in some animals. Due to the lack of efficient recombination and genome engineering tools, the production of mutant strains for the identification of virulence factors and the development of improved vaccine strains remains a bottleneck for many mycoplasma species. In this study, we demonstrate the efficacy of a CRISPR-derived genetic tool to introduce targeted mutations in three major pathogenic species that span the phylogenetic diversity of these bacteria: the avian pathogen *Mycoplasma gallisepticum* and the two most important bovine mycoplasmas, *Mycoplasma bovis* and *Mycoplasma mycoides* subsp. *mycoides*. As a proof of concept, an inducible dCas9-cytidine deaminase system was successfully used to disrupt some main virulence factors in these pathogens. Different induction times and inductor concentrations were evaluated to optimize the editing efficacy. The optimized system was found powerful enough to disrupt 54 out of 55 insertion sequence transposases in a single step. Whole genome sequencing showed that off-target mutations are limited and suggest that most variations detected in the edited genomes are Cas9-independent. This effective, fast and easy-to-use genetic tool opens a new avenue for the study of these important animal pathogens, and most likely for the whole class *Mollicutes*.

Significance

Mycoplasmas are minimal wall-less pathogenic bacteria infecting a wide range of hosts including human, livestock and wildlife animals. Main pathogenic species cause acute to chronic infections involving still poorly characterized virulent factors. The lack of precise genome editing tool has hampered functional studies in many of them, leaving unanswered many questions about the molecular basis of their pathogenicity. Here, we developed a CRISPR-derived base editor tool to efficiently modify the genome of three major pathogenic species, *Mycoplasma gallisepticum*, *Mycoplasma bovis* and *Mycoplasma mycoides* subsp. *mycoides*. Some virulence factors were successfully targeted and we edited up to 54 targets sites in a single step. Availability of this efficient and easy-to-use genetic tool will greatly facilitate functional studies in these economically important mycoplasmas.

Introduction

Mycoplasmas are minimal pathogens belonging to the class *Mollicutes* (1). They are characterized by a streamlining evolution from a Gram-positive ancestor marked by a drastic genome reduction (2–4). This particular evolution led to the loss of diverse cellular functions such as cell-wall production, to a limited number of metabolic pathways and to the lack of an efficient recombination machinery (5). These minimal bacteria are found in a wide range of animals including humans. Many mycoplasmas are etiologic agents of diseases that considerably reduce the animal's production yields and inflate the veterinary healthcare costs (3, 6, 7). *Mycoplasma gallisepticum* (*Mgal*), *Mycoplasma bovis* (*Mbov*), *Mycoplasma mycoides* subsp

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

mycoides (*Mmm*) are three major pathogenic mycoplasmas in terms of economic impact and recent outbreaks. *Mgal* is an animal pathogen listed by the World Organization for Animal Health (OIE) and responsible for respiratory disease in poultry farms worldwide. In the USA alone, its economic annual impact on meat and egg productions are estimated at ~\$780 M and \$150 M, respectively (8–10). *Mgal* also emerged in 1994 in a new host with a jump that was traced back from domestic poultry into North American house finches (*Haemorhous mexicanus*) that is particularly alarming for birds ecology (11). *Mbov* is responsible for a bovine respiratory disease (BRD), but also mastitis and reproductive disease (12). *Mbov* infections in cattle are problematic due to the loss of production yields, lowered fertility rates, and costs of control strategies. Retaining or reaching the status of BRD-free zone is key to many countries, as exemplified by the eradication program currently operated in New Zealand to contain the 2018 epidemics (13), with an estimated cost of more than \$400 M. *Mmm* is a member of the so-called mycoides cluster that gathers five significant ruminant pathogens. It is the causative agent of contagious bovine pleuropneumonia (CBPP), an OIE-listed disease that can take chronic, acute or hyperacute forms. In the latter case, CBPP symptoms includes pericardial effusion, fever and death. During the 20th century, CBPP was eradicated from most developed country but veterinary controls are still required to prevent new outbreaks. In African countries where CBPP is often endemic, the annual cost for mortality, reduced livestock production and disease control efforts is estimated at €44,8 millions (14). The control strategies of the mycoplasmoses varies geographically and includes methods of prophylaxis, the use of antibiotics, of vaccines and ultimately slaughter. In a context of increasing antibiotic resistance (15) and the relative lack of efficacy of vaccines, a better knowledge on the molecular basis of virulence and host-pathogen interactions is required to improve curative protocols and design more efficient vaccines (14, 16). However, the lack of efficient genome engineering tools for many mycoplasmas, including the three species listed above limits the functional genomics approaches and hinders efforts toward the production of rationally designed vaccines. During the past decade, cutting-edge genome engineering methods have been developed that rely on the cloning of mycoplasma genome in yeast before editing with various genetic tools and back transplantation into a recipient cell (17) (18). However, while opening undreamed possibilities to investigate and redesign complete genomes, these in-yeast approaches can be difficult to adapt to a new species. They are still restricted to a small number of species and they have not been developed for any of the three pathogens considered here. Intriguingly, efforts to adapt these synthetic biology approaches to *Mmm* have been unsuccessful so far, while they are now available for all other members of the mycoides cluster (19). For *Mgal*, *Mbov* and *Mmm*, only replicative (*oriC*) plasmids and transposon-based mutagenesis are currently accessible (20–22). Considering the importance of these three mycoplasmas and, more generally of mollicutes for both disease control and fundamental science, there is a need for precise and efficient tools for targeted mutagenesis.

Since 2012, CRISPR-based genetic tools have revolutionized the field of genome engineering in eukaryotes and prokaryotes. The typical CRISPR-Cas9 tool produces double strand breaks in target DNA molecules that are catalyzed by the Cas9 nuclease, guided to specific loci by single guide RNAs (sgRNAs) (23). Following DNA cleavage, double-strand breaks are often repaired by different cellular mechanisms including Non-homologous end joining (NHEJ) and Homology Directed Recombination (HDR). Imperfect chromosome repair by NHEJ is used to disrupt genes whereas HDR enables deletion, insertion or gene exchange in presence of specifically designed DNA repair templates. However, many bacteria, including mycoplasmas, lack efficient NHEJ and HDR repair machineries. In this context, double strand breaks introduced by Cas9 are lethal to the cell and the CRISPR-Cas9 tools can only be used as a counter-selection method (24). Recently developed Base Editor systems (BEs) have offered ways to overcome this problem and introduce single nucleotide variations (SNVs) using a derivative of CRISPR-Cas9 system. Instead of inducing DNA cleavage and, subsequently, DNA repair systems, BEs combine an inactivated Cas9 (dCas9) merged with a cytosine deaminase (CBE) or an adenosine deaminase (ABE) and an uracil glycosylase

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

inhibitor (UGI). An R-loop is opened in the target site by the dCas9-sgRNA complex, the single strand DNA is made accessible in a precise editing window to the attached deaminase (Figure S1). CBE catalyzes deamination of cytosine into uracil, which will be recognized as thymine after replication (Figure S1) ; ABE converts adenine into inosil that is further recognized as guanine after replication (25–27). Base excision repair on edited nucleotides will be prevented by UGI and this improved editing efficiency (28). Therefore, CBE and ABE induce transitions C:G to T:A and A:T to G:C, respectively. Evaluation of CBEs and ABEs in different organisms have shown their high efficiency and low off-target mutations frequencies and toxicity (25, 29). They have been successfully used in various eukaryotic systems including plants and human cells (30, 31) but also in several bacteria including *Klebsiella* (32), *Pseudomonas* (33), *Clostridium* (34), *Streptomyces* (35) or *Agrobacterium* (36).

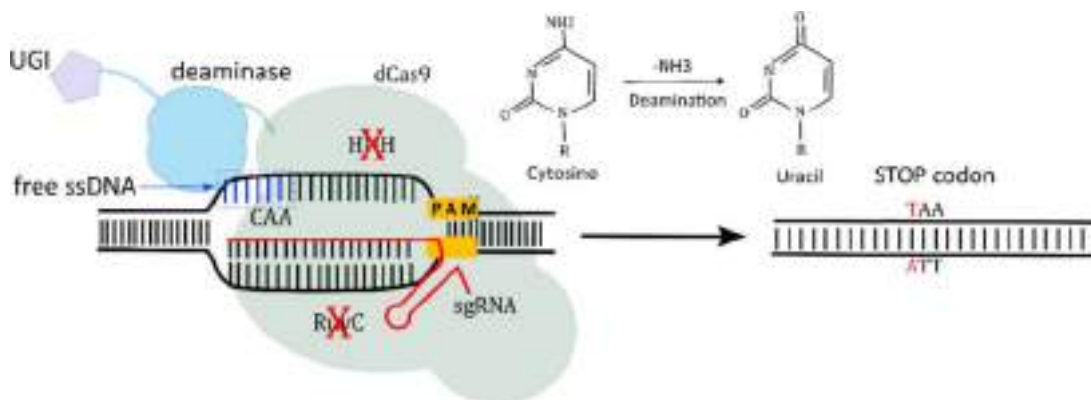


Figure S1: Schematic representation of CBE system. R-loop mediated by the complex dCas9 (Green) – sgRNA (red) is represented. UGI (uracil glycosylase inhibitor; Purple) and deaminase protein (light blue) are fused to dCas9 by amino-acid linker. Editing window is represented as free ssDNA (dark blue) and corresponded to the only nucleotides positions where deaminase can act. Cytosine deamination reaction is represented.

In this study, we have developed CRISPR Cytosine Base-Editor system (CBEs) for three main pathogenic mycoplasmas (Figure S1). First, we have demonstrated the functionality of a dead SpCas9-pmCDA1 fusion protein in *Mgal* using a transposon to introduce the CBE and the sgRNA into the bacterial cells. After optimization, three virulence genes located at various positions on the *Mgal* chromosome were successfully disrupted and individual mutants were isolated. Next, a replicative plasmid and another transposon were used as vectors to evaluate the CBE in *Mbov* and *Mmm*, demonstrating the high efficacy of the system for single gene targeting in ruminant pathogens. Finally we showed the ability of the system for multi-target mutagenesis and successfully targeted 54 insertion sequences in *Mmm* using one single sgRNA.

Results

Design and construction of a base editor genetic tool for mycoplasmas

Two different cytosine deaminases commonly used for base editing, rAPOBEC1 from *Rattus norvegicus* (25), and pmCDA1 from *Petromyzan marinus* (25, 26) (SI, Figure 1A), were evaluated as editing tools in mycoplasmas. The genes encoding these two proteins were codon optimized, chemically synthesized and fused to an inactive version (dead) of Cas9 (dCas9) from *Streptococcus pyogenes* (*Spyo*) using different linkers and positions as described in the Material and Methods section. Briefly, rAPOBEC1 was merged with the N-terminus of *Spyo*dCas9 using the linker XTEN (35). By contrast, pmCDA1 was fused to the C-terminus of dCas9 with a 36 amino acid-linker (37). The uracyl-glycosylase inhibitor from *Bacillus subtilis* phage (35, 38) was also fused at the C-terminus of

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

the resulting hybrid protein (Figure S1) and was used to block a potential base excision repair activity of the bacterial cell. The tetracycline-inducible promoter *PxyI/tetO2* (Breton 2010) was used to control the expression of these two hybrid proteins. The PS promoter, which is considered as a constitutive promoter in many mollicutes species, was used to control the expression of the sgRNA (Figure S2). All genetic elements were assembled within a *Tn4001*-derived transposon (39), resulting in plasmids pTI4.0_SpCas9_pmcDA1 (Figure S2B) and pTI4.0_SpCas9_rAPOBEC1 (Figure S3A, SI).

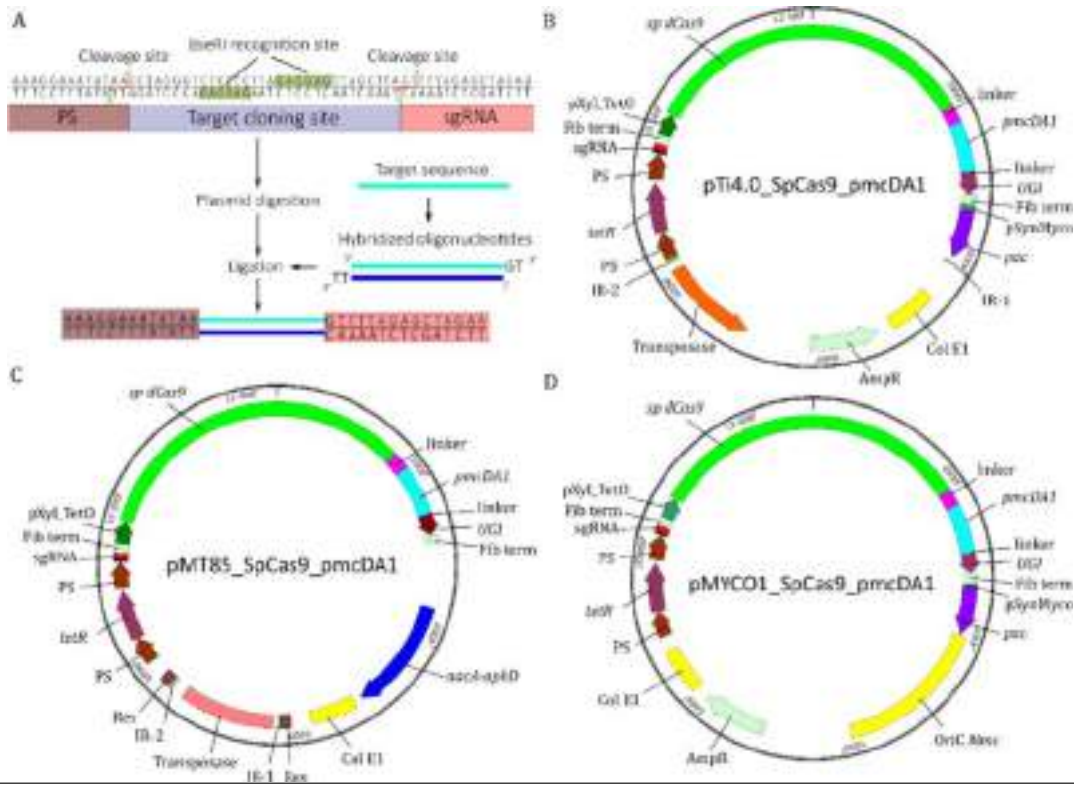


Figure S2. Design of three plasmids for genome editing in mycoplasmas. A- Schematics of the method used for adding a 20 nucleotides spacer defining the target site of the sgRNA, using double BseRI sites. B- Representation of the three plasmid backbones used for base editing experiments in mycoplasmas. Antibiotic resistance markers *pac* (Puromycin) or *aacA-aphD* (Gentamicin) are represented in dark purple and blue, respectively. The pTI4.0_SpCas9_pmcDA1 plasmid (for *Mgal* experiments) contains inverted repeats (IR-1 and IR-2, green) flanking the CBE cassette and allows its integration in the genome using the dedicated transposase (orange). The pMT85_SpCas9_pmcDA1 plasmid (for *Mbov* experiments), has inverted repeat sequences and contained two resolvase sequences (*Res*). These *Res* sequences would allow the elimination of the deaminase cassette and the antibiotic resistance marker by using an *OriC* plasmid that encode for resolvase activity (Janis *et al.* 2008). The pMYCO1_SpCas9_pmcDA1 (for *Mmm* experiments), is a replicative plasmid containing an *oriC* region from *Mycoplasma mycoides* subsp. *capri* (*oriC Mmc*, yellow).

In order to evaluate the ability of these recombinant CBEs to produce targeted mutations in mycoplasmas, the *ksgA* gene of *Mgal* was selected as a first target. *ksgA* is a non-essential gene that encodes a RNA methyltransferase that renders the bacteria sensitive to kasugamycin (40). A 20 nucleotides sequence (TGACCAAATATTATTAATA), upstream of the *spCas9* compatible PAM sequence AGG (Figure 1B), was chosen as a target for *ksgA* inactivation. On the basis of previous reports, the two cytosines in position 16 and 17 upstream of the PAM sequence are in the theoretical editing window of the two CBE systems (41). Deamination of C₁₇ is expected to change a Gln CAA codon into a stop codon TAA. After transformation of *Mgal* with either pTI4.0_SpCas9_pmcDA1 or pTI4.0_SpCas9_rAPOBEC1 and induction with 0.5 µg.mL⁻¹ anhydrotetracycline (aTC) in broth medium (24), the impact of CBEs on cytosine deamination within the target site was analyzed by PCR and Sanger

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

sequencing on the global population of transformants (Figure 1B, Figure S3A). For the rAPOBEC1 transformants (Figure S3), only limited base editing was observed for the two cytosines at positions 16 and 17 (2%) whereas base editing was highly prevalent (59% of C to T conversion in cytosine 16, 50% in cytosine 17) for the pmcDA1-transformants (Figure 1). In the control without the sgRNA, no conversion was observed. **We concluded that the deaminase pmcDA1 can be used to design a CBE that is functional in *Mgal*.**

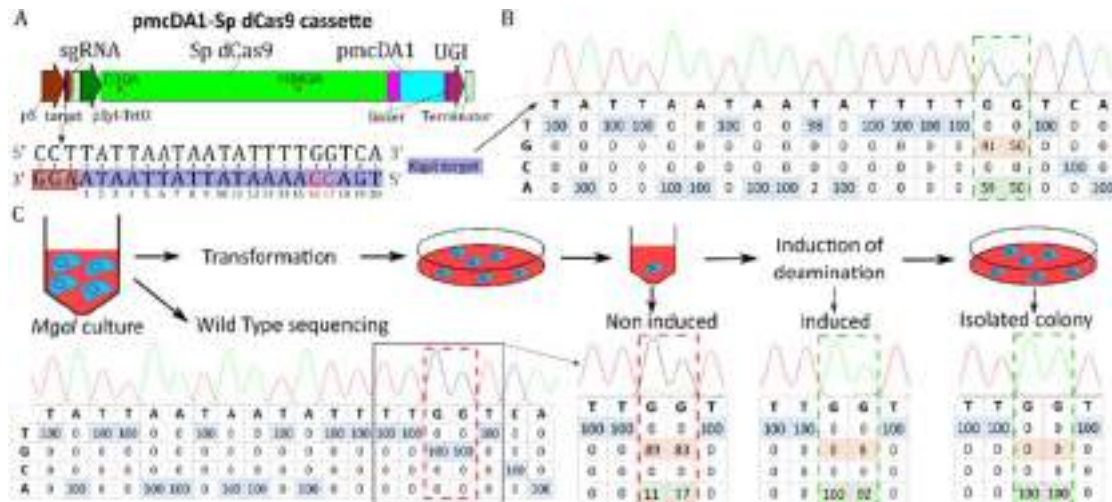


Figure 1. Targeting *ksgA* gene in *Mgal* gene using pmcDA1 deaminase-encoding plasmid. A- Schematics of the CBE expression cassette. The synthetic cassette is composed of (1) a sgRNA (red) including the 20 nucleotides target spacer (dark blue) and under the control of the spiralin promoter (brown) and (2) the codon-optimized hybrid protein composed of *S. pyogenes* dead Cas9 (green), linkers (purple), pmcDA1 deaminase protein (light blue) and UGI (Uracil Glycosylase Inhibitor, dark purple), under control of the *PxyI/tetO2* inducible promoter (dark green). Fibril terminators from *S. citri* (light green) are added downstream of the sgRNA and hybrid protein expression cassettes. The sequence of the *ksgA* target is represented here. Cytosine residues that are susceptible to be deaminated are colored in red. Positions of bases in the target are indicated below each base with numbering starting at the PAM sequence. B-. Percentage of bases found in population of transformants were determined from Sanger sequencing chromatograms using the EditR software and are represented in the table for each nucleotide position in the target sequence. C- Schematic representation of the experiment of base editing and indication of different checkpoints (Sanger results and EditR analysis) for screening until obtaining an isolated mutants.

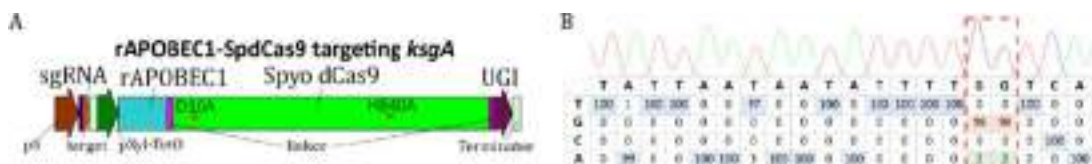


Figure S3. Targeting *ksgA* gene in *Mgal* gene using rAPOBEC1 deaminase-encoding plasmid. A- Diagram of the CBE based on the rAPOBEC1 deaminase. The sgRNA (red) expression cassette includes a 20 nucleotides target spacer (dark blue) under the control of the spiralin promoter (in brown). The *PxyI/tetO2* inducible promoter (dark green) drives the expression of a codon optimized hybrid protein merging *S. pyogenes* dead Cas9 (green) (inactivated by 2 mutations in position 10 and 840), linkers (purple) with rAPOBEC1 deaminase protein (light blue) and UGI (Uracil Glycosylase Inhibitor (dark purple)). Fibril terminators from *S. citri* (light green) are present downstream of the sgRNA and CBE encoding gene. B-- Percentage of bases found in population were determined from Sanger sequencing chromatograms using the EditR software and are represented in the table for each nucleotide position in the target sequence.

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

Optimization of the CBE system and isolation of *ksgA* mutants in *M. gallisepticum*

Given our previous results, the CBE system based on the *pmcDA1* (pTI4.0_spCas9_pmcDA1) was selected for further optimization. First, we studied the inducer concentration. Using the same experimental strategy and *ksgA* as a target, deamination level of targeted cytosines was estimated after an overnight induction using 0.1, 0.25, 0.5, 1, 2.5 or 5 $\mu\text{g}\cdot\text{mL}^{-1}$ of aTC (Figure S4A). An important expression leakage was recorded in the absence of inducer, with up to 25% of conversion at position C16. In presence of the inducer, we determined that the optimal concentration for efficient base-editing was in the range of 0.1 to 0.5 $\mu\text{g}\cdot\text{mL}^{-1}$ of aTC (>50% efficiency) (Figure S4A). With up to 1 $\mu\text{g}\cdot\text{mL}^{-1}$ of aTC, no detectable growth defect was observed in *Mgal*. However, the addition of 5 $\mu\text{g}\cdot\text{mL}^{-1}$ of aTC resulted in a significant growth reduction as evidenced by almost no pH shift of the broth medium after aTC induction (pH= 6.6 at induction point versus pH= 6.54, pH= 5.41 and pH= 5.3 after one night induction with 5 $\mu\text{g}\cdot\text{mL}^{-1}$, 0.5 $\mu\text{g}\cdot\text{mL}^{-1}$ and no inducer respectively). This result suggested that aTC was toxic at high concentration for *Mgal*. We also observed that the use of freshly prepared (<24h) inducer was necessary for maximum efficacy of base editing (data not shown).

Second, we studied the induction time required for a maximum efficacy of the CBE system. Inducer concentrations used for this experiment were 0.25 and 0.5 $\mu\text{g}\cdot\text{mL}^{-1}$ and base conversion were observed at two time points: at two hours or after an overnight induction (Figure S4B). After two hours, significant base conversion were already evident compared to the non-induced condition, with up to 30% of base conversion in the two cytosine positions. However, the overnight induction resulted in the best efficacy, with a 2-fold increase compared to the two-hour induction time (around 60%). **In accordance with these results, we concluded that the best induction conditions for a maximum efficacy of the mycoplasma CBE were 0.1 to 0.5 $\mu\text{g}\cdot\text{mL}^{-1}$ of aTC and an overnight incubation time.**

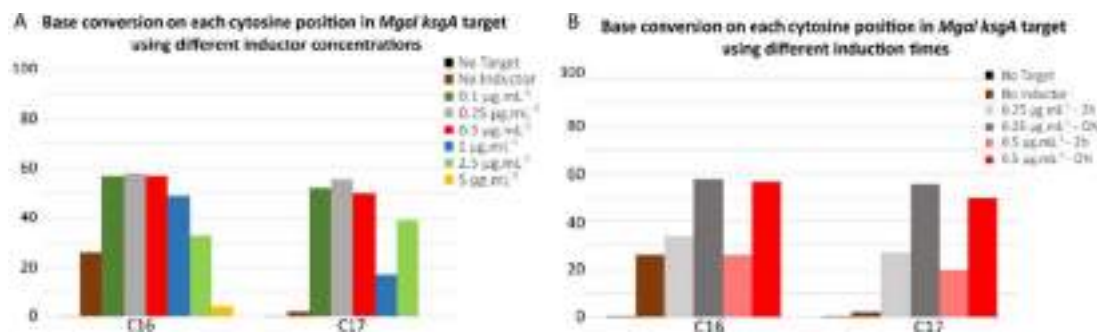


Figure S4. Optimization of CBE induction conditions for base editing in *Mgal*. A- Identification of optimal inducer concentrations for induction of base editing. The percentages of base conversion C to T at each cytosine position 16 and 17 in the target were measured after an overnight induction at 0, 0.1, 0.25, 0.5, 1, 2.5 or 5 $\mu\text{g}\cdot\text{mL}^{-1}$ of aTC. B- Identification of the optimal induction times for maximum efficacy. The percentage of base conversion C to T in each cytosine position 16 and 17 in the target is shown after two hours of induction or an overnight induction.

Finally, in order to get isolated *Mgal* mutants using the pTI4.0_SpCas9_pmcDA1 construct, we performed a third experiment using the newly defined conditions (Figure 1C). During this experiment, targeted *ksgA* sites were analyzed at four points: before transformation (Wild type), after transformation, after induction and on isolated colonies selected on puromycin selective plates. Without induction, 11 and 17% of cytosine 16 and 17 were edited, respectively (Figure 1C), suggesting an important leakage of the inducible promoter, as previously shown. After induction, the deamination in the edited population was nearly complete with 100% conversion of cytosine 16 and 92% conversion of cytosine 17 to thymine. Further screening of five isolated colonies confirmed that four of them carried mutations on the two targeted cytosines while the last one presented an incomplete deamination profile. Resistance to kasugamycin was confirmed by plating the isolated mutant on

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

Hayflick plate, supplemented or not with the antibiotic (Figure S5). We concluded that the pmcDA1-based CBE tool promote base conversion and gene inactivation with a high efficacy in *Mgal*.

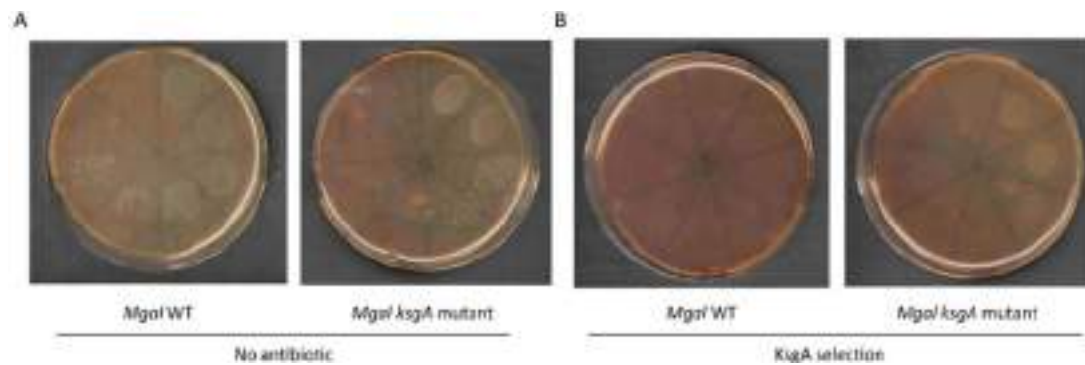


Figure S5. Kasugamycin resistance phenotypic assay on a *Mgal_ksgA* mutant. A- Growth of *Mgal* WT or *Mgal_ksgA* mutant was evaluated in absence (A) or presence of kasugamycin (B) using diluted cell suspensions (10^{-1} to 10^{-7}).

Determination of the editing window for maximum CBE efficacy in *M. gallisepticum*

To investigate the base editing window of the CBE system, three virulence genes of *Mgal* were targeted: *crmA* (GCW_RS01080), *gapA* (GCW_RS01075) and *cysP* (GCW_RS01695) (Figure 2). The *crmA* and *gapA* genes encode two primary cytoadherence proteins that are recognized virulence factors (42, 43). The cysteine protease encoded by *cysP* was shown to cleave chicken immunoglobulin G, and is therefore identified as a potential virulence factor (44). In the *crmA* target (Figure 2A) (ATTCAATATCCTTCGTTGTT upstream of a TGG PAM sequence), cytosine residues C₇, C₁₀, C₁₁ and C₁₇ were potential sites for deamination. After induction, C₁₇ was the only base deaminated at 57% at the population level (before the second plating, Figure 1C). Screening of isolated colonies resulted in 5 out of 10 clones showing a C₁₇ to T₁₇ transition. In *gapA* target (Figure 2B) (TGAACAACAAGGTTCTGCTAA/CGG), C₄, C₇, C₁₄ and C₁₆ were potential sites for deamination and after induction, C₁₆ was the sole deaminated position, with a 54% conversion estimated in the transformants population. Here, 4 isolated clones out of 10 showed a mutation C₁₆ to T₁₆. Finally, in the *cysP* target (Figure 2C) (CAGCAATGAGGAGGAATTTT/CGG), C₁₇ and C₂₀ were potential sites for deamination. After induction and population analyses, the conversion levels were 76% and 77% for these two cytosines, respectively. All (*i.e.* 10 out of 10) isolated clones showed mutations at the two positions. **From these results, we concluded that the editing window of the mycoplasma CBE system ranged from positions 16 to 20. Its high efficacy was confirmed in *Mgal* and mutants of three important virulence factors were isolated.** Interestingly, during this work, we realized that induction can be performed immediately following plasmid transformation (*i.e.* right after the 2h cells recovery in Hayflick medium) instead of being carried out after colonies recovery and regrowth. **Given that *Mgal* colonies generally appear in one week, this improvement allowed us to save days of experiments and obtain the expected mutants in less than 3 weeks.**

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

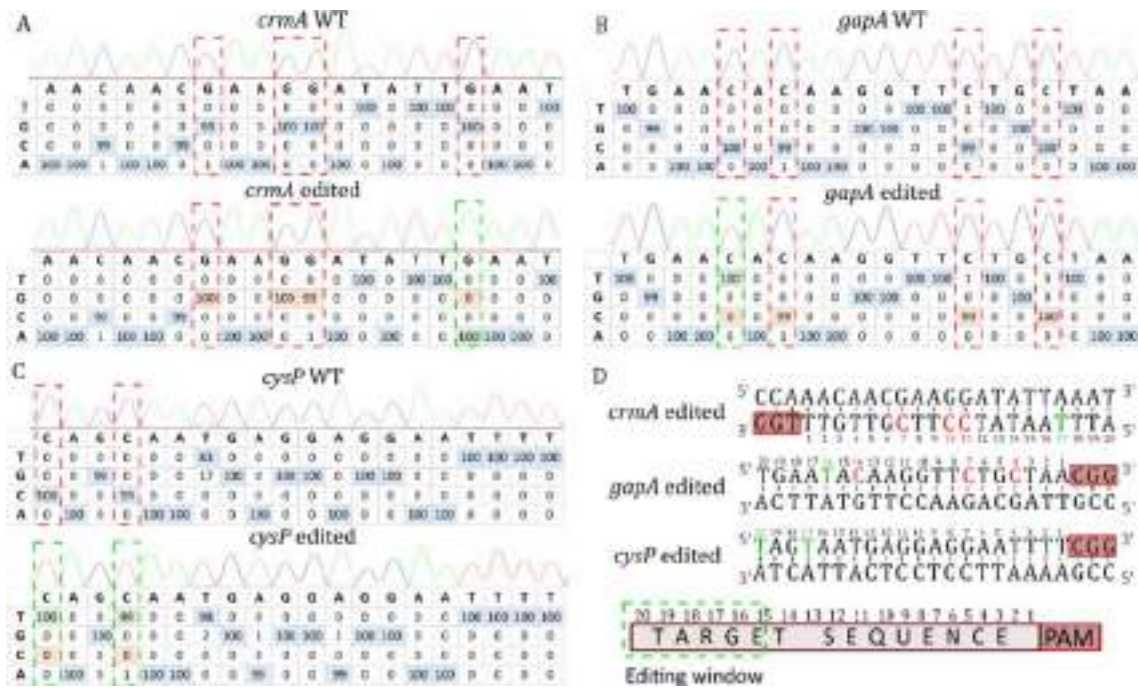


Figure 2. Targeting three virulence factors of *Mgal* to explore the editing window of the CBE system in *Mgal*. The 20 nucleotides target sites were sequenced and are represented here for *crmA* (A), *gapA* (B) and *cysP* (C) genes before and after the base-editing experiments. Cytosines susceptible to deamination (or guanine in the reverse strand) are framed with red dotted lines and deaminated ones with green dotted lines. Percentage of each base is represented in the tables; they were determined using Sanger sequencing and the EditR software. Edited bases are highlighted in green. D- The 20 nucleotide target sites are shown for the three targets. The position of each base in the target is indicated below each nucleotide and corresponds to the nucleotide position in the target upstream of the PAM sequence. Red letter represents un-deaminated cytosines and green letter represents converted thymine. A schema deduced from those three experiments was drawn to shine a light on the editing window. As shown in the schema, cytosines located at position 16 to 20 can be CBE-targeted.

Development of the CBE system for *Mycoplasma bovis*

In a first attempt, the pTI4.0_spCas9_pmcDA1 that had been successfully used in *Mgal* was evaluated in *Mbov*. However, we failed to obtain *Mbov* transformants using this CBE-carrying plasmid. This could be linked to the *pac* resistance marker that was not described in the literature for *Mbov*. Therefore, the inducible CBE system was inserted in another transposon-based plasmid carrying a gentamycin resistance marker to generate the pMT85_spCas9_pmcDA1 (Figure S2). The resulting construction was evaluated in *Mbov*, using the *mnuA* gene (MBOVPG45_0215) as a target (Figure 3). MnuA is a major membrane nuclease that has been suggested to play a key role in *Mbov* virulence by degrading the neutrophil extracellular traps (NETs) produced by the host in response to the pathogen (45). In the target spacer (AACC¹⁷AAAAATATGACTTAGT/AGG), C₁₇ and C₁₈ stand as two potential sites for deamination. Conversion of C₁₇ would change a CAA (Gln) codon into a TAA stop codon. *Mbov* cells were transformed with the pMT85_spCas9_pmcDA1 targeting the *mnuA* gene and transformants were grown in gentamycin selective medium (liquid) for three passages before induction and plating. The expression of the CBE system was induced overnight with 0.5 µg.mL⁻¹ of aTC and *Mbov* transformants were plated on selective medium to isolate colonies. Target sites were analyzed in the global population before transformation, after each passage following transformation (P1, P2 and P3) and after induction (on the cell suspension and on isolated colonies) (Figure 3). As previously reported in *Mgal* and *Mmm*, an important leakage of the inducible promoter was observed in the *Mbov* population right after transformation with ~50% of the cytosines converted to thymines at positions 17 and 18 at

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

P1, P2 and P3. Overnight induction with aTC increased the conversion level to 90%. Screening of three independent colonies showed that the *Mbov_mnuA* mutants had both cytosines (C17 and C18) converted to thymines. In order to demonstrate that the mutations introduced in the *mnuA* gene lead to its inactivation, a nuclease phenotypic assay was carried out. (46). This nuclease assay indicated that the tested *Mbov_mnuA* mutant was not able to hydrolyze DNA when the WT strain degraded it all (Figure S6). This work demonstrates the portability of this CBE-based method to produce targeted mutations in *Mbov*.

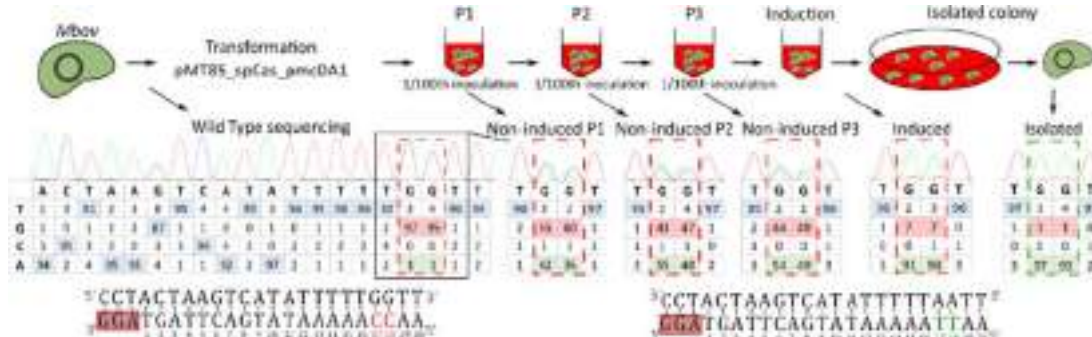


Figure 3. Disruption of the MnuA nuclease encoding gene by base editing in *Mbov*. After transformation of *Mbov* with the plasmid pMT85_spCas9_pmcDA1, cells were propagated in liquid media supplemented with gentamycin. After three passages in liquid medium (P1 to P3) the inducer was added to the cell suspension. At each passage and after induction, aliquots were collected in order to monitor target sites sequences using Sanger sequencing and the EditR software analysis. Sequencing chromatograms are presented as well as base percentage for each position of the target sequence. Cytosines susceptible to be deaminated are framed in red color and edited bases in green color.

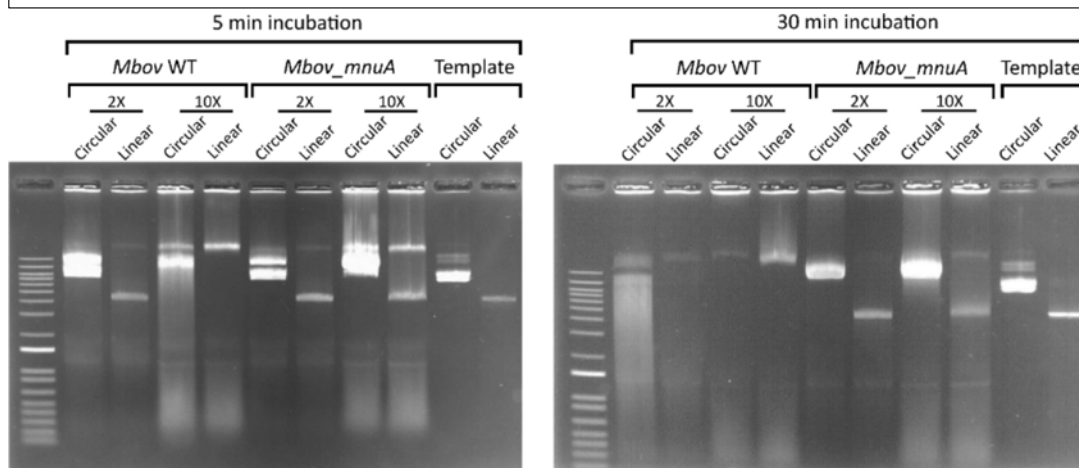


Figure S6. *Mbov_mnuA* DNase phenotypic assay. *Mbov* Wt and *Mbov_mnuA* mutant were incubated with 2 μ g of linear or circular DNA. After 5 min (left) or 30 min (right) at 37°C, aliquots were removed and run on 1% agarose gels. DNA digestion results were observed after staining the gels with ethidium bromide. Two concentrations of *Mbov* cells (2X and 10X) were used as described in the Materials and Methods section.

Development of the CBE system for editing *Mycoplasma mycoides* subsp. *mycoides* genome

As a new proof of the flexibility of the mycoplasma CBE system, we introduced the sgRNA and CBE expression cassettes in the replicative *oriC* plasmid pMYCO1. This latter is routinely used in several species of the mycoides cluster (22), including *Mmm*, an important ruminant pathogen. The resulting pMYCO1_spCas9_pmcDA1 plasmid (Figure S2) was used to target the non-essential gene *glpO* (CAACAACAATACGATAACAT/TGG) that encodes a metabolic enzyme, (the L- α -glycerophosphate oxidase) possibly involved mycoplasma virulence (47). Indeed, GlpO catalyzes the oxidation of glycerol-3-phosphate that comes with the release of hydrogen peroxide (H₂O₂), a product known to contribute

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

to cytopathic effects in host tissue. Cytosines C₃, C₉, C₁₄, C₁₇ and C₂₀ were potential sites for deamination, with conversion of the last three positions leading to three TAA stop codons. After transformation of *Mmm*, different inducer concentrations and induction times were tested, and conversion of the cytosines was surveyed in the cell population using Sanger sequencing and EditR analysis (Figure S7). Before the induction of the CBE expression, conversion of cytosines was observed as in *Mgal* and *Mbov*, reaching up to 30% in the position C₁₇. After induction, best results were obtained with 0.5 µg.mL⁻¹ of aTC applied overnight (just like *Mgal*). We observed cytosine conversion for each cytosine within the target region but also for the cytosine located 22 nt upstream from the PAM sequence, suggesting an extended editing window between positions C₁₄ and C₂₂ in *Mmm*. Low conversion frequencies (20-30%) were observed at the editing window extremities whereas higher efficiencies were recorded between C₁₇ and C₂₀ (60-70%). After subcloning on agar plates, mutants showed different editing profiles on the four cytosines positions (mixed population and different fully deaminated cytosines combination). Nevertheless, two out of ten isolated clones displayed the expected all four mutations (C₁₄, C₁₇ and C₂₀ and C₂₂). A phenotypic assay for evaluating the H₂O₂ production confirmed the *glpO* gene inactivation in these clones (Figure S7-E). Finally, after three passages in liquid without selective pressure and one passage in solid medium for isolation, plasmids-free *Mmm* mutants edited on the *glpO* were recovered. To our knowledge, they are the first site-specific *Mmm* mutants. .

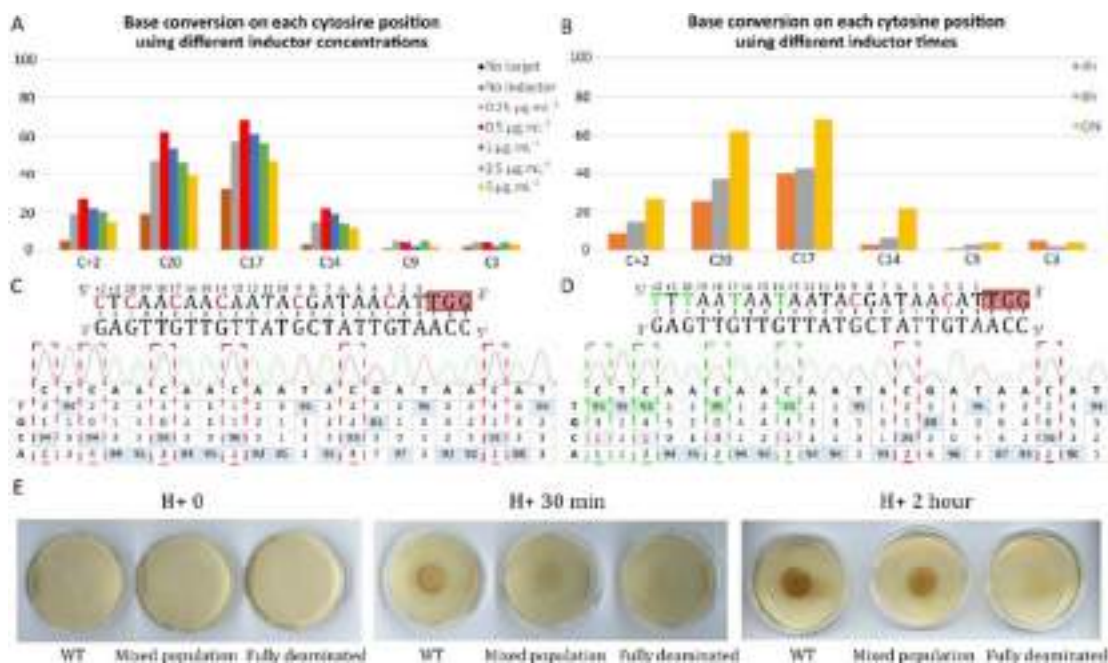


Figure S7. Targeting the *glpO* gene in *Mmm* using the mycoplasma CBE system. A- Identification of optimal inducer concentration for induction of base editing. Percentages of base conversion C to T were calculated at each cytosines position in the target sequence using overnight induction with 0, 0.25, 0.5, 1, 2.5, 5 µg.mL⁻¹ of aTC. B- Cytosine deamination efficacy after 4h, 8h or overnight induction time using 0.5 µg/mL of aTC. C, D- Results of *glpO* editing. Target sequences used are represented; cytosines susceptible to deamination are represented in red and position of the nucleotide before PAM sequence is indicated above each nucleotide. These results represent Sanger sequencing chromatograms and EditR analyses before induction (C) and after clone isolation (D). E- H₂O₂ detection using a DAB (3,3'-Diaminobenzidine) assay on three *Mmm* clones, wild type (WT), mixed population of *Mmm* cells with incomplete deamination and a clone fully deaminated at positions C₁₄, C₁₇ and C₂₀ and C₂₂. The results were observed at different times after adding the reagent: 0 minute, 30 minutes and two hours of incubation.

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

Inducing multiple mutations in *Mmm* with one single sgRNA

To test the limits of the mycoplasma CBE system, we chose to target transposase encoding genes associated with two families of insertion sequence IS1634 and IS3 that are presents 58 and 26 times in the *Mmm* genome, respectively (Figure 4, Figure S8). As transposase gene sequences are highly conserved within each family, we tried to target a maximum number of copies using one single sgRNA per family. For the IS1634 transposase, a 20-nucleotide spacer was designed to target the inactivation of 55 copies out of 58 (AGACCAGATTGTTATAGGTA/TGG). Deamination of C₁₆ on the reverse strand would change a CAG (Gln) codon into a TAG stop codon at position 204 of the protein (Figure 4A). For IS3 transposase, two 20 nucleotides spacers were designed to build sgRNAs, the first one targeting the start codon of the transposase (sgRNA1, CATATAAAAACCCCATTTCC/TGG) on all 26 copies and the second one allowing the introduction of stop codon on 22 out of the 26 copies (sgRNA2, ACAAGTGGAACTATAAGT/TGG) (Figure S8). After transformation and induction, all target sites were surveyed by PCR and Sanger sequencing (Figure 4B, Figure S7B). A low base-editing efficiency on IS1634 target was observed after the first induction, with 5% conversion in C₁₆ and 4% in C₁₇. Two additional induction steps were conducted, resulting in an increased base conversion level of C₁₆ and C₁₇ with 37% and 40% after the second induction and 72% and 75% after the third induction. After spreading on plates and PCR/sequencing screening of 20 isolated clones, two of them presented sequencing profiles suggesting that all IS1634 sites had been mutated (Figure 4C). Finally, full genome sequencing indicated that 52 and 54 transposases out of 55 were inactivated in these two clones. For the IS3 inactivation, 66% and 77% of conversion were observed in population analyses after three inductions using sgRNAs with spacers one and two, respectively (Figure S8B). After spreading on plate and PCR screening of 10 clones, one fully deaminated clone was rescued for each of the two sgRNAs (Figure S8C). Full genome sequencing confirmed mutations on the 22/26 and 22/22 target sites with sgRNA1 and sgRNA2, respectively (Table S3). **These results demonstrate the ability of the mycoplasma CBS system to edit numerous target sites in a single step.**

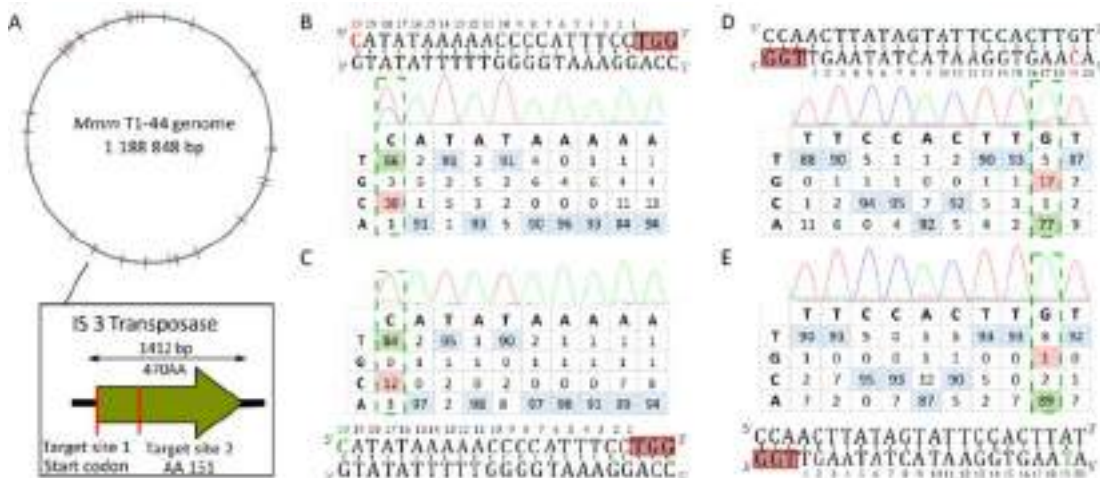


Figure S8. Multitargeting of IS3 copies in *Mmm* genome. A- Schematic representation of *Mmm* T1/44 genome with 30 complete or truncated copies of the IS3 transposases represented as solid lines. In the right panel, the targeted site within IS3 transposase gene is represented as a red bar. B, D- Population screening was performed by PCR and Sanger sequencing after three steps of inductions and using the sgRNA1 (B) or the sgRNA2 (D). Targeted cytosines are indicated in red. Percentages of each base at each position are represented in tables. C, E- Sequence profiles of fully mutated clones on site targeted with sgRNA1 (C) and sgRNA2 (E). Mutated positions are in green frame and are represented in green letters in the target sequences.

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

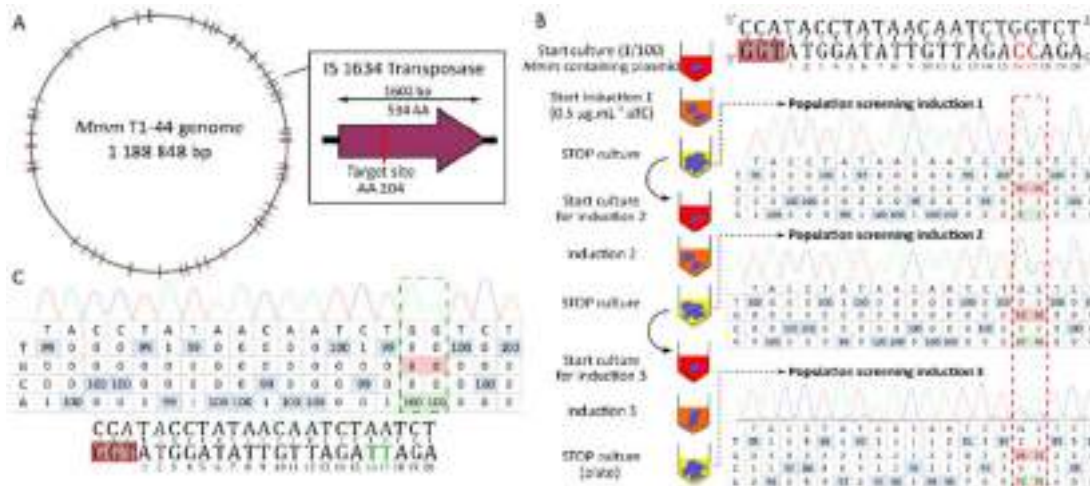


Figure 4. Multitargeting of IS1634 copies in *Mmm* genome. **A-** Schematic representation of *Mmm* T1/44 genome with 58 complete or truncated copies of the IS1634 transposases represented as solid lines. In the right panel, the targeted site within IS1634 transposase gene is represented as a red bar. **B-** Schematic representation of the three induction steps of the mycoplasma CBE system for targeting 55 IS1634 using one single gRNA. Nucleotide sequence of the target is represented with the PAM sequence framed in a red rectangle; cytosines susceptible to be deaminated are in red font. Fresh cultures of *Mmm* cells harboring the pMYCO1_spCas_pmcDA1 plasmid are represented as red colored tubes. Cultures in exponential and stationary growth phases are indicated by orange and yellow colored tubes. Three consecutive culture steps were performed (including the start culture) by 1/100 dilution in fresh medium. At each step, induction was performed by adding aTC and incubating the cells until stationary phase is reached. Deamination of target cytosines was evaluated using PCR and Sanger sequencing. Percentages of base at each position were estimated from chromatograms using the EditR software. The two positions of interest are highlighted in a red dotted rectangle. **C-** Sanger sequencing results and base distribution in a selected isolated clone. Fully mutated positions are indicated by green frame on the table and green font in the target sequence.

Evaluation of off-target mutations by whole genome sequencing

To detect potential undesired mutations and evaluate the off-target activity of the CBE system in mycoplasmas, whole genome sequencing was conducted on seven edited clones using both short reads (Illumina) and long reads (Oxford Nanopore Technologies) sequencing platforms: *Mgal_ksgA* mutant, *Mbov_0215* mutant, *Mmm_glpO* mutant, *Mmm_IS1634_cl3.1.6* mutant, *Mmm_IS1634_cl3.1.11* mutant, *Mmm_IS3_cl4.1.2* mutant and *Mmm_IS3_cl5.1.18* mutant (Table S3). Five to 36 undesired mutations were detected in the sequenced genomes. Mutations were further analyzed and classified into three categories (Figure S9A – Table S3): 1- sgRNA-guided off-target mutations possibly generated by the CBE system, if the targeted sequence was similar to the sgRNA spacer sequence and adjacent to a NGG PAM sequence ; 2- Potentially unguided spurious deaminations of pmcDA1 (29, 48) for all C-T or G-A mutations in genome regions with no similarity to the sgRNA spacer sequence ; 3- Other mutations that cannot be induced by the CBE system and probably occurred during passages. Based on these criteria, 0 to 5 sgRNA-guided off-target mutations were predicted, representing 0 to 17% of the total mutations, with the exception of *Mmm_glpO* for which 2 mutations out of 5 were classified in this category. Two examples of off-target interaction (1) of the *ksgA* sgRNA targeting DNA in *Mgal* (localization in the genome: 1313) and the *glpO* sgRNA targeting DNA in *Mmm* (localization in the genome: 220 532) are presented (Figure S9B – Figure S9C – Table S3a). Spurious deaminations represented 20-84% whereas other variations represented 11-50% of the mutations. Interestingly, the numbers and percentages of spurious deamination when using the mycoplasma CBE system in *Mmm* were remarkably higher when a three-inductions protocol was used to maximize IS1634 and IS3 targeting efficiencies (11-24 mutations), compared to the single induction

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

protocol used to target *glpO* (1 mutation). We concluded that sgRNA-guided off-target mutations were rather rare whereas spurious deaminations accumulate upon extended CBE activity.

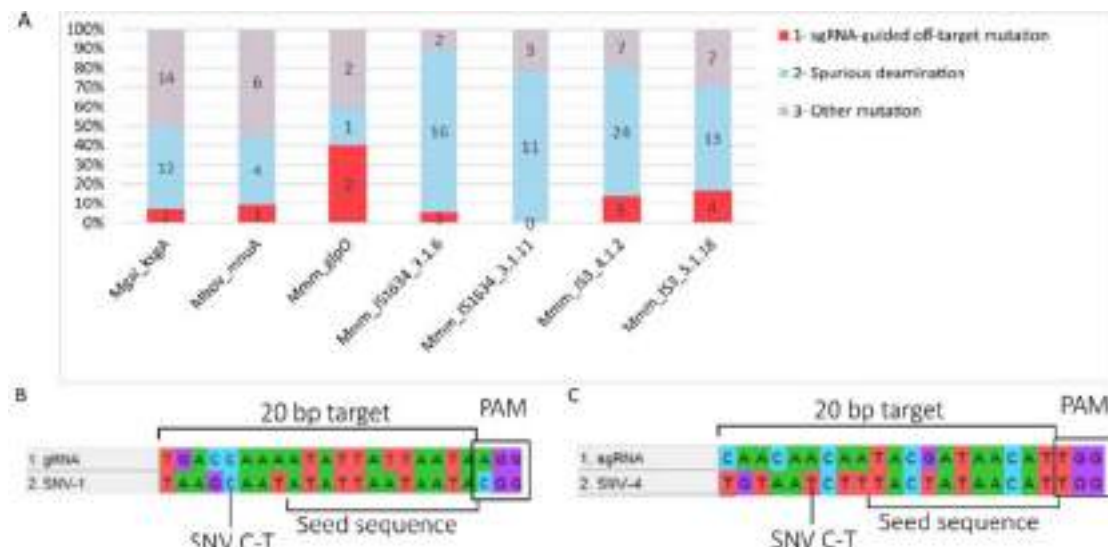


Figure S9. Analysis of undesired mutations from whole genome sequences. A- Histogram summarizing the percentages of different categories of mutations predicted from analyses of whole genome sequencing data on seven mycoplasma mutants. Numbers of mutations are indicated within bars. Blue, sgRNA-guided off-target mutations; orange, spurious deaminations; grey, other mutations. B- Example of off-target interaction of the *ksgA* sgRNA targeting position 13113 of the *Mgal* genome. C- Example of off-target interaction of the *glpO* sgRNA targeting position 220 532 in the *Mmm* genome.

Discussion

Despite being known as minimal bacteria, the functional genomics of mollicutes has long been hampered by a lack of efficient genetic tools to generate targeted mutations. While in-yeast genome engineering methods developed during the past decade represent a real breakthrough for several species from or related to the mycoides cluster, most species, including significant pathogens of human, animals and plants remain unamenable to targeted mutagenesis. In this study, we developed a new genetic tool based on CRISPR for direct mutagenesis in mollicutes and evaluated its efficacy in three economically relevant animal pathogens.

Since their first publication in 2016, cytosine base editor systems have been proven highly efficient in a diversity of eukaryotic and prokaryotic organisms. CBE systems take advantage of a deactivated Cas9 that retains the ability to be driven to target by a guide RNA but does not induce dsDNA breaks, coupled with a cytosine deaminase, generally rAPOBEC1 or pmcDA1. At first, we designed and constructed two mycoplasma CBE systems, but our preliminary results suggested that only the pmcDA1-based CBE was active in *Mgal* (Figure 1 – Figure S3). Both cytosine deaminases have been successfully used in bacteria (35, 37) and more investigation would be required to understand why the rAPOBEC1-based CBE was not efficient in *Mgal*. By contrast, the pmcDA1-based CBE was shown highly efficient for generating point mutations in *Mgal*, *Mbov* and *Mmm* (Figure 1 – Figure 3 – Figure S7). While belonging to different phylogenetic groups of mollicutes, the same promoters driving the expression of the sgRNA (PS) and the CBE system (*Pxyl/tetO2*) were used, suggesting that both expression cassettes could be used without modification in various species of mycoplasmas and possibly other mollicutes, including *Spiroplasma*, *Acholeplasma* or *Ureaplasma*. But these promoters could also be replaced if required. Three different plasmid backbones were evaluated including two Tn4001-derived transposons and one replicative *oriC* plasmid (Figure S2). Transposons and *oriC* plasmids are currently the most widespread

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

genetic tools available in mollicutes. Transposon mutagenesis with Tn4001 derivatives has been used in 15 species and *oriC* plasmids are available for 14 species. Therefore, the mycoplasma CBE system could be easily evaluated in many species, either directly or after cloning the sgRNA and CBE expression cassettes in a compatible vector.

The standard usage of Cas9 relies on targeted DNA cleavage and requires efficient DNA repair mechanisms of double strand breaks to avoid cell death. By contrast, CBE-mediated mutagenesis does not lead to chromosome cleavage. Therefore, the cell toxicity of the sgRNA/dCas9- pmcDA1-UGI complex is expected to be low, and indeed, no particular difficulty was encountered when expressing the CBE system in the three mycoplasmas tested here. Taking advantage of this specificity, we succeeded in targeting up to 54 targets in the *Mmm* chromosome, with one sgRNA (Figure 4). Multiple targeting of repeated sequences has been previously reported in eukaryotic and prokaryotic systems (37, 38, 49), but to the best of our knowledge, this is the first time so many targets sites were modified in bacteria. This application of CBE opens new possibilities for targeting multigene families with a limited number of sgRNA. In mycoplasmas, such families are often predicted to encode surface proteins suspected to be involved in host-pathogen interaction but current mutagenesis methods, including in-yeast genome engineering, were unable to disrupt all genes of a family in a reasonable time.

In order to evaluate potential off-target mutations induced by the mycoplasma CBE system, whole genome sequencing was achieved for seven mutants selected over the study (Figure S9 – Table 3). From our analyses of the detected mutations, only few of them possibly resulted from sgRNA-guided off-target deamination events. Concerning spurious deamination of pmcDA1 that preferentially deaminate TC motif (29), 12, 4 and 1 sites were observed in *Mgal_ksgA* mutant, *Mbov_0215* mutant and *Mmm_glpO* mutant clone with nine, two and one TC motifs. Spurious deamination was suggested to increase with extended induction periods, suggesting a control of CBE expression might be key to reduce the frequency of undesired mutations. Such spurious deamination induced by CBE systems have already been reported in other bacteria, including *Corynebacterium glutamicum* or *Bacillus subtilis* where 9 and 19 SNV respectively might be attributed to deaminase activity (50, 51). In the CBE system developed here, the expression of the dCas9- pmcDA1-UGI hybrid protein is driven by the promoter *Pxyl/tetO2* which can be induced in presence of aTC. While a clear increase of deaminated base frequencies was shown in the three studied mycoplasmas after induction, the conversion process was already observed before induction (Figure 1 – Figure 3 – Figure S7). This indicated some promoter leakage in the three species, in accordance with another report in *M. pneumoniae* (52). Several strategies could be envisaged to limit the background of undesired mutations, including limiting the induction time and using an improved inducible promoter as the one recently developed for *M. pneumoniae* (52). In addition, the use of high fidelity Cas9 variants or CBE variants (41) may also improve the specificity of the genetic tool. Finally, elimination of the CBE system after mutagenesis is also key to avoid accumulation of undesired mutations over time. This can be achieved for the CBE systems based on the pMT85_2Res and pMYCO1 *oriC* backbones. Indeed, all genetic elements flanked by the *res* sequences can be removed from the chromosome using a dedicated resolvase activity (53) and the *oriC* plasmids are generally lost after a few passages in non-selective medium (54). Alternatively, CBE constructions can be enhanced with the CRE-Lox system that is functional in some mycoplasma species (52).

One of the limits of CBE systems is the number of targets that could lead to gene disruption by generation of a stop codon. Indeed, the targeted cytosine has to be present within the editing window of the CBE system, which we estimated was located 16-22 nt upstream of an NGG PAM sequence. The number of targets in bacteria with AT-rich genomes like mycoplasmas could be increased by using the

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

VQR variant of SpCas9 which PAM recognition specificity has been changed to NGA (41, 55). Other deaminase-based tools were also recently developed, including an adenosine base editor system (ABE) and a double ABE-CBE system called SPACE, that offer other possibilities of base change (27, 56).

In conclusion, the mycoplasma CBE system is an easy-to-use genetic tool that was shown to be highly efficient to target individual or multiple genes in three mycoplasma species for which no targeted mutagenesis methods was available. This tool is independent from cell repair mechanisms and enables to obtain an isolated mutant of *Mgal*, *Mbov* or *Mmm* in less than a month. During this study, we already generated individual mutants of potential virulence genes for each species, demonstrating the efficacy and versatility of this new genetic tool.

Materials and methods

Oligonucleotides and plasmids

All oligonucleotides used for this study were supplied by Eurogentec and are described in Table S1. All plasmids constructed and used in this study are listed in Table S2. Detailed protocols for plasmid constructions, DNA sequences of genes and plasmids used are provided as Supplementary Information (SI-2).

Bacterial strains and culture

Mgal strain S6 (Tax ID: 1006581) was cultivated at 37°C in Hayflick modified medium (57), under a 5% CO₂ atmosphere. Puromycin 10 µg.mL⁻¹ and kasugamycin 400 µg.mL⁻¹ were used in selective conditions. *Mbov* PG45 (Tax ID: 289397) was cultivated in SP4 medium (57). Gentamycin 100 µg.mL⁻¹ was used in selective conditions. *Mmm* T1/44 (Tax ID: 2103) was cultivated in SP5 medium (19); puromycin 8 µg.mL⁻¹ was used in selective conditions (58). Phenol red was used as a pH indicator in mycoplasmas media. *Escherichia coli* 5α (NEB, C2987H) was used for plasmid constructions and was cultivated in LB Broth (ThermoFisher: 12795027). Ampicillin 100 µg.mL⁻¹ or kanamycin 50 µg.mL⁻¹ were added in selective conditions. Protocols for *Mgal*, *Mbov* and *Mmm* transformation, transformants screening and characterization are available as SI.

In silico design of codon optimized base editor systems:

DNA sequence encoding dead SpCas9, pmcDA1, rAPOBEC1, UGI as well as the puromycin resistance marker (linked to SynMYCO promoter) were codon optimized with the Optimizer online tool (<http://genomes.urv.es/OPTIMIZER/>) using the codon usage of *Mgal* from Kazusa database (<https://www.kazusa.or.jp/codon/>). The sgrNA encoding gene was cloned between the promoter of the spiralin gene and the terminator of the Fibril gene from *Spiroplasma citri* (Figure S1). Sp dCas9 was fused to rAPOBEC1 sequence by 48 bp linker used in BE4 system (28). Sp dCas9 was fused to pmcDA1 sequence using the same 207 bp linker as in the hAID system (26). UGI was fused to the N-terminal sequence using a 12 bp sequence linker (28). The inducible promoter *Pxyl/tetO2* (59) and the terminator of the Fibril gene were added respectively upstream and downstream of the CBE sequence. Manual modifications to avoid regions with AT-rich tracks or unwanted BseRI restriction sites have been performed before synthesis of the DNA fragments by Twist Bioscience.

Plasmids construction:

The Mini transposon vector derived from Tn4001 (60) and the synthesized fragments provided by Twist Biosciences were used as PCR template. Primers A1/A2, A3/A4 and A5/A6 (SI) and the Q5 High-Fidelity DNA Polymerase kit (NEB, M0491) were respectively used to amplify (i) the plasmid backbone encoding the transposon and the tetracycline repressor, (ii) the synthesized fragment encoding the sgrNA and the first 665 amino-acids of the hybrid protein dCas9-deaminase-UGI, (iii) the synthesized fragment encoding the last 1067 amino-acids of the dCas9-deaminase-UGI hybrid protein and the puromycin

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

resistance marker. PCR products were incubated with the DpnI restriction enzyme (NEB, R0176S) following the manufacturer's recommendations. Both products were purified using the GFX™ PCR DNA or Gel Band Purification Kit (GE-healthcare). The NEBuilder HiFi DNA Assembly Cloning Kit (NEB, E5520S) was used to assemble DNA fragments and create the circular plasmids pTi4.0_spCas9_rAPOBEC1 and pTi4.0_SpCas9_pmcDA1. Two μL of assemblies were used to transform *E. coli* NEB 5 α (NEB, C2987H). Transformants were screened after DNA extraction with the NucleoSpin Plasmid kit (Macherey-nagel, 740588.50) and enzymatic digestion. Sanger sequencing of the targeted locus was performed (Genewiz) for final verification.

For the plasmid pMT85_spCas9-pmcDA1, the pMT85-2Res-Genta backbone (53) was used instead of the Mini transposon backbone. Primers G38/G39 were used to amplify a fragment containing the sgRNA, deaminase complex and Tetracyclin repressor; and primers G36/G37 were used to amplify the pMT85-2Res-Genta backbone. Cloning was performed using the same procedure as for pTi4.0 (see above).

For the plasmid pMYCO1_spCas9-pmcDA1, the pMYCO1 backbone (*oriC* plasmid (54)) was used instead of the transposon backbone. Primers D27/D28 were used to amplify a fragment containing the sgRNA, deaminase complex, Tetracyclin repressor and puromycin resistance marker; primers D29/D30 were used to amplify the pMYCO1 backbone. Cloning was performed using the same procedure as for pTi4.0 (see above).

For all CBE constructs, 20 nucleotide targets were added using the same process. Targets were designed as oligonucleotides R and F. 10 μL of each primer (100 μM) were heated with 2 μL of Adv2 polymerase buffer (TAKARA, 639232) at 95°C during 5 min and cooled down slowly (-0,1°C/sec) to room temperature. Plasmids were digested with BseRI (NEB, R0581S) 2 h at 37°C. After purification, target sequences and plasmids were ligated using the T4 DNA ligase (Promega) overnight at 4°C. Transformation of 2 μL of the ligation mix was performed in NEB 5 α competent *E. coli* (NEB, C2987H). Colonies were screened by digestion of the plasmids with BseRI and KpnI restriction enzyme and target sequences verified by Sanger sequencing (Genewiz).

Transformation of *Mycoplasma* species

Mycoplasma transformation was performed using PEG mediated protocol (61, 62). Late-log phase *Mycoplasma* cultures were transformed with plasmid DNA (20 μg). After transformation, cells were resuspended in 1 mL of appropriate medium, incubated 2 hours at 37°C and plated onto selective solid medium. After incubation at 37 °C for 3 to 10 days, single colonies containing deaminase construct were obtained.

Induction of CBE:

For each *Mycoplasma* species, transformants were picked and propagated in selective liquid media during three passages (one passage is equivalent to a 1:100 dilution). At passage 3, fresh aTC (Abcam, ab145350) (in ETOH 50%) was added to the culture at early logarithmic growth phase (~10 h for *Mbov* and *Mmm*, ~24 h for *Mgal*), and let in contact with the cells until the stationary phase was reached (12 h for *Mbov*, 18 h for *Mmm*, 24 h for *Mgal*). A concentration of 0.5 $\mu\text{g}\cdot\text{mL}^{-1}$ of aTC was used in these three *Mycoplasma* species. After induction, cells were plated onto selective solid medium. After incubation at 37 °C for 3 to 10 days, isolated clones were obtained.

Alternatively, induction was performed right after transformation. In this case, after the 2 h incubation step at 37°C (recovery of the cells), antibiotics (puromycin or gentamicin) were added to the media for two hours. Then, the base editor system was induced using fresh aTC (0.5 $\mu\text{g}\cdot\text{mL}^{-1}$) for 12 h to 15 h (overnight). Induced cultures were plated on selective media and incubated at 37°C with 5% CO₂.

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

PCR screening of transformants

PCR screening was performed on all transformants before induction, after induction or on isolated clones, using Advantage HF 2 PCR Kit (Takara, 639123). List of all the primers used for sites deamination are detailed in SI.

EditR analysis

The EditR 1.0.10 software (https://moriaritylab.shinyapps.io/editr_v10/) was used to analyze and quantify base editing at each position on the 20 nucleotides target (63). The analysis was performed using the .ab files available after Sanger sequencing,

Whole genome sequencing of mycoplasmas

Genomic DNA of *Mgal* was extracted from a 10 mL culture using the Qiagen Genomic-Tips 100/G kit. Genome sequencing was performed by the Genome Transcriptome Facility of Bordeaux. Long reads and short reads were produced using a GridION (Oxford Nanopore) and a MiSeq (Illumina), respectively.

For *Mmm_glpO* mutant cl18_4, ONT sequencing generated 24,400 reads (Mean read length: 29,377 bp) and Illumina 1,044,350 read pairs. Analyses were performed on Galaxy (<https://usegalaxy.eu/>). Mutations were detected after mapping on the *Mmm T1/44* genome (CP014346.1). Illumina reads were trimmed using Trimmomatic (V 0.38.1; Sliding Window 10, 20 ; Drop read below Minimal length 250), mapped using BWA-MEM (V 0.7.17.1), Samtools sort (V 2.0.3), MPileup (V 2.1.1), and variants were detected using VarScan mpileup (V 2.4.3.1; Minimum coverage 30, Minimum supporting read 20, Minimum Base quality 30, Minimum variant allele frequency 0,8, Minimum homozygous variants 0,75). Mutations are showed in Table S3. Genome assembly was performed using the following steps: ONT reads were filtered using Filter FASTQ (V 1.1.5, Minimum size 45,000 bp), assembled using Flye Assembly (V 2.6), polished using 4 rounds of Pilon (1.20.1) combined with Illumina short reads. Assembled genome was compared to the *Mmm T1/44* (CP014346.1) reference genome using MAUVE software (64).

For the *Mgal_ksgA* mutant, the same procedure was used with 45,406 ONT long-reads (Mean read length: 23,295 bp) and 851,806 Illumina short reads pairs using *Mgal_S6* reference genome (CP006916.3). For detection of SNV in *Mgal*, results were obtained by comparison with a previously sequenced clone (accession number PRJNA769398) to eliminate mutations that are present in the laboratory clone.

For the *Mbov_WT*, the same procedure was used with 28,729 ONT long-reads (Mean read length: 27,460 bp) and 881,705 Illumina short reads pairs. Mutations of our lab strain were detected after mapping *Mbov_PG45* reference genome (CP002188.1).

For the *Mbov_mnuA* mutant cl3, the same procedure was used with 42,506 ONT long-reads (Mean read length: 24,459 bp) and 910,742 Illumina short reads pairs. Mutations were detected after mapping using the *Mbov_PG45* reference genome (CP002188.1). For detection of SNV in *Mbov* linked to our CBE, the genome sequence of the *Mbov_WT* was used to eliminate mutations that are present in the laboratory strain.

For the four *Mmm_IS* mutants the same procedure was used. *Mmm_IS3* mutant cl 5.1.18 was sequenced with 54,523 ONT long-reads (Mean read length: 22,424 bp) and 941,944 Illumina short reads pairs. *Mmm_IS3* mutant cl 4.1.2 was sequenced with 57,466 ONT long-reads (Mean read length: 20,074 bp) and 778,172 Illumina short reads pairs. *Mmm_IS1634* mutant cl 3.1.11 was sequenced with 67,316 ONT long-reads (Mean read length: 22,647 bp) and 724,029 Illumina short reads pairs. *Mmm_IS1634* mutant cl 3.1.6 was sequenced with 31,875 ONT long-reads (Mean read length: 24,551

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

bp) and 1,199,725 Illumina short reads pairs. For those four clones, mutations were detected after mapping using *Mmm* T1/44 reference genome (CP014346.1).

SNP analyses

SNP found in all sequenced clones were analyzed manually. A summary can be found in Table S3. Blast of the sgRNA sequence against DNA region around the SNP was performed. Two conditions were established to consider SNV as sgRNA dependent off-target: (1) Presence of PAM sequence; (2) 70% of sequence similarity on the first 12 nucleotides (considered as Seed-sequence).

Phenotypic assay in *Mgal*: kasugamycin resistance

Mgal S6 and *Mgal_ ksgA* mutant were cultured during 36h. Serial dilution were performed until 10^{-7} . 20 μ L of each dilution were plated in Hayflick medium with or without kasugamycin and incubated for 10 days at 37°C (5% CO₂).

Phenotypic assay in *Mbov*: nuclease activity

The nuclease activity assays on WT *M. bovis* PG45 and *Mbov_mnuA* mutant were performed as described previously in Sharma *et al.* 2015 (46) with some modifications. Briefly, for each strain, 2 mL of culture were prepared in SP4 medium. At late log phase, cultures were divided in 2x 1 mL and centrifuged 10 min. at 7,000 x g, 10°C. A 1 mL sample was mixed with in 500 μ L of nuclease buffer (25 mM Tris-HCl, pH 8.8, 10 mM CaCl₂, 10 mM MgCl₂) and another 1 mL sample was mixed with 100 μ L of nuclease buffer in order to concentrate the mycoplasma cells (2 and 10 times respectively). Then, 2 μ g of circular plasmid DNA (pMT85_SpCas9_pmcDA1) or 500 ng of double stranded linear DNA (PCR fragment) were incubated with 50 μ L of each sample at 37°C during 5 or 60 minutes. At each time, aliquots of 10 μ L were removed and the reaction was stopped by addition of EDTA to a final concentration of 20 mM. Aliquots were mixed with 6X loading buffer (Promega) and immediately loaded on 1% agarose gels made in 1X TAE buffer.

Phenotypic assay in *Mmm*: H₂O₂ production

Mmm T1/44 WT and *glpO* mutants were grown until pH=7. A drop of 20 μ L of each culture was then loaded on PPLO media plate and incubated for 48 hours. GlpO activity was evaluated using a qualitative "on-the-plate" H₂O₂ test that allows to detect the production of H₂O₂ in response to the addition of a glycerol-containing reaction mix (65). Briefly, 40 μ L of glycerol reaction mix (H₂O qsp 1,6 mL + DAB (3,3'-Diaminobenzidine ; 9,6 mg) + 64 μ L HCl (1N) + horseradish peroxidase (1,6 mg) + 64 μ L of PDS stock (CaCl₂ (1.5 mg.mL⁻¹), KCl (1.5 mg.mL⁻¹), NaHCO₃ (0.625 mg.mL⁻¹), NaCl (56.25 mg.mL⁻¹)) are spread carefully onto bacterial lawns and colored reaction are observed after 30 min and 2 hours of incubation at 37°C. *Mycoplasma* colonies producing H₂O₂ adopt a red-brown colour.

Acknowledgements

Genome sequencing was performed by the Genome Transcriptome Facility of Bordeaux (<https://pgtb.cgfb.u-bordeaux.fr>, Grants from Investissements d'avenir, Convention attributive d'aide EquipEx Xyloforest ANR-10-EQPX-16-01).

References

1. M. May, M. F. Balish, A. Blanchard, *The Order Mycoplasmatales* (2014) https://doi.org/https://doi.org/10.1007/978-3-642-30120-9_289.
2. P. Sirand-Pugnet, C. Citti, A. Barré, A. Blanchard, Evolution of mollicutes: down a bumpy road with twists and turns. *Res. Microbiol.* **158**, 754–766 (2007).
3. S. Razin, D. Yogev, Y. Naot, Molecular biology and pathogenicity of Mycoplasmas. *Microbiol. Mol. Biol. Rev.* **165**, 199–221 (1998).
4. C. Citti, E. Dordet-Frisoni, L. X. Nouvel, C. H. Kuo, E. Baranowski, Horizontal gene transfers in mycoplasmas (Mollicutes). *Curr. Issues Mol. Biol.* **29**, 3–22 (2018).
5. H. Grosjean, *et al.*, Predicting the Minimal Translation Apparatus: Lessons from the Reductive Evolution of Mollicutes. *PLoS Genet.* **10** (2014).

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

6. C. Citti, A. Blanchard, Mycoplasmas and their host: Emerging and re-emerging minimal pathogens. *Trends Microbiol.* **21**, 196–203 (2013).
7. R. Rosengarten, *et al.*, The changing image of mycoplasmas: from innocent bystanders to emerging and reemerging pathogens in human and animal diseases. *Contrib. Microbiol.* **8**, 166–185 (2001).
8. S. L. Hennigan, *et al.*, Detection and differentiation of avian mycoplasmas by surface-enhanced raman spectroscopy based on a silver nanorod array. *Appl. Environ. Microbiol.* **78**, 1930–1935 (2012).
9. E. D. Peebles, S. L. Branton, *Mycoplasma gallisepticum* in the commercial egg-laying hen: A historical perspective considering the effects of pathogen strain, age of the bird at inoculation, and diet on performance and physiology. *J. Appl. Poult. Res.* **21**, 897–914 (2012).
10. CABl, Avian mycoplasmosis (2019).
11. N. F. Delaney, *et al.*, Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS Genet.* **8** (2012).
12. K. Dudek, E. Szacawa, *Mycoplasma bovis* infections: Occurrence, pathogenesis, diagnosis and control, including prevention and therapy. *Pathogens* **9**, 1–3 (2020).
13. A. G. Jordan, *et al.*, *Mycoplasma bovis* outbreak in New Zealand cattle: An assessment of transmission trends using surveillance data (2020) <https://doi.org/10.1111/tbed.13941>.
14. G. Di Teodoro, *et al.*, Contagious Bovine Pleuropneumonia: A Comprehensive Overview. *Vet. Pathol.* **57**, 476–489 (2020).
15. A. V. Gautier-Bouchardon, Antimicrobial Resistance in *Mycoplasma* spp. *Microbiol. Spectr.* **6** (2018).
16. M. Ishfaq, *et al.*, Current status of vaccine research, development, and challenges of vaccines for *Mycoplasma gallisepticum*. *Poult. Sci.* **99**, 4195–4202 (2020).
17. E. Ruiz, *et al.*, CReasPy-Cloning: A Method for Simultaneous Cloning and Engineering of Megabase-Sized Genomes in Yeast Using the CRISPR-Cas9 System. *ACS Synth. Biol.* (2019) <https://doi.org/10.1021/acssynbio.9b00224>.
18. C. Lartigue, *et al.*, Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science (80-)*. **325**, 1693–1696 (2009).
19. F. Labroussaa, *et al.*, Impact of donor-recipient phylogenetic distance on bacterial genome transplantation. *Nucleic Acids Res.* **44**, 8501–8511 (2016).
20. S. W. Lee, G. F. Browning, P. F. Markham, Development of a replicable oriC plasmid for *Mycoplasma gallisepticum* and *Mycoplasma imitans*, and gene disruption through homologous recombination in *M. gallisepticum*. *Microbiology* (2008) <https://doi.org/10.1099/mic.0.2008/019208-0>.
21. R. Chopra-Dewasthaly, M. Marena, R. Rosengarten, W. Jechlinger, C. Citti, Construction of the first shuttle vectors for gene cloning and homologous recombination in *Mycoplasma agalactiae*. *FEMS Microbiol. Lett.* **253**, 89–94 (2005).
22. C. Janis, *et al.*, Versatile Use of oriC Plasmids for Functional Genomics of *Mycoplasma capricolum* subsp. *capricolum*†. *Society* **71**, 2888–2893 (2005).
23. M. Jinek, *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-)*. **337**, 816–821 (2012).
24. C. Piñero-Lambea, *et al.*, *Mycoplasma pneumoniae* Genome Editing Based on Oligo Recombineering and Cas9-Mediated Counterselection. *ACS Synth. Biol.* **9**, 1693–1704 (2020).
25. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
26. K. Nishida, *et al.*, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science (80-)*. **353** (2016).
27. N. M. Gaudelli, *et al.*, Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
28. A. C. Komor, *et al.*, Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to:T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, 1–10 (2017).
29. A. V. Anzalone, L. W. Koblan, D. R. Liu, Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
30. Y. Wu, *et al.*, Increasing cytosine base editing scope and efficiency with engineered Cas9-PMCD1 fusions and the modified sgRNA in rice. *Front. Genet.* **10**, 1–10 (2019).
31. H. S. Kim, Y. K. Jeong, J. K. Hur, J. S. Kim, S. Bae, Adenine base editors catalyze cytosine conversions in human cells. *Nat. Biotechnol.* **37**, 1145–1148 (2019).
32. Y. Wang, *et al.*, CRISPR-Cas9 and CRISPR-Assisted Cytidine Deaminase Enable Precise and Efficient Genome Editing in *Klebsiella pneumoniae*. *Appl. Environ. Microbiol.* **84**, 1–15 (2018).
33. J. Sun, L. B. Lu, T. X. Liang, L. R. Yang, J. P. Wu, CRISPR-Assisted Multiplex Base Editing System in *Pseudomonas putida* KT2440. *Front. Bioeng. Biotechnol.* **8**, 1–13 (2020).
34. Q. Li, *et al.*, CRISPR–Cas9 D10A nickase-assisted base editing in the solvent producer *Clostridium beijerinckii*. *Biotechnol. Bioeng.* **116**, 1475–1483 (2019).
35. Y. Tong, *et al.*, Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20366–20375 (2019).
36. S. D. Rodrigues, *et al.*, Efficient CRISPR-mediated base editing in agrobacterium spp. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–8 (2021).

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

37. S. Banno, K. Nishida, T. Arazoe, H. Mitsunobu, A. Kondo, Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **3**, 423–429 (2018).
38. S. J. Bae, B. G. Park, B. G. Kim, J. S. Hahn, Multiplex Gene Disruption by Targeted Base Editing of *Yarrowia lipolytica* Genome Using Cytidine Deaminase Combined with the CRISPR/Cas9 System. *Biotechnol. J.* **15**, 1–46 (2019).
39. G. G. Mahairas, F. C. Minion, Random Insertion of the Gentamicin Resistance Transposon Tn4001 in *Mycoplasma pulmonis*. *Plasmid* (1989).
40. A. M. Mariscal, *et al.*, Tuning Gene Activity by Inducible and Targeted Regulation of Gene Expression in Minimal Bacterial Cells. *ACS Synth. Biol.* **7**, 1538–1552 (2018).
41. T. P. Huang, G. A. Newby, D. R. Liu, Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* **16**, 1089–1128 (2021).
42. L. Papazisi, *et al.*, GapA and CrmA coexpression is essential for *Mycoplasma gallisepticum* cytoadherence and virulence. *Infect. Immun.* **70**, 6839–6845 (2002).
43. S. P. Mugunthan, M. C. Harish, Multi-epitope-Based Vaccine Designed by Targeting Cytoadherence Proteins of *Mycoplasma gallisepticum*. *ACS Omega* **6**, 13742–13755 (2021).
44. I. Cizelj, *et al.*, *Mycoplasma gallisepticum* and *Mycoplasma synoviae* express a cysteine protease CysP, which can cleave chicken IgG into Fab and Fc. *Microbiology* **157**, 362–372 (2011).
45. F. Mitiku, *et al.*, The major membrane nuclease MnuA degrades neutrophil extracellular traps induced by *Mycoplasma bovis*. *Vet. Microbiol.* **218**, 13–19 (2018).
46. S. Sharma, K. A. Tivendale, P. F. Markham, G. F. Browning, Disruption of the membrane nuclease gene (MBOVPG45_0215) of *Mycoplasma bovis* greatly reduces cellular nuclease activity. *J. Bacteriol.* **197**, 1549–1558 (2015).
47. P. Pilo, *et al.*, A metabolic enzyme as a primary virulence factor of *Mycoplasma mycoides* subspecies *mycoides* small colony. *J. Bacteriol.* (2005) <https://doi.org/10.1128/JB.187.19.6824-6831.2005>.
48. Y. Yu, *et al.*, Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* **11**, 1–10 (2020).
49. Z. Zhong, *et al.*, Base editing in *Streptomyces* with Cas9-deaminase fusions. *BioRxiv*, 630137 (2019).
50. Y. Wang, *et al.*, MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method. *Metab. Eng.* **47**, 200–210 (2018).
51. S. Yu, *et al.*, CRISPR-dCas9 mediated cytosine deaminase base editing in *Bacillus subtilis*. *ACS Synth. Biol.* (2020) <https://doi.org/10.1021/acssynbio.0c00151>.
52. A. M. Mariscal, L. González-González, E. Querol, J. Piñol, All-in-one construct for genome engineering using Cre-lox technology. *DNA Res.* **23**, 263–270 (2016).
53. C. Janis, *et al.*, Unmarked insertional mutagenesis in the bovine pathogen *Mycoplasma mycoides* subsp. *mycoides* SC. *Microbiology* **154**, 2427–2436 (2008).
54. C. Lartigue, A. Blanchard, J. Renaudin, F. Thiaucourt, P. Sirand-Pugnet, Host specificity of mollicutes oriC plasmids: Functional analysis of replication origin. *Nucleic Acids Res.* **31**, 6610–6618 (2003).
55. M. Evanoff, A. C. Komor, Base editors: Modular tools for the introduction of point mutations in living cells. *Emerg. Top. Life Sci.* **3**, 483–491 (2019).
56. J. Grünewald, *et al.*, A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat. Biotechnol.* **38**, 861–864 (2020).
57. E. A. Freundt, “CULTURE MEDIA FOR CLASSIC MYCOPLASMAS” in *Methods in Mycoplasmaology*, (1983) <https://doi.org/10.1016/b978-0-12-583801-6.50029-9>.
58. M. A. Algire, *et al.*, New selectable marker for manipulating the simple genomes of *Mycoplasma* species. *Antimicrob. Agents Chemother.* **53**, 4429–4432 (2009).
59. M. Breton, *et al.*, First report of a tetracycline-inducible gene expression system for mollicutes. *Microbiology* **156**, 198–205 (2010).
60. Clyde A. Hutchison, *et al.*, Global Transposon Mutagenesis and a Minimal *Mycoplasma* Genome. *Science (80-.)*. **286** (1999).
61. K. W. King, K. Dybvig, Transformation of *Mycoplasma capricolum* and examination of DNA restriction modification *Mycoplasma capricolum* and *Mycoplasma mycoides* subsp. *mycoides*. *Plasmid* **31**, 308–311 (1994).
62. X. Zhu, *et al.*, Mbov_0503 encodes a novel cytoadhesin that facilitates *Mycoplasma bovis* interaction with tight junctions. *Microorganisms* **8** (2020).
63. M. G. Kluesner, *et al.*, EditR: A Method to Quantify Base Editing from Sanger Sequencing. *Cris. J.* **1**, 239–250 (2018).
64. A. C. E. Darling, B. Mau, F. R. Blattner, N. T. Perna, Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
65. P. Rice, B. M. Houshaymi, E. A. M. Abu-Groun, R. A. J. Nicholas, R. J. Miles, Rapid screening of H₂O₂ production by *Mycoplasma mycoides* and differentiation of European subsp. *mycoides* SC (small colony) isolates. *Vet. Microbiol.* **78**, 343–351 (2001).

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

3. Données supplémentaires de la publication

Mutant strain	<i>Mgal_ksqA</i>	<i>Mbov_mnuA</i>	<i>Mmm_glpO</i>	<i>Mmm_IS1634_3.1.6</i>	<i>Mmm_IS1634_3.1.11</i>	<i>Mmm_IS3_4.1.2</i>	<i>Mmm_IS3_5.1.18</i>
Edited sites/targeted sites	2/2	2/2	3/3	54/55	52/55	22/22	22/25
1- sgRNA-guided off-target mutation	2	1	2	1	0	5	4
2- Spurious deamination	12	4	1	16	11	24	13
TC motifs on spurious deamination sites	9/12	2/4	1/1	14/16	7/11	14/24	10/13
3- Other mutation	14	6	2	2	3	7	7
Total number of mutations	28	11	5	19	14	36	24

Reference genome	Localisation	Reference	Alternative	Gene	Codon	Amino-Acid	Sequence	
Wanted mutation								
	10130	C	T	GCW_RS00055: KsqA				Tot=2
	10131	C	T	GCW_RS00055: KsqA				
sgRNA guided off-targets: TGACCAAAATATTATAAAGG								
	1313	C	T	Intergenic	-	-	TAAGCAATATATAATAATACGG	Tot=2
	705465	G	A	GCW_RS02960: PdxIFFG protein	GCA-ACA	Ala-Thr	TGATGCAACTAATATAAAGG	
Spurious deamination								
	55373	C	T	GCW_RS00215: PstA	ACA-ATA	Thr-Ile	TTACATTATTTGTACGTTT	Tot=12 TC=9/12
	55583	C	T	GCW_RS00215: PstA	TCA-TTA	Ser-Leu	CTTCATGAAATCTGATTACA	
	275592	G	A	GCW_RS01220: Hypothetical	TGC-TAC	Cys-Tyr	GTGCAAGACGACAAAATAT	
	387640	C	T	IG	-	-	TTTCATATTAAGAAATG	
	539449	G	A	GCW_RS02315: hypothetical protein	TTC-TTT	Phe	TCATTATTAGTAAAAAATG	
	609868	C	T	GCW_RS02580: arginine deiminase	GAT-AAT	Asp-Asn	CATCTAAAACCTAACTGCTAAG	
	622663	G	A	GCW_RS02635: ribonuclease R	CCA-TCA	Pro-Ser	GAACCAACCTTTAGTCAAGAG	
	639577	G	A	IG	-	-	TATCTAAATGATTGATTT	
	656814	G	A	GCW_RS02775: Cation transport protein	CTG-TTG	Leu	ATTCTGATTGAAGTTGGGAGT	
	663798	G	A	GCW_RS02800: truncated protein	-	-	TCCCTAATCTGAAAAATCTACGG	
	845264	C	T	GCW_RS03495: YitI family protein	GAA-AAA	Glu-Lys	TTTCACAGCTGTTGAGTAGA	
	858949	C	T	threonine kinase	GAT-AAT	Asp-Asn	AATCATTGAGTTGGTTGAATCTG	
Others mutation								
	136847	GAAAAACA	G	IG	-	-		
	259371	CGAAGAAGAAGAA	C	Promotor: GCW_RS01185 VlhA	-	-		
	464694	CGAAGAA	C	Promotor: GCW_RS01965 VlhA	-	-		
	486678	T	TAAC	GCW_RS02080: ABC transporter ATP-binding protein	-	-		
	586131	GAGTTGT	G	GCW_RS02465: VlhA	-	-		
	586403	TTC	T	Promotor: GCW_RS02465 VlhA	-	-		
	804250	A	ACTT	Promotor: GCW_RS03335 VlhA	-	-		
	10132	A	T	GCW_RS00055: KsqA	TAA-TTA	STOP-leu		
	107452	G	C	GCW_RS00475: elongation factor Ts	TCT-TGT	Ser-Cys		
	198639	G	T	GCW_RS00920: methionine--tRNA ligase	CAC-CAA	His-Glu		
	252513	C	A	GCW_RS01165: amino acid permease	ACG-ATG	Thr-Met		
	442414	T	A	Promotor: GCW_RS01895	-	-		
	588853	T	G	GCW_RS02465: hypothetical protein	ACA-TCA	Thr-Ser		
	622188	G	T	IG	-	-		

Reference genome	Localisation	Reference	Alternative	Gene	Codon	Amino-Acid	Sequence	
Wanted mutation								
	187515	C	T	MBOVPG45_0215: MnuA				Tot=2
	187516	C	T	MBOVPG45_0215: MnuA				
sgRNA guided off-targets: ACATTACAATTTGGTCTTCAGG								
	974237	G	A	MBOVPG45_0405: MutM	CTG-CTA	Leu	AAATGTAACCTTGTACTAACAGG	Tot=1
Spurious deamination								
	112595	G	A	MBOVPG45_0957: 23S ribosomal RNA	-	-	CCACGCTTCCTAGAAAAGTT	Tot=4
	181758	G	A	MBOVPG45_0955: tRNA Gly	-	-	GTTCAATGGTAGAACACAG	
	488806	C	T	MBOVPG45_0825: lipoprotein	GAT-AAT	Asp-Asn	CATCAATAAAAAGATTITTC	
	553569	C	T	MBOVPG45_0765: truncated protein	-	-	AAACCACATAATCGCGACTA	
Others mutation								
	232528	AC	A	MBOVPG45_0178: truncated protein	-	-		Tot=6
	244083	C	CGA	Promoter: MBOVPG45_0168 type III restriction-modification system	-	-		
	574591	GC	G	MBOVPG45_0242: IS transposase	-	-		
	577563	G	GC	MBOVPG45_0743: truncated protein	-	-		
	580909	A	AC	MBOVPG45_0740: truncated protein	-	-		
	867346	A	AC	MBOVPG45_0498: truncated protein	-	-		

Reference genome	Localisation	Reference	Alternative	Gene	Codon	Amino-Acid	Sequence	
Wanted mutation								
	1029926	C	T	GlpO	CAA-TAA	Glu-STOP		Tot=3
	1029929	C	T	GlpO	CAA-TAA	Glu-STOP		
	1029932	C	T	GlpO	CAA-TAA	Glu-STOP		
sgRNA guided off-targets: CAACAACAATACGATAACATTGG								
	220532	C	T	MSCT144_02030: UDP-glucose 4-epimerase ctt-ttt galE	Ctt-ttt	Leu-Phe	TGTAATCTTTACTATAACATTGG	Tot=2
	228315	C	T	MSCT144_02090: UDP-glucose 4-epimerase ctt-ttt galE	Ctt-ttt	Leu-Phe	TGTAATCTTTACTATAACATTGG	
Spurious deamination								
	1109604	C	T	MSCT144_10690: ABC transporter ATP-binding protein	GTT-ATT	Val-Ile	CTTCTATTAGAAATATTAGT	Tot=1
Others mutation								
	423841	TA	T	Promoter ISMmy: MSCT144_03990	-	-		Tot=2
	1190819	A	AC	Intergenic	-	-		

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

Table S3d. Mutations detected in Mmm_IS1634_3.1.6 edited genome after WGS						
Reference genome	Localisation	Reference	Alternative	Gene	Sequence	
sgRNA guided off-targets: AGACCAGATTGTTATAGGTATGG						
Mmm_T1/44	347732	G	A	PtsG: MSCT144_03240	AGCAATTATTGATATAGTTAAGG	Tot=1
	Spurious deamination					
	12705	G	A	sugar transporter: MSCT144_00090	AATCCAATAGGAATCATTGC	Tot=16 TC=14/16
	87643	C	T	16S ribosomal RNA: MSCT144_00760	CTTTCGTCATTGTGAAAAA	
	179131	C	T	rpoC: MSCT144_01620	CTTCTAATAATCTATTAGCT	
	186427	G	A	VlhA: MSCT144_01670	TTCTAATGTAATAGATATGG	
	194514	C	T	Transglutaminase: MSCT144_01750	CATCTTAAATGACATAGTAT	
	227160	C	T	Promoter: glycosyltransferase: MSCT144_01860	AATCAGTCAGATGATAGTAA	
	294349	G	A	IS1634: MSCT144_02680	ATTCCITTTTCATCAGTTGC	
	354549	C	T	Promoter nrdF: MSCT144_03310	TATCAAAITTTAATAATAATGAG	
	361261	C	T	purB: MSCT144_03380	TCAAAGTTAAGAAATCACAG	
	395743	C	T	hypothetical: MSCT144_03770	CTTCATTACTAATATAAA	
	407045	C	T	VlhA: MSCT144_03850	AATCATAATTTATCATCATGAGA	
	602028	C	A	Virulence signal recognition: MSCT144_05830	ACCCACTGTGCTTTAAAAAAT	
	782061	G	A	smc3: MSCT144_07480	TATCTTTTATTTGTTGATCA	
	839757	C	T	rnhB: MSCT144_07950	TAACCTTTATGAGATTTAAA	
	860886	C	T	der: MSCT144_08210	TATCATATTTTATAATTGCT	
	1018982	G	A	pdhB: MSCT144_09740	TTATCAAGGTGTACAGGTTATGA	
	Others mutation					
423841	TA	T	Promoter ISMmy: MSCT144_03990		Tot=2	
626488	T	TA	HrcA: MSCT144_05990			

Table S3e. Mutations detected in Mmm_IS1634_3.1.11 edited genome after WGS							
Reference genome	Localisation	Reference	Alternative	Gene	Sequence		
sgRNA guided off-targets: AGACCAGATTGTTATAGGTATGG							
-							
Spurious deamination							
Mmm_T1/44	43037	G	A	ftsX: MSCT144_00340	TATCAGTTATTAGCATAATT	Tot=11 TC=7/11	
	107097	C	T	secA: MSCT144_00940	ATCCATTTTTCTTTATTGT		
	112963	C	T	hypothetical: MSCT144_00990	AATCATATTCAGAAATAT		
	260537	C	T	hypothetical: MSCT144_02350	AACCTATAAAAAACATAATA		
	591883	G	A	Intergenic	GATCAATATATTAATTTAAA		
	689817	G	A	nanK: MSCT144_06640	CATCTTTTATTGGAATGTT		
	764572	C	T	cutC: MSCT144_07300	GCTCTATTAAGTAATTTT		
	869142	C	T	hypothetical: MSCT144_08280	ATTATCTAGAGTTATTATTTGA		
	886219	C	T	hypothetical: MSCT144_08480	CTACAAATAATTATAAATAA		
	1077987	G	A	Intergenic	ATCAAAAAGTACTTGAAATT		
	1172420	G	A	glcB: MSCT144_11280	TTTACCAATTGCAGCTTTATTAG		
	Others mutation						
	411886	G	GA	cof_1: MSCT144_04020			Tot=3
626488	T	TA	HrcA: MSCT144_05990				
1088107	T	TA	Promoter hypothetical protein: MSCT144_10520				

CHAPITRE 4 : Construction d'un outil universel pour des bactéries à faible efficacité de recombinaison : les systèmes Base-Editor

Table S3f. Mutations detected in Mmm_IS3_4.1.2 edited genome after WGS					
Reference genome	Localisation	Reference	Alternative	Gene	Sequence
Mmm_T1/44	sgRNA guided off-targets: ACAAGTGGAACTACTATAAGTTGG				
	755798	G	A	lpp: MSCT144_07230	TATCGTAATTATAAGTTGGA
	852246	C	T	Promoter recU/gpsB	CATCTTTTTCTAACTACTATAATAAGG
	852249	C	T	Promoter recU/gpsB	CATCTTTTTCTAACTACTATAATAAGG
	852255	C	T	Promoter recU/gpsB	CATCTTTTTCTAACTACTATAATAAGG
	1085088	G	A	Intergenic	AACTTGGAATACTATTGGTAGG
	Spurious deamination				
	30920	G	A	rpsF: MSCT144_00280	CCATTTATTAAGAGAAGAAG
	79992	C	T	Intergenic	CTCATAAGTTATGTTTTTATAA
	79994	C	T	Intergenic	CTCATAAGTTATGTTTTTATAA
	171467	G	A	Intergenic	TTTTCTAACTAATATAGG
	185504	C	T	rplJ: MSCT144_01660	TATCAACAACCTTACCTTCA
	327516	G	A	atpD: MSCT144_03030	ATCCATACCTAAAATTGCAATGA
	383347	C	T	pdp: MSCT144_03580	GTAACCTCACATTTTCATCCTGG
	412720	G	A	p37: MSCT144_03910	CCTCATCTTGGTTTTTAAT
	430810	G	A	ABC: MSCT144_04060	CAATTTAATTAGACTTGCTTGA
	449546	G	A	VlhA: MSCT144_04210	CAGCAACTCATCCTAATTTA
	451269	C	T	Intergenic	AATCATAATAAGATCAATT
	505056	C	T	hypothetical: MSCT144_04900	AAACCAAATACAAATAATAA
	604117	C	T	lpp: MSCT144_05840	CATCATTAACTAGATTAAT
	622242	G	A	Pyrophosphatase: MSCT144_05950	AACCTTTAGGAATAAAAATA
	623615	G	A	YidA: MSCT144_05970	TATCATTATGATTATCTCCA
	648639	G	A	FtsX: MSCT144_06170	ATATCTGAAATATTAGAACTAGG
	685555	C	T	Intergenic	TCACCACATAGGCGAGTCTAGA
	703516	C	T	pdhD: MSCT144_06760	TATCACCAATTGCATAAATA
	765868	C	T	metK: MSCT144_07310	CATCTCTTGCAACTTGTTC
	778493	G	A	hypothetical: MSCT144_07460	ATATCTTTGAACTATATTTTGG
	991514	C	T	engB: MSCT144_09520	CTTCATCAGCAATAAAAAGC
	1033238	G	A	YrrK: MSCT144_09850	CAGCTAATTTAGAACTATA
	1042195	C	T	Intergenic	TATCTACATTATCTGTATT
	1071422	C	T	Promoter defB : MSCT144_10280	CATCCATATTATAGCAAAATAGA
	Others mutation				
	748443	C	CT	Promoter dUTPase :MSCT144_07160	
	1190819	A	AC	Intergenic	
	351768	T	C	smpB	
	540973	T	C	MutM	
	627026	T	C	hrcA	
	789929	A	G	Hypothetical	
	843507	T	C	rlmH	
	Tot=24 TC=14/24				

Table S3g. Mutations detected in Mmm_IS3_5.1.18 edited genome after WGS					
Reference ge	Localisation	Reference	Alternative	Gene	Sequence
Mmm_T1/44	sgRNA guided off-targets: CATATAAAAACCCATTTCCTGG				
	193168	C	T	Close IS3	GGACATATAAAAACCCATT
	193217	G	A	Close IS3	CCCCCGAACCCCATATTCGGAC
	193219	G	A	Close IS3	CCCCCGAACCCCATATTCGGAC
	193220	G	A	Close IS3	CCCCCGAACCCCATATTCGGAC
	Spurious deamination				
	1523	G	A	Intergenic	TTCTATTTCTATTTCTAAAT
	45788	C	T	Permease: MSCT144_00360	CAACTCCACCAGCAACAGTTGTAG
	154903	C	T	IS1634: MSCT144_08700	AATCATAATACATCATTTTT
	339948	G	A	Intergenic	TATCTTATTTTATAATCATT
	407651	C	T	hypothetical: MSCT144_03860	TACTAGAGAAGACAAAGATTGA
	451269	C	T	Intergenic	AATCATAATAAGATCAATT
	473936	C	T	MalR: MSCT144_04490	AATCTAGTACAAAGAATAAATGA
	680010	C	T	Hypothetical: MSCT144_06490	GATCTTATACAAATATCAAA
	747201	C	T	treB: MSCT144_04500	TCTCGCCTTAGATTAGTTTT
	827223	C	T	nadD: MSCT144_07800	GATCAAACTACCACCAAT
	1003678	C	T	thil: MSCT144_09610	TTATCATAAGTAATTACAGG
	1093554	G	A	hypothetical: MSCT144_10560	GAACCATTTCTGGTTTATAAGG
	1096091	C	T	fic: MSCT144_10590	TATCATCTTGACTAATTTTT
	Others mutation				
	748443	C	CT	Promoter dUTPase :MSCT144_07160	
	351768	T	C	ssrA	
	540973	T	C	mutM	
	627026	T	C	hrcA	
	679866	C	A	Hypothetical	
789929	A	G	hypothetical		
843507	T	C	rlmH		
Tot=13 TC=10/13					

4. Données et discussions supplémentaires non publiées

Cette dernière étude répond de manière directe à la nécessité d'un outil universel permettant la modification précise de l'ADN pour de nombreuses espèces de mycoplasmes. Les systèmes de bases éditions sont les outils chirurgicaux parfaits pour une expérimentation rapide et facile d'emploi. Les nombreuses mises en évidence de leur efficacité dans notre étude sont autant de preuves de concept importantes pour l'ingénierie génomique de trois espèces majeures de mycoplasmes pathogènes d'animaux d'élevage. Cet outil permet donc l'inactivation rapide de gènes d'intérêts sous condition de la présence d'une séquence cible dans celui-ci. Ceci ouvre ainsi la porte pour des études fonctionnelles chez ces trois mycoplasmes, mais aussi vraisemblablement chez d'autres espèces plus ou moins proches.

Au cours de ce travail, des essais ont également été réalisés avec la Cas9 endogène de *Mgal* comme système d'édition de base. Ce remplacement de la composante Cas9 du système CBE permettrait, par exemple, d'étendre grandement le nombre de cibles potentielles que l'on trouverait dans un génome de mycoplasme. Des résultats très variables ont été obtenus lors de nos essais, avec une cible sur cinq montrant une désamination effective. Le système actuel ne semble donc pas assez efficace pour induire l'édition de base. Une alternative pour augmenter le nombre de cibles possibles serait l'utilisation des variants de Cas9 tel que le variant VQR spécifique des PAM NGA qui est décrite comme fonctionnelle en système de base édition et qui nécessite peu de modifications sur la séquence de spCas9 (Anzalone, Koblan and Liu, 2020). Les systèmes d'adénine désaminase pourront également agrandir la boîte à outils pour de l'édition de base chez les mycoplasmes (Gaudelli *et al.*, 2017).

Finalement au laboratoire, cet outil est d'ores et déjà en pleine application chez plusieurs espèces, dont *M. agalactiae* chez qui il est également fonctionnel. Ceci clôture la partie développement d'outils génétiques de mon travail et répond de manière satisfaisante aux objectifs qui fixés au début de ma thèse.

DISCUSSION GÉNÉRALE ET CONCLUSION

1. DISCUSSION GENERALE

Ma thèse s'est articulée autour de deux orientations distinctes mais liées :

- L'étude des systèmes CRISPR chez les mollicutes et plus spécifiquement chez *M. gallisepticum*
- Le développement d'outils d'ingénierie de génome chez les mycoplasmes

Afin d'apporter un nouvel éclairage sur mon travail, je vais revenir, dans une première partie de cette discussion, sur les systèmes de défenses de mycoplasmes et leurs probables rôles dans un environnement naturel en me basant principalement sur *Mgal* mais en ouvrant également le propos à d'autres espèces. Cette discussion présente une vision générale et pose un certain nombre d'hypothèses qui pourraient faire l'objet de futurs axes de recherche. Dans une seconde partie, je m'efforcerais de décrire une méthodologie possible dans l'optique de réaliser de l'ingénierie de génome chez les différentes espèces de mycoplasmes.

1. Les systèmes de défense chez les mycoplasmes

Malgré une évolution marquée par des pertes importantes de gènes, partiellement compensée par des acquisitions issues de transferts horizontaux, il semble que bon nombre de mollicutes ont conservé un système CRISPR dont la fonction primaire est la défense contre l'invasion par des ADN exogènes. En effet, de nombreuses espèces possèdent encore ce système et on observe, chez certaines espèces ou souches particulières, une dégradation du système voire même son élimination totale qui peut être interprétée comme une diminution de la pression évolutive sur ce système (Delaney *et al.*, 2012; Ipoutcha *et al.*, 2019).

L'analyse approfondie du système CRISPR de *Mgal*, nous a permis de voir à quel point le système est dynamique chez ce mycoplasme. Chez les souches de rosélin, on constate : (i) le changement de spécificité de Cas9 via l'évolution de sa séquence protéique et (ii) l'inactivation du système chez les souches post-2007 avec une perte totale du système ou une mutation dans le gène *cas9* qui entraîne l'apparition d'un codon-stop. Ceci tend à montrer un système labile, dont l'évolution peut être rapide, au moins dans certains contextes de pression évolutive. De ce point de vue, il est intéressant de noter que, malgré l'inactivité probable du système dans toutes les souches de rosélin (HF) isolées après

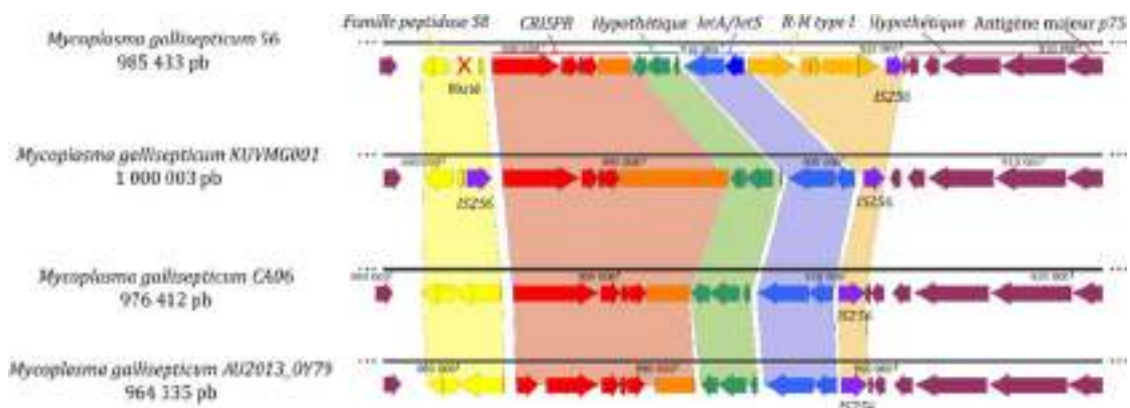


Figure 22: Ilot de défense chez *Mycoplasma gallisepticum*. Représentation de la région autour du locus CRISPR dans le génome de différentes souches représentatives de *Mgal*. Une ancienne souche de poulet (S6), une récente souche de poulet (KUVMG), une souche de roselin près-2007 (CA06) et une souche de roselin post-2007 (AU2013_OY79).

DISCUSSION GÉNÉRALE ET CONCLUSION : Discussion générale

2007, la non-dégradation du locus chez la plupart des isolats laisse à penser qu'il pourrait être, malgré tout, soumis à une pression de sélection pour être maintenu « en attente de réactivation » par sélection purifiante s'il venait à être à nouveau essentiel. Chez les souches de poulet, le système semble persister à l'état actif, comme en témoigne l'analyse du génome d'une souche de Corée du Sud isolée de poulet en 2019 (KUVMG). Une étude plus large des souches circulant actuellement dans les volailles permettrait de confirmer cette différence majeure d'évolution du système CRISPR selon l'hôte de *Mgal*.

Finalement ce contraste entre évolution et stabilité semble démontrer qu'il existe des pressions de sélection qui ont poussé le système à rester actif ou à s'inactiver au cours du temps selon l'environnement de la bactérie. **L'ensemble de nos observations suggère donc que le système CRISPR de *Mgal* évolue de manière rapide, avec un changement de spécificité de Cas9 et une dynamique conjointe d'inactivation et de maintien du système CRISPR au fur et à mesure de l'évolution de la bactérie dans le nouvel hôte.**

Pour aller plus loin, nous nous sommes intéressés aux autres systèmes de défense de *Mgal*. Parmi les deux systèmes de Restrictions-Modifications connus, l'un d'eux est retrouvé co-localisé avec le système CRISPR. Or, il est maintenant admis que les systèmes de défense sont le plus souvent rassemblés dans des îlots de défense, ce qui facilite leurs échanges entre bactéries (Bernheim and Sorek, 2020). La plupart des articles cités en introduction à propos des nouveaux systèmes de défense sont tous basés sur une même méthode de détection de nouveaux systèmes. Cela consiste à regarder les gènes autour de systèmes de défense bactériens connus. En regardant de plus près la région génomique du locus CRISPR des différentes souches de *Mgal* (Figure 22), on peut observer deux autres groupes de gènes intéressants. Le premier correspond au système *letA-letS* récemment découvert (Gao *et al.*, 2020) et caractérisé comme un système Toxine-Antitoxine fonctionnel chez *Mmc* (Hill *et al.*, 2021). Le second code une sérine peptidase putative de la famille S8, non caractérisée et tronquée. Si l'on va plus loin dans le temps et que l'on compare avec une souche récemment isolée de poulet KUVMG001 (Song and Lee, 2021), le gène de la sérine peptidase S8 a été éliminé et remplacé par une transposase associée à la famille des IS256. En revanche chez la quasi-totalité des souches de roselin (excepté celles ayant perdu le locus CRISPR), le gène semble complet, sans mutation et la peptidase pourrait donc bien être fonctionnelle. Ces protéines sont répandues dans le monde du vivant, chez les plantes par exemple, elles sont connues comme étant des

systèmes de défenses en réponse à des stress environnementaux (Figueiredo, Sousa Silva and Figueiredo, 2018). Chez certaines bactéries elles vont avoir un rôle dans la spécificité de liaison à un substrat (H.-J. Li *et al.*, 2016) ou même dans l'évasion à l'immunité innée de l'hôte (McKenna *et al.*, 2021). Dernièrement, leurs rôles dans des systèmes de défenses bactériens ont également été démontrés, comme c'est le cas pour *IetS* qui code une sérine peptidase de la famille S8 (Gao *et al.*, 2020). En comparant également le système de Restriction-Modification, chez la majeure partie des souches autres que la souche S6 (Voir partie résultat), on observe son élimination et son remplacement par une transposase associée à la famille des IS256 (observé également pour le second système de restriction). **Ces éléments suggèrent qu'il existe un îlot de défense chez *Mgal*, et que celui-ci est particulièrement dynamique, avec des profils d'activation ou d'inactivation des systèmes différents entre les souches. Cette analyse plus globale permet de replacer l'évolution du système CRISPR dans un contexte plus large, impliquant une dynamique évolutive globale des systèmes de défense de *Mgal*, marquée par des événements successifs d'activation et d'inactivation de ces systèmes.**

Cet îlot de défense chez *Mgal* comprendrait donc à minima trois systèmes déjà caractérisés comme système de défense chez d'autres systèmes bactériens. Un système R-M, un système CRISPR et un des nouveaux systèmes caractérisés *IetA/IetS*. Cet exemple illustre bien le fait qu'il existe plusieurs types de systèmes de défense chez les mycoplasmes, ces systèmes n'étant pas mutuellement exclusifs. Les systèmes R-M sont retrouvés en très grand nombre chez les mycoplasmes qui représentent le deuxième groupe de bactéries en termes de nombre de systèmes R-M après *Helicobacter* (Oliveira, Touchon and Rocha, 2014). Le très récent travail de l'équipe du Pr. J. Jores de l'Université de Berne a montré la présence de trois systèmes de types Toxines-Antitoxines fonctionnels chez *Mmc* et présents chez différents membres du cluster Mycoïdes (Hill *et al.*, 2021). Ces derniers pourraient avoir un rôle dans la défense contre les phages comme cela a pu être montré chez d'autres bactéries avec le système ToxIN-ToxN chez *E. coli* où l'antitoxine est activée lors du blocage de la transcription de l'hôte par le phage et la toxine va donc agir en dégradant l'ARN (Guegler and Laub, 2021). Nous nous sommes alors intéressés aux systèmes présents plus largement dans les différentes espèces de mycoplasmes, en utilisant notamment l'outil PADLOC récemment publié par le JacksonLab de l'Université de Otago (Payne *et al.*, 2021). Ce dernier permet d'analyser un génome par la recherche de systèmes de défense connus comprenant l'ensemble des

DISCUSSION GÉNÉRALE ET CONCLUSION : Discussion générale

nouveaux systèmes récemment identifiés. Il nous a, par exemple, permis d'identifier le système letA-letS (Gao *et al.*, 2020) retrouvé chez *Mgal*, mais aussi un système de type CBASS (Cohen *et al.*, 2019) chez *Mycoplasma canis*, Tmn (Gao *et al.*, 2020) chez *M. capricolum* subsp. *capricolum*, ou encore LmuA (Doron *et al.*, 2018) chez *Mycoplasma conjunctivae*.

Si on poursuit l'étude, au sein de ces clusters de systèmes de défenses, on retrouve également très souvent des gènes codant des protéines de surface variables qui peuvent avoir un rôle dans la variation antigénique contre le système immunitaire de l'hôte (Citti, Nouvel and Baranowski, 2010). La présence de tels gènes dans ces loci très dynamiques peut être associée à une tendance commune à être transmis par HGT. Une autre hypothèse est toutefois possible : il se pourrait qu'en premier lieu les protéines variables de surface des mycoplasmes puissent avoir un rôle dans la défense contre les bactériophages en éliminant ou en masquant le récepteur d'entrée du bactériophage. En effet, les récepteurs bactériens permettant l'entrée des phages sont divers et variés (Hantke, 2020) et des lipoprotéines de surfaces ont déjà été identifiées comme des récepteurs permettant l'entrée de certains phages (Gueguen *et al.*, 2014). Ainsi une double fonction de ces protéines de surface pourrait être envisagée, leur évolution rapide permettant alors au mycoplasme de déjouer les attaques de bactériophages et d'échapper au système immunitaire de l'hôte. Une telle hypothèse serait en accord avec le fait que chez *Mgal*, la variation de surface est observée bien avant l'apparition d'anticorps contre la bactérie (Pflaum *et al.*, 2015), ce qui sous-entend que ce n'est pas la simple pression de sélection par le système immunitaire qui régit la sélection des variants de surface. De manière intéressante, on pourrait même relier l'absence de paroi chez les mollicutes au mécanisme hors du commun de *Streptomyces* MBT86 qui est capable de perdre sa paroi transitoirement en réponse à une attaque de phages pour augmenter la survie de la population bactérienne (Ongena *et al.*, 2021). **L'ensemble de ces données semblent montrer que de nombreux systèmes de défense bactériens divers et variés sont présents au sein des espèces de mycoplasmes et pas seulement chez *Mgal*.** Sur un plan expérimental, ces systèmes pourraient aussi expliquer certains problèmes de transformations observés au laboratoire pendant la thèse dans certaines espèces ou souches particulières de mycoplasmes.

DISCUSSION GÉNÉRALE ET CONCLUSION : Discussion générale

Si l'on imagine que l'environnement phagique au moment de l'émergence de *Mgal* dans le nouvel hôte posait problème à la bactérie, il serait normal que ces bactéries développent plusieurs mécanismes de résistances aux bactériophages. Cela va de pair avec le changement de spécificité de MgCas9 et son maintien actif durant les premières années de l'émergence chez les rosélins, mais aussi avec la réactivation de la peptidase S8 présente dans un ilot de défense et avec la tendance des premières souches de *Mgal* récoltées lors du début de l'épidémie chez le rosélin à envahir plus facilement les fibroblastes (Dowling, Hill and Bonneaud, 2020). En diversifiant les stratégies de défenses, voire leur capacité à devenir intracellulaire, ces souches ont alors pu augmenter leurs chances de persister et d'émerger dans ce nouvel hôte. De manière plus générale, **le « combat » entre phages et bactéries pourrait imposer aux mycoplasmes une pression évolutive sans doute bien plus importante qu'envisagée et la présence de nombreux systèmes de défense chez ces bactéries minimales tend à montrer qu'elles sont certainement exposées à de nombreux phages et autres éléments mobiles.** C'est une idée qui est en accord avec les récents travaux chez *M. bovis* dans lesquelles sont retrouvés des ICEs et des traces de prophages intégrés aux génomes de différentes souches à différents endroits du génome (García-Galán *et al.*, 2021).

En résumé, le changement de spécificité de MgCas9 de *Mgal* en lien avec un évènement de changement d'hôte interpelle. Si on note la présence de nombreux systèmes de défense, il est probable que de nombreux phages et éléments mobiles soient présents dans l'environnement des mycoplasmes. Lorsque la balance bénéfice-coût d'un système CRISPR n'est plus en faveur de son maintien (Koonin, 2019), il peut être rapidement éliminé. Cette pression est sans doute particulièrement forte chez des bactéries comme les mycoplasmes qui tendent globalement à la réduction de génome. L'évolution de la spécificité plutôt que son élimination pourrait alors s'expliquer par un bénéfice important pour *Mgal*, dans le contexte de l'adaptation à un nouvel hôte. Plusieurs questions peuvent alors être posées :

- Est-ce que *Mgal* s'est adapté à un nouvel environnement phagique rapidement par évolution de son système CRISPR après le changement d'hôte ou est-ce que l'avantage conféré par une spécificité de MgCas9 modifiée avant le changement d'hôte a permis l'établissement dans cette nouvelle niche écologique ?

DISCUSSION GÉNÉRALE ET CONCLUSION : Discussion générale

- Si de nombreux phages et éléments mobiles sont présents dans une niche écologique, est-ce qu'ils ne peuvent pas freiner l'émergence et la persistance d'une bactérie ?
- De manière plus extrême, les mycoplasmes possédant déjà de très nombreux facteurs de virulence permettant de résister à un hôte, l'environnement phagique serait-il l'élément central pour que ces bactéries sans paroi s'établissent dans un nouvel hôte proche ? De manière plus générale, la virulence contre l'hôte est un facteur important pour la survie de la bactérie, mais ne négligeons-nous pas l'importance du « combat » bactérie-phage qui pourrait inhiber l'établissement initial de la bactérie dans un hôte ?

L'infection d'un hôte par une espèce de mycoplasme pourrait alors être le résultat d'une interaction biologique tripartite : Hôtes – Bactéries – Phages.

Pour aller plus loin, les HGT sont des moteurs d'évolution importants chez les bactéries et chez les mycoplasmes. Ils pourraient être en partie responsables de la dynamique des systèmes de défense comme chez de nombreuses bactéries (Bernheim and Sorek, 2020). Chez *Mgal*, ils ont également permis l'acquisition de facteurs de virulence importants provenant de *Mycoplasma synoviae* un autre mycoplasme aviaire, comme la protéase CysP (Cizelj *et al.*, 2011), des protéines variables de surfaces (Sirand-Pugnet, Lartigue, *et al.*, 2007) et le système de clivage des anticorps MIB-MIP (Arfi *et al.*, 2016). Pourtant la présence de nombreux systèmes de défense devrait limiter ceux-ci. Des interrupteurs naturels sont capables d'activer ou inactiver ces systèmes de défenses. Par exemple, chez *Neisseria gonorrhoeae* des systèmes de restriction qui seraient associés à une fonction de régulation du flux d'ADN entrant (Sánchez-Busó *et al.*, 2019). Ce genre de système de restriction est retrouvé chez *M. pulmonis* (Dybvig, Sitaraman and French, 1998), *Mmc* (Algire *et al.*, 2012) ainsi que d'autres mycoplasmes. Ces variations d'activité permettent une adaptation rapide aux changements environnementaux. Sans ces interrupteurs, l'avantage qu'un élément mobile doit conférer à la bactérie pour rendre la balance bénéfice-coût favorable à l'élément mobile, doit être extrêmement important. Finalement, on peut imaginer que l'introduction de mutations rendant inactif un système de défense comme on l'a vu dans *Mgal*, permet une diversité des systèmes de défenses dans une population bactérienne (actifs, inactifs, éliminés). Ceci pourrait être vue comme un interrupteur naturel favorisant ou non les événements de HGT dans certaines cellules qui sont par la suite sélectionnées lors de changements environnementaux. L'apparition d'une résistance chez le nouvel hôte roselin pour lutter contre la bactérie *Mgal* en 2007

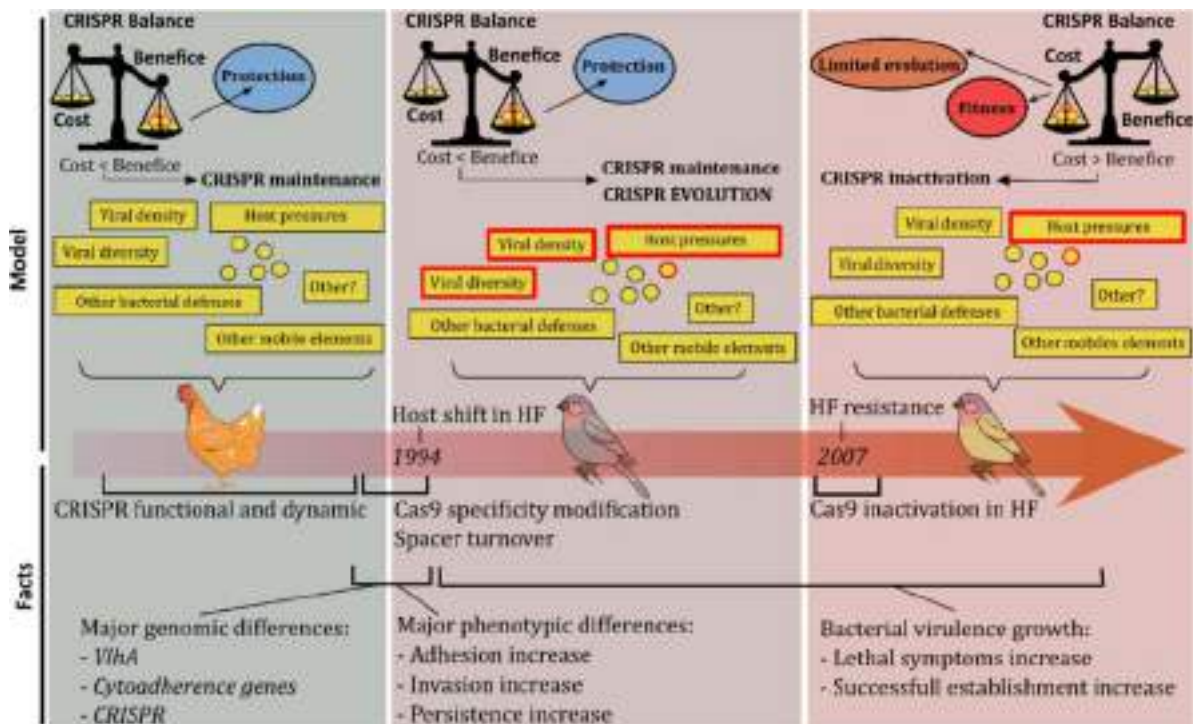


Figure 23 : Modèle d'évolution du système CRISPR dans le contexte de l'émergence et l'adaptation de *Mgal* chez le Roselin. Les principales étapes de l'émergence de *Mgal* chez le roselin et les données documentées dans la littérature and dans cette thèse sont indiquées. Le modèle proposé intègre la balance bénéfice-coût du système CRISPR et les différents paramètres qui vont intervenir dans la modulation de celle-ci.

pourrait être l'un de ces facteurs environnementaux qui aurait éteint l'interrupteur du système CRISPR.

Les systèmes de défense pourraient donc être à la fois une défense importante contre les phages avec un rôle dans la colonisation d'une niche écologique nouvelle mais aussi un système de régulation des éléments mobiles (Figure 23) :

- Les systèmes seraient stables lors d'une situation écologique pour laquelle la bactérie est adaptée depuis un long moment (Exemple de la bactérie *Mgal* chez le poulet avec un système CRISPR actif, un système IetA-IetS actif, un système putatif parfois inactif composé d'une peptidase S8).
- Lors d'une modification majeure de l'environnement écologique, ils peuvent être conservés si la balance Bénéfice-Coût est toujours en leur faveur (Exemple de l'émergence de *Mgal* dans le nouvel hôte où il semblerait même qu'une évolution de la spécificité de la Cas9 soit nécessaire).
- Mais en condition de faible pression de sélection imposée par les phages et de forte pression imposée par l'évolution des systèmes de défense immunitaires de l'hôte, les systèmes de défense de la bactérie pourraient être un frein à une évolution plus poussée de la bactérie via l'acquisition de matériel génétique (Exemple de l'adaptation du roselin en 2007 lié à la perte de fonctionnalité du système CRISPR en 2007 chez *Mgal*).

Pour le moment il n'existe aucune preuve, mais les éléments que l'on a pu voir suggèrent fortement un rôle très important des systèmes de défenses pour la protection et l'évolution des mycoplasmes et leur émergence dans de nouveaux hôtes. L'axe de recherche serait intéressant à approfondir ; il s'agirait d'une part d'identifier des îlots de défense et de caractériser leur fonctionnement et la synergie des éléments qu'ils rassemblent. Pour rappel, les mollicutes ont un code génétique atypique limitant les transferts horizontaux vers d'autres classes bactériennes ce qui implique un isolement génétique important et la présence possible de protéines uniques aux fonctions inédites. D'autre part, il serait intéressant de concevoir des expérimentations chez l'animal pour étudier s'il existe ou non, un lien entre le « défensome » du mycoplasme et sa virulence dans son contexte écologique naturel.

2. La boîte à outils : développement de nouveaux outils et de nouveaux modèles

Les mollicutes sont des organismes dont l'étude reste ralentie par de nombreux obstacles qui persistent pour manipuler leur génome. Durant ma thèse, le deuxième objectif majeur concernait le développement de la boîte à outils pour l'ingénierie des mycoplasmes. Comme j'ai pu l'expliquer précédemment, trois outils majeurs ont été mis au point. De nombreux paramètres sont à prendre en compte lorsque l'on veut réaliser de l'ingénierie génétique dans une espèce en particulier. Dans cette partie de la discussion, je vais repositionner les outils développés durant cette thèse, essayer de décrire ce qu'il reste à faire pour compléter la boîte à outils chez les mollicutes, et présenter un « mode d'emploi » pour l'ingénierie de génome avec une réflexion sur l'ensemble des paramètres à prendre en compte.

2.1. Les outils développés

Finalement durant cette thèse, trois principaux outils d'ingénierie ont été construits :

- Le premier est l'outil de contre-sélection basé sur la protéine endogène de mycoplasme MgCas9. Celle-ci semble fonctionnelle dans la plupart des espèces de mycoplasmes testés (*Mcap*, *Mmm*, *Mgal*, *Mpulm*), qu'elles soient phylogénétiquement éloignées ou non. Plusieurs plasmides « All-in-one » sont maintenant disponibles pour la communauté dans le but de sélectionner des événements génétiques rares tels que l'excision d'une ICE.
- Le système de recombinaison RecET basé sur des systèmes SSAPs que j'ai développé chez *Mgal* est le second outil de ce type à être utilisé chez les mycoplasmes, après le développement réalisé chez *M. pneumoniae* (Piñero-Lambea *et al.*, 2020). Malgré une efficacité limitée, il permet ainsi de développer les possibilités d'ingénierie de génome chez *Mgal* et il serait intéressant de tester ce système dans les différentes espèces en utilisant les vecteurs et les promoteurs adaptés.
- Finalement, l'outil d'édition de base, qui permet la modification précise de l'ADN et ainsi l'inactivation ou la réactivation de gènes codant une protéine d'intérêt. Il est efficace et rapide d'utilisation et semble fonctionnel dans un large panel d'espèce de mycoplasmes phylogénétiquement éloignées (*Mgal*, *Mmm*, *Mbov* et *M. agalactiae*). Trois plasmides différents contenant le système seront ainsi offerts à la communauté scientifique et peuvent permettre des essais très rapide dans l'espèce de leur choix.

2.2. Amélioration des outils construits

L'utilisation de l'outil de contre-sélection par l'endonucléase Cas9 endogène de *Mgal* pourrait être optimisée via l'utilisation d'un promoteur inductible. Celui que nous avons essayé, le *p_{xyl-tetO2}* (Breton *et al.*, 2010) n'est pas suffisamment réprimé en absence d'inducteur puisqu'une fuite importante est observée, qui se traduit par un effet toxique de contre sélection de la population bactérienne. Le promoteur optimisé de 2016 (Mariscal *et al.*, 2016) est une solution envisageable, mais il serait intéressant d'en développer un nouveau pour avoir plusieurs choix possibles. Le contrôle dans le temps et l'induction à volonté de la coupure du chromosome permettrait d'imaginer de plus nombreuses applications ainsi que de laisser plus de temps à la bactérie pour des évènements rares tels que la recombinaison.

Concernant l'utilisation d'un système de recombinaison exogène, quelques améliorations sont possibles. Tester d'autres systèmes éloignés phylogénétiquement ou découvrir un système efficace proche, devraient être les leviers principaux. Le système que nous avons développé est fonctionnel chez *Mgal* et permet toutes sortes de modifications, mais au vu du faible nombre de recombinants obtenus lors de nos expériences, une simple perte d'efficacité de transformation pourrait empêcher l'obtention de recombinants. Cette limite pourrait ainsi empêcher l'utilisation du système dans d'autres espèces. L'essai d'une utilisation du système RecET exprimé à partir d'un plasmide *oriC* pour conserver une expression stable entre les clones, ou d'un plasmide *oriC* portant une matrice de recombinaison, pourrait aider à l'obtention de plus nombreux recombinants. Pour effectuer de plus grandes modifications, la méthode RAGE appliquée à *M. pneumoniae* (Garcia-Morales *et al.*, 2020), pourrait être une solution intéressante. Le système RecET servirait alors à l'ajout des sites lox dans le génome de la bactérie.

Concernant l'outil le plus universel, le système CBE d'édition de base, nombreuses améliorations sont possibles. Tout comme pour le système de contre-sélection, un promoteur inductible sans fuite réduirait le bruit de fond de mutations off-target de type « spurious deamination ». L'utilisation d'un variant HF (High Fidelity) de spCas9 limiterait également les off-targets dépendantes du sgRNA. Finalement la principale limite de ce système est le nombre de cibles disponibles étant donné que le système est restreint à la présence d'une séquence PAM combinée à un codon pouvant devenir un codon-stop dans la bonne fenêtre d'action. Un travail spécifique sur l'utilisation de la protéine endogène

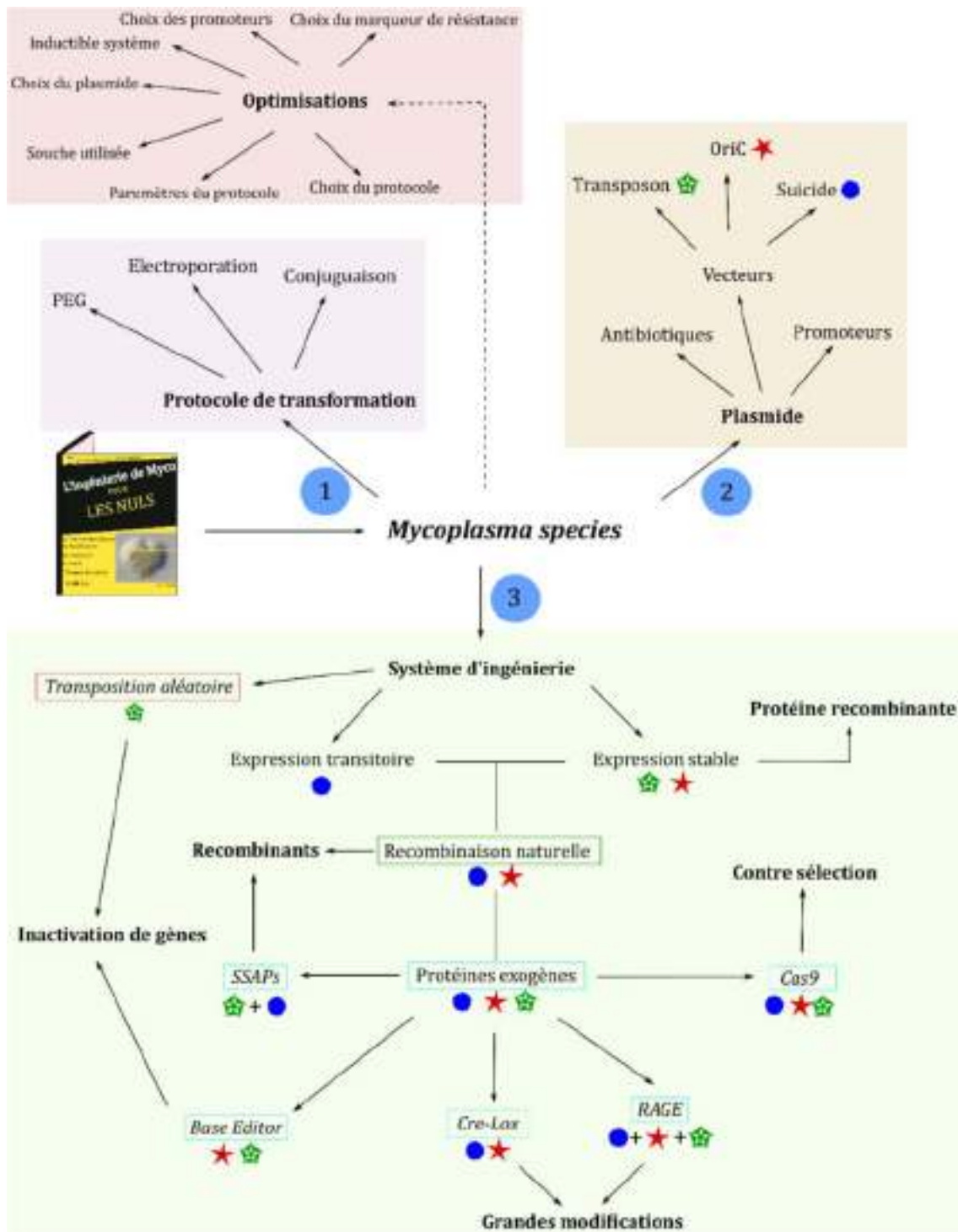


Figure 24 : L'ingénierie des mycoplasmes pour les Nuls. Description des trois étapes nécessaires vers l'utilisation d'outils d'ingénierie chez une espèce de mycoplasme. Les plasmides déjà utilisés pour un système d'ingénierie sont représentés : plasmides suicides (Bleu), transposons (Vert), OriC (Rouge).

MgCas9 pour de l'édition de base ou des variants de spCas9 avec une spécificité de PAM NGA pourrait permettre d'accroître fortement le nombre de cibles possibles pour les espèces ayant un génome riche en AT comme les mycoplasmes. La dernière amélioration majeure de cette boîte à outils concerne l'essai de l'adénosine désaminase ecTadA permettant un autre type de mutation, de A:T vers G:C.

2.3. Guide pour l'ingénierie d'une espèce de mycoplasme

Finalement, lorsqu'on veut faire de l'ingénierie de génome de mycoplasmes, ou plus largement, de mollicutes, plusieurs paramètres sont à prendre en compte. Ceux-ci seront détaillés ci-dessous et leur identification provient souvent des questions que nous avons dû nous poser au laboratoire pour pallier les problèmes rencontrés (Figure 24). Pour toute ingénierie de génome, chez les mycoplasmes, l'étape de transformation bactérienne est essentielle. Trois protocoles ont été décrits dans la littérature chez les mollicutes : la conjugaison bactérienne, l'électroporation et la transformation au PEG.

Ainsi, la première étape sera de développer, s'il n'existe pas déjà, un protocole de transformation pour l'espèce, voire la souche d'intérêt. L'électroporation et la transformation au PEG sont les deux méthodes les plus utilisées dans de nombreuses espèces de mycoplasmes (Table 4). L'optimisation du protocole de transformation pour chaque bactérie est une étape importante et peut permettre de rendre possible certaines applications. La présence de nombreux systèmes de défense contre l'ADN et notamment de systèmes de Restriction-Modifications peut s'avérer être une barrière à la transformation. La présence d'un motif de coupure sur un plasmide à transformer peut s'avérer être délétère par exemple. Il existe des méthodes de méthylation *in vitro* de l'ADN, à partir d'extraits cellulaires de la bactérie qui peuvent permettre de dépasser cette barrière (Lartigue *et al.*, 2009). Une autre barrière importante est la sélection antibiotique utilisée. En effet, il s'est avéré qu'entre deux souches d'une même espèce, une différence importante de la CMI (Concentration Minimale Inhibitrice) pouvait être observée et que la calibration de la concentration de l'antibiotique utilisée permettait d'augmenter de manière importante l'efficacité de transformation de la souche considérée. L'utilisation d'un autre antibiotique, d'un autre promoteur régissant l'expression du gène de résistance ou d'une autre souche de l'espèce sont aussi des facteurs pouvant moduler significativement le résultat de la transformation. Réaliser une mesure de CMI pour différents antibiotiques et identifier un promoteur fort fonctionnel dans l'espèce choisie

DISCUSSION GÉNÉRALE ET CONCLUSION : Discussion générale

(comme le promoteur *Tuf*) sont des étapes préliminaires à ne pas négliger. Le choix du plasmide utilisé est important, d'une part pour le développement du protocole de transformation et d'autre part, comme vecteur, selon l'application envisagée. Des différences d'efficacité importantes, selon l'utilisation d'un plasmide *oriC* ou d'un transposon, peuvent être observées d'une espèce ou d'une souche à l'autre. Il est donc parfois nécessaire d'essayer les différents transposons disponibles chez les mycoplasmes ainsi que de construire et essayer un plasmide *oriC* spécifique de l'espèce ou provenant d'une espèce très proche. Si malgré tout cela, la transformation reste inefficace, la dernière solution reste la plus difficile à mettre en place, mais la plus puissante, qui implique la transplantation de génome. Il s'agit d'abord de cloner le génome de la bactérie dans la levure par les nombreuses méthodes actuellement disponibles, puis de le modifier dans la levure et, finalement, le transplanter dans une cellule receveuse adaptée.

Si un protocole de transformation est disponible, il s'agira alors de passer à l'étape suivante et choisir une des stratégies d'ingénierie. Des promoteurs fonctionnels dans les espèces ciblées sont nécessaires pour l'expression de gènes d'intérêt. Chez les mycoplasmes, il existe les promoteurs utilisés couramment, comme le promoteur du gène de la spiruline de *S. citri* ou le promoteur synthétique SynMyco qui sont fonctionnels chez de nombreuses espèces. L'utilisation de promoteurs endogènes comme le promoteur du gène *tuf* est également de plus en plus courante. Ils permettront d'exprimer les différents outils qui seront utilisés.

Le but de l'ingénierie oriente ensuite vers le système à tester ainsi que le choix du vecteur de transport :

- Pour l'inactivation d'un gène chez une espèce d'intérêt, l'utilisation d'un plasmide *oriC* contenant une matrice de recombinaison peut être expérimentée plutôt facilement, mais pour une obtention rapide et efficace, l'utilisation d'un système Base Editor semble être la meilleure option.
- Pour de plus larges modifications avec insertions, délétions ou multiples modifications de nucléotides, la recombinaison naturelle avec des plasmides *oriC* peut s'avérer efficace chez certaines espèces (Table 4), mais elles sont rares. L'apport d'une protéine RecA hétérologue sur ce même plasmide *oriC* a également été décrite dans d'autres espèces (Table 4). Mais l'utilisation d'un système de type RecET est sans doute une option plus intéressante. Actuellement cette méthode n'est fonctionnelle

DISCUSSION GÉNÉRALE ET CONCLUSION : Discussion générale

qu'à partir d'un transposon qui intègre le système dans le génome et permet l'obtention d'une souche exprimant le système. Une seconde transformation par une matrice présente sur un plasmide suicide, permet alors la recombinaison. Les efficacités de l'ensemble de ces stratégies étant limitées, elles pourraient s'avérer inadéquat pour des recombinaisons de fragments de plus de 1 ou 2 kb.

- La méthode RAGE permet d'obtenir des réarrangements plus importants (Garcia-Morales *et al.*, 2020). La première étape consiste à insérer, à l'aide d'une des méthodes citées précédemment, des « landing pad » qui sont de petites séquences d'ADN portant des sites lox du système Cre-Lox. L'ajout de deux sites lox encadrant la région à modifier permet l'échange avec de grands fragments d'ADN qui ont pu être modifiés au préalable, par exemple dans la levure ou *E. coli*.

Parmi ces méthodes d'ingénierie, certaines nécessitent une expression transitoire d'une protéine d'intérêt dans l'optique par exemple d'utiliser le système Cre-lox pour l'élimination des marqueurs de résistance qui pourraient être gênants pour des applications à visées biotechnologiques. Dans ce cas, un plasmide suicide peut être utilisé. Il ne comporte ni système d'intégration, ni origine de réplication, mais seulement un marqueur de résistance à l'antibiotique. Le maintien du plasmide dans les transformants est assuré transitoirement par pression de sélection en culture liquide puis un étalement sans pression de sélection permet d'obtenir des clones qui devront ensuite être criblés pour le génotype ou le phénotype recherché. D'autres systèmes vont nécessiter une expression stable d'une protéine comme l'expression de protéines fluorescentes pour de la microscopie. Des plasmide *oriC* ou intégratif sont alors des options à privilégier par rapport au plasmide suicide.

Finalement, le type d'ingénierie de génome désiré doit dans tous les cas régir le choix du système, la méthode de transformation et le vecteur de transport. Ceci sera déterminant dans la fonctionnalité du système utilisé qui peuvent être plus ou moins efficaces et sensibles.

2. CONCLUSION

En conclusion, l'étude des systèmes CRISPR s'est avérée révélatrice d'une histoire évolutive remarquable pour la bactérie *M. gallisepticum*. A plus grande échelle, de nombreuses hypothèses sur l'écologie des mycoplasmes pourront faire l'objet d'étude avec notamment le lien entre ces bactéries, leurs hôtes mais aussi les phages et les éléments mobiles qui les entourent. De plus, nous sommes désormais capables d'utiliser un large panel d'outils d'ingénierie de génomes pour plusieurs espèces de mycoplasmes. Que ce soit pour des applications de contre-sélection, des inactivations ou des remplacements de gènes, différents choix sont désormais offerts à la communauté et ces outils ouvrent un champ d'étude important pour ces bactéries étonnantes que sont les mycoplasmes.

“A pessimist sees the difficulty in every opportunity; an optimist sees the opportunity in every difficulty”.

Winston Churchill, 1922

RÉFÉRENCES BIBLIOGRAPHIQUES

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abudayyeh, O. O. *et al.* (2017) 'RNA targeting with CRISPR-Cas13', *Nature*, 550(7675), pp. 280–284. doi: 10.1038/nature24049.
- Aelst, K. Van *et al.* (2019) 'The effect of DNA topology on observed rates of R-loop formation and DNA strand cleavage by CRISPR Cas12a', *Genes*, 10(2), pp. 1–14. doi: 10.3390/genes10020169.
- Agar, N. (1995) 'DESIGNING BABIES: MORALLY PERMISSIBLE WAYS TO MODIFY THE HUMAN GENOME', *Bioethics*, 9(1), pp. 1–15. doi: 10.1111/j.1467-8519.1995.tb00297.x.
- Algire, M. A. *et al.* (2009) 'New selectable marker for manipulating the simple genomes of *Mycoplasma* species', *Antimicrobial Agents and Chemotherapy*, 53(10), pp. 4429–4432. doi: 10.1128/AAC.00388-09.
- Algire, M. A. *et al.* (2012) 'A Type III restriction - Modification system in *Mycoplasma mycoides* subspecies *capri*', *Open Biology*, 2(OCT). doi: 10.1098/rsob.120115.
- Allam, A. B. *et al.* (2010) 'Enhancement of targeted homologous recombination in *Mycoplasma mycoides* subspecies *capri* by Inclusion of heterologous *recA*', *Applied and Environmental Microbiology*, 76(20), pp. 6951–6954. doi: 10.1128/AEM.00056-10.
- Altae-Tran, H. *et al.* (2021) 'The widespread IS200/605 transposon family encodes diverse programmable RNA-guided endonucleases', *Science*.
- Altenbuchner, J. (2016) 'Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system', *Applied and Environmental Microbiology*. doi: 10.1128/AEM.01453-16.Editor.
- Amanfu, W. *et al.* (2006) 'Comparison between c-ELISA and CFT in Detecting Antibodies to *Mycoplasma mycoides mycoides* Biotype SC in Cattle Affected by CBPP in Botswana', *Annals of the New York Academy of Sciences*, 916(1), pp. 364–369. doi: 10.1111/j.1749-6632.2000.tb05314.x.
- Anders, C., Bargsten, K. and Jinek, M. (2016) 'Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9', *Molecular Cell*, 61(6), pp. 895–902. doi: 10.1016/j.molcel.2016.02.020.
- Anzalone, A. V. *et al.* (2019) 'Search-and-replace genome editing without double-strand breaks or donor DNA', *Nature*. doi: 10.1038/s41586-019-1711-4.
- Anzalone, A. V., Koblan, L. W. and Liu, D. R. (2020) 'Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors', *Nature Biotechnology*, 38(7), pp. 824–844. doi: 10.1038/s41587-020-0561-9.
- Arber, B. W. and N, S. L. I. N. (1969) 'DNA modification and restriction', *Annual Review of Biochemistry*, pp. 467–500.
- Arfi, Y. *et al.* (2016) 'MIB-MIP is a mycoplasma system that captures and cleaves immunoglobulin G', *Proceedings of the National Academy of Sciences of the United States of America*, 113(19), pp. 5406–5411. doi: 10.1073/pnas.1600546113.
- Arfi, Y. *et al.* (2021) 'Beware of Mycoplasma Anti-immunoglobulin Strategies', *mBio*. Edited by J. Yount. doi: 10.1128/mBio.01974-21.
- Arndt, D. *et al.* (2016) 'PHASTER: a better, faster version of the PHAST phage search tool', *Nucleic acids research*, 44(W1), pp. W16–W21. doi: 10.1093/nar/gkw387.
- Askar, H. *et al.* (2021) 'Immune Evasion of *Mycoplasma bovis*', *Pathogens*, 10(3), p. 297. doi: 10.3390/pathogens10030297.
- Aubrey, B. J. *et al.* (2015) 'An Inducible Lentiviral Guide RNA Platform Enables the Identification of Tumor-Essential Genes and Tumor-Promoting Mutations InVivo', *Cell Reports*, 10(8), pp. 1422–1432. doi: 10.1016/j.celrep.2015.02.002.
- Babu, K. *et al.* (2019) 'Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance', *Biochemistry*, 58(14), pp. 1905–1917. doi: 10.1021/acs.biochem.8b01241.
- Baby, V., Labroussaa, F., *et al.* (2018) 'Cloning and Transplantation of the *Mesoplasma florum* Genome', *ACS Synthetic Biology*, 7(1), pp. 209–217. doi: 10.1021/acssynbio.7b00279.
- Baby, V., Lachance, J.-C., *et al.* (2018) 'Inferring the Minimal Genome of *Mesoplasma florum* by Comparative Genomics and Transposon Mutagenesis', *mSystems*. Edited by D. F. Savage, 3(3). doi: 10.1128/mSystems.00198-17.
- Bae, S., Park, J. and Kim, J. (2014) 'Sequence analysis Cas-OFFinder : a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases', *Bioinformatics*, 30(10), pp. 1473–1475. doi: 10.1093/bioinformatics/btu048.
- Balish, M. *et al.* (2019) 'Recommended rejection of the names malacoplasma gen. nov., mesomycoplasma gen. nov., metamyoplasma gen. nov., metamyoplasmataceae fam. nov., mycoplasmoidaceae fam. nov., mycoplasmoidales ord. nov., mycoplasmoides gen. nov., mycoplasmosis gen. nov. [gup]', *International Journal of Systematic and Evolutionary*

RÉFÉRENCES BIBLIOGRAPHIQUES

- Microbiology*, 69(11), pp. 3650–3653. doi: 10.1099/ijsem.0.003632.
- Balish, M. F. (2006) 'Subcellular structures of mycoplasmas', *Frontiers in Bioscience*, 11(1), p. 2017. doi: 10.2741/1943.
- Balish, M. F. (2014) '*Mycoplasma pneumoniae*, an Underutilized Model for Bacterial Cell Biology', *Journal of Bacteriology*, 196(21), pp. 3675–3682. doi: 10.1128/JB.01865-14.
- Baquero, M. *et al.* (2021) 'Effects of inflammatory stimuli on responses of macrophages to *Mycoplasma bovis* infection', *Veterinary Microbiology*, 262, p. 109235. doi: 10.1016/j.vetmic.2021.109235.
- Barrangou, R. *et al.* (2007) 'CRISPR Provides Against Viruses in Prokaryotes', *Science*, 315(March), pp. 1709–1712.
- Barrangou, R. and Horvath, P. (2012) 'CRISPR: New horizons in phage resistance and strain identification', *Annual Review of Food Science and Technology*, 3(1), pp. 143–162. doi: 10.1146/annurev-food-022811-101134.
- Barrangou, R. and Horvath, P. (2017) 'A decade of discovery : CRISPR functions and applications', *Nature Publishing Group*, 2(June), pp. 1–9. doi: 10.1038/nmicrobiol.2017.92.
- Barrangou, R. and Oost, J. Van Der (2015) 'Bacteriophage exclusion, a new defense system', *EMBO Journal*.
- Beaman, K. D. and Pollack, J. D. (1983) 'Synthesis of adenylate nucleotides by Mollicutes (Mycoplasmas)', *Journal of General Microbiology*, 129(10), pp. 3103–3110. doi: 10.1099/00221287-129-10-3103.
- Becker, J. and Wittmann, C. (2015) 'Advanced biotechnology: Metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products', *Angewandte Chemie - International Edition*, 54(11), pp. 3328–3350. doi: 10.1002/anie.201409033.
- Beloglazova, N. *et al.* (2008) 'A Novel Family of Sequence-specific Endoribonucleases Associated with the Clustered Regularly Interspaced Short Palindromic Repeats', *Journal of Biological Chemistry*, 283(29), pp. 20361–20371. doi: 10.1074/jbc.M803225200.
- Berg, D. E. *et al.* (1975) 'Transposition of R factor genes to bacteriophage λ ', *Proceedings of the National Academy of Sciences of the United States of America*, 72(9), pp. 3628–3632. doi: 10.1073/pnas.72.9.3628.
- Bernheim, A. *et al.* (2017) 'Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria', *Nature Communications*, 8(1), pp. 25–28. doi: 10.1038/s41467-017-02350-1.
- Bernheim, A. *et al.* (2021) 'Prokaryotic viperins produce diverse antiviral molecules', *Nature*, 589(January). doi: 10.1038/s41586-020-2762-2.
- Bernheim, A. and Sorek, R. (2020) 'The pan-immune system of bacteria: antiviral defence as a community resource', *Nature Reviews Microbiology*, 18(February), pp. 113–119. doi: 10.1038/s41579-019-0278-2.
- Bertin, C. *et al.* (2013) 'Characterization of Free Exopolysaccharides Secreted by *Mycoplasma mycoides* Subsp. *mycoides*', *PLoS ONE*, 8(7). doi: 10.1371/journal.pone.0068373.
- Betzig, E. *et al.* (2006) 'Imaging intracellular fluorescent proteins at nanometer resolution', *Science*, 313(5793), pp. 1642–1645. doi: 10.1126/science.1127344.
- Béven, L. *et al.* (2012) 'Specific evolution of F 1-like ATPases in mycoplasmas', *PLoS ONE*, 7(6). doi: 10.1371/journal.pone.0038793.
- Bikard, D. *et al.* (2012) 'CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during In Vivo Bacterial Infection', *cell host and microbe*. doi: 10.1016/j.chom.2012.06.003.
- Bikard, D. *et al.* (2013) 'Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system', *Nucleic Acids Research*, 41(15), pp. 7429–7437. doi: 10.1093/nar/gkt520.
- Bikard, D. *et al.* (2014) 'Exploiting CRISPR-cas nucleases to produce sequence-specific antimicrobials', *Nature Biotechnology*, 32(11), pp. 1146–1150. doi: 10.1038/nbt.3043.
- Bischof, D. F. *et al.* (2008) 'Cytotoxicity of *Mycoplasma mycoides* subsp. *mycoides* Small Colony Type to Bovine Epithelial Cells', *Infection and Immunity*, 76(1), pp. 263–269. doi: 10.1128/IAI.00938-07.
- Blanchard, A. (1990) '*Ureaplasma urealyticum* urease genes; use of a UGA tryptophan codon', *Molecular Microbiology*, 4(4), pp. 669–676. doi: 10.1111/j.1365-2958.1990.tb00636.x.
- Blötz, C. *et al.* (2018) 'Development of a replicating plasmid based on the native oriC in *Mycoplasma pneumoniae*', *Microbiology (United Kingdom)*, 164(11), pp. 1372–1382. doi: 10.1099/mic.0.000711.
- Blötz, C. and Stülke, J. (2017) 'Glycerol metabolism and its implication in virulence in *Mycoplasma*', *FEMS Microbiology Reviews*, 41(5), pp. 640–652. doi: 10.1093/femsre/fux033.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Bolotin, A. *et al.* (2005) 'Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin', *Microbiology*, 151(8), pp. 2551–2561. doi: 10.1099/mic.0.28048-0.
- Bonneaud, C. *et al.* (2011) 'Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird', *Proceedings of the National Academy of Sciences of the United States of America*, 108(19), pp. 7866–7871. doi: 10.1073/pnas.1018580108.
- Bonneaud, C. *et al.* (2018) 'Rapid Antagonistic Coevolution in an Emerging Pathogen and Its Vertebrate Host', *Current Biology*, 28(18), pp. 2978–2983.e5. doi: 10.1016/j.cub.2018.07.003.
- Bonnefois, T. *et al.* (2016) 'Development of fluorescence expression tools to study host-mycoplasma interactions and validation in two distant mycoplasma clades', *Journal of Biotechnology*, 236, pp. 35–44. doi: 10.1016/j.jbiotec.2016.08.006.
- Bové, J. M. (1993) 'Molecular Features of Mollicutes', *Clinical Infectious Diseases*, 17(Suppl 1), pp. S10–S31. doi: 10.1093/clinids/17.Supplement_1.S10.
- Bratovič, M. *et al.* (2020) 'Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches', *Nature Chemical Biology*, 16(5), pp. 587–595. doi: 10.1038/s41589-020-0490-4.
- Breton, M. *et al.* (2010) 'First report of a tetracycline-inducible gene expression system for mollicutes', *Microbiology*, 156(1), pp. 198–205. doi: 10.1099/mic.0.034074-0.
- Broniewski, J. M. *et al.* (2020) 'The effect of phage genetic diversity on bacterial resistance evolution', *The ISME Journal*, 14(3), pp. 828–836. doi: 10.1038/s41396-019-0577-7.
- Brouns, S. J. J. *et al.* (2008) 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes', *Science*. doi: 10.1126/science.1159689.Small.
- Budhathoki, J. B. *et al.* (2020) *Real-time observation of CRISPR spacer acquisition by Cas1–Cas2 integrase*, *Nature Structural and Molecular Biology*. Springer US. doi: 10.1038/s41594-020-0415-7.
- Burgos, R. *et al.* (2012) 'RecA mediates MgpB and MgpC phase and antigenic variation in *Mycoplasma genitalium*, but plays a minor role in DNA repair', *Molecular Microbiology*, 85(4), pp. 669–683. doi: 10.1111/j.1365-2958.2012.08130.x.
- Calcutt, M. J., Lewis, M. S. and Wise, K. S. (2002) 'Molecular genetic analysis of ICEF, an integrative conjugal element that is present as a repetitive sequence in the chromosome of *Mycoplasma fermentans* PG18', *Journal of Bacteriology*, 184(24), pp. 6929–6941. doi: 10.1128/JB.184.24.6929-6941.2002.
- Cao, J. *et al.* (2016) 'An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting', *Nucleic Acids Research*, 44(19), pp. 1–10. doi: 10.1093/nar/gkw660.
- Cao, J., Kapke, P. A. and Minton, F. C. (1994) 'Transformation of *Mycoplasma gallisepticum* with Tn916, Tn4001, and integrative plasmid vectors', *Journal of Bacteriology*, 176(14), pp. 4459–4462. doi: 10.1128/jb.176.14.4459-4462.1994.
- Carvalho, F. M. *et al.* (2005) 'DNA repair in reduced genome: The Mycoplasma model', *Gene*, 360(2), pp. 111–119. doi: 10.1016/j.gene.2005.06.012.
- Charbonneau, M. R. *et al.* (2020) 'Developing a new class of engineered live bacterial therapeutics to treat human diseases', *Nature Communications*, 11(1), pp. 1–11. doi: 10.1038/s41467-020-15508-1.
- Chatterjee, P. *et al.* (2020) 'A Cas9 with PAM recognition for adenine dinucleotides', *Nature Communications*, 11(1), pp. 1–6. doi: 10.1038/s41467-020-16117-8.
- Chen, B. *et al.* (2013) 'Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System', *Cell*. doi: 10.1016/j.cell.2013.12.001.Dynamic.
- Chen, H. *et al.* (2012) 'Identification of biofilm formation by *Mycoplasma gallisepticum*', *Veterinary Microbiology*, 161(1–2), pp. 96–103. doi: 10.1016/j.vetmic.2012.07.013.
- Chen, T. *et al.* (2017) 'Chemically Controlled Epigenome Editing Through an Inducible dCas9 System', *J Am Chem Soc*. doi: 10.1021/jacs.7b06555.
- Cheng, A. W. *et al.* (2013) 'Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system', *Cell Research*, 23(10), pp. 1163–1171. doi: 10.1038/cr.2013.122.
- Chernova, O. A. *et al.* (2016) 'Mycoplasmas and Their Antibiotic Resistance: The Problems and Prospects in Controlling Infections', *acta naturae*, 8(29), pp. 24–34.
- Chevallereau, A. *et al.* (2021) 'Interactions between bacterial and phage communities in natural environments', *Nature Reviews Microbiology*, 0123456789. doi: 10.1038/s41579-021-00602-y.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Cho, S., Shin, J. and Cho, B. K. (2018) 'Applications of CRISPR/Cas system to bacterial metabolic engineering', *International Journal of Molecular Sciences*, 19(4). doi: 10.3390/ijms19041089.
- Cho, S. W. *et al.* (2013) 'Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease', *Nature Biotechnology*, 31(3), pp. 230–232. doi: 10.1038/nbt.2507.
- Chopra-Dewasthaly, R. *et al.* (2005) 'First steps towards the genetic manipulation of *Mycoplasma agalactiae* and *Mycoplasma bovis* using the transposon Tn4001mod', *International Journal of Medical Microbiology*, 294(7), pp. 447–453. doi: 10.1016/j.ijmm.2004.09.010.
- Choudhary, E. *et al.* (2015) 'Gene silencing by CRISPR interference in mycobacteria', *Nature Communications*, 6, pp. 1–11. doi: 10.1038/ncomms7267.
- Chowdhury, S. *et al.* (2017) 'Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex', *Cell*. doi: 10.1016/j.cell.2017.03.012.Structure.
- Chylinski, K. *et al.* (2014) 'Classification and evolution of type II CRISPR-Cas systems', *Nucleic Acids Research*, 42(10), pp. 6091–6105. doi: 10.1093/nar/gku241.
- Chylinski, K. *et al.* (2019) 'CRISPR-Switch regulates sgRNA activity by Cre recombination for sequential editing of two loci', *Nature Communications*, 10(1). doi: 10.1038/s41467-019-13403-y.
- Ciccarelli, F. D. *et al.* (2006) 'Toward Automatic Reconstruction of a Highly Resolved Tree of Life Francesca', *SCIENCE*, (March), pp. 1–5. doi: 10.1126/science.1121745.
- Citti, C. *et al.* (2018) 'Horizontal gene transfers in mycoplasmas (Mollicutes)', *Current Issues in Molecular Biology*, 29, pp. 3–22. doi: 10.21775/cimb.029.003.
- Citti, C. *et al.* (2020) 'Genomic islands in mycoplasmas', *Genes*, 11(8), pp. 1–16. doi: 10.3390/genes11080836.
- Citti, C. and Blanchard, A. (2013) 'Mycoplasmas and their host: Emerging and re-emerging minimal pathogens', *Trends in Microbiology*, 21(4), pp. 196–203. doi: 10.1016/j.tim.2013.01.003.
- Citti, C., Nouvel, L. X. and Baranowski, E. (2010) 'Phase and antigenic variation in mycoplasmas', *Future Microbiology*, 5(7), pp. 1073–1085. doi: 10.2217/fmb.10.71.
- Cizelj, I. *et al.* (2011) '*Mycoplasma gallisepticum* and *Mycoplasma synoviae* express a cysteine protease CysP, which can cleave chicken IgG into Fab and Fc', *Microbiology*, 157(2), pp. 362–372. doi: 10.1099/mic.0.045641-0.
- Clampitt, J. M., Madsen, M. L. and Minion, F. C. (2021) 'Construction of *Mycoplasma hyopneumoniae* P97 Null Mutants', *Frontiers in Microbiology*, 12(April), pp. 1–12. doi: 10.3389/fmicb.2021.518791.
- Clewell, D. B. and Gawron-burke, C. (1986) 'CONJUGATIVE TRANSPOSONS AND THE DISSEMINATION OF ANTIBIOTIC RESISTANCE IN STREPTOCOCCI', *Annual Review of Microbiology*, pp. 635–659.
- Clyde A. Hutchison *et al.* (1999) 'Global Transposon Mutagenesis and a Minimal Mycoplasma Genome', *Science*, 286(December).
- Cobb, R. E., Wang, Y. and Zhao, H. (2015) 'High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System', *ACS Synthetic Biology*, 4(6), pp. 723–728. doi: 10.1021/sb500351f.
- Cohen, D. *et al.* (2019) 'Cyclic GMP–AMP signalling protects bacteria against viral infection', *Nature*, 574(7780), pp. 691–695. doi: 10.1038/s41586-019-1605-5.
- Cole, R. M. *et al.* (1973) 'Morphology, ultrastructure, and bacteriophage infection of the helical mycoplasma like organism (*Spiroplasma citri* gen. nov., sp. nov.) cultured from "stubborn" disease of citrus', *Journal of Bacteriology*, 115(1), pp. 367–386. doi: 10.1128/jb.115.1.367-386.1973.
- Collias, D. *et al.* (2020) 'A positive, growth-based PAM screen identifies noncanonical motifs recognized by the *S. pyogenes* Cas9', *Science Advances*, (July).
- Collins, F. S. *et al.* (2021) 'Human Molecular Genetics and Genomics — Important Advances and Exciting Possibilities', *The NEW ENGLAND JOURNAL of MEDICINE*.
- Cong, L. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas Systems Le', *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.Multiplex.
- Cordova, C. M. M. *et al.* (2002) 'Identification of the Origin of Replication of the *Mycoplasma pulmonis* Chromosome and Its Use in oriC Replicative Plasmids', *Journal of Bacteriology*, 184(19), pp. 5426–5435. doi: 10.1128/JB.184.19.5426.
- Corts, A. D. *et al.* (2019) 'Efficient and Precise Genome Editing in *Shewanella* with Recombineering and CRISPR/Cas9-

RÉFÉRENCES BIBLIOGRAPHIQUES

- Mediated Counter-Selection', *ACS Synthetic Biology*, 8(8), pp. 1877–1889. doi: 10.1021/acssynbio.9b00188.
- Csörgő, B. *et al.* (2020) 'A compact Cascade-Cas3 system for targeted genome engineering', *Nature Methods*, 17(12), pp. 1183–1190. doi: 10.1038/s41592-020-00980-w.
- Cui, L. *et al.* (2018) 'A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9', *Nature Communications*, 9(1), pp. 1–10. doi: 10.1038/s41467-018-04209-5.
- Daniloski, Z. *et al.* (2021) 'Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells', *Cell*, 184(1), pp. 92–105.e16. doi: 10.1016/j.cell.2020.10.030.
- Darling, A. C. E. *et al.* (2004) 'Mauve: Multiple alignment of conserved genomic sequence with rearrangements', *Genome Research*, 14(7), pp. 1394–1403. doi: 10.1101/gr.2289704.
- Datta, S. *et al.* (2008) 'Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages', *Proceedings of the National Academy of Sciences of the United States of America*, 105(5), pp. 1626–1631. doi: 10.1073/pnas.0709089105.
- Dieu, L., Chapey, E. and Balcer-Rodrigues, V. (2005) 'Mycoplasma mycoides spp. mycoides biotype small colony-secreted components induce apoptotic cell death in bovine leucocytes', *Scandinavian Journal of Immunology*, 62(6), pp. 528–538. doi: 10.1111/j.1365-3083.2005.01690.x.
- Delaney, N. F. *et al.* (2012) 'Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*', *PLoS Genetics*, 8(2). doi: 10.1371/journal.pgen.1002511.
- Depardieu, F. and Bikard, D. (2019) 'Gene silencing with CRISPRi in bacteria and optimization of dCas9 expression levels', *Methods*, (April), pp. 1–15. doi: 10.1016/j.jymeth.2019.07.024.
- Deveau, H. *et al.* (2008) 'Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*', *Journal of Bacteriology*, 190(4), pp. 1390–1400. doi: 10.1128/JB.01412-07.
- Deveau, H., Garneau, J. E. and Moineau, S. (2010) 'CRISPR/Cas system and its role in phage-bacteria interactions', *Annual Review of Microbiology*, 64, pp. 475–493. doi: 10.1146/annurev.micro.112408.134123.
- Dhandayuthapani, S. *et al.* (2001) 'Peptide methionine sulfoxide reductase (MsrA) is a virulence determinant in *Mycoplasma genitalium*', *Journal of Bacteriology*, 183(19), pp. 5645–5650. doi: 10.1128/JB.183.19.5645-5650.2001.
- Dhandayuthapani, S., Rasmussen, W. G. and Baseman, J. B. (1999) 'Disruption of gene mg218 of *Mycoplasma genitalium* through homologous recombination leads to an adherence-deficient phenotype', *Proceedings of the National Academy of Sciences of the United States of America*, 96(9), pp. 5227–5232. doi: 10.1073/pnas.96.9.5227.
- Dhondt, A. A. *et al.* (2005) 'Dynamics of a novel pathogen in an avian host: Mycoplasmal conjunctivitis in house finches', *Acta Tropica*, 94(1), pp. 77–93. doi: 10.1016/j.actatropica.2005.01.009.
- Dhondt, A. A. *et al.* (2014) 'Diverse Wild Bird Host Range of *Mycoplasma gallisepticum* in Eastern North America', *PLoS ONE*. Edited by K. McGraw, 9(7), p. e103553. doi: 10.1371/journal.pone.0103553.
- Dhondt, A. A., Tessaglia, D. L. and Slothower, R. L. (1998) 'Epidemic From Conjunctivitis America in House Finches', *Journal of Wildlife Diseases*, 34(2), pp. 265–280.
- DiCarlo, J. E. *et al.* (2013) 'Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems', *Nucleic Acids Research*, 41(7), pp. 4336–4343. doi: 10.1093/nar/gkt135.
- Dong, C. *et al.* (2018) 'Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-04901-6.
- Dong, X. *et al.* (2017) 'CRISPR/dCas9-mediated inhibition of gene expression in *Staphylococcus aureus*', *Journal of Microbiological Methods*, 139(May), pp. 79–86. doi: 10.1016/j.mimet.2017.05.008.
- Dordet-Frisoni, E. *et al.* (2014) 'Chromosomal transfers in mycoplasmas: When minimal genomes go mobile', *mBio*, 5(6), pp. 1–11. doi: 10.1128/mBio.01958-14.
- Dordet-Frisoni, E. *et al.* (2019) 'Mycoplasma Chromosomal Transfer: A Distributive, Conjugative Process Creating an Infinite Variety of Mosaic Genomes', *Frontiers in Microbiology*, 10(OCT), pp. 1–16. doi: 10.3389/fmicb.2019.02441.
- Dordet Frisoni, E. *et al.* (2013) 'ICEA of *Mycoplasma agalactiae*: A new family of self-transmissible integrative elements that confers conjugative properties to the recipient strain', *Molecular Microbiology*, 89(6), pp. 1226–1239. doi: 10.1111/mmi.12341.
- Doron, S. *et al.* (2018) 'Systematic discovery of anti-phage defense systems in the microbial pan-genome', *Science*. doi: 10.1126/science.aar4120.Systematic.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Doudna, J. A. (2020) 'The promise and challenge of therapeutic genome editing', *Nature*, 578(7794), pp. 229–236. doi: 10.1038/s41586-020-1978-5.
- Doudna, J. A. and Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213). doi: 10.1126/science.1258096.
- Dowling, A., Hill, G. E. and Bonneaud, C. (2020) 'Multiple differences in pathogen- host cell interactions following a bacterial host shift', *Scientific Reports*, pp. 1–12. doi: 10.1038/s41598-020-63714-0.
- Dudek, K. and Szacawa, E. (2020) 'Mycoplasma bovis infections: Occurrence, pathogenesis, diagnosis and control, including prevention and therapy', *Pathogens*, 9(12), pp. 1–3. doi: 10.3390/pathogens9120994.
- Dugar, G. et al. (2018) 'CRISPR RNA-Dependent Binding and Cleavage of Endogenous RNAs by the Campylobacter jejuni Cas9', *Molecular Cell*, 69(5), pp. 893-905.e7. doi: 10.1016/j.molcel.2018.01.032.
- Dupuis, M.-E. et al. (2013) 'CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance', *Nature Communications*. doi: 10.1038/ncomms3087.
- Dupuy, V. et al. (2012) 'Evolutionary History of Contagious Bovine Pleuropneumonia Using Next Generation Sequencing of *Mycoplasma mycoides* Subsp. *mycoides* "Small Colony"', *PLoS ONE*. Edited by O. Neyrolles, 7(10), p. e46821. doi: 10.1371/journal.pone.0046821.
- Duret, S. et al. (1999) 'Gene disruption through homologous recombination in *Spiroplasma citri*: An scm1-disrupted motility mutant is pathogenic', *Journal of Bacteriology*, 181(24), pp. 7449–7456. doi: 10.1128/jb.181.24.7449-7456.1999.
- Dybvig, K. et al. (2010) 'Fewer essential genes in mycoplasmas than previous studies suggest', *FEMS Microbiology Letters*, 311(1), pp. 51–55. doi: 10.1111/j.1574-6968.2010.02078.x.
- Dybvig, K. and Cassell, G. H. (1987) 'Transposition of Gram-Positive Transposon Tn916 in *Acholeplasma laidlawii* and *Mycoplasma pulmonis*', *Science*, 235(11), pp. 14–16.
- Dybvig, K., French, C. T. and Voelker, L. R. L. (2000) 'Construction and use of derivatives of transposon Tn4001 that function in *Mycoplasma pulmonis* and *Mycoplasma arthritidis*', *Journal of Bacteriology*, 182(15), pp. 4343–4347. doi: 10.1128/JB.182.15.4343-4347.2000.
- Dybvig, K., Sitaraman, R. and French, C. T. (1998) 'A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements', *Proceedings of the National Academy of Sciences of the United States of America*, 95(23), pp. 13923–13928. doi: 10.1073/pnas.95.23.13923.
- Dybvig, K. and Voelker, L. L. (1996) 'MOLECULAR BIOLOGY OF MYCOPLASMAS', *Microbiology*, pp. 25–57.
- Dybvig, K. and Woodard, A. (1992) 'Construction of recA mutants of *Acholeplasma laidlawii* by insertional inactivation with a homologous DNA fragment', *Plasmid*, 28(3), pp. 262–266. doi: 10.1016/0147-619X(92)90058-1.
- Edward and Freundt, E. A. (1967) 'PROPOSAL FOR MOLLICUTES AS NAME OF THE CLASS ESTABLISHED FOR THE ORDER MYCOPLASMATALES', *International Journal of Systematic and Evolutionary Microbiology*, 17(3), pp. 267–268.
- Enghiad, B. et al. (2021) 'Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination', *Nature Communications*, 12(1). doi: 10.1038/s41467-021-21275-4.
- Erdmann, S. et al. (2014) 'Inter-viral conflicts that exploit host CRISPR immune systems of *Sulfolobus*', *Molecular Microbiology*, 91(5020), pp. 900–917. doi: 10.1111/mmi.12503.
- Esvelt, K. M. et al. (2013) 'Orthogonal Cas9 proteins for RNA-guided gene regulation and editing', *Nature Methods*, 10(11), pp. 1116–1123. doi: 10.1038/nmeth.2681.
- Evanoff, M. and Komor, A. C. (2019) 'Base editors: Modular tools for the introduction of point mutations in living cells', *Emerging Topics in Life Sciences*, 3(5), pp. 483–491. doi: 10.1042/ETLS20190088.
- Faure, G. et al. (2019) 'CRISPR–Cas in mobile genetic elements: counter-defence and beyond', *Nature Reviews Microbiology*, 17(8), pp. 513–525. doi: 10.1038/s41579-019-0204-7.
- Di Federico, M. et al. (2020) 'Pro-Inflammatory Response of Bovine Polymorphonuclear Cells Induced by *Mycoplasma mycoides* subsp. *mycoides*', *Frontiers in Veterinary Science*, 7(March), pp. 1–7. doi: 10.3389/fvets.2020.00142.
- Figueiredo, J., Sousa Silva, M. and Figueiredo, A. (2018) 'Subtilisin-like proteases in plant defence: the past, the present and beyond', *Molecular Plant Pathology*, 19(4), pp. 1017–1028. doi: 10.1111/mpp.12567.
- Filsinger, G. T. et al. (2021) 'Characterizing the portability of phage-encoded homologous recombination proteins', *Nature Chemical Biology*, 17(4), pp. 394–402. doi: 10.1038/s41589-020-00710-5.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Finger-Bou, M. *et al.* (2020) 'CRISPR with a Happy Ending: Non-Templated DNA Repair for Prokaryotic Genome Engineering', *Biotechnology Journal*, 15(7), pp. 1–7. doi: 10.1002/biot.201900404.
- Fischer, J. R. *et al.* (1997) 'Mycoplasmal Conjunctivitis in Wild Songbirds : The Spread of a New Contagious Disease in a Mobile Host Population', *emerging infectious diseases*, 3(1), pp. 69–72. doi: 10.3201/eid0301.970110.
- Fonfara, I. *et al.* (2014) 'Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems', *Nucleic Acids Research*, 42(4), pp. 2577–2590. doi: 10.1093/nar/gkt1074.
- Fontana, J. *et al.* (2020) 'Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements', *Nature Communications*, 11(1), pp. 1–11. doi: 10.1038/s41467-020-15454-y.
- Frangoul, H. *et al.* (2020) 'CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia', *The new england journal of medicine*.
- Fraser, C. M. *et al.* (1995) 'The Minimal Gene Complement of *Mycoplasma genitalium*', *Science*, 270(5235), pp. 397–404. doi: 10.1126/science.270.5235.397.
- Fu, P. *et al.* (2014) 'Development of a direct competitive ELISA for the detection of *Mycoplasma bovis* infection based on a monoclonal antibody of P48 protein', *BMC Veterinary Research*, 10(1), pp. 1–8. doi: 10.1186/1746-6148-10-42.
- Fu, Y. *et al.* (2013) 'High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells', *Nature Biotechnology*. doi: 10.1038/nbt.2623.High.
- Fu, Y. *et al.* (2014) 'Improving CRISPR-Cas nuclease specificity using truncated guide RNAs', *Nature Biotechnology*. doi: 10.1038/nbt.2808.Improving.
- Fujita, T., Yuno, M. and Fujii, H. (2018) 'EnChIP systems using different CRISPR orthologues and epitope tags', *BMC Research Notes*, 11(1), pp. 1–7. doi: 10.1186/s13104-018-3262-4.
- Ganter, S. *et al.* (2019) 'Proteases as secreted exoproteins in mycoplasmas from ruminant lungs and their impact on surface-exposed proteins', *Applied and Environmental Microbiology*, 85(23), pp. 1–19. doi: 10.1128/AEM.01439-19.
- Gao, L. *et al.* (2020) 'Diverse Enzymatic Activities Mediate Antiviral Immunity in Prokaryotes', *Science*. doi: 10.1126/science.aba0372.Diverse.
- García-Galán, A. *et al.* (2021) 'Genome mosaicism in field strains of *Mycoplasma bovis* as footprints of in-host horizontal chromosomal transfer', *Applied and Environmental Microbiology*, (October). doi: 10.1128/aem.01661-21.
- Garcia-Morales, L. *et al.* (2020) 'A RAGE Based Strategy for the Genome Engineering of the Human Respiratory Pathogen *Mycoplasma pneumoniae*', *ACS Synthetic Biology*, 9(10), pp. 2737–2748. doi: 10.1021/acssynbio.0c00263.
- Garrido, V. *et al.* (2021) 'Engineering a genome-reduced bacterium to eliminate *Staphylococcus aureus* biofilms in vivo', *Molecular Systems Biology*, 17(10). doi: 10.15252/msb.202010145.
- Gasiunas, G. *et al.* (2020) 'A catalogue of biochemically diverse CRISPR-Cas9 orthologs', *Nature Communications*. doi: 10.1038/s41467-020-19344-1.
- Gasparich, G. E. (2010) 'Spiroplasmas and phytoplasmas: Microbes associated with plant hosts', *Biologicals*, 38(2), pp. 193–203. doi: 10.1016/j.biologicals.2009.11.007.
- Gaudelli, N. M. *et al.* (2017) 'Programmable base editing of T to G C in genomic DNA without DNA cleavage', *Nature*, 551(7681), pp. 464–471. doi: 10.1038/nature24644.
- Gautier-Bouchardon, A. V. (2018) 'Antimicrobial Resistance in *Mycoplasma* spp', *Microbiology Spectrum*, 6(4). doi: 10.1128/microbiolspec.arba-0030-2018.
- Gibson, D. G. *et al.* (2010) 'Creation of a bacterial cell controlled by a chemically synthesized genome', *Science*, 329(5987), pp. 52–56. doi: 10.1126/science.1190719.
- Gilbert, L. A. *et al.* (2013) 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes Luke', *Cell*. doi: 10.1016/j.cell.2013.06.044.CRISPR-Mediated.
- Glass, J. I. *et al.* (2006) 'Essential genes of a minimal bacterium', *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), pp. 425–430. doi: 10.1073/pnas.0510013103.
- Gleditsch, D. *et al.* (2019) 'PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures', *RNA Biology*, 16(4), pp. 504–517. doi: 10.1080/15476286.2018.1504546.
- Glew, M. D. *et al.* (1998) 'Expression of the pMGA genes of *Mycoplasma gallisepticum* is controlled by variation in the GAA trinucleotide repeat lengths within the 5' noncoding regions', *Infection and Immunity*, 66(12), pp. 5833–5841. doi:

RÉFÉRENCES BIBLIOGRAPHIQUES

10.1128/jiai.66.12.5833-5841.1998.

Glew, M. D. *et al.* (2000) 'pMGA phenotypic variation in *Mycoplasma gallisepticum* occurs in vivo and is mediated by trinucleotide repeat length variation', *Infection and Immunity*, 68(10), pp. 6027–6033. doi: 10.1128/IAI.68.10.6027-6033.2000.

Goh, M. S. *et al.* (1998) 'Molecular and biochemical analysis of a 105 kDa *Mycoplasma gallisepticum* cytoadhesin (GapA)', *Microbiology*, 144(11), pp. 2971–2978. doi: 10.1099/00221287-144-11-2971.

Goh, Y. J. and Barrangou, R. (2021) 'Portable CRISPR-Cas9 N System for Flexible Genome Engineering', *Aem*, 87(6), pp. 1–16.

Goldfarb, T. *et al.* (2014) 'BREX is a novel phage resistance system widespread in microbial genomes', *EMBO Journal*.

Gómez-Martín, Á. *et al.* (2013) 'Contagious agalactia due to *Mycoplasma* spp. in small dairy ruminants: Epidemiology and prospects for diagnosis and control', *Veterinary Journal*, 198(1), pp. 48–56. doi: 10.1016/j.tvjl.2013.04.015.

Gootenberg, J. S. *et al.* (2017) 'Nucleic acid detection with CRISPR-Cas13a/C2c2', *Science*, 356(6336), pp. 438–442. doi: 10.1126/science.aam9321.

Goret, J. *et al.* (2017) 'Interaction of *Mycoplasma hominis* PG21 with Human Dendritic Cells: Interleukin-23-Inducing Mycoplasma Lipoproteins and Inflammasome Activation of the Cell', *Journal of Bacteriology*. Edited by T. J. Silhavy, 199(15), pp. 1–15. doi: 10.1128/JB.00213-17.

Gourgues, G. *et al.* (2016) 'Complete genome sequence of *Mycoplasma mycoides* subspecies *mycoides* T1/44, a vaccine strain against contagious bovine pleuropneumonia', *Genome Announcements*, 4(2), pp. 2–3. doi: 10.1128/genomeA.00263-16.

Grosjean, H. *et al.* (2014) 'Predicting the Minimal Translation Apparatus: Lessons from the Reductive Evolution of Mollicutes', *PLoS Genetics*, 10(5). doi: 10.1371/journal.pgen.1004363.

Grover, R. K. *et al.* (2014) 'A Structurally Distinct Human *Mycoplasma* Protein that Generically Blocks Antigen-Antibody Union', *Science*, 343(6171), pp. 656–661. doi: 10.1126/science.1246135.

Grünewald, J. *et al.* (2020) 'A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing', *Nature Biotechnology*, 38(7), pp. 861–864. doi: 10.1038/s41587-020-0535-y.

Guegler, C. K. and Laub, M. T. (2021) 'Shutoff of host transcription triggers a toxin- antitoxin system to cleave phage RNA and abort infection', *Molecular Cell*, pp. 1–13. doi: 10.1016/j.molcel.2021.03.027.

Gueguen, E. *et al.* (2014) 'Transcriptional Frameshifting Rescues *Citrobacter rodentium* Type VI Secretion by the Production of Two Length Variants from the Prematurely Interrupted *tssM* Gene', *PLoS Genetics*. Edited by L. Søgaard-Andersen, 10(12), p. e1004869. doi: 10.1371/journal.pgen.1004869.

Guk, K. *et al.* (2017) 'A facile, rapid and sensitive detection of MRSA using a CRISPR-mediated DNA FISH method, antibody-like dCas9/sgRNA complex', *Biosensors and Bioelectronics*, 95(February), pp. 67–71. doi: 10.1016/j.bios.2017.04.016.

Gupta, R. S. *et al.* (2018) 'Phylogenetic framework for the phylum Tenericutes based on genome sequence data: proposal for the creation of a new order Mycoplasmoidales ord. nov., containing two new families Mycoplasmoidaceae fam. nov. and Metamycoplasmataceae fam. nov. harbouring Epe', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 111(9), pp. 1583–1630. doi: 10.1007/s10482-018-1047-3.

Haft, D. H. *et al.* (2005) 'A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes', *PLoS Computational Biology*, 1(6), pp. 0474–0483. doi: 10.1371/journal.pcbi.0010060.

Hale, H. H. *et al.* (1962) 'Bovine mastitis caused by a *Mycoplasma* species.', *The Cornell veterinarian*, 52, pp. 582–91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/13952069>.

Hall, S. D. and Kolodner, R. D. (1994) 'Homologous pairing and strand exchange promoted by the *Escherichia coli* RecT protein', *Biochemistry*, 91, pp. 3205–3209. doi: 10.1073/pnas.91.8.3205.

Hamilton, T. A. *et al.* (2019) 'Efficient inter-species conjugative transfer of a CRISPR nuclease for targeted bacterial killing', *Nature Communications*, 10(1), pp. 1–9. doi: 10.1038/s41467-019-12448-3.

Hantke, K. (2020) 'Compilation of *Escherichia coli* K-12 outer membrane phage receptors – their function and some historical remarks', *FEMS Microbiology Letters*, 367(2), pp. 1–9. doi: 10.1093/femsle/fnaa013.

Hatfull, G. F. and Hendrix, R. W. (2011) 'Bacteriophages and their genomes', *Current Opinion in Virology*, 1(4), pp. 298–303. doi: 10.1016/j.coviro.2011.06.009.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Hedreyda, C. T. and Krause, D. C. (1995) 'Identification of a possible cytoadherence regulatory locus in *Mycoplasma pneumoniae*', *Infection and Immunity*, 63(9), pp. 3479–3483. doi: 10.1128/iai.63.9.3479-3483.1995.
- Heler, R. *et al.* (2015) 'Cas9 specifies functional viral targets during CRISPR-Cas adaptation', *Nature*, 519(7542), pp. 199–202. doi: 10.1038/nature14245.
- Hennigan, S. L. *et al.* (2012) 'Detection and differentiation of avian mycoplasmas by surface-enhanced raman spectroscopy based on a silver nanorod array', *Applied and Environmental Microbiology*, 78(6), pp. 1930–1935. doi: 10.1128/AEM.07419-11.
- Hernalsteens, J. *et al.* (1978) 'Mutagenesis by insertion of the drug resistance transposon Tn7 applied to the Ti plasmid of *Agrobacterium tumefaciens*', *Plasmid*, 1(2), pp. 218–225. doi: 10.1016/0147-619X(78)90040-9.
- Hernalsteens, J. P. *et al.* (no date) 'A technique for mutagenesis by transposon insertion, applicable to most Gram-negative bacteria', *Arch Int Physiol Biochim*.
- Hill, V. *et al.* (2021) 'Minimalistic mycoplasmas harbor different functional toxin-antitoxin systems', *PLOS Genetics*. Edited by J. Stülke, 17(10), p. e1009365. doi: 10.1371/journal.pgen.1009365.
- Hirano, S. *et al.* (2019) 'Structural basis for the promiscuous PAM recognition by *Corynebacterium diphtheriae* Cas9', *Nature Communications*, 10(1), pp. 1–11. doi: 10.1038/s41467-019-09741-6.
- Horvath, P. *et al.* (2008) 'Diversity, Activity, and Evolution of CRISPR Loci in *Streptococcus thermophilus*', *Journal of Bacteriology*, 190(4), pp. 1401–1412. doi: 10.1128/JB.01415-07.
- Horvath, P. *et al.* (2009) 'Comparative analysis of CRISPR loci in lactic acid bacteria genomes', *International Journal of Food Microbiology*, 131(1), pp. 62–70. doi: 10.1016/j.ijfoodmicro.2008.05.030.
- Horvath, P. and Barrangou, R. (2010) 'CRISPR/Cas, the immune system of Bacteria and Archaea', *Science*, 327(5962), pp. 167–170. doi: 10.1126/science.1179555.
- Houte, S. Van *et al.* (2016) 'The diversity-generating benefits of a prokaryotic adaptive immune system', *Nature*. doi: 10.1038/nature17436.The.
- Hsu, P. D. *et al.* (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology*. doi: 10.1038/nbt.2647.DNA.
- Hsu, Patrick D., Lander, E. S. and Zhang, F. (2014) 'Development and Applications of CRISPR-Cas9 for Genome Engineering', *Cell*. doi: 10.2307/4589497.
- Hsu, Patrick D., Lander, E. S. and Zhang, F. (2014) 'Development and Applications of CRISPR-Cas9 for Genome Engineering Patrick', *Cell*. doi: 10.1016/j.cell.2014.05.010.
- Hu, X. *et al.* (2018) 'Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice', *Plant Biotechnology Journal*, 16(1), pp. 292–297. doi: 10.1111/pbi.12771.
- Huang, H., Song, X. and Yang, S. (2019) 'Development of a RecE/T-Assisted CRISPR–Cas9 Toolbox for *Lactobacillus*', *Biotechnology Journal*, 14(7), pp. 1–12. doi: 10.1002/biot.201800690.
- Huang, T. P., Newby, G. A. and Liu, D. R. (2021) 'Precision genome editing using cytosine and adenine base editors in mammalian cells', *Nature Protocols*, 16(2), pp. 1089–1128. doi: 10.1038/s41596-020-00450-9.
- Huang, Y. *et al.* (2017) 'Recombineering using RecET in *Corynebacterium glutamicum* ATCC14067 via a self-excisable cassette', *Scientific Reports*, 7(1), pp. 1–8. doi: 10.1038/s41598-017-08352-9.
- Hudson, P. *et al.* (2006) 'Identification of a virulence-associated determinant, dihydrolipoamide dehydrogenase (lpd), in *Mycoplasma gallisepticum* through in vivo screening of transposon mutants', *Infection and Immunity*, 74(2), pp. 931–939. doi: 10.1128/IAI.74.2.931-939.2006.
- Hutchison, C. A. *et al.* (2016) 'Design and synthesis of a minimal bacterial genome', *Science*, 351(6280). doi: 10.1126/science.aad6253.
- Hutchison, C. A. *et al.* (2019) 'Polar Effects of Transposon Insertion into a Minimal Bacterial Genome', *Journal of Bacteriology*.
- Hwang, W. Y. *et al.* (2013) 'Efficient In Vivo Genome Editing Using RNA-Guided Nucleases Woong', *Nature biotechnology*, 31(3), pp. 227–229. doi: 10.1038/nbt.2501.Efficient.
- Ibraheim, R. *et al.* (2018) 'All-in-one adeno-associated virus delivery and genome editing by *Neisseria meningitidis* Cas9 in vivo', *Genome Biology*, 19(1), pp. 1–11. doi: 10.1186/s13059-018-1515-0.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Indiková, I. *et al.* (2013) 'Role of the gapa and crma cytoadhesins of *Mycoplasma gallisepticum* in promoting virulence and host colonization', *Infection and Immunity*, 81(5), pp. 1618–1624. doi: 10.1128/IAI.00112-13.
- Indikova, I., Vronka, M. and Szostak, M. P. (2014) 'First identification of proteins involved in motility of *Mycoplasma gallisepticum*', *Veterinary Research*, 45(1), pp. 1–13. doi: 10.1186/s13567-014-0099-2.
- Ipoutcha, T. *et al.* (2019) 'Multiple Origins and Specific Evolution of CRISPR / Cas9 Systems in Minimal Bacteria (Mollicutes)', *Frontiers in microbiology*, 10(November), pp. 1–14. doi: 10.3389/fmicb.2019.02701.
- Iranzo, J. *et al.* (2013) 'Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context', *Journal of Bacteriology*, 195(17), pp. 3834–3844. doi: 10.1128/JB.00412-13.
- Isaev, A. B., Musharova, O. S. and Severinov, K. V. (2021) 'Microbial Arsenal of Antiviral Defenses – Part I', *Biochemistry*, 86(3).
- Ishag, H. Z. A. *et al.* (2016) 'A replicating plasmid-based vector for GFP expression in *Mycoplasma hyopneumoniae*', *Genetics and Molecular Research*, 15(2), pp. 1–8. doi: 10.4238/gmr.15027832.
- Ishag, H. Z. A. *et al.* (2017a) 'Development of oriC-plasmids for use in *Mycoplasma hyorhinis*', *Scientific Reports*, 7(1), pp. 1–10. doi: 10.1038/s41598-017-10519-3.
- Ishag, H. Z. A. *et al.* (2017b) 'E. coli recA gene improves gene targeted homologous recombination in *Mycoplasma hyorhinis*', *Journal of Microbiological Methods*, 136, pp. 49–56. doi: 10.1016/j.mimet.2017.03.004.
- Ishfaq, M. *et al.* (2020) 'Current status of vaccine research, development, and challenges of vaccines for *Mycoplasma gallisepticum*', *Poultry Science*, 99(9), pp. 4195–4202. doi: 10.1016/j.psj.2020.06.014.
- Ishino, Y. *et al.* (1987) 'Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product', *Journal of Bacteriology*, 169(12), pp. 5429–5433. doi: 10.1128/jb.169.12.5429-5433.1987.
- Ivanov, I. E. *et al.* (2020) 'Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling', *Proceedings of the National Academy of Sciences of the United States of America*, pp. 1–8. doi: 10.1073/pnas.1913445117.
- Iyer, L. M., Koonin, E. V. and Aravind, L. (2002) 'Classification and evolutionary history of the single-strand annealing proteins, RecT, Red β , ERF and RAD52', *BMC Genomics*, 3, pp. 1–11. doi: 10.1186/1471-2164-3-8.
- Janis, C. *et al.* (2005) 'Versatile Use of oriC Plasmids for Functional Genomics of *Mycoplasma capricolum* subsp. *capricolum*', *Society*, 71(6), pp. 2888–2893. doi: 10.1128/AEM.71.6.2888.
- Janis, C. *et al.* (2008) 'Unmarked insertional mutagenesis in the bovine pathogen *Mycoplasma mycoides* subsp. *mycoides* SC', *Microbiology*, 154(Pt 8), pp. 2427–2436. doi: 10.1099/mic.0.2008/017640-0.Unmarked.
- Jansen, R. *et al.* (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes', *Molecular Microbiology*, 43(6), pp. 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x.
- Jensen, J. S., Blom, J. and Lind, K. (1994) 'Intracellular location of *Mycoplasma genitalium* in cultured Vero cells as demonstrated by electron microscopy', *International journal of experimental pathology*, 75(2), pp. 91–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8199010> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2002111>.
- Jiang, F. *et al.* (2016) 'Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage', *Science*, 351(6275), pp. 867–871. doi: 10.1126/science.aad8282.
- Jiang, W., Maniv, I. *et al.* (2013) 'Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids', *PLoS Genetics*, 9(9). doi: 10.1371/journal.pgen.1003844.
- Jiang, W., Bikard, D. *et al.* (2013) 'RNA-guided editing of bacterial genomes using CRISPR-Cas systems', *Nature Biotechnology*, 31(3). doi: 10.1038/nbt.2508.
- Jiang, Y. *et al.* (2017) 'CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*', *Nature Communications*, 8(May), pp. 1–11. doi: 10.1038/ncomms15179.
- Jimbo, S. *et al.* (2017) 'Effect of *Mycoplasma bovis* on bovine neutrophils', *Veterinary Immunology and Immunopathology*, 188, pp. 27–33. doi: 10.1016/j.vetimm.2017.04.011.
- Jinek, M. *et al.* (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.
- Jinek, M. *et al.* (2013) 'RNA-programmed genome editing in human cells', *eLife*, 2013(2), pp. 1–9. doi: 10.7554/eLife.00471.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Jinek, M. *et al.* (2014) 'Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation', *Science*, 343(6176), pp. 1–28.
- Jo, D. H. *et al.* (2019) 'Long-Term Effects of In Vivo Genome Editing in the Mouse Retina Using *Campylobacter jejuni* Cas9 Expressed via Adeno-Associated Virus', *Molecular Therapy*, 27(1), pp. 130–136. doi: 10.1016/j.jymthe.2018.10.009.
- Jordan, A. *et al.* (2021) '*Mycoplasma bovis* outbreak in New Zealand cattle: An assessment of transmission trends using surveillance data', *Transboundary and Emerging Diseases*, 68(6), pp. 3381–3395. doi: 10.1111/tbed.13941.
- Jores, J., Schieck, E., *et al.* (2019) 'In vivo role of capsular polysaccharide in *Mycoplasma mycoides*', *The Journal of Infectious Diseases*, 219(10), pp. 1559–1563. doi: 10.1093/infdis/jiy713.
- Jores, J., Ma, L., *et al.* (2019) 'Removal of a Subset of Non-essential Genes Fully Attenuates a Highly Virulent *Mycoplasma* Strain', *Frontiers in Microbiology*, 10(APR), pp. 1–13. doi: 10.3389/fmicb.2019.00664.
- Jores, J. *et al.* (2020) 'Contagious Bovine and Caprine Pleuropneumonia: a research community's recommendations for the development of better vaccines', *npj Vaccines*, 5(1). doi: 10.1038/s41541-020-00214-2.
- Jores, J., Mariner, J. C. and Naessens, J. (2013) 'Development of an improved vaccine for contagious bovine pleuropneumonia: An African perspective on challenges and proposed actions', *Veterinary Research*, 44(1), pp. 1–5. doi: 10.1186/1297-9716-44-122.
- Ka, D. *et al.* (2020) 'Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD⁺ degradation', *Nature Communications*. doi: 10.1038/s41467-020-16703-w.
- Karas, B. J. *et al.* (2013) 'Direct transfer of whole genomes from bacteria to yeast', *Nature Methods*, 10(5), pp. 410–412. doi: 10.1038/nmeth.2433.
- Karikó, K. (2021) 'Modified uridines are the key to a successful message', *Nature Reviews Immunology*, 21(10), pp. 619–619. doi: 10.1038/s41577-021-00608-w.
- Karvelis, T. *et al.* (2015) 'Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements', *Genome Biology*, 16(1), pp. 1–13. doi: 10.1186/s13059-015-0818-7.
- Karvelis, T., Gasiunas, G. and Siksnys, V. (2013) 'Programmable DNA cleavage in vitro by Cas9', *Biochemical Society Transactions*, 41(6), pp. 1401–1406. doi: 10.1042/BST20130164.
- Karvelis, T., Young, J. K. and Siksnys, V. (2019) *A pipeline for characterization of novel Cas9 orthologs*. 1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc. doi: 10.1016/bs.mie.2018.10.021.
- Kasman, L. and Porter, L. (2021) *Bacteriophages*.
- Kazlauskienė, M. *et al.* (2017) 'A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems', *Science*, 0100(June), pp. 1–11.
- Kim, S. K. *et al.* (2017) 'CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in *Escherichia coli*', *Microbial Cell Factories*, 16(1), p. 188. doi: 10.1186/s12934-017-0802-x.
- King, K. W. and Dybvig, K. (1991) 'Plasmid transformation of *Mycoplasma mycoides* subspecies *mycoides* is promoted by high concentrations of polyethylene glycol', *Plasmid*, 26(2), pp. 108–115. doi: 10.1016/0147-619X(91)90050-7.
- Kleinstiver, Benjamin P. *et al.* (2015) 'Broadening *Staphylococcus aureus* Cas9 Targeting Range by Modifying PAM Recognition', *Nature Biotechnology*, 33(12), pp. 1293–1298. doi: 10.1038/nbt.3404.Broadening.
- Kleinstiver, Benjamin P. *et al.* (2015) 'Engineered CRISPR-Cas9 nucleases with altered PAM specificities', *Nature*, 523(7561), pp. 481–485. doi: 10.1038/nature14592.
- Kleven, S. H. (2008) 'Control of Avian *Mycoplasma* Infections in Commercial Poultry', *Avian Diseases Digest*, 3(3), pp. e1–e1. doi: 10.1637/8424.1.
- Knudtson, K. L. and Minion, F. C. (1993) 'Construction of Tn4001lac derivatives to be used as promoter probe vectors in mycoplasmas', *Gene*, 137(2), pp. 217–222. doi: 10.1016/0378-1119(93)90009-R.
- Koblan, L. W. *et al.* (2021) 'Base editor treats progeria in mice', *Nature*. doi: 10.1038/d41586-021-01114-8.
- Koike-Yusa, H. *et al.* (2014) 'Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library', *Nature Biotechnology*, 32(3), pp. 267–273. doi: 10.1038/nbt.2800.
- Komor, A. C. *et al.* (2016) 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature*, 533, pp. 420–424. doi: 10.1038/nature17946.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Koonin, E. V. (2019) 'CRISPR: a new principle of genome engineering linked to conceptual shifts in evolutionary biology', *Biology and Philosophy*, 34(1). doi: 10.1007/s10539-018-9658-7.
- Koonin, E. V., Makarova, K. S. and Zhang, F. (2017) 'Diversity, classification and evolution of CRISPR-Cas systems', *current opinion in Microbiology*, 176(12), pp. 139–148. doi: 10.1016/j.mib.2017.05.008.Diversity.
- Koonin, E. V. et al. (2017) 'Comparative genomics of defense systems in archaea and bacteria', *Annual Review of Microbiology*, 41(6), pp. 4360–4377. doi: 10.1093/nar/gkt157.
- Krishnakumar, R. et al. (2010) 'Targeted chromosomal knockouts in *Mycoplasma pneumoniae*', *Applied and Environmental Microbiology*, 76(15), pp. 5297–5299. doi: 10.1128/AEM.00024-10.
- Krishnakumar, R. et al. (2013) 'Transfer RNA Misidentification Scrambles Sense Codon Recoding', *ChemBioChem*, 14(15), pp. 1967–1972. doi: 10.1002/cbic.201300444.
- Kronheim, S. et al. (2018) 'A chemical defence against phage infection', *Nature*. doi: 10.1038/s41586-018-0767-x.
- Kumar, S. et al. (2018) 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.
- Kundert, K. et al. (2019) 'Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs', *Nature Communications*, 10(1), pp. 1–11. doi: 10.1038/s41467-019-09985-2.
- Labroussaa, F. et al. (2016) 'Impact of donor-recipient phylogenetic distance on bacterial genome transplantation', *Nucleic Acids Research*, 44(17), pp. 8501–8511. doi: 10.1093/nar/gkw688.
- Lam, K. M. (2005) 'Chemotaxis in *Mycoplasma gallisepticum*', *Avian Diseases*, 49(1), pp. 152–154. doi: 10.1637/7232-070604R.
- Lartigue, C. et al. (2002) 'New plasmid vectors for specific gene targeting in *Spiroplasma citri*', *Plasmid*. doi: 10.1016/S0147-619X(02)00121-X.
- Lartigue, C. et al. (2003) 'Host specificity of mollicutes oriC plasmids: Functional analysis of replication origin', *Nucleic Acids Research*, 31(22), pp. 6610–6618. doi: 10.1093/nar/gkg848.
- Lartigue, C. et al. (2007) 'Genome transplantation in bacteria: Changing one species to another', *Science*, 317(5838), pp. 632–638. doi: 10.1126/science.1144622.
- Lartigue, C. et al. (2009) 'Creating bacterial strains from genomes that have been cloned and engineered in yeast', *Science*, 325(5948), pp. 1693–1696. doi: 10.1126/science.1173759.
- Lee, J. K. J. et al. (2018) 'Directed evolution of CRISPR-Cas9 to increase its specificity', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-05477-x.
- Lee, S. W., Browning, G. F. and Markham, P. F. (2008) 'Development of a replicable oriC plasmid for *Mycoplasma gallisepticum* and *Mycoplasma imitans*, and gene disruption through homologous recombination in *M. gallisepticum*', *Microbiology*. doi: 10.1099/mic.0.2008/019208-0.
- Leenay, R. T. et al. (2017) 'Identifying and visualizing functional PAM diversity across CRISPR-Cas systems', *Molecular Cell*, 62(1), pp. 137–147. doi: 10.1016/j.molcel.2016.02.031.Identifying.
- Levisohn, S. and Kleven, S. H. (2000) 'Avian mycoplasmosis (*Mycoplasma gallisepticum*)', *Scientific and Technical Review of the Office International des Epizooties*, 19(2), pp. 425–442.
- Levy, A. et al. (2015) 'CRISPR adaptation biases explain preference for acquisition of foreign DNA', *Nature*, 520(7548), pp. 505–510. doi: 10.1038/nature14302.CRISPR.
- Ley, D. H. (2013) *Diseases Poultry, Diseases of Poultry*.
- Li, H.-J. et al. (2016) 'Characterization of a New S8 serine Protease from Marine Sedimentary Photobacterium sp. A5-7 and the Function of Its Protease-Associated Domain', *Frontiers in Microbiology*, 7(DEC), pp. 1–11. doi: 10.3389/fmicb.2016.02016.
- Li, J. et al. (2013) 'Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9', *Nat Biotechnol*, 31(8), pp. 688–691. doi: 10.1038/nbt.2654.Multiplex.
- Li, K. et al. (2018) 'Development of an Efficient Genome Editing Tool in Bacillus', *Applied and Environmental Microbiology*, 84(6), pp. 1–15.
- Li, L. et al. (2013) 'Leucyl-tRNA synthetase editing domain functions as a molecular rheostat to control codon ambiguity in *Mycoplasma* pathogens', *Proceedings of the National Academy of Sciences*, 110(10), pp. 3817–3822. doi:

RÉFÉRENCES BIBLIOGRAPHIQUES

10.1073/pnas.1218374110.

Li, Q. *et al.* (2019) 'CRISPR-Cas9 D10A nickase-assisted base editing in the solvent producer *Clostridium beijerinckii*', *Biotechnology and Bioengineering*, 116(6), pp. 1475–1483. doi: 10.1002/bit.26949.

Li, Y. *et al.* (2015) 'Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 mediated genome editing', *Metabolic Engineering*, 31, pp. 13–21. doi: 10.1016/j.mbs.2015.06.006.

Li, Y. *et al.* (2016) 'Changes in pathogenicity and immunogenicity of *Mycoplasma mycoides subsp. mycoides* strains revealed by comparative genomics analysis', *Scientific Reports*, 6(1), p. 19081. doi: 10.1038/srep19081.

Lino, C. A. *et al.* (2018) 'Delivering crispr: A review of the challenges and approaches', *Drug Delivery*, 25(1), pp. 1234–1257. doi: 10.1080/10717544.2018.1474964.

Liu, L. *et al.* (2000) 'GAA trinucleotide repeat region regulates M9/pMGA gene expression in *Mycoplasma gallisepticum*', *Infection and Immunity*, 68(2), pp. 871–876. doi: 10.1128/IAI.68.2.871-876.2000.

Liu, L. *et al.* (2019) 'Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race', *Molecular Cell*, 73(3), pp. 611–620.e3. doi: 10.1016/j.molcel.2018.11.011.

Liu, M. *et al.* (2008) 'Displaying the protein of *Mycoplasma gallisepticum* agglutinin on the cell surface of *Bacillus thuringiensis* with the S-layer protein', *Veterinary Microbiology*, 130(1–2), pp. 99–106. doi: 10.1016/j.vetmic.2007.12.007.

Liu, M. *et al.* (2019) 'Methodologies for improving HDR efficiency', *Frontiers in Genetics*, 10(JAN), pp. 1–9. doi: 10.3389/fgene.2018.00691.

Liu, T. Y. *et al.* (2021) 'Accelerated RNA detection using tandem CRISPR nucleases', *Nature Chemical Biology*, 17(9), pp. 982–988. doi: 10.1038/s41589-021-00842-2.

Liu, T. Y., Iavarone, A. T. and Doudna, J. A. (2017) 'RNA and DNA Targeting by a Reconstituted *Thermus thermophilus* Type III-A CRISPR-Cas System', *PLoS ONE*, pp. 1–20. doi: 10.1371/journal.pone.0170552.

Liu, W. *et al.* (2019) 'Surface proteins mhp390 (P68) contributes to cilium adherence and mediates inflammation and apoptosis in *Mycoplasma hyopneumoniae*', *Microbial Pathogenesis*, 126, pp. 92–100. doi: 10.1016/j.micpath.2018.10.035.

Lluch-Senar, M. *et al.* (2015) 'Defining a minimal cell: essentiality of small <scp>ORF</scp> s and nc <scp>RNA</scp> s in a genome-reduced bacterium', *Molecular Systems Biology*, 11(1), p. 780. doi: 10.15252/msb.20145558.

Lo, S. C. *et al.* (1992) 'Mycoplasma penetrans sp. nov., from the urogenital tract of patients with AIDS', *International Journal of Systematic Bacteriology*, 42(3), pp. 357–364. doi: 10.1099/00207713-42-3-357.

Lo, W.-S. *et al.* (2013) 'Comparison of Metabolic Capacities and Inference of Gene Content Evolution in Mosquito-Associated *Spiroplasma diminutum* and *S. taiwanense*', *Genome Biology and Evolution*, 5(8), pp. 1512–1523. doi: 10.1093/gbe/evt108.

Loo, C. Y., Corliss, D. A. and Ganeshkumar, N. (2000) 'Streptococcus gordonii biofilm formation: Identification of genes that code for biofilm phenotypes', *Journal of Bacteriology*, 182(5), pp. 1374–1382. doi: 10.1128/JB.182.5.1374-1382.2000.

Luo, M. L. *et al.* (2015) 'Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression', *Nucleic Acids Research*, 43(1), pp. 674–681. doi: 10.1093/nar/gku971.

Ma, D. *et al.* (2019) 'Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information', *Nature Communications*, 10(1). doi: 10.1038/s41467-019-08395-8.

Maglennon, G. A. *et al.* (2013) 'Development of a self-replicating plasmid system for *Mycoplasma hyopneumoniae*', *Veterinary Research*, 44(1), pp. 1–10. doi: 10.1186/1297-9716-44-63.

Mahairas, G. G., Jian, C. and Minion, F. C. (1990) 'Development of a cloning system in *Mycoplasma pulmonis*', *Gene*.

Mahairas, G. G. and Minion, F. C. (1989) 'Random Insertion of the Gentamicin Resistance Transposon Tn4001 in *Mycoplasma pulmonis*', *Plasmid*.

Mahendra, C. *et al.* (2020) 'Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer', *Nature Microbiology*, 5(March). doi: 10.1038/s41564-020-0692-2.

Maji, B. *et al.* (2017) 'Multidimensional chemical control of CRISPR-Cas9 Basudeb', *Nature Chemical Biology*. doi: 10.1016/j.natcb.2017.03.040.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Makarova, K. S. *et al.* (2006) 'A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action', *Biology Direct*, 1, pp. 1–26. doi: 10.1186/1745-6150-1-7.
- Makarova, K. S. *et al.* (2011) 'Evolution and classification of the CRISPR-Cas systems', *Nature Reviews Microbiology*, 9(6), pp. 467–477. doi: 10.1038/nrmicro2577.
- Makarova, K. S. *et al.* (2015) 'An updated evolutionary classification of CRISPR-Cas systems', *Nature Reviews Microbiology*, 13(11), pp. 722–736. doi: 10.1038/nrmicro3569.
- Makarova, K. S. *et al.* (2020) 'Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants', *Nature Reviews Microbiology*, 18(2), pp. 67–83. doi: 10.1038/s41579-019-0299-x.
- Makarova, K. S., Wolf, Y. I. and Koonin, E. V. (2013) 'Comparative genomics of defense systems in archaea and bacteria', *Nucleic Acids Research*, 41(8), pp. 4360–4377. doi: 10.1093/nar/gkt157.
- Malember-Maher, S. *et al.* (2020) 'When a Palearctic bacterium meets a Nearctic insect vector: Genetic and ecological insights into the emergence of the grapevine Flavescence dorée epidemics in Europe', *PLoS Pathogens*. Edited by S. A. Hogenhout, 16(3), p. e1007967. doi: 10.1371/journal.ppat.1007967.
- Mali, P. *et al.* (2013) 'RNA-guided human genome engineering via Cas9', *Science*, 339(6121), pp. 823–826. doi: 10.1126/science.1232033.
- Maniloff, J. (2002) 'Phylogeny and evolution', *Nematology Monographs and Perspectives*, (chapter 2), pp. 693–733.
- Manna, D. *et al.* (2019) 'A Singular System with Precise Dosing and Spatiotemporal Control of CRISPR-Cas9', *Angewandte Chemie - International Edition*, 58(19), pp. 6285–6289. doi: 10.1002/anie.201900788.
- Manso-Silván, L. *et al.* (2009) '*Mycoplasma leachii* sp. nov. as a new species designation for *Mycoplasma* sp. bovine group 7 of Leach, and reclassification of *Mycoplasma mycoides* subsp. *mycoides* LC as a serovar of *Mycoplasma mycoides* subsp. *capri*', *International Journal of Systematic and Evolutionary Microbiology*, 59(6), pp. 1353–1358. doi: 10.1099/ijs.0.005546-0.
- Manso-Silván, L., Perrier, X. and Thiaucourt, F. (2007) 'Phylogeny of the *Mycoplasma mycoides* cluster based on analysis of five conserved protein-coding sequences and possible implications for the taxonomy of the group', *International Journal of Systematic and Evolutionary Microbiology*, 57(10), pp. 2247–2258. doi: 10.1099/ijs.0.64918-0.
- Marenda, M. *et al.* (2006) 'A new integrative conjugative element occurs in *Mycoplasma agalactiae* as chromosomal and free circular forms', *Journal of Bacteriology*, 188(11), pp. 4137–4141. doi: 10.1128/JB.00114-06.
- Marino, N. D. *et al.* (2020) 'Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies', *Nature Methods*, 17(5), pp. 471–479. doi: 10.1038/s41592-020-0771-6.
- Mariscal, A. M. *et al.* (2016) 'All-in-one construct for genome engineering using Cre-lox technology', *DNA Research*, 23(3), pp. 263–270. doi: 10.1093/dnares/dsw015.
- Mariscal, A. M. *et al.* (2018) 'Tuning Gene Activity by Inducible and Targeted Regulation of Gene Expression in Minimal Bacterial Cells', *ACS Synthetic Biology*, 7(6), pp. 1538–1552. doi: 10.1021/acssynbio.8b00028.
- Marraffini, L. A. and Sontheimer, E. J. (2008) 'CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA', *Science*, 322(5909), pp. 1843–1845. doi: 10.1126/science.1165771.
- Marraffini, L. A. and Sontheimer, E. J. (2009) 'Invasive DNA, Chopped and in the CRISPR', *Structure*, 17(6), pp. 786–788. doi: 10.1016/j.str.2009.05.002.
- Masukagami, Y. *et al.* (2013) 'The *Mycoplasma gallisepticum* virulence factor lipoprotein MslA is a novel polynucleotide binding protein', *Infection and Immunity*, 81(9), pp. 3220–3226. doi: 10.1128/IAI.00365-13.
- Matteau, D. *et al.* (2017) 'Inferring the Minimal Genome of *Mesoplasma florum* by Comparative Genomics and Transposon Mutagenesis', *Aem*, 83(7), pp. 1–16.
- Maunsell, F. P. and Chase, C. (2019) '*Mycoplasma bovis*: Interactions with the Immune System and Failure to Generate an Effective Immune Response', *Veterinary Clinics of North America - Food Animal Practice*, 35(3), pp. 471–483. doi: 10.1016/j.cvfa.2019.08.003.
- May, M., Balish, M. F. and Blanchard, A. (2014) *The Order Mycoplasmatales*. doi: https://doi.org/10.1007/978-3-642-30120-9_289.
- Mazin, P. V. *et al.* (2014) 'Transcriptome analysis reveals novel regulatory mechanisms in a genome-reduced bacterium', *Nucleic Acids Research*. doi: 10.1093/nar/gku976.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Mbulu, R. S. *et al.* (2004) 'Contagious bovine pleuropneumonia (CBPP) caused by vaccine strain T1/44 of *Mycoplasma mycoides subsp. mycoides* SC', *Veterinary Microbiology*, 98(3-4), pp. 229-234. doi: 10.1016/j.vetmic.2003.11.007.
- McAuliffe, L. *et al.* (2006) 'Biofilm formation by mycoplasma species and its role in environmental persistence and survival', *Microbiology*, 152(4), pp. 913-922. doi: 10.1099/mic.0.28604-0.
- McAuliffe, L. *et al.* (2008) 'Biofilm-grown *Mycoplasma mycoides subsp. mycoides* SC exhibit both phenotypic and genotypic variation compared with planktonic cells', *Veterinary Microbiology*, 129(3-4), pp. 315-324. doi: 10.1016/j.vetmic.2007.11.024.
- McKenna, S. *et al.* (2021) 'The Role of Streptococcal Cell-Envelope Proteases in Bacterial Evasion of the Innate Immune System', *Journal of Innate Immunity*, pp. 1-20. doi: 10.1159/000516956.
- Meaden, S. *et al.* (2021) 'High viral abundance and low diversity are associated with increased CRISPR-Cas prevalence across microbial ecosystems', *Current Biology*, pp. 1-8. doi: 10.1016/j.cub.2021.10.038.
- Millman, A. *et al.* (2020) 'Bacterial Retrons Function In Anti-Phage Defense Bacterial Retrons Function In Anti-Phage Defense', *Cell*, pp. 1-11. doi: 10.1016/j.cell.2020.09.065.
- Mir, A. *et al.* (2018) 'Type II-C CRISPR-Cas9 Biology, Mechanism, and Application', *ACS Chemical Biology*, 13(2), pp. 357-365. doi: 10.1021/acscchembio.7b00855.
- Mitiku, F. *et al.* (2018) 'The major membrane nuclease MnuA degrades neutrophil extracellular traps induced by *Mycoplasma bovis*', *Veterinary Microbiology*, 218(March), pp. 13-19. doi: 10.1016/j.vetmic.2018.03.002.
- Mojica, F. J. M. *et al.* (2005) 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements', *Journal of Molecular Evolution*, 60(2), pp. 174-182. doi: 10.1007/s00239-004-0046-3.
- Mojica, F. J. M. *et al.* (2009) 'Short motif sequences determine the targets of the prokaryotic CRISPR defence system', *Microbiology*, 155(3), pp. 733-740. doi: 10.1099/mic.0.023960-0.
- Montero-Blay, A. *et al.* (2019) 'SynMyco transposon: engineering transposon vectors for efficient transformation of minimal genomes', *DNA research : an international journal for rapid publication of reports on genes and genomes*, 26(4), pp. 327-339. doi: 10.1093/dnares/dsz012.
- Morley, D. *et al.* (2016) 'Host diversity limits the evolution of parasite local adaptation', *molecular ecology*. doi: 10.1111/mec.13917.
- Morowitz, H. J. and Wallace, D. C. (1973) 'Genome Size and Life Cycle of the Mycoplasma', *Annals of the New York Academy of Sciences*, 225(1), pp. 62-73. doi: 10.1111/j.1749-6632.1973.tb45637.x.
- Mougiakos, I. *et al.* (2017) 'Efficient Genome Editing of a Facultative Thermophile Using Mesophilic spCas9', *ACS Synthetic Biology*, 6(5), pp. 849-861. doi: 10.1021/acssynbio.6b00339.
- Murphy, K. C. (2016) 'λ Recombination and Recombineering', *EcoSal Plus*, 7(1). doi: 10.1128/ecosalplus.esp-0011-2015.
- Murray, S. M. and McKay, P. F. (2021) 'Chlamydia trachomatis: Cell biology, immunology and vaccination', *Vaccine*, 39(22), pp. 2965-2975. doi: 10.1016/j.vaccine.2021.03.043.
- Mushegian, A. R. and Koonin, E. V. (1996) 'A minimal gene set for cellular life derived by comparison of complete bacterial genomes.', *Proceedings of the National Academy of Sciences*, 93(19), pp. 10268-10273. doi: 10.1073/pnas.93.19.10268.
- Mutaqin, K. *et al.* (2011) 'Selection and characterization of Spiroplasma citri mutants by random transposome mutagenesis', *canadian journal of microbiology*, 532, pp. 525-532. doi: 10.1139/W11-026.
- Muto, A. and Osawa, S. (1987) 'The guanine and cytosine content of genomic DNA and bacterial evolution', *Proceedings of the National Academy of Sciences of the United States of America*, 84(1), pp. 166-169. doi: 10.1073/pnas.84.1.166.
- Nakane, D. and Miyata, M. (2009) 'Cytoskeletal Asymmetrical Dumbbell Structure of a Gliding Mycoplasma, *Mycoplasma gallisepticum*, Revealed by Negative-Staining Electron Microscopy', *Journal of Bacteriology*, 191(10), pp. 3256-3264. doi: 10.1128/JB.01823-08.
- Neimark, H. (1983) 'Evolution of mycoplasmas and genome losses', *Yale Journal of Biology and Medicine*, 56(5-6), pp. 377-383.
- Nekrasov, V. *et al.* (2013) 'Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9', *Nature Biotechnology*, 31(8), pp. 688-691. doi: 10.1038/nbt.2654.
- Nelson, K. E., Richardson, D. L. and Dougherty, B. A. (1997) 'Tn 916 Transposition in *Haemophilus influenzae*, Preferential Insertion into Noncoding DNA', *Microbial & Comparative Genomics*, 2(4), pp. 313-321. doi: 10.1089/omi.1.1997.2.313.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Nicholas, R. A. J. (2011) 'Bovine mycoplasmosis: Silent and deadly', *Veterinary Record*, 168(17), pp. 459–462. doi: 10.1136/vr.d2468.
- Nishida, K. *et al.* (2016) 'Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems', *Science*, 353(6305). doi: 10.1126/science.aaf8729.
- Nishimasu, H. *et al.* (2014) 'Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA Hiroshi', *Cell*. doi: 10.1016/j.cell.2014.02.001.Crystal.
- Nishimasu, H. *et al.* (2015) 'Crystal Structure of Staphylococcus aureus Cas9', *Cell*, 162(5), pp. 1113–1126. doi: 10.1016/j.cell.2015.08.007.
- Nolan, P. M., Hill, G. E. and Stoehr, A. M. (1998) 'Sex, size, and plumage redness predict house finch survival in an epidemic', *Proceedings of the Royal Society B: Biological Sciences*, 265(1400), pp. 961–965. doi: 10.1098/rspb.1998.0384.
- Nottelet, P. *et al.* (2021) 'The mycoplasma surface proteins MIB and MIP promote the dissociation of the antibody-antigen interaction', *Science Advances*, 7(10), pp. 1–15. doi: 10.1126/sciadv.abf2403.
- Nouvel, L. X. *et al.* (2010) 'Comparative genomic and proteomic analyses of two *Mycoplasma agalactiae* strains: Clues to the macro- and micro-events that are shaping mycoplasma diversity', *BMC Genomics*, 11(1). doi: 10.1186/1471-2164-11-86.
- Nuñez, J. K. *et al.* (2014) 'Cas1 – Cas2 complex formation mediates spacer acquisition during CRISPR – Cas adaptive immunity', *Nature Structural and Molecular Biology*, (May). doi: 10.1038/nsmb.2820.
- O'Brien, R. E. *et al.* (2020) 'Structural basis for assembly of non-canonical small subunits into type I-C Cascade', *Nature Communications*. doi: 10.1038/s41467-020-19785-8.
- Oakes, B. L. *et al.* (2019) 'CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification', *Cell*, 176(1–2), pp. 254–267.e16. doi: 10.1016/j.cell.2018.11.052.
- Ofir, G. *et al.* (2018) 'DISARM is a widespread bacterial defence system with broad anti-phage activities', *Nature Microbiology*, 3(1), pp. 90–98. doi: 10.1038/s41564-017-0051-0.DISARM.
- OIE (2009) 'Contagious bovine pleuroneumonia (CBPP) - Factsheet', pp. 1–4.
- Oliveira, P. H., Touchon, M. and Rocha, E. P. C. (2014) 'The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts', *Nucleic Acids Research*, 42(16), pp. 10618–10632. doi: 10.1093/nar/gku734.
- Omar O. Abudayyeh *et al.* (2016) 'C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector', *Science*, 5573(June), p. 16. Available at: <http://science.sciencemag.org/content/early/2016/06/01/science.aaf5573>.
- Omelchenko, M. V. *et al.* (2003) 'Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ', *Genome biology*, 4(9). doi: 10.1186/gb-2003-4-9-r55.
- Ongena, V. *et al.* (2021) 'Reversible bacteriophage resistance by shedding the bacterial cell wall', *bioRxiv*. doi: <https://doi.org/10.1101/2021.11.17.468999>.
- Paez-Espino, D. *et al.* (2015) 'CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*', *mBio*, 6(2), pp. 1–9. doi: 10.1128/mBio.00262-15.
- Page, R. and Peti, W. (2016) 'Toxin-antitoxin systems in bacterial growth arrest and persistence', *Nature Chemical Biology*, 12(4), pp. 208–214. doi: 10.1038/nchembio.2044.
- Papazisi, L. *et al.* (2000) 'Analysis of cytoadherence-deficient, GapA-negative *Mycoplasma gallisepticum* strain R', *Infection and Immunity*, 68(12), pp. 6643–6649. doi: 10.1128/IAI.68.12.6643-6649.2000.
- Papazisi, L. *et al.* (2002) 'GapA and CrmA coexpression is essential for *Mycoplasma gallisepticum* cytoadherence and virulence', *Infection and Immunity*, 70(12), pp. 6839–6845. doi: 10.1128/IAI.70.12.6839-6845.2002.
- Pardee, K. *et al.* (2016) 'Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components', *Cell*, 165(5), pp. 1255–1266. doi: 10.1016/j.cell.2016.04.059.
- Pattanayak, V. *et al.* (2013) 'High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity', *Nature Biotechnology*. doi: 10.1038/nbt.2673.High-throughput.
- Payne, L. J. *et al.* (2021) 'Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types', *Nucleic acids research*, pp. 1–11.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Peebles, E. D. and Branton, S. L. (2012) 'Mycoplasma gallisepticum in the commercial egg-laying hen: A historical perspective considering the effects of pathogen strain, age of the bird at inoculation, and diet on performance and physiology', *Journal of Applied Poultry Research*, 21(4), pp. 897–914. doi: 10.3382/japr.2012-00555.
- Pereyre, S. et al. (2009) 'Life on arginine for *Mycoplasma hominis*: Clues from its minimal genome and comparison with other human urogenital mycoplasmas', *PLoS Genetics*, 5(10). doi: 10.1371/journal.pgen.1000677.
- Pereyre, S. and Tardy, F. (2021) 'Integrating the Human and Animal Sides of Mycoplasmas Resistance to Antimicrobials', *antibiotics*, pp. 1–23.
- Perez-Casal, J. (2020) 'Pathogenesis and Virulence of *Mycoplasma bovis*', *Veterinary Clinics of North America - Food Animal Practice*, 36(2), pp. 269–278. doi: 10.1016/j.cvfa.2020.02.002.
- Perez-Pinera, P. et al. (2013) 'RNA-guided gene activation by CRISPR-Cas9-based transcription factors', *Nature Methods*. doi: 10.1016/B0-12-370879-6/00389-6.
- Perilla-Henao, L. M. and Casteel, C. L. (2016) 'Vector-Borne Bacterial Plant Pathogens: Interactions with Hemipteran Insects and Plants', *Frontiers in Plant Science*, 7(AUG2016), pp. 1–15. doi: 10.3389/fpls.2016.01163.
- Persson, A. et al. (2002) 'Variable Surface Protein Vmm of *Mycoplasma mycoides subsp. mycoides* Small Colony Type', *Journal of Bacteriology*, 184(13), pp. 3712–3722. doi: 10.1128/JB.184.13.3712-3722.2002.
- Peters, J. M. et al. (2016) 'A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria Jason', *Cell*. doi: 10.1016/j.cell.2016.05.003.
- Peterson, S. N. and Fraser, C. M. (2001) 'The complexity of simplicity', *Genome Biology*, 2(2), pp. 1–7. doi: 10.1186/gb-2001-2-2-comment2002.
- Pflaum, K. et al. (2015) 'Global changes in *Mycoplasma gallisepticum* phase-variable lipoprotein gene vlhA expression during in vivo infection of the natural chicken host', *Infection and Immunity*, 84(1), pp. 351–355. doi: 10.1128/IAI.01092-15.
- Pflaum, K. et al. (2018) 'Variable Lipoprotein Hemagglutinin A Gene (vlhA) Expression in variant *Mycoplasma gallisepticum* Strains In vivo', *American society for microbiology*, pp. 1–9.
- Pflaum, K. et al. (2020) 'The influence of host tissue on *M. gallisepticum* vlhA gene expression', *Veterinary Microbiology*, 251, p. 108891. doi: 10.1016/j.vetmic.2020.108891.
- Pich, O. Q. et al. (2006) 'Comparative analysis of antibiotic resistance gene markers in *Mycoplasma genitalium*: Application to studies of the minimal gene complement', *Microbiology*, 152(2), pp. 519–527. doi: 10.1099/mic.0.28287-0.
- Piergentili, R. et al. (2021) 'CRISPR-Cas and Its Wide-Ranging Applications: From Human Limitations, Hazards and Bioethical Issues', *cell*.
- Pilo, P. et al. (2005) 'A metabolic enzyme as a primary virulence factor of *Mycoplasma mycoides* subspecies *mycoides* small colony', *Journal of Bacteriology*. doi: 10.1128/JB.187.19.6824-6831.2005.
- Piñero-Lambea, C. et al. (2020) '*Mycoplasma pneumoniae* Genome Editing Based on Oligo Recombineering and Cas9-Mediated Counterselection', *ACS Synthetic Biology*, 9(7), pp. 1693–1704. doi: 10.1021/acssynbio.0c00022.
- Pingoud, A., Wilson, G. G. and Wende, W. (2014) 'Type II restriction endonucleases -- a historical perspective and more', *Nucleic Acids Research*, 42(12), pp. 7489–7527. doi: 10.1093/nar/gku447.
- Pinilla-redondo, R. et al. (2020) 'Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements', *Nature Communications*. doi: 10.1038/s41467-020-19415-3.
- Pinilla-Redondo, R. et al. (2021) 'CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids', *Nucleic Acids Research*. Available at: <https://www.biorxiv.org/content/10.1101/2021.06.04.447074v1>.
- Piotrowska, M. and Popowska, M. (2015) 'Insight into the mobilome of *Aeromonas strains*', *Frontiers in Microbiology*, 6(MAY). doi: 10.3389/fmicb.2015.00494.
- Poteete, A. R. (2008) 'Involvement of DNA replication in phage lambda Red-mediated homologous recombination', *Molecular Microbiology*, 68(1), pp. 66–74. doi: 10.1111/j.1365-2958.2008.06133.x.
- Pyne, M. E. et al. (2016) 'Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*', *Scientific Reports*, 6(December 2015), pp. 1–15. doi: 10.1038/srep25666.
- Qi, L. S. et al. (2013) 'Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression', *Cell*, 152(5), pp. 1173–1183. doi: 10.1016/j.cell.2013.02.022.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Qin, L., Chen, Y. and You, X. (2019) 'Subversion of the immune response by human pathogenic mycoplasmas', *Frontiers in Microbiology*, 10(AUG), pp. 1–12. doi: 10.3389/fmicb.2019.01934.
- Ran, F. A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature protocols*, pp. 1-. doi: 10.1038/nprot.2013.143.Genome.
- Ran, F. A. *et al.* (2015) 'In vivo genome editing using Staphylococcus aureus Cas9 F.', *Nature*. doi: 10.1038/nature14299.
- Raymond, B. B. A. *et al.* (2018) 'Mycoplasma hyopneumoniae resides intracellularly within porcine epithelial cells', *Scientific Reports*, 8(1), p. 17697. doi: 10.1038/s41598-018-36054-3.
- Razin, S. and Hayflick, L. (2010) 'Highlights of mycoplasma research-An historical perspective', *Biologicals*, 38(2), pp. 183–190. doi: 10.1016/j.biologicals.2009.11.008.
- Razin, S., Yogev, D. and Naot, Y. (1998) 'Molecular biology and pathogenicity of Mycoplasmas', *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, 165(2), pp. 199–221.
- Renaudin, J. *et al.* (1995) 'Integrative and free Spiroplasma citri oriC plasmids: Expression of the Spiroplasma phoeniceum spiralin in Spiroplasma citri', *Journal of Bacteriology*, 177(10), pp. 2870–2877. doi: 10.1128/jb.177.10.2870-2877.1995.
- Renaudin, J., Breton, M. and Citti, C. (2014) *Molecular Genetic Tools for Mollicutes*.
- Rideau, F. *et al.* (2019) 'Random transposon insertion in the Mycoplasma hominis minimal genome', *Scientific Reports*, 9(1), pp. 1–11. doi: 10.1038/s41598-019-49919-y.
- Roberts, M. C. and Kenny, G. E. (1987) 'Conjugal transfer of transposon Tn916 from Streptococcus faecalis to Mycoplasma hominis', *Journal of Bacteriology*, 169(8), pp. 3836–3839. doi: 10.1128/jb.169.8.3836-3839.1987.
- Roberts, R. J. *et al.* (2003) 'A nomenclature for restriction enzymes , DNA methyltransferases , homing endonucleases and their genes', *Nucleic Acids Research*, 31(7). doi: 10.1093/nar/gkg274.
- Roberts, R. J. *et al.* (2010) 'REBASE — a database for DNA restriction and modification : enzymes , genes and genomes', *Nucleic Acids Research*, 38(October 2009), pp. 2009–2011. doi: 10.1093/nar/gkp874.
- Rocha, E. P. C. and Bikard, D. (2021) 'Many defense systems in microbial genomes, but which is defending whom from what?', *EcoEvoRxiv*, pp. 1–12.
- Rodrigues, S. D. *et al.* (2021) 'Efficient CRISPR-mediated base editing in agrobacterium spp', *Proceedings of the National Academy of Sciences of the United States of America*, 118(2), pp. 1–8. doi: 10.1073/pnas.2013338118.
- Röske, K. *et al.* (2010) 'A versatile palindromic amphipathic repeat coding sequence horizontally distributed among diverse bacterial and eucaryotic microbes', *BMC Genomics*, 11(1). doi: 10.1186/1471-2164-11-430.
- Rouch, D. A. *et al.* (1987) 'The aacA-aphD Gentamicin and Kanamycin Resistance Determinant of Tn4001 from Staphylococcus aureus: Expression and Nucleotide Sequence Analysis', *Microbiology*, 133(11), pp. 3039–3052. doi: 10.1099/00221287-133-11-3039.
- Rousset, F. *et al.* (2021) 'Prophage-encoded hotspots of bacterial immune systems', *BioRxiv*.
- Le Roy, C. *et al.* (2016) 'Fluoroquinolone-resistant Mycoplasma genitalium, Southwestern France', *Emerging Infectious Diseases*, 22(9), pp. 1677–1679. doi: 10.3201/eid2209.160446.
- Ruffin, D. C. *et al.* (2000) 'Transposon mutagenesis of Mycoplasma gallisepticum by conjugation with Enterococcus faecalis and determination of insertion site by direct genomic sequencing', *Plasmid*, 44(2), pp. 191–195. doi: 10.1006/plas.2000.1485.
- Ruiz, E. *et al.* (2019) 'CReasPy-Cloning: A Method for Simultaneous Cloning and Engineering of Megabase-Sized Genomes in Yeast Using the CRISPR-Cas9 System', *ACS Synthetic Biology*. doi: 10.1021/acssynbio.9b00224.
- Sacchini, F. *et al.* (2012) 'Plasma levels of TNF- α , IFN- γ , IL-4 and IL-10 during a course of experimental contagious bovine pleuropneumonia', *BMC Veterinary Research*, 8(1), p. 44. doi: 10.1186/1746-6148-8-44.
- Sagné, E., Citti, C. and Dordet-Frisoni, E. (2021) 'Bacterial Conjugation Protocol for Ruminant Mycoplasmas', *Bio-Protocol*, 11(2), pp. 0–10. doi: 10.21769/bioprotoc.3893.
- Saha, C. *et al.* (2020) 'Guide-free Cas9 from pathogenic Campylobacter jejuni bacteria causes severe damage to DNA', *Science Advances*, 6(25), pp. 1–13. doi: 10.1126/sciadv.aaz4849.
- Sampson, T. R. and Weiss, D. S. (2013) 'Alternative Roles for CRISPR / Cas Systems in Bacterial Pathogenesis', *plos pathogens*, 9(10), pp. 9–11. doi: 10.1371/journal.ppat.1003621.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Sánchez-Busó, L. *et al.* (2019) 'Genetic variation regulates the activation and specificity of Restriction-Modification systems in *Neisseria gonorrhoeae*', *Scientific Reports*, 9(1), pp. 1–11. doi: 10.1038/s41598-019-51102-2.
- Sander, J. D. and Joung, J. K. (2014) 'CRISPR-Cas systems for genome editing, regulation and targeting', *Nature Biotechnology*, 32(4), pp. 347–355. doi: 10.1038/nbt.2842.CRISPR-Cas.
- Sapranaukas, R. *et al.* (2011) 'The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*', *Nucleic Acids Research*, 39(21), pp. 9275–9282. doi: 10.1093/nar/gkr606.
- Schieck, E. *et al.* (2016) 'Galactofuranose in *Mycoplasma mycoides* is important for membrane integrity and conceals adhesins but does not contribute to serum resistance', *Molecular Microbiology*, 99(1), pp. 55–70. doi: 10.1111/mmi.13213.
- Seib, K. L. *et al.* (2020) 'Epigenetic Regulation of Virulence and Immuno-evasion by Phase-Variable Restriction-Modification Systems in Bacterial Pathogens', *Annual Review of Microbiology*, 74(1), pp. 655–671. doi: 10.1146/annurev-micro-090817-062346.
- Semenova, E. *et al.* (2011) 'Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence', *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), pp. 10098–10103. doi: 10.1073/pnas.1104144108.
- Serfiotis-mitsa, D. *et al.* (2010) 'The structure of the KlcA and ArdB proteins reveals a novel fold and antirestriction activity against Type I DNA restriction systems in vivo but not in vitro', *Nucleic Acids Research*, 38(5), pp. 1723–1737. doi: 10.1093/nar/gkp1144.
- Shahid, M. A. *et al.* (2014) 'Development of an oriC vector for use in *Mycoplasma synoviae*', *Journal of Microbiological Methods*, 103, pp. 70–76. doi: 10.1016/j.mimet.2014.05.014.
- Shalem, O. *et al.* (2014) 'Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells', *Science*, 343(6166), pp. 84–87. doi: 10.1126/science.1247005.Genome-Scale.
- Sharma, S., Citti, C., *et al.* (2015) 'Development and host compatibility of plasmids for two important ruminant pathogens, *Mycoplasma bovis* and *Mycoplasma agalactiae*', *PLoS ONE*, 10(3), pp. 1–17. doi: 10.1371/journal.pone.0119000.
- Sharma, S., Tivendale, K. A., *et al.* (2015) 'Disruption of the membrane nuclease gene (MBOVPG45_0215) of *Mycoplasma bovis* greatly reduces cellular nuclease activity', *Journal of Bacteriology*, 197(9), pp. 1549–1558. doi: 10.1128/JB.00034-15.
- Sharma, S., Markham, P. F. and Browning, G. F. (2014) 'Genes found essential in other mycoplasmas are dispensable in *Mycoplasma bovis*', *PLoS ONE*, 9(6). doi: 10.1371/journal.pone.0097100.
- Shaw, D., Serrano, L. and Lluch-Senar, M. (2021) 'Lox'd in translation: Contradictions in the nomenclature surrounding common lox-site mutants and their implications in experiments', *Microbiology (United Kingdom)*, 167(1), pp. 1–11. doi: 10.1099/mic.0.000997.
- Shen, B. *et al.* (2013) 'Generation of gene-modified mice via Cas9/RNA-mediated gene targeting', *Cell Research*, 23(5), pp. 720–723. doi: 10.1038/cr.2013.46.
- Shibata, M. *et al.* (2017) 'Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy', *Nature Communications*, pp. 1–9. doi: 10.1038/s41467-017-01466-8.
- Shimizu, T., Kida, Y. and Kuwano, K. (2005) 'A Dipalmitoylated Lipoprotein from *Mycoplasma pneumoniae* Activates NF- κ B through TLR1, TLR2, and TLR6', *The Journal of Immunology*, 175(7), pp. 4641–4646. doi: 10.4049/jimmunol.175.7.4641.
- Shmakov, S. A. *et al.* (2017) 'The CRISPR spacer space is dominated by sequences from species-specific mobilomes', *mBio*, 8(5), pp. 1–18. doi: 10.1128/mBio.01397-17.
- Siguier, P. *et al.* (2006) 'ISfinder: the reference centre for bacterial insertion sequences.', *Nucleic acids research*, 34(Database issue), pp. 32–36. doi: 10.1093/nar/gkj014.
- Sinkunas, T. *et al.* (2013) 'In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*', *EMBO Journal*, 32(3), pp. 385–394. doi: 10.1038/emboj.2012.352.
- Sirand-Pugnet, P., Lartigue, C., *et al.* (2007) 'Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome', *PLoS Genetics*, 3(5), pp. 744–758. doi: 10.1371/journal.pgen.0030075.
- Sirand-Pugnet, P., Citti, C., *et al.* (2007) 'Evolution of mollicutes: down a bumpy road with twists and turns', *Research in Microbiology*, 158(10), pp. 754–766. doi: 10.1016/j.resmic.2007.09.007.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Song, X. *et al.* (2017) 'CRISPR-Cas9D10A Nickase-Assisted Genome Editing in *Lactobacillus casei*', *Aem*, 83(22), pp. 1–13.
- Song, Y. and Lee, S.-W. (2021) 'Complete Genome Sequence of *Mycoplasma gallisepticum* Strain KUVMG001, an Isolate from South Korea', *Genome Announcements*. doi: 10.1128/genomeA.00712-15.
- Spergser, J. *et al.* (2013) 'Emergence, re-emergence, spread and host species crossing of *Mycoplasma bovis* in the Austrian Alps caused by a single endemic strain', *Veterinary Microbiology*, 164(3–4), pp. 299–306. doi: 10.1016/j.vetmic.2013.02.007.
- Staley, M. *et al.* (2018) 'Bacterial pathogen emergence requires more than direct contact with a novel passerine host', *Infection and Immunity*, 86(3), pp. 1–9. doi: 10.1128/IAI.00863-17.
- Staley, M. and Bonneaud, C. (2015) 'Immune responses of wild birds to emerging infectious diseases', *Parasite Immunology*, 37(5), pp. 242–254. doi: 10.1111/pim.12191.
- Stern, A. and Sorek, R. (2011) 'The phage-host arms-race : Shaping the evolution of microbes', *Bioessays*, 33(1), pp. 43–51. doi: 10.1002/bies.201000071.The.
- Strich, J. R. and Chertow, D. S. (2018) 'CRISPR-cas biology and its application to infectious diseases', *Journal of Clinical Microbiology*, 57(4), pp. 1–14. doi: 10.1128/JCM.01307-18.
- Sun, C. L. *et al.* (2013) 'Phage mutations in response to CRISPR diversification in a bacterial population', *Environmental Microbiology*, 15(2), pp. 463–470. doi: 10.1111/j.1462-2920.2012.02879.x.
- Sun, J. *et al.* (2020) 'CRISPR-Assisted Multiplex Base Editing System in *Pseudomonas putida* KT2440', *Frontiers in Bioengineering and Biotechnology*, 8(July), pp. 1–13. doi: 10.3389/fbioe.2020.00905.
- Sun, Z. *et al.* (2015) 'A high-efficiency recombineering system with PCR-based ssDNA in *Bacillus subtilis* mediated by the native phage recombinase GP35', *Applied Microbiology and Biotechnology*, 99(12), pp. 5151–5162. doi: 10.1007/s00253-015-6485-5.
- Sung, L. Y. *et al.* (2019) 'Combining orthogonal CRISPR and CRISPRi systems for genome engineering and metabolic pathway modulation in *Escherichia coli*', *Biotechnology and Bioengineering*. doi: 10.1002/bit.26915.
- Swarts, D. C. *et al.* (2014) 'The evolutionary journey of Argonaute proteins', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2879.The.
- Swarts, D. C. *et al.* (2017) 'Structural Basis for Guide RNA Processing and Seed- Dependent DNA Targeting by CRISPR-Cas12a Article Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a', *Molecular Cell*, 66(2), pp. 221-233.e4. doi: 10.1016/j.molcel.2017.03.016.
- Swingle, B. *et al.* (2010) 'Recombineering using recTE from *Pseudomonas syringae*', *Applied and Environmental Microbiology*, 76(15), pp. 4960–4968. doi: 10.1128/AEM.00911-10.
- Szczelkun, M. D. *et al.* (2014) 'Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes', *Proceedings of the National Academy of Sciences of the United States of America*, 111(27). doi: 10.1073/pnas.1402597111.
- Szczepanek, S. M. *et al.* (2010) 'Comparative genomic analyses of attenuated strains of *Mycoplasma gallisepticum*', *Infection and Immunity*, 78(4), pp. 1760–1771. doi: 10.1128/IAI.01172-09.
- Szczepanek, S. M. *et al.* (2014) 'Hydrogen peroxide production from glycerol metabolism is dispensable for virulence of *Mycoplasma gallisepticum* in the tracheas of chickens', *Infection and Immunity*, 82(12), pp. 4915–4920. doi: 10.1128/IAI.02208-14.
- Tadic, V. *et al.* (2019) 'CRISPR/Cas9-based epigenome editing: An overview of dCas9-based tools with special emphasis on off-target activity', *Methods*. doi: <https://doi.org/10.1016/j.jymeth.2019.05.003>.
- Tardy, F. *et al.* (2015) 'Integrative conjugative elements are widespread in field isolates of *Mycoplasma* species pathogenic for ruminants', *Applied and Environmental Microbiology*, 81(5), pp. 1634–1643. doi: 10.1128/AEM.03723-14.
- Taylor-Robinson, D. and Bébéar, C. (1997) 'Antibiotic susceptibilities of mycoplasmas and treatment of mycoplasmal infections', *Journal of Antimicrobial Chemotherapy*, 40(5), pp. 622–630. doi: 10.1093/jac/40.5.622.
- Di Teodoro, G. *et al.* (2018) 'Respiratory explants as a model to investigate early events of contagious bovine pleuropneumonia infection', *Veterinary Research*, 49(1), pp. 1–12. doi: 10.1186/s13567-017-0500-z.
- Di Teodoro, G. *et al.* (2020) 'Contagious Bovine Pleuropneumonia: A Comprehensive Overview', *Veterinary Pathology*, 57(4), pp. 476–489. doi: 10.1177/0300985820921818.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Terahara, N., Tulum, I. and Miyata, M. (2017) 'Transformation of crustacean pathogenic bacterium *Spiroplasma eriocheiris* and expression of yellow fluorescent protein', *Biochemical and Biophysical Research Communications*, 487(3), pp. 488–493. doi: 10.1016/j.bbrc.2017.03.144.
- Tesson, F. *et al.* (2021) 'Systematic and quantitative view of the antiviral arsenal of prokaryotes', *bioRxiv*, p. 2021.09.02.458658. doi: 10.1101/2021.09.02.458658.
- Thi Nhu Thao, T. *et al.* (2020) 'Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform', *Nature*, 582(7813), pp. 561–565. doi: 10.1038/s41586-020-2294-9.
- Thiaucourt, F. *et al.* (1998) 'Vaccination against contagious bovine pleuropneumonia and the use of molecular tools in epidemiology', *Annals of the New York Academy of Sciences*, pp. 146–151. doi: 10.1111/j.1749-6632.1998.tb11043.x.
- Thiaucourt, F. *et al.* (2011) 'Mycoplasma mycoides, from "mycoides Small Colony" to "capri". A microevolutionary perspective', *BMC Genomics*, 12(1), p. 114. doi: 10.1186/1471-2164-12-114.
- Thomas, A. *et al.* (2003) 'Adherence to various host cell lines of *Mycoplasma bovis* strains differing in pathogenic and cultural features', *Veterinary Microbiology*, 91(2–3), pp. 101–113. doi: 10.1016/S0378-1135(02)00303-6.
- Thomason, L. C., Costantino, N. and Court, D. L. (2016) 'Examining a DNA replication requirement for bacteriophage λ red and rac prophage recET-promoted recombination in *Escherichia coli*', *mBio*, 7(5), pp. 1–10. doi: 10.1128/mBio.01443-16.
- Thresher, R. J. *et al.* (1995) *Electron Microscopic Visualization of RecT Protein and its Complexes with DNA*, *J. Mol. Biol.*
- Tong, Y. *et al.* (2019) 'Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST', *Proceedings of the National Academy of Sciences of the United States of America*, 116(41), pp. 20366–20375. doi: 10.1073/pnas.1913493116.
- Torres-Puig, S. *et al.* (2018) 'Activation of r20-dependent recombination and horizontal gene transfer in *Mycoplasma genitalium*', *DNA Research*, 25(4), pp. 383–393. doi: 10.1093/dnares/dsy011.
- Tsarpopoulou, I. *et al.* (2016) 'In-Yeast Engineering of a Bacterial Genome Using CRISPR/Cas9', *ACS Synthetic Biology*, 5(1), pp. 104–109. doi: 10.1021/acssynbio.5b00196.
- Tseng, C. W. *et al.* (2013) 'MalF is essential for persistence of *Mycoplasma gallisepticum* in vivo', *Microbiology (United Kingdom)*, 159(PART7), pp. 1459–1470. doi: 10.1099/mic.0.067553-0.
- Tseng, C. W. *et al.* (2017) 'The oppD Gene and Putative Peptidase May Be Required for Virulence in *Mycoplasma gallisepticum*', *Infection and Immunity*, pp. 1-.
- Tucker, A. T. *et al.* (2019) 'RecET-Mediated Recombineering in *Acinetobacter baumannii*', *Methods in Molecular Biology*, 1946, pp. 107–113. doi: 10.1007/978-1-4939-9118-1_11.
- Tully, J. G. *et al.* (1993) 'Revised Taxonomy of the Class Mollicutes: Proposed Elevation of a Monophyletic Cluster of Arthropod-Associated Mollicutes to Ordinal Rank (Entomoplasmatales ord. nov.), with Provision for Familial Rank To Separate Species with Nonhelical Morphology (Entom)', *International Journal of Systematic Bacteriology*, 43(3), pp. 630–630. doi: 10.1099/00207713-43-3-630a.
- Tully, J. G. *et al.* (1995) '*Spiroplasma ixodetis* sp. nov., a New Species from Ixodes pacificus Ticks Collected in Oregon', *International Journal of Systematic Bacteriology*, 45(1), pp. 23–28. doi: 10.1099/00207713-45-1-23.
- Tulman, E. R. *et al.* (2012) 'Extensive variation in surface lipoprotein gene content and genomic changes associated with virulence during evolution of a novel North American house finch epizootic strain of *Mycoplasma gallisepticum*', *Microbiology (United Kingdom)*, 158(8), pp. 2073–2088. doi: 10.1099/mic.0.058560-0.
- Del Val, E. *et al.* (2019) 'RecA and DNA recombination: A review of molecular mechanisms', *Biochemical Society Transactions*, 47(5), pp. 1511–1531. doi: 10.1042/BST20190558.
- Valentine-King, M. A. *et al.* (2020) 'Efficacy data of halogenated phenazine and quinoline agents and an NH125 analogue to veterinary mycoplasmas', *BMC Veterinary Research*, 16(1), pp. 1–11. doi: 10.1186/s12917-020-02324-4.
- Vasconcelos, A. T. R. *et al.* (2005) 'Swine and Poultry Pathogens: the Complete Genome Sequences of Two Strains of', *Microbiology*, 187(16), pp. 5568–5577. doi: 10.1128/JB.187.16.5568.
- Vashee, S. *et al.* (2017) 'Cloning, Assembly, and Modification of the Primary Human Cytomegalovirus Isolate Toledo by Yeast-Based Transformation-Associated Recombination', *mSphere*, 2(5), pp. 1–17. doi: 10.1128/mspheredirect.00331-17.
- Vashee, S., Arfi, Y. and Lartigue, C. (2020) 'Budding yeast as a factory to engineer partial and complete microbial genomes', *Current Opinion in Systems Biology*, 24(January), pp. 1–8. doi: 10.1016/j.coisb.2020.09.003.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Vento, J. M., Crook, N. and Beisel, C. L. (2019) 'Barriers to genome editing with CRISPR in bacteria', *Journal of Industrial Microbiology and Biotechnology*, 46(9–10), pp. 1327–1341. doi: 10.1007/s10295-019-02195-1.
- Viana, D. *et al.* (2015) 'A single natural nucleotide mutation alters bacterial pathogen host tropism', *Nature Genetics*, 47(4), pp. 361–366. doi: 10.1038/ng.3219.
- Vigouroux, A. and Bikard, D. (2020) 'CRISPR Tools To Control Gene Expression in Bacteria', *Microbiology and Molecular Biology Reviews*, 84(2), pp. 1–18. doi: 10.1128/mnbr.00077-19.
- Violle, C. *et al.* (2020) 'Plasmid transfer by conjugation in gram-negative bacteria: From the cellular to the community level', *Genes*, 11(11), pp. 1–33. doi: 10.3390/genes11111239.
- Voelker, L. L. and Dybvig, K. (1996) 'Gene transfer in *Mycoplasma arthritidis*: Transformation, conjugal transfer of Tn916, and evidence for a restriction system recognizing AGCT', *Journal of Bacteriology*, 178(20), pp. 6078–6081. doi: 10.1128/jb.178.20.6078-6081.1996.
- Waites, K. B. and Talkington, D. (2005) 'New Developments in Human Disease Due to Mycoplasmas', in *Mycoplasmas: Molecular Biology, Pathogenicity, and Strategies for Control*.
- Wan, H. *et al.* (2019) 'Probing the Behaviour of Cas1-Cas2 upon Protospacer Binding in CRISPR-Cas Systems using Molecular Dynamics Simulations', *Scientific Reports*, 9(1), pp. 1–16. doi: 10.1038/s41598-019-39616-1.
- Wang, H. *et al.* (2013) 'One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering', *Cell*. doi: 10.1016/j.cell.2013.04.025.One-Step.
- Wang, J. *et al.* (2015) 'Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems Article Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems', *Cell*, 163(4), pp. 840–853. doi: 10.1016/j.cell.2015.10.008.
- Wang, L. *et al.* (2018) 'DNA phosphorothioate modification — a new multi-functional epigenetic system in bacteria', *FEMS Microbiology*. doi: 10.1093/femsre/fuy036.
- Wang, S. *et al.* (2017) 'Genome Editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 System', *Applied and Environmental Microbiology*. Edited by R. M. Kelly, 83(10), pp. 1–16. doi: 10.1128/AEM.00233-17.
- Wang, X. *et al.* (2020) 'Type VII Toxin / Antitoxin Classification System for Antitoxins that Enzymatically Neutralize Toxins', *Trends in Microbiology*, xx(xx), pp. 1–6. doi: 10.1016/j.tim.2020.12.001.
- Wang, Y. *et al.* (2015) 'Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system', *Journal of Biotechnology*, 200, pp. 1–5. doi: 10.1016/j.jbiotec.2015.02.005.
- Wang, Y. *et al.* (2016) 'Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "clean" Mutant Selection in *Clostridium beijerinckii* as an Example', *ACS Synthetic Biology*, 5(7), pp. 721–732. doi: 10.1021/acssynbio.6b00060.
- Wang, Y. *et al.* (2017) 'Identification of genes involved in *Mycoplasma gallisepticum* biofilm formation using mini-Tn4001-SGM transposon mutagenesis', *Veterinary Microbiology*, 198, pp. 17–22. doi: 10.1016/j.vetmic.2016.11.021.
- Wang, Y. *et al.* (2018) 'CRISPR-Cas9 and CRISPR-Assisted Cytidine Deaminase Enable Precise and Efficient Genome Editing in *Klebsiella pneumoniae*', *Applied and Environmental Microbiology*, 84(23), pp. 1–15. doi: <https://doi.org/10.1128/AEM.01834-18>.
- Washburn, L. *et al.* (2004) '*Mycoplasma arthritidis* bacteriophage MAV1 prophage integration, deletions, and strain-related polymorphisms', *Plasmid*, 52(1), pp. 31–47. doi: 10.1016/j.plasmid.2004.04.004.
- Washburn, L. R., Bird, D. W. and Dybvig, K. (2003) 'Restoration of cytoadherence to an adherence-deficient mutant of *Mycoplasma arthritidis* by genetic complementation', *Infection and Immunity*, 71(2), pp. 671–675. doi: 10.1128/IAI.71.2.671-675.2003.
- Wei, Y., Terns, R. M. and Terns, M. P. (2015) 'Cas9 function and host genome sampling in Type II-A CRISPR – Cas adaptation', *Genes*, pp. 356–361. doi: 10.1101/gad.257550.114.
- Weinberger, A. D. *et al.* (2012) 'Viral Diversity Threshold for Adaptive Immunity in Prokaryotes', *mBio*. doi: 10.1128/mBio.00456-12.
- Weisburg, W. G. *et al.* (1989) 'A Phylogenetic Analysis of the Mycoplasmas: Basis for Their Classification', *Journal of Bacteriology*, 171(12), pp. 6455–6467.
- Westra, E. R. *et al.* (2015) 'Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense', *Current Biology*, 25(8), pp. 1043–1049. doi: 10.1016/j.cub.2015.01.065.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Westra, E. R. and Levin, B. R. (2020) 'It is unclear how important CRISPR-Cas systems are for protecting natural populations of bacteria against infections by mobile genetic elements', *Proceedings of the National Academy of Sciences of the United States of America*, 117(45), pp. 27777–27785. doi: 10.1073/pnas.1915966117.
- Whithear, K. G. (1996) 'Control of avian mycoplasmoses by vaccination', *OIE Revue Scientifique et Technique*, 15(4), pp. 1527–1553. doi: 10.20506/rst.15.4.985.
- Wijesurendra, D. S. *et al.* (2015) 'Development of a *Mycoplasma gallisepticum* infection model in turkeys', *Avian Pathology*, 44(1), pp. 35–42. doi: 10.1080/03079457.2014.992390.
- Wimmer, F. and Beisel, C. L. (2020) 'CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers', *Frontiers in Microbiology*, 10(January), pp. 1–17. doi: 10.3389/fmicb.2019.03078.
- Winner, F., Rosengarten, R. and Citti, C. (2000) 'In vitro cell invasion of *Mycoplasma gallisepticum*', *Infection and Immunity*, 68(7), pp. 4238–4244. doi: 10.1128/IAI.68.7.4238-4244.2000.
- Wise, K. S. *et al.* (2006) 'Distinctive repertoire of contingency genes conferring mutation-based phase variation and combinatorial expression of surface lipoproteins in *Mycoplasma capricolum subsp. capricolum* of the *Mycoplasma mycoides* phylogenetic cluster', *Journal of Bacteriology*, 188(13), pp. 4926–4941. doi: 10.1128/JB.00252-06.
- Woese, C. R., Maniloff, J. and Zablen, L. B. (1980) 'Phylogenetic analysis of the mycoplasmas', *Proceedings of the National Academy of Sciences of the United States of America*, 77(1), pp. 494–498. doi: 10.1073/pnas.77.1.494.
- Woese, C. R., Stackebrandt, E. and Ludwig, W. (1985) 'What are mycoplasmas: The relationship of tempo and mode in bacterial evolution', *Journal of Molecular Evolution*, 21(4), pp. 305–316. doi: 10.1007/BF02115648.
- Workman, R. E. *et al.* (2020) 'A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression', *Cell*. doi: 10.1101/2020.05.21.102756.
- Wozniak, R. A. F. and Waldor, M. K. (2010) 'Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow', *Nature Reviews Microbiology*, 8(8), pp. 552–563. doi: 10.1038/nrmicro2382.
- Wright, D. G. *et al.* (2017) '*Mycobacterium tuberculosis* and *Mycobacterium marinum* non-homologous end-joining proteins can function together to join DNA ends in *Escherichia coli*', *Mutagenesis*, 32(2), pp. 245–256. doi: 10.1093/mutage/gew042.
- Wu, X. *et al.* (2021) '*Mycoplasmas bovis* P48 induces apoptosis in EBL cells via an endoplasmic reticulum stress-dependent signaling pathway', *Veterinary Microbiology*, 255(October 2020), p. 109013. doi: 10.1016/j.vetmic.2021.109013.
- Wu, Y. *et al.* (2019) 'Increasing cytosine base editing scope and efficiency with engineered Cas9-PMCD1 fusions and the modified sgRNA in rice', *Frontiers in Genetics*, 10(APR), pp. 1–10. doi: 10.3389/fgene.2019.00379.
- Wu, Z. Y. *et al.* (2019) 'Reversal of carbapenem-resistance in *Shewanella algae* by CRISPR/Cas9 genome editing', *Journal of Advanced Research*, 18, pp. 61–69. doi: 10.1016/j.jare.2019.01.011.
- Wysocka, J. and Herr, W. (2003) 'The herpes simplex virus VP16-induced complex: The makings of a regulatory switch', *Trends in Biochemical Sciences*, 28(6), pp. 294–304. doi: 10.1016/S0968-0004(03)00088-4.
- Xie, K. and Yang, Y. (2013) 'RNA-Guided genome editing in plants using a CRISPR-Cas system', *Molecular Plant*, 6(6), pp. 1975–1983. doi: 10.1093/mp/sst119.
- Xiong, X. *et al.* (2020) 'SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities', *Nature Microbiology*. doi: 10.1038/s41564-020-0700-6.
- Xu, X. *et al.* (2021) 'Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing', *Molecular Cell*, 81(20), pp. 4333–4345.e4. doi: 10.1016/j.molcel.2021.08.008.
- Xu, Z. and Brown, W. R. A. (2016) 'Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*', *BMC Biotechnology*, 16(1), pp. 1–10. doi: 10.1186/s12896-016-0241-5.
- Yadav, J. P. *et al.* (2021) 'Insights on *Mycoplasma gallisepticum* and *Mycoplasma synoviae* infection in poultry: a systematic review', *Animal Biotechnology*, 0(0), pp. 1–10. doi: 10.1080/10495398.2021.1908316.
- Yamada, M. *et al.* (2017) 'Crystal Structure of the Minimal Cas9 from *Campylobacter jejuni* Reveals the Molecular Diversity in the CRISPR-Cas9 Systems', *Molecular Cell*, 65(6), pp. 1109–1121.e3. doi: 10.1016/j.molcel.2017.02.007.
- Yamano, T. *et al.* (2017) 'Structural basis for the canonical and non-canonical PAM recognition by CRISPR-Cpf1', *Molecular Cell*. doi: 10.1016/j.molcel.2017.06.035.Structural.
- Yamao, F. *et al.* (1985) 'UGA is read as tryptophan in *Mycoplasma capricolum*', *Proceedings of the National Academy of*

RÉFÉRENCES BIBLIOGRAPHIQUES

Sciences of the United States of America, 82(8), pp. 2306–2309. doi: 10.1073/pnas.82.8.2306.

Yang, D. *et al.* (2014) 'Effective gene targeting in rabbits using RNA-guided Cas9 nucleases', *Journal of molecular cell biology*. doi: 10.1093/jmcb/mjt047

Yang, P., Wang, J. and Qi, Q. (2015) 'Prophage recombinases-mediated genome engineering in *Lactobacillus plantarum*', *Microbial Cell Factories*, 14(1), pp. 1–11. doi: 10.1186/s12934-015-0344-z.

Yaya, A. *et al.* (2008) 'Genotyping of *Mycoplasma mycoides subsp. mycoides* SC by multilocus sequence analysis allows molecular epidemiology of contagious bovine pleuropneumonia', *Veterinary Research*, 39(2). doi: 10.1051/vetres:2007052.

Yin, C. *et al.* (2017) 'In Vivo Excision of HIV-1 Provirus by saCas9 and Multiplex Single-Guide RNAs in Animal Models', *Molecular Therapy*, 25(5), pp. 1168–1186. doi: 10.1016/j.jymthe.2017.03.012.

Yin, J. *et al.* (2015) 'A new recombineering system for Photorhabdus and Xenorhabdus', *Nucleic Acids Research*, 43(6), p. e36. doi: 10.1093/nar/gku1336.

Yin, J. *et al.* (2019) 'Single-Stranded DNA-Binding Protein and Exogenous RecBCD Inhibitors Enhance Phage-Derived Homologous Recombination in Pseudomonas', *iScience*, 14, pp. 1–14. doi: 10.1016/j.isci.2019.03.007.

Yiwon, C. *et al.* (2021) 'Infection strategies of mycoplasmas: Unraveling the panoply of virulence factors', *Virulence*, 12(1), pp. 788–817. doi: 10.1080/21505594.2021.1889813.

Yoneji, T. *et al.* (2021) 'Grand scale genome manipulation via chromosome swapping in *Escherichia coli* programmed by three one megabase chromosomes', *Nucleic Acids Research*, 49(15), pp. 8407–8418. doi: 10.1093/nar/gkab298.

Yourik, P. *et al.* (2019) 'Staphylococcus aureus Cas9 is a multiple-turnover enzyme', *RNA*, 25(1), pp. 35–44. doi: 10.1261/rna.067355.118.

Zander, D. V. (1961) 'Origin of S6 Strain Mycoplasma', *Avian Diseases*, 5(2), p. 154. doi: 10.2307/1587617.

Zeng, Y. *et al.* (2018) 'The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex', *Nucleic Acids Research*, 46(1), pp. 350–361. doi: 10.1093/nar/gkx1117.

Zhang, H. *et al.* (2016) '*Mycoplasma bovis* MBOV_RS02825 Encodes a Secretory Nuclease Associated with Cytotoxicity', *International Journal of Molecular Sciences*, 17(5), p. 628. doi: 10.3390/ijms17050628.

Zhang, Jingfang *et al.* (2019) 'Drug Inducible CRISPR/Cas Systems', *Computational and Structural Biotechnology Journal*, 17, pp. 1171–1177. doi: 10.1016/j.csbj.2019.07.015.

Zhang, X. *et al.* (2017) 'Multiplex gene regulation by CRISPR-ddCpf1', *Cell Discovery*, 3, pp. 1–9. doi: 10.1038/celldisc.2017.18.

Zhang, X. *et al.* (2020) 'Dual base editor catalyzes both cytosine and adenine base conversions in human cells', *Nature Biotechnology*, 38(7), pp. 856–860. doi: 10.1038/s41587-020-0527-y.

Zhao, C., Shu, X. and Sun, B. (2017) 'Construction of a Gene Knockdown System Based on Catalytically Inactive ("Dead") Cas9 (dCas9) in Staphylococcus aureus Changlong', *Applied and Environmental Microbiology*.

Zhao, G. *et al.* (2021) 'Novel Secreted Protein of *Mycoplasma bovis* MbovP280 Induces Macrophage Apoptosis Through CRYAB', *Frontiers in Immunology*, 12(February), pp. 1–15. doi: 10.3389/fimmu.2021.619362.

Zhao, N. *et al.* (2020) 'Multiplex gene editing and large DNA fragment deletion by the CRISPR/Cpf1-RecE/T system in *Corynebacterium glutamicum*', *Journal of Industrial Microbiology and Biotechnology*, 47(8), pp. 599–608. doi: 10.1007/s10295-020-02304-5.

Zheng, X. *et al.* (2017) 'An efficient system for deletion of large DNA fragments in *Escherichia coli* via introduction of both Cas9 and the non-homologous end joining system from *Mycobacterium smegmatis*', *Biochemical and Biophysical Research Communications*, 485(4), pp. 768–774. doi: 10.1016/j.bbrc.2017.02.129.

Zhou, Y. *et al.* (2016) 'P19 contributes to *Mycoplasma mycoides subsp. mycoides* adhesion to EBL cells', *Microbial Pathogenesis*, 93, pp. 13–21. doi: 10.1016/j.micpath.2016.01.011.

Zimmerman, C. U. and Herrmann, R. (2005) 'Synthesis of a small, cysteine-rich, 29 amino acids long peptide in *Mycoplasma pneumoniae*', *FEMS Microbiology Letters*, 253(2), pp. 315–321. doi: 10.1016/j.femsle.2005.09.054.

Zubair, M. *et al.* (2020) 'Identification of 60 secreted proteins for *Mycoplasma bovis* with secretome assay', *Microbial Pathogenesis*, 143, p. 104135. doi: 10.1016/j.micpath.2020.104135.

Zuo, E. *et al.* (2017) 'CRISPR/Cas9-mediated targeted chromosome elimination', *Genome Biology*, 18(1), pp. 1–18. doi:

RÉFÉRENCES BIBLIOGRAPHIQUES

10.1186/s13059-017-1354-4.

Zuris, J. A. *et al.* (2015) 'Efficient Delivery of Genome-Editing Proteins In Vitro and In Vivo', *Nature Biotechnology*. doi: 10.1038/nbt.3081.Efficient.

Références Table 4 :

1. K. Dybvig, G. H. Cassell, Transposition of Gram-Positive Transposon Tn916 in *Acholeplasma laidlawii* and *Mycoplasma pulmonis*. *Science* (80-.). **235**, 14–16 (1987).
2. K. Dybvig, Transformation of *Acholeplasma laidlawii* with streptococcal plasmids pVA868 and pVA920. *Plasmid* **21**, 155–160 (1989).
3. T. K. Sundström, Å. Wieslander, Plasmid transformation and replica filter plating of *Acholeplasma laidlawii*. *FEMS Microbiol. Lett.* **72**, 147–151 (1990).
4. K. Dybvig, A. Woodard, Construction of recA mutants of *Acholeplasma laidlawii* by insertional inactivation with a homologous DNA fragment. *Plasmid* **28**, 262–266 (1992).
5. K. L. Knudtson, F. C. Minion, Construction of Tn4001 lac derivatives to be used as promoter probe vectors in mycoplasmas. *Gene* **137**, 217–222 (1993).
6. P. A. Kapke, K. L. Knudtson, F. C. Minion, Transformation of Mollicutes with Single-Stranded Tn4001 DNA. *Plasmid* **32**, 85–88 (1994).
7. F. C. Minion, P. A. Kapke, Transformation of mycoplasmas. *Methods Mol. Biol.* **104**, 227–234 (1998).
8. B. De Barbeyrac, M. Dupon, P. Rodriguez, H. Renaudin, C. Bébéar, A Tn1545-like transposon carries the *tet(M)* gene in tetracycline resistant strains of *Bacteroides ureolyticus* as well as *Ureaplasma urealyticum* but not *Neisseria gonorrhoeae*. *J. Antimicrob. Chemother.* **37**, 223–232 (1996).
9. A. F. Aboklaish, *et al.*, Random insertion and gene disruption via transposon mutagenesis of *Ureaplasma parvum* using a mini-transposon plasmid. **304**, 1218–1225 (2014).
10. J. Renaudin, *et al.*, Integrative and free *Spiroplasma citri* oriC plasmids: Expression of the *Spiroplasma phoeniceum* spiralin in *Spiroplasma citri*. *J. Bacteriol.* **177**, 2870–2877 (1995).
11. S. Duret, J. L. Danet, M. Garnier, J. Renaudin, Gene disruption through homologous recombination in *Spiroplasma citri*: An scm1-disrupted motility mutant is pathogenic. *J. Bacteriol.* **181**, 7449–7456 (1999).
12. C. Lartigue, S. Duret, M. Garnier, J. Renaudin, New plasmid vectors for specific gene targeting in *spiroplasma citri*. *Plasmid* (2002) [https://doi.org/10.1016/S0147-619X\(02\)00121-X](https://doi.org/10.1016/S0147-619X(02)00121-X).
13. C. Lartigue, A. Blanchard, J. Renaudin, F. Thiaucourt, P. Sirand-Pugnet, Host specificity of mollicutes oriC plasmids: Functional analysis of replication origin. *Nucleic Acids Res.* **31**, 6610–6618 (2003).
14. M. Breton, *et al.*, First report of a tetracycline-inducible gene expression system for mollicutes. *Microbiology* **156**, 198–205 (2010).
15. K. Mutaqin, J. L. Comer, A. C. Wayadande, U. Melcher, Selection and characterization of *Spiroplasma citri* mutants by random transposome mutagenesis. *Can. J. Microbiol.* **532**, 525–532 (2011).
16. N. Terahara, I. Tulum, M. Miyata, Transformation of crustacean pathogenic bacterium *Spiroplasma eriocheiris* and expression of yellow fluorescent protein. *Biochem. Biophys. Res. Commun.* **487**, 488–493 (2017).
17. D. Matteau, M. Pepin, V. Baby, S. Gauthier, T. F. Knight, Inferring the Minimal Genome of *Mesoplasma florum* by Comparative Genomics and Transposon Mutagenesis. *Aem* **83**, 1–16 (2017).
18. V. Baby, *et al.*, Cloning and Transplantation of the *Mesoplasma florum* Genome. *ACS Synth. Biol.* **7**, 209–217 (2018).
19. K. W. King, K. Dybvig, Transformation of *Mycoplasma capricolum* and examination of DNA restriction modification *Mycoplasma capricolum* and *Mycoplasma mycoides* subsp. *mycoides*. *Plasmid* **31**, 308–311 (1994).
20. C. Janis, *et al.*, Versatile Use of oriC Plasmids for Functional Genomics of *Mycoplasma capricolum* subsp. *capricolum*. *Society* **71**, 2888–2893 (2005).
21. A. B. Allam, L. Reyes, N. G. Assad-Garcia, J. I. Glass, M. B. Brown, Enhancement of targeted homologous recombination in *Mycoplasma mycoides* subspecies *capri* by Inclusion of heterologous recA. *Appl. Environ. Microbiol.* **76**, 6951–6954 (2010).
22. J. C. Whitley, L. R. Finch, Location of sites of transposon Tn916 insertion in the *Mycoplasma mycoides* genome. *J. Bacteriol.* **171**, 6870–6872 (1989).
23. K. W. King, K. Dybvig, Plasmid transformation of *Mycoplasma mycoides* subspecies *mycoides* is promoted by high concentrations of polyethylene glycol. *Plasmid* **26**, 108–115 (1991).
24. C. Janis, *et al.*, Unmarked insertional mutagenesis in the bovine pathogen *Mycoplasma mycoides* subsp. *mycoides* SC. *Microbiology* **154**, 2427–2436 (2008).
25. T. Bonnefois, *et al.*, Development of fluorescence expression tools to study host-mycoplasma interactions and validation in two distant mycoplasma clades. *J. Biotechnol.* **236**, 35–44 (2016).
26. A. M. Mariscal, *et al.*, Tuning Gene Activity by Inducible and Targeted Regulation of Gene Expression in Minimal Bacterial Cells. *ACS Synth. Biol.* **7**, 1538–1552 (2018).
27. C. A. Hutchison, *et al.*, Polar Effects of Transposon Insertion into a Minimal Bacterial Genome. *J. Bacteriol.* (2019).
28. A. Montero-Blay, S. Miravet-Verde, M. Lluch-Senar, C. Piñero-Lambea, L. Serrano, SynMyco transposon: engineering transposon vectors for efficient transformation of minimal genomes. *DNA Res.* **26**, 327–339 (2019).
29. M. C. Roberts, G. E. Kenny, Conjugal transfer of transposon Tn916 from *Streptococcus faecalis* to *Mycoplasma hominis*. *J. Bacteriol.* **169**, 3836–3839 (1987).
30. F. Rideau, *et al.*, Random transposon insertion in the *Mycoplasma hominis* minimal genome. *Sci. Rep.* **9**, 1–11 (2019).
31. L. L. Voelker, K. Dybvig, Gene transfer in *Mycoplasma arthritidis*: Transformation, conjugal transfer of Tn916, and evidence for a restriction system recognizing AGCT. *J. Bacteriol.* **178**, 6078–6081 (1996).
32. K. Dybvig, C. T. French, L. R. L. Voelker, Construction and use of derivatives of transposon Tn4001 that function in *Mycoplasma pulmonis* and *Mycoplasma arthritidis*. *J. Bacteriol.* **182**, 4343–4347 (2000).
33. L. R. Washburn, D. W. Bird, K. Dybvig, Restoration of cytoadherence to an adherence-deficient mutant of *Mycoplasma arthritidis* by genetic complementation. *Infect. Immun.* **71**, 671–675 (2003).
34. K. Dybvig, J. Alderete, Transformation of *Mycoplasma pulmonis* and *Mycoplasma hyorhinae*: Transposition of Tn916 and formation of cointegrate structures. *Plasmid* **20**, 33–41 (1988).
35. H. Z. A. Ishag, Q. Xiong, M. Liu, Z. Feng, G. Shao, Development of oriC-plasmids for use in *Mycoplasma hyorhinae*. *Sci. Rep.* **7**, 1–10 (2017).
36. H. Z. A. Ishag, Q. Xiong, M. Liu, Z. Feng, G. Shao, E. coli recA gene improves gene targeted homologous recombination in *Mycoplasma hyorhinae*. *J. Microbiol. Methods* **136**, 49–56 (2017).
37. G. G. Mahairas, C. Jian, F. C. Minion, Development of a cloning system in *Mycoplasma pulmonis*. *Gene* (1990).
38. G. G. Mahairas, F. C. Minion, Random Insertion of the Gentamicin Resistance Transposon Tn4001 in *Mycoplasma pulmonis*. *Plasmid* (1989).
39. C. M. M. Cordova, C. Lartigue, P. Sirand-pugnet, R. A. F. Cunha, A. Blanchard, Identification of the Origin of Replication of the *Mycoplasma pulmonis* Chromosome and Its Use in oriC Replicative Plasmids. *J. Bacteriol.* **184**, 5426–5435 (2002).

40. G. A. Maglennon, *et al.*, Development of a self-replicating plasmid system for *Mycoplasma hyopneumoniae*. *Vet. Res.* **44**, 1–10 (2013).
41. G. A. Maglennon, *et al.*, Transposon mutagenesis in *Mycoplasma hyopneumoniae* using a novel mariner-based system for generating random mutations. *Vet. Res.* **44**, 1–11 (2013).
42. H. Z. A. Ishag, *et al.*, A replicating plasmid-based vector for GFP expression in *Mycoplasma hyopneumoniae*. *Genet. Mol. Res.* **15**, 1–8 (2016).
43. J. M. Clampitt, M. L. Madsen, F. C. Minion, Construction of *Mycoplasma hyopneumoniae* P97 Null Mutants. *Front. Microbiol.* **12**, 1–12 (2021).
44. R. Chopra-Dewasthaly, M. Zimmermann, R. Rosengarten, C. Citti, First steps towards the genetic manipulation of *Mycoplasma agalactiae* and *Mycoplasma bovis* using the transposon Tn4001mod. *Int. J. Med. Microbiol.* **294**, 447–453 (2005).
45. R. Chopra-Dewasthaly, M. Marena, R. Rosengarten, W. Jechlinger, C. Citti, Construction of the first shuttle vectors for gene cloning and homologous recombination in *Mycoplasma agalactiae*. *FEMS Microbiol. Lett.* **253**, 89–94 (2005).
46. S. Sharma, K. A. Tivendale, P. F. Markham, G. F. Browning, Disruption of the membrane nuclease gene (MBOVPG45_0215) of *Mycoplasma bovis* greatly reduces cellular nuclease activity. *J. Bacteriol.* **197**, 1549–1558 (2015).
47. S. Hegde, R. Rosengarten, R. Chopra-Dewasthaly, Disruption of the pdhB Pyruvate Dehydrogenase Gene Affects Colony Morphology, In Vitro Growth and Cell Invasiveness of *Mycoplasma agalactiae*. *PLoS One* **10**, e0119706 (2015).
48. B. N. Kent, M. F. Foecking, M. J. Calcutt, Development of a novel plasmid as a shuttle vector for heterologous gene expression in *Mycoplasma yeastsii*. *J. Microbiol. Methods* **91**, 121–127 (2012).
49. M. A. Shahid, M. S. Marena, P. F. Markham, A. H. Noormohammadi, Development of an oriC vector for use in *Mycoplasma synoviae*. *J. Microbiol. Methods* **103**, 70–76 (2014).
50. S. W. Lee, G. F. Browning, P. F. Markham, Development of a replicable oriC plasmid for *Mycoplasma gallisepticum* and *Mycoplasma imitans*, and gene disruption through homologous recombination in *M. gallisepticum*. *Microbiology* (2008) <https://doi.org/10.1099/mic.0.2008/019208-0>.
51. J. Cao, P. A. Kapke, F. C. Minton, Transformation of *Mycoplasma gallisepticum* with Tn916, Tn4001, and integrative plasmid vectors. *J. Bacteriol.* **176**, 4459–4462 (1994).
52. D. C. Ruffin, *et al.*, Transposon mutagenesis of *Mycoplasma gallisepticum* by conjugation with *Enterococcus faecalis* and determination of insertion site by direct genomic sequencing. *Plasmid* **44**, 191–195 (2000).
53. I. Pour-Ei, C. Adams, F. C. Minion, Construction of mini-Tn4001tet and its use in *Mycoplasma gallisepticum*. *Plasmid* **47**, 129–137 (2002).
54. P. L. Whetzel, L. L. Hnatow, C. L. Keeler, J. E. Dohms, Transposon mutagenesis of *Mycoplasma gallisepticum*. *Plasmid* **49**, 34–43 (2003).
55. S. Mudahi-Orenstein, S. Levisohn, S. J. Geary, D. Yogeve, Cytadherence-deficient mutants of *Mycoplasma gallisepticum* generated by transposon mutagenesis. *Infect. Immun.* **71**, 3812–3820 (2003).
56. P. Hudson, *et al.*, Identification of a virulence-associated determinant, dihydroliipoamide dehydrogenase (lpd), in *Mycoplasma gallisepticum* through in vivo screening of transposon mutants. *Infect. Immun.* **74**, 931–939 (2006).
57. H. J. Chen, *et al.*, Construction of mini-Tn4001 transposon vector for *Mycoplasma gallisepticum*. *Sci. China Life Sci.* **53**, 1340–1345 (2010).
58. I. Indikova, M. Vronka, M. P. Szostak, First identification of proteins involved in motility of *Mycoplasma gallisepticum*. *Vet. Res.* **45**, 1–13 (2014).
59. C. T. Hedreyda, K. K. Lee, D. C. Krause, Transformation of *Mycoplasma pneumoniae* with Tn4001 by Electroporation. *Plasmid* **30**, 170–175 (1993).
60. C. T. Hedreyda, D. C. Krause, Identification of a possible cytodherence regulatory locus in *Mycoplasma pneumoniae*. *Infect. Immun.* **63**, 3479–3483 (1995).
61. D. C. Krause, *et al.*, Transposon mutagenesis reinforces the correlation between *Mycoplasma pneumoniae* cytoskeletal protein HMW2 and cytodherence. *J. Bacteriol.* **179**, 2668–2677 (1997).
62. C. U. Zimmerman, R. Herrmann, Synthesis of a small, cysteine-rich, 29 amino acids long peptide in *Mycoplasma pneumoniae*. *FEMS Microbiol. Lett.* **253**, 315–321 (2005).
63. B. M. Hasselbring, C. A. Page, E. S. Sheppard, D. C. Krause, Transposon mutagenesis identifies genes associated with *Mycoplasma pneumoniae* gliding motility. *J. Bacteriol.* **188**, 6335–6345 (2006).
64. R. Krishnakumar, *et al.*, Targeted chromosomal knockouts in *Mycoplasma pneumoniae*. *Appl. Environ. Microbiol.* **76**, 5297–5299 (2010).
65. C. Blötz, *et al.*, Development of a replicating plasmid based on the native oriC in *Mycoplasma pneumoniae*. *Microbiol. (United Kingdom)* **164**, 1372–1382 (2018).
66. C. Piñero-Lambea, *et al.*, *Mycoplasma pneumoniae* Genome Editing Based on Oligo Recombineering and Cas9-Mediated Counterselection. *ACS Synth. Biol.* **9**, 1693–1704 (2020).
67. L. Garcia-Morales, *et al.*, A RAGE Based Strategy for the Genome Engineering of the Human Respiratory Pathogen *Mycoplasma pneumoniae*. *ACS Synth. Biol.* **9**, 2737–2748 (2020).
68. S. P. Reddy, W. G. Rasmussen, J. B. Baseman, Isolation and characterization of transposon Tn4001-generated, cytodherence-deficient transformants of *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *FEMS Immunol. Med. Microbiol.* **15**, 199–211 (1996).
69. S. Dhandayuthapani, W. G. Rasmussen, J. B. Baseman, Disruption of gene mg218 of *Mycoplasma genitalium* through homologous recombination leads to an adherence-deficient phenotype. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5227–5232 (1999).
70. O. Q. Pich, R. Burgos, R. Planell, E. Querol, J. Piñol, Comparative analysis of antibiotic resistance gene markers in *Mycoplasma genitalium*: Application to studies of the minimal gene complement. *Microbiology* **152**, 519–527 (2006).
71. R. Burgos, *et al.*, *Mycoplasma genitalium* P140 and P110 cytoadhesins are reciprocally stabilized and required for cell adhesion and terminal-organelle development. *J. Bacteriol.* **188**, 8627–8637 (2006).
72. J. Li, *et al.*, Development of a replicative plasmid for gene expression in *Mycoplasma bovis*. *J. Microbiol. Methods* **108**, 12–18 (2015).
73. A. M. Mariscal, L. González-González, E. Querol, J. Piñol, All-in-one construct for genome engineering using Cre-lox technology. *DNA Res.* **23**, 263–270 (2016).
74. X. Foissac, C. Saillard, J. M. Bové, Random Insertion of Transposon Tn4001 in the Genome of *Spiroplasma citri* Strain G13. **86**, 80–86 (1997).
75. M. Breton, S. Duret, N. Arricau-Bouvery, L. Béven, J. Renaudin, Characterizing the replication and stability regions of *Spiroplasma citri* plasmids identifies a novel replication protein and expands the genetic toolbox for plant-pathogenic spiroplasmas. *Microbiology* **154**, 3232–3244 (2008).

CRISPR/Cas9 chez les Mollicutes : Du système naturel aux outils d'ingénierie de génome

Les systèmes CRISPR/Cas sont largement représentés chez les bactéries et les archées et sont à la base de mécanismes de défense adaptatifs contre l'ADN et l'ARN exogène. De par son intérêt comme outil génétique, le système CRISPR de type II de *Streptococcus pyogenes* a été particulièrement étudié. Il est composé principalement d'une endonucléase Cas9 et d'un sgRNA qui a pour rôle de guider Cas9 jusqu'à une séquence cible, la spécificité de reconnaissance dépendant également d'une séquence NGG située en aval appelée PAM. Les mollicutes sont des bactéries minimales, dépourvues de paroi et qui utilisent, pour la plupart, un code génétique alternatif. Elles possèdent de petits génomes riches en AT (~70-75%) allant de 0,58 à 2,2 Mpb. La plupart d'entre elles ont un mode de vie parasitique et sont des pathogènes d'une grande diversité d'hôtes (plantes et animaux y compris humains). Chez la plupart des mollicutes, les possibilités d'ingénierie génomique restent limitées car la recombinaison homologue est peu efficace. Le but de ma thèse a été (1) de caractériser les systèmes CRISPR/Cas9 des mollicutes, (2) de développer un outil endogène CRISPR/Cas9 pour l'ingénierie de génome de mollicutes et (3) d'étendre cette boîte à outils chez certaines espèces d'intérêt pour lesquelles aucun outil génétique de précision n'était disponible.

La première partie de mon travail de thèse était l'étude globale des systèmes CRISPR chez les mollicutes. Une reconstruction phylogénétique a permis de mettre en évidence une origine commune pour tous les systèmes CRISPR de mycoplasmes et deux origines pour les systèmes de spiroplasmales. L'étude structurale de la protéine Cas9 des mollicutes montre une forte proximité structurale avec la protéine Cas9 de *Staphylococcus aureus* à l'exception du domaine PI qui intervient dans l'interaction avec la séquence PAM. Ceci suggère que les séquences PAM des systèmes de mollicutes sont sans doute très différentes. Le système CRISPR/Cas9 de *Mycoplasma gallisepticum* (*Mgal*), un pathogène d'oiseau, a ensuite été l'objet d'une étude approfondie. Le système naturel a été démontré comme fonctionnel *in vivo* et le système minimal composé de la *MgalCas9* et d'un sgRNA a permis d'induire un clivage de l'ADN *in vitro*. Une différence de séquence PAM consensus a été observée entre deux souches de *Mgal* isolées de poulet et de pinson et a pu être mise en corrélation avec le récent changement d'hôte du pathogène et l'émergence de celui-ci dans les populations de pinsons d'Amérique. Un plasmide « all-in-one » contenant le système a été conçu, et la fonctionnalité du système comme outil précis de clivage a été démontré dans différentes espèces, notamment *Mgal*, *Mycoplasma capricolum* (*Mcap*), *Mycoplasma pulmonis* et *Mycoplasma mycoides*. Cet outil de contre sélection a été appliqué à *Mcap* afin de curer son génome d'un élément mobile intégré de 23 kpb (ICE). Finalement, pour pallier la faible efficacité de recombinaison des mycoplasmes, deux nouveaux outils d'ingénierie génomique ont été mis au point. Un outil « CRISPR Base Editor », composé d'une protéine Cas9 désactivée et fusionnée à une protéine de type désaminase, qui permet de faire des modifications de base de manière très ciblée. La fonctionnalité de cet outil, très efficace et simple d'utilisation, a été validée dans 3 espèces pathogènes majeures de mycoplasmes. Le second outil est basé sur l'importation d'un système de recombinaison RecET issu de phages. Chez *Mgal*, cet outil d'ingénierie de génome permet ainsi pour la première fois de réaliser des modifications précises à plus grande échelle (délétions, insertions et remplacements).

En conclusion, nos travaux ont concerné l'étude fonctionnelle des systèmes CRISPR des mycoplasmes et le développement d'outils d'ingénierie de génomes chez ces mêmes bactéries. Ce travail ouvre de manière significative les possibilités d'études et d'utilisation des mycoplasmes pour des applications biotechnologiques.

Mots clés : CRISPR, Ingénierie de génome, Mycoplasme, *Mycoplasma gallisepticum*

UMR 1332 Biologie du Fruit et Pathologie - Equipe Mollicutes

Centre INRA Bordeaux-Aquitaine, Bâtiment IBVM 71 Avenue Edouard Bourlaux, CS 20032
33882 Villenave d'Ornon Cedex, France