UNIVERSITÉ DE TOURS

ÉCOLE DOCTORALE : SSBCV

UMR Infectiologie et Santé Publique

Equipe Bactéries et Risque Materno-Foetal

THÈSE présentée par :

Adeline PASTUSZKA

soutenue le : **15 septembre 2021**

pour obtenir le grade de : **Docteur de l'université de Tours**

Discipline / Spécialité : Biologie / Microbiologie

LE SYSTEME CRISPR-CAS DE TYPE II-A DES SOUCHES HYPERVIRULENTES DE *STREPTOCOCCUS AGALACTIAE* : ROLE ET SPECIFICITES

THÈSE dirigée par :

Monsieur LANOTTE Philippe	Professeur, Université de Tours

RAPPORTEURS :

Monsieur BIKARD David	Directeur de Recherche, Institut Pasteur, Paris
Madame PAYOT-LACROIX Sophie	Directrice de Recherche, INRAE Nancy

PRESIDENT :

Madame POYART Claire	Professeur, Université de Paris

JURY :

Monsieur BIKARD David	Directeur de Recherche, Institut Pasteur, Paris
Madame HERVE Virginie	Maître de Conférences, Université de Tours
Monsieur HORVATH Philippe	Senior Scientist, IFF, Dangé-Saint-Romain
Monsieur LANOTTE Philippe	Professeur, Université de Tours
Monsieur MOINEAU Sylvain	Professeur, Université Laval, Québec
Madame PAYOT-LACROIX Sophie	Directrice de Recherche, INRAE Nancy
Madame POYART Claire	Professeur, Université de Paris

Remerciements

Madame le Professeur Claire Poyart,

Vous me faites l'honneur de présider ce jury de thèse et de juger ce travail. Soyez assurée de ma gratitude et de mon profond respect.

Madame le Docteur Sophie Payot-Lacroix,

Vous avez accepté de rapporter et de juger mon travail. Soyez assurée de ma sincère reconnaissance.

Monsieur le Docteur David Bikard,

Je vous remercie d'avoir accepté d'être rapporteur de cette thèse et d'en juger le contenu. Soyez assuré de mon profond respect.

Madame le Docteur Virginie Hervé,

Un grand merci pour avoir accepté de participer à ce jury de thèse. Soyez assurée de ma gratitude.

Monsieur le Docteur Philippe Horvath,

Vous me faites l'honneur de juger ce travail. Je vous remercie pour votre accueil et la dispensation de vos précieux conseils, à différents moments clés de ce travail. Soyez assuré de mon profond respect.

Monsieur le Professeur Sylvain Moineau,

Je suis honorée que tu aies accepté de participer à ce jury. Ce fut une grande chance pour moi d'être accueillie dans ton équipe, tant professionnellement que personnellement. Souhaitons que notre collaboration future soit fructueuse.

Monsieur le Professeur Philippe Lanotte,

Je vous remercie pour l'encadrement de ce travail de thèse, vous m'avez fait découvrir un sujet qui me passionne (toujours !) aujourd'hui. J'espère avoir fait honneur à la confiance que vous m'avez accordée. Je vous remercie pour la richesse de nos échanges, le partage de vos précieux conseils et votre bienveillance. Soyez assuré de ma sincère reconnaissance et de mon profond respect.

Un merci particulier à Monsieur le Professeur Laurent Mereghetti. Merci de m'avoir accueillie au sein du laboratoire de Bactériologie et dans l'équipe Bactéries et Risque Materno-Fœtal. Au cours de ces années, j'ai grandement apprécié travailler à vos côtés et la qualité de nos échanges, toujours dans la bonne humeur. Je vous suis reconnaissante pour votre disponibilité et votre implication à mon égard.

Je tiens à remercier l'ensemble de l'équipe BRMF, anciens et nouveaux membres, Aurélia, Daniel, Frédérique, Philippe, Kévin, Pauline, le groupe des « thésardes »… pour leur accueil chaleureux, leur aide quotidienne et leur rigueur scientifique. Je remercie particulièrement Emilie, qui a suivi mes travaux de près avec beaucoup de bienveillance et qui a toujours su trouver un peu de temps pour répondre à mes (nombreuses) interrogations. Un immense merci à Vanessa, pour m'avoir initiée au monde de la recherche et ses techniques, pour son soutien et son amitié précieuse.

Je remercie également toute l'équipe du laboratoire de Bactériologie-Virologie de l'hôpital Bretonneau. Je remercie l'ensemble de mes collègues biologistes, pour leur accueil amical et le partage de leurs connaissances, permettant un apprentissage dans la bonne humeur. Merci aussi à toute l'équipe des techniciens et aux internes pour les très bons moments passés « à la paillasse », mais aussi (et surtout !) en dehors.

Merci à l'ensemble de l'équipe du laboratoire de Sylvain Moineau pour leur accueil, leur altruisme, le partage de leur expertise et les échanges conviviaux. Merci à Geneviève pour l'encadrement amical et à Cécile pour m'avoir fait découvrir le Québec.

Un grand merci à Cathy Schouler et Éric Morello pour la qualité des échanges scientifiques que nous avons eu lors des Comités de Thèse.

Merci au Docteur Arnaud Firon pour la mise à disposition des souches et des résultats concernant le système CovRS et pour les échanges fructueux qui ont suivi.

Merci à l'équipe Signalisation, Portage et Virulence Bactérienne et particulièrement à Isabelle Virlogeux-Payant pour son aide pour la réalisation du travail sur l'analyse protéique de Cas9.

Je remercie Elodie, Clémence et Clément, pour la mise à disposition des résultats de leurs travaux.

Merci à l'ensemble des étudiants, Marion, Vincent, Philippe, Charlotte… avec qui j'ai eu le plaisir de travailler et qui ont participé, de près ou de loin à l'accomplissement de ce travail expérimental. En particulier, merci à Coralie pour son aide, sa gentillesse et sa bonne humeur.

Merci à mes amis de Tours, de Poitiers, de Mimizan… pour les moments de joie et de partage passés ensemble et pour ceux qui viendront.

Enfin, un grand merci à toute ma famille, mon frère et surtout mes parents. Merci de m'avoir soutenue depuis toujours, de m'avoir écoutée parler de mon travail et surtout de m'avoir comprise quand j'en parlais ! Cette fois-ci Papa, c'est une « vraie » thèse, elle est pour toi ;) Un dernier merci particulier à Guillaume, qui est entré dans ma vie pendant cette thèse et qui a survécu, tant bien que mal, à la douloureuse période de rédaction. Merci pour ta patience et ton soutien.

Résumé

Les systèmes CRISPR-Cas représentent un mécanisme important de défense développé par les bactéries, composés de deux grandes régions : le locus CRISPR, constitué d'unités repeat-spacer et l'opéron *cas*, codant les protéines effectrices du système. *S. agalactiae*, pathogène humain majeur de la période néonatale et pathogène émergent chez l'adulte, possède un système CRISPR1, de type II-A, composé des gènes *cas9*, *cas1*, *cas2* et *csn2*. Des travaux préliminaires ont amorcé l'ébauche d'un lien entre la virulence de cette bactérie, associée principalement aux souches de ST17 et la fonctionnalité de son système CRISPR1-Cas. En effet, ces souches hypervirulentes possèdent significativement moins de spacers que les souches des autres lignées phylogénétiques. Afin d'expliciter ces différences, nous avons centré notre travail sur l'étude de la région promotrice en amont de l'opéron *cas*. L'alignement des séquences de cette région a mis en évidence une délétion de 103 paires de bases chez les souches de ST17, entraînant la disparition du promoteur de l'opéron *cas*. L'expérience de fusion transcriptionnelle confirme l'absence d'activité de ce promoteur chez ces souches, corroborée par l'analyse par RT-PCR quantitative, qui montre une expression très faible de *cas9*. Alors que ces résultats confortent l'hypothèse d'un système CRISPR1-Cas moins actif chez les souches de ST17, l'expérience d'interférence réalisée par la suite, a montré que le système semble tout de même conserver sa capacité de mécanisme de défense vis à vis d'une séquence nucléotidique déjà acquise. De même, nous avons démontré pour la première fois *in vitro*, la capacité d'intégration de nouveaux spacers au sein du locus CRISPR1 des souches de ST17. Nous avons attribué la fonctionnalité du système CRISPR1-Cas chez les souches hypervirulentes à la présence d'un autre promoteur, P_{srn036}, situé en amont de l'opéron *cas*, suffisant et essentiel pour sa transcription. De plus, nous avons également montré que ce promoteur était régulé par CovR, le régulateur majeur de virulence chez *S. agalactiae*, pouvant potentiellement expliquer les différences observées entre le faible niveau de transcription des gènes *cas* et l'efficacité du système chez les souches hypervirulentes du ST17. Il semble donc exister des mécanismes de régulation du système CRISPR1-Cas de *S. agalactiae*, spécifiques des lignées phylogénétiques, pouvant participer à l'évolution de l'espèce et à son adaptation à l'Homme, en contrôlant par exemple la diversité génétique liée aux éléments génétiques mobiles.

Mots clés : *Streptococcus agalactiae*, CRISPR-Cas, ST17, hypervirulence, fonctionnalité, régulation

Abstract

CRISPR-Cas systems are an important defense mechanism developed by bacteria, composed of two major regions: the CRISPR locus, which is an alternation of repeats and spacers, and the *cas* operon, encoding the effector proteins of the system. *S. agalactiae*, a major human pathogen responsible for invasive neonatal infections and an emerging pathogen in adults, possesses a type II-A system, CRISPR1, characterized by *cas9*, *cas1*, *cas2* and *csn2* genes. Previous work has established a link between the virulence of this bacterium, mainly associated to ST17 strains, and the functionality of the CRISPR1-Cas system. Indeed, the hypervirulent strains possess significantly fewer spacer in their CRISPR1 locus, than the strains belonging to other phylogenetic lineages. In order to investigate these differences, we focused on studying the promoter region upstream *cas* operon. Sequence alignment performed on this region, revealed a 103 bp deletion in the ST17 strains, resulting in P_{cas} deletion. We performed transcriptional fusion experiments which have confirmed the absence of promoter activity, corroborated by RT-qPCR analysis, which revealed very low expression of *cas9*. While these results support the hypothesis that the ST17 CRISPR1-Cas system might not be properly functioning, we have demonstrated that the system remains capable of cleaving an already encountered foreign element. Moreover, we have demonstrated, for the first time *in vitro*, the capacity of the ST17 system to acquire new spacers in its CRISPR1 locus. We attributed the functionality of the CRISPR1-Cas system of the hypervirulent strains to the presence of another promoter, P_{srn036}, located upstream *cas* operon, essential and sufficient for the transcription of *cas* genes. In addition, we demonstrated that this promoter is regulated by CovR, the major regulator of *S. agalactiae* virulence. This regulation may explain the differences observed between *cas9* low levels of transcription and the efficiency of the CRISPR1-Cas system in hypervirulent strains. Some mechanisms seem able to regulate CRISPR1-Cas expression, particularly in specific phylogenetic lineages, thus participating in species evolution and adaptation to humans by controlling genetic diversity linked to mobile genetic elements.

Keywords: *Streptococcus agalactiae*, CRISPR-Cas, ST17, hypervirulence, functionality, regulation

Table des matières

Résumé 5
Abstract 6
Table des matières 7
Liste des tableaux 10
Liste des figures 11
Liste des abréviations 14
Revue de la littérature 15
1. *Streptococcus agalactiae* 16
1.1. Pouvoir pathogène 16
1.1.1. Réservoir 16
1.1.2. Infections chez le nouveau-né 17
1.1.3. Infections chez l'adulte 20
1.1.4. Virulence de *S. agalactiae* 22
1.2. Epidémiologie et diversité génétique de *S. agalactiae* 23
1.2.1. Sérotypage capsulaire 24
1.2.2. Multi Locus Sequence Typing 25
1.2.3. Le clone hypervirulent CC17 26
1.2.4. Variations épidémiologiques 28
1.2.5. Autres méthodes de typage 29
2. Les systèmes CRISPR-Cas 31
2.1. Historique 32
2.2. Organisation des systèmes CRISPR-Cas 33
2.2.1. Le locus CRISPR 34
2.2.2. Le leader 34
2.2.3. Les gènes *cas* 35
2.3. Classification des systèmes CRISPR-Cas 35
2.4. Mode d'action des systèmes CRISPR-Cas 38
2.4.1. Adaptation 39
2.4.2. Expression 40
2.4.3. Interférence 41
2.5. Applications dérivées des systèmes CRISPR-Cas 43
2.5.1. « Genome editing » 43
2.5.2. Typage moléculaire 44
2.5.3. Les systèmes CRISPR-Cas comme antibactériens 45

2.6. CRISPR-Cas et virulence bactérienne 46
2.6.1. Limitation des transferts horizontaux de gènes par les systèmes CRISPR-Cas 46
2.6.2. Régulation de gènes bactériens par les systèmes CRISPR-Cas 48
2.7. La régulation des systèmes CRISPR-Cas 50
3. Le système CRISPR-Cas de type II-A de *S. agalactiae* 54
3.1. Organisation 54
3.1.1. Organisation générale des systèmes de type II-A 54
3.1.2. Organisation du système CRISPR1-Cas de *S. agalactiae* 54
3.2. Fonctionnalité 56
3.3. Travaux préliminaires 57
Objectifs 59
Travail expérimental 61
Partie 1 : Etude de la fonctionnalité du système CRISPR-Cas de type II-A des souches hypervirulentes de *S. agalactiae* 62
1. Matériel et Méthodes 62
1.1. Souches étudiées 62
1.2. Analyse *in silico* : alignement et comparaison nucléotidique 63
1.3. Extraction de l'ADN génomique bactérien 63
1.4. Réactions d'amplification 64
1.5. Séquençage 64
1.6. Préparation de bactéries compétentes et transformation 65
1.7. Extraction de l'ADN plasmidique 65
1.8. Préparation des ARNs totaux de *S. agalactiae* 65
1.9. 5'-RACE PCR 66
1.10. Fusion transcriptionnelle 67
1.11. PCR quantitative en temps réel 70
1.12. Mutagénèse chez *S. agalactiae* 72
1.13. Expériences d'interférence 75
1.14. Acquisition de spacers 77
2. Résultats 80
2.1. Etude des régions promotrices du système CRISPR1-Cas de *S. agalactiae* 80
2.2. Etude de l'activité promotrice par fusion transcriptionnelle 85
2.3. Etude de la transcription de *cas9* par qPCR 87
2.4. Démonstration de la fonctionnalité du système 87
2.5. Implication de Cas9 dans la fonctionnalité du système 93
2.6. Identification du promoteur de l'opéron *cas* chez les souches du ST17 94

2.7. Spécificités du système CRISPR1-Cas des souches hypervirulentes..... 98
2.8. Etude des SNPs et du « terminateur de transcription » de *srn036*..... 99
2.9. Valorisation..... 101
PUBLICATION 103
Partie 2 : Etude de la régulation de l'expression du système CRISPR1-Cas..... 115
1. Matériel et Méthodes 115
1.1. Souches étudiées..... 115
1.2. Etude de la fonctionnalité du système CRISPR1-Cas..... 115
1.3. Détermination de l'activité promotrice de P_{srn036} dans le mutant BM110Δ*covR*..... 116
1.4. Identification de conditions régulant l'expression du système CRISPR1-Cas 117
1.5. Etude de la transcription de *cas9* par PCR quantitative en temps réel..... 118
1.6. Etude de l'expression de la protéine Cas9 119
2. Résultats 123
2.1. Régulation du système CRISPR1-Cas par le régulateur majeur de virulence CovR..... 123
2.2. Etude de la transcription de *cas9* à travers l'application de divers stimuli..... 127
2.3. Etude de l'expression de Cas9..... 132
Discussion..... 135
Conclusion..... 147
Références bibliographiques..... 151
Valorisation de la recherche..... 165

Liste des tableaux

Tableau I : Répartition des sérotypes de 651 souches de *S. agalactiae*, responsables d'infections invasives chez le nouveau-né et l'adulte, isolées en France entre 2006 et 2009. 27
Tableau II : Principales caractéristiques des souches de *S. agalactiae* étudiées. 62
Tableau III : Distribution of the 331 strains on which the alignment of the *cas* promoter region was performed. 63
Tableau IV : Amorces utilisées pour la PCR de vérification de l'absence de contamination des ARNs par de l'ADNg. 66
Tableau V : Amorces utilisées pour la réaction de 5'-RACE PCR. 67
Tableau VI : Amorces utilisées pour la construction des fusions transcriptionnelles. 68
Tableau VII : Amorces utilisées pour la réaction de PCR quantitative en temps réel. 71
Tableau VIII : Amorces utilisées pour la construction des différents mutants de *S. agalactiae*. 74
Tableau IX : Amorces utilisées pour amplifier les spacers d'intérêt avant leur insertion sur le vecteur pG1^{ts}. 77
Tableau X : Amorces permettant l'amplification du locus CRISPR1 des différentes souches de *S. agalactiae* d'intérêt. 79
Tableau XI : Résumé des différents SNPs observés dans la séquence du leader, leur fréquence relative et leur distribution au sein des CCs. 82
Tableau XII : Amorces utilisées pour le séquençage de la région intergénique. 84
Tableau XIII : Amorces utilisées pour quantifier la transcription de *srn036* par PCR quantitative en temps réel. 117
Tableau XIV : Amorces utilisées pour les réactions de PCR quantitative en temps réel. 119
Tableau XV : Amorces utilisées pour le marquage de la protéine Cas9 chez les différentes souches de *S. agalactiae*. 120

Liste des figures

Figure 1 : Physiopathologie des infections néonatales à *S. agalactiae* 18
Figure 2 : Incidence des infections invasives néonatales précoces et tardives à Streptocoque du groupe B, France métropolitaine 1996-2018. 20
Figure 3 : Incidence pour 100 000 habitants des infections invasives (méningites et bactériémies) à *S. agalactiae* par groupe d'âge, Epibac, France métropolitaine, 2018. 21
Figure 4 : Comparaison de la fréquence des différentes manifestations cliniques des infections invasives de l'adulte à SGB entre 2016 et 2017. 21
Figure 5 : Distribution mondiale des sérotypes dans les infections invasives néonatales à *S. agalactiae* 24
Figure 6 : Répartition phylogénétique mondiale des souches de *S. agalactiae* humaines. 26
Figure 7 : Fréquence des sérotypes capsulaires des souches de SGB responsables d'infections invasives chez l'adulte en fonction de l'année. 28
Figure 8 : Répartition au sein des CCs des souches de *S. agalactiae* isolées au CHRU de Tours entre 2000 et 2018. 29
Figure 9 : Corrélation entre la composition en spacers du locus CRISPR1 et le CC chez *S. agalactiae*. 30
Figure 10 : Les différentes étapes du cycle d'infection phagique. 31
Figure 11 : Nombre de publications indexées dans Pubmed, citant le mot CRISPR, entre 2002 et le 10 juin 2021. 33
Figure 12 : Organisation générale d'un système CRISPR-Cas. 33
Figure 13 : Organisation des deux classes de systèmes CRISPR-Cas. 35
Figure 14 : Classification des systèmes de classe 1. 36
Figure 15 : Classification des systèmes CRISPR-Cas de classe 2. 37
Figure 16 : Récapitulatif du mode d'action des systèmes CRISPR-Cas. 38
Figure 17 : Description de l'étape d'adaptation. 39
Figure 18 : Représentation des étapes d'expression et d'interférence des systèmes CRISPR-Cas de type II-A. 40
Figure 19 : Implication de différentes protéines Cas en fonction des étapes et en fonction du type de système CRISPR-Cas. 41
Figure 20 : Récapitulatif des protéines Cas effectrices et de leur mode d'action en fonction des types de système. 42
Figure 21 : Le PAM permet la bonne orientation des spacers au sein du locus CRISPR. 43
Figure 22 : Principe du « genome editing » dérivé des systèmes CRISPR-Cas de type II-A. 44
Figure 23 : Principe de l'utilisation du polymorphisme des loci CRISPR comme méthode de typage moléculaire. 45
Figure 24 : Principe de l'utilisation des systèmes CRISPR-Cas comme antimicrobiens. 46
Figure 25 : Association entre le nombre de spacers et la capacité des souches de STEC à engendrer des épidémies (A) ou des pathologies sévères (B) ou avec la présence des gènes *stx* (C). 47
Figure 26 : Corrélation inverse entre le nombre de prophages et le nombre de spacers par génome chez *S. pyogenes*. 48
Figure 27 : Les deux complexes ribonucléoprotéiques de *F. novicida*. 49
Figure 28 : Le quorum sensing régule l'expression des systèmes CRISPR-Cas de *Serratia* sp. 51
Figure 29 : Etude de l'étape d'adaptation en fonction de la température chez *P. aeruginosa*. 51

Figure 30 : Représentation schématique de la réponse BaeSR suite à un stress au niveau de l'enveloppe. 52
Figure 31 : Le *tracr-L* guide Cas9 pour autoréguler l'expression du système CRISPR-Cas. 53
Figure 32 : Organisation générale des systèmes de type II-A 54
Figure 33 : Organisation du système CRISPR1-Cas de *S. agalactiae*. 55
Figure 34 : Nombre de spacers par locus en fonction du complexe clonal. 58
Figure 35 : Organisation du plasmide pTCV-*lac*. 67
Figure 36 : Organisation du plasmide pG1^{ts}. 72
Figure 37 : Représentation schématique de la construction des plasmides vecteurs de protospacer à partir du locus CRISPR1 des souches *S. agalactiae* NEM316 (ST23) et BM110 (ST17). 76
Figure 38 : Principe des expériences d'acquisition. 78
Figure 39 : Deletion of the regular *cas* promoter in the ST17 lineage. 81
Figure 40 : Nucleotidic sequence of the intergenic region upstream *cas9* of the three *S. agalactiae* strains NEM316 (ST23/CC23), 2603V/R (ST110/CC19) and BM110 (ST17/CC17). 83
Figure 41 : Identification d'un promoteur alternatif par 5'-RACE PCR chez *S. agalactiae* BM110. 84
Figure 42 : The large deletion in ST17 strains reduces *cas* promoter activity and *cas9* transcription. 86
Figure 43 : Résultats attendus lors des tests d'interférence. 88
Figure 44 : CRISPR interference in ST17 strain. 89
Figure 45 : Perte de la capacité d'interférence face à un protospacer homologue d'un spacer anciennement acquis. 90
Figure 46 : Adaptation-positive clones over time in response to plasmid challenges. 91
Figure 47 : Représentation schématique des loci CRISPR1 des souches BM110 et A909 après les tests d'acquisition. 92
Figure 48 : CRISPR adaptation in ST17 strain. 93
Figure 49 : Cas9 is essential for adaptation. 94
Figure 50 : *srn036* plays a key role in the functionality of the CRISPR1-Cas system in ST17 lineage. .. 95
Figure 51 : Représentation de la séquence promotrice de *srn036* chez les souches sauvages (WT) et son pendant chez les souches mutées (ΔP_{srn036}). 96
Figure 52 : Migration électrophorétique des produits de PCR obtenus après amplification du locus CRISPR1 après « challenge » avec pNZ123. 97
Figure 53 : Représentation graphique du mutant NEM316ΔP_{cas}. 98
Figure 54 : P_{cas} is essential for CRISPR interference in ST23. 99
Figure 55 : Représentation schématique des régions intergéniques en amont de *cas9* des mutants BM110::$IG_{NEM316\Delta Pcas}$ et NEM316ΔP_{cas}::IG_{BM110}. 100
Figure 56 : Détermination de la fonctionnalité du système CRISPR1-Cas de NEM316ΔP_{cas}::IG_{BM110} et BM110::$IG_{NEM316\Delta Pcas}$. 101
Figure 57 : Schématisation de la position du 3Xflag sur le génome de *S. agalactiae* NEM316. 119
Figure 58 : P_{srn036} CovR binding sites detected by DNase I footprint. 123
Figure 59 : L'absence de CovR améliore la capacité d'interférence de *S. agalactiae* BM110 (ST17). 124
Figure 60 : Absence d'impact du système CovRS sur la capacité d'acquisition des souches du ST23. 125
Figure 61 : Restauration de la capacité d'acquisition du système CRISPR1-Cas de *S. agalactiae* BM110 (ST17) en l'absence de CovR. 126
Figure 62 : Effet répresseur de CovR sur P_{srn036}. 127
Figure 63 : Expression relative du gène *cas9* après application d'un « stress acide ». 128
Figure 64 : Evaluation de l'impact de la présence d'un plasmide sur la transcription de *cas9*. 129

Figure 65 : Effet des antibiotiques sur la transcription de *cas9*..... 130
Figure 66 : Expression relative du gène *cas9* chez A909 et BM110 par rapport à NEM316..... 131
Figure 67 : Migration des extraits protéiques colorée au bleu de Neuhoff..... 132
Figure 68 : Evaluation de l'expression de la protéine 3Xflag-Cas9 chez *S. agalactiae*. 133
Figure 69 : L'ajout du 3Xflag entraîne une perte de fonctionnalité du système CRISPR1-Cas. 134

Liste des abréviations

ADN : Acide désoxyribonucléique
ADNc : ADN complémentaire
ADNg : ADN génomique
ARN : Acide ribonucléique
BLAST : Basic Local Alignment Search Tool
Cas : CRISPR-*associated*
CC : Complexe clonal
CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats
crRNA : ARN CRISPR
crRNP : Complexe CRISPR ribonucléoprotéique
dNTP : Désoxyribonucléoside 5' triphosphate
DO : Densité optique
DR : Direct repeat
DRT : DR terminale
DTT : dithiothréitol
EGM : élément génétique mobile
EOD / LOD : Early-onset disease / Late-onset disease
HCl : Acide chlorhydrique
ICE : élément intégratif conjugatif
INT / INP : Infection néonatale tardive / précoce
LB : Lysogeny Broth
MLST : MultiLocus Sequence Typing
NCBI : National Center for Biotechnology Information
PAM : Protospacer adjacent motif
PIM : Plasmid interfering mutant
pb : paires de base
PCR : Polymerase chain reaction
RACE (PCR) : Rapid amplification of cDNA Ends
RBS : Ribosome Binding Site
RNAi : ARN interférence
rpm : rotations par minute
RT-qPCR : Reverse Transcription quantitative Polymerase chain reaction
SGB/GBS : Streptocoque du groupe B
sgRNA : single guide ARN
SNP : single nucleotide polymorphism
ST : Sequence type
TH : Todd Hewitt
tracrRNA : *trans*-activating CRISPR RNA
tracr-L : tracr-Long
TBS : Tris Buffered Saline
TCS : système à deux composants
TSH / TSS : trypticase soja horse blood / trypticase soja sheep blood
TSS : Transcription start site

Revue de la littérature

1. *Streptococcus agalactiae*

Streptococcus agalactiae est une bactérie de la famille des *Streptococcaceae,* appartenant au genre *Streptococcus*. Le genre *Streptococcus* fut décrit pour la première fois en 1874 par Billroth dans des lésions d'érysipèle. Depuis, plus de 100 espèces ont été décrites, dont la grande majorité sont associées à un pouvoir pathogène chez l'Homme ou les animaux. Morphologiquement, les streptocoques sont des cocci à Gram positif, qui se présentent sous la forme de diplocoques ou de chaînettes de longueur variable (Hardie et Whiley, 1995). Historiquement, les streptocoques ont été classés par Rebecca Lancefield, selon les propriétés antigéniques du polyoside C, composant de la paroi bactérienne (Lancefield, 1933). *S. agalactiae* appartient au groupe B de cette classification de Lancefield (Lancefield et Freimer, 1966).

1.1. Pouvoir pathogène

S. agalactiae a été initialement décrit en médecine vétérinaire et incriminé notamment dans les mammites bovines (Nocard et Mollereau, 1887). Cette affection, responsable d'une diminution de la quantité et de la qualité du lait, lui a valu son nom d'« *agalactiae* », qui signifie littéralement « absence de lait ». Son émergence en pathologie humaine remonte, quant à elle, aux années 1960, se manifestant par des infections néonatales opportunistes (Eickhoff *et al*., 1964). Cette bactérie est, par la suite, rapidement devenue une des causes principales d'infections materno-fœtales (Baker *et al*., 1973 ; Barton, Feigin, et Lins, 1973 ; Franciosi, Knostman, et Zimmerman, 1973 ; McCracken, 1973). Alors que l'incidence des infections néonatales semble s'être stabilisée, cette bactérie a émergé dans les infections de l'adulte depuis les années 2000.

1.1.1. Réservoir

La possible colonisation d'une large gamme d'hôtes, entraînant des manifestations cliniques variées, reflète la grande capacité d'adaptation à son environnement de cette bactérie. Le réservoir naturel de cette bactérie est à la fois humain et animal. Depuis sa découverte chez les bovins, *S. agalactiae* a ensuite été isolé chez d'autres espèces de mammifères, où les cas sont restés sporadiques, avant d'apparaître comme un pathogène émergent chez diverses espèces de poissons d'élevage. Il est impliqué dans des épidémies d'infections invasives dans les fermes piscicoles, responsables d'une importante morbi-mortalité chez ces animaux (Mian *et al*., 2009).

Chez l'Homme, le réservoir naturel de *S. agalactiae* est le tractus gastro-intestinal. Il est aussi la source de la colonisation des voies génito-urinaires. En France, environ 30% des adultes sont colonisés, que ce soit au niveau du tube digestif ou des voies génitales, et plus rarement au niveau de la flore oropharyngée ou de la peau (van der Mee-Marquet *et al*., 2008). La colonisation par *S. agalactiae* et

plus particulièrement, la colonisation vaginale peut être transitoire, chronique ou intermittente (Hansen *et al*., 2004). Ainsi, le statut d'une femme enceinte vis-à-vis du portage de *S. agalactiae* peut changer au cours de la grossesse (Verani *et al.*, 2010). En revanche, lorsque plusieurs dépistages itératifs sont positifs, il s'agit généralement de la même souche (Beauruelle *et al*., 2017 ; Hansen *et al*., 2004). Selon les études, la prévalence de la colonisation vaginale est variable, avec une moyenne de 10 à 30 % de femmes colonisées. En Europe, la prévalence s'étend de 6,5 % jusqu'à 36 % selon les pays (Barcaite *et al*., 2008) et en France, le portage vaginal de *S. agalactiae* chez la femme enceinte est compris, en moyenne, entre 10 et 15 % (Jauréguy *et al*., 2003).

1.1.2. Infections chez le nouveau-né

S. agalactiae reste actuellement la première cause bactérienne d'infections néonatales dans les pays développés (Edmond *et al*., 2012 ; Seale *et al*., 2017). Celles-ci se manifestent par deux grands types de syndromes : un syndrome précoce, ou Early-onset disease (EOD) et un syndrome tardif, ou Late-onset disease (LOD).

✓ Le syndrome précoce

Les infections néonatales précoces surviennent durant les 7 premiers jours de vie et le plus souvent, dans les 24 à 48 premières heures (Joubrel *et al*., 2015). Elles résultent généralement d'une transmission verticale de la bactérie de la mère à l'enfant, lors de l'accouchement ou juste avant. Par voie ascendante, la bactérie, présente au niveau vaginal, peut contaminer le liquide amniotique au moment du début du travail, suite à la rupture des membranes (Verani *et al*., 2010), voire même lorsque les membranes sont intactes (Katz et Bowes, 1988). Durant l'accouchement par voie basse, le nouveau-né peut également se contaminer par inhalation de sécrétions vaginales contenant *S. agalactiae*. La bactérie va alors atteindre le système respiratoire, conduisant aux manifestations cliniques principales des infections précoces, les pneumonies, pouvant s'accompagner d'une bactériémie, après translocation à travers la barrière pulmonaire, et plus rarement d'une méningite, après franchissement de la barrière hémato-encéphalique (Figure 1 ; Madrid *et al*., 2017).

Plusieurs facteurs de risque d'infections néonatales précoces ont pu être identifiés au cours de diverses études, comme la prématurité, un petit poids de naissance, ou plus classiquement, la rupture prématurée des membranes. Le facteur de risque le plus important reste la colonisation vaginale à *S. agalactiae* de la mère, qui à elle seule, multiplie le risque par 25 (Boyer et Gotoff, 1985). Environ 50 % des enfants nés de mères colonisées le seront également à la naissance, mais seulement 1 à 2 % d'entre eux développeront des symptômes d'infection. La plupart des autres resteront colonisés au niveau du tube digestif ou de l'arbre respiratoire, sans pour autant développer de manifestations cliniques dans l'immédiat (Verani *et al*., 2010).

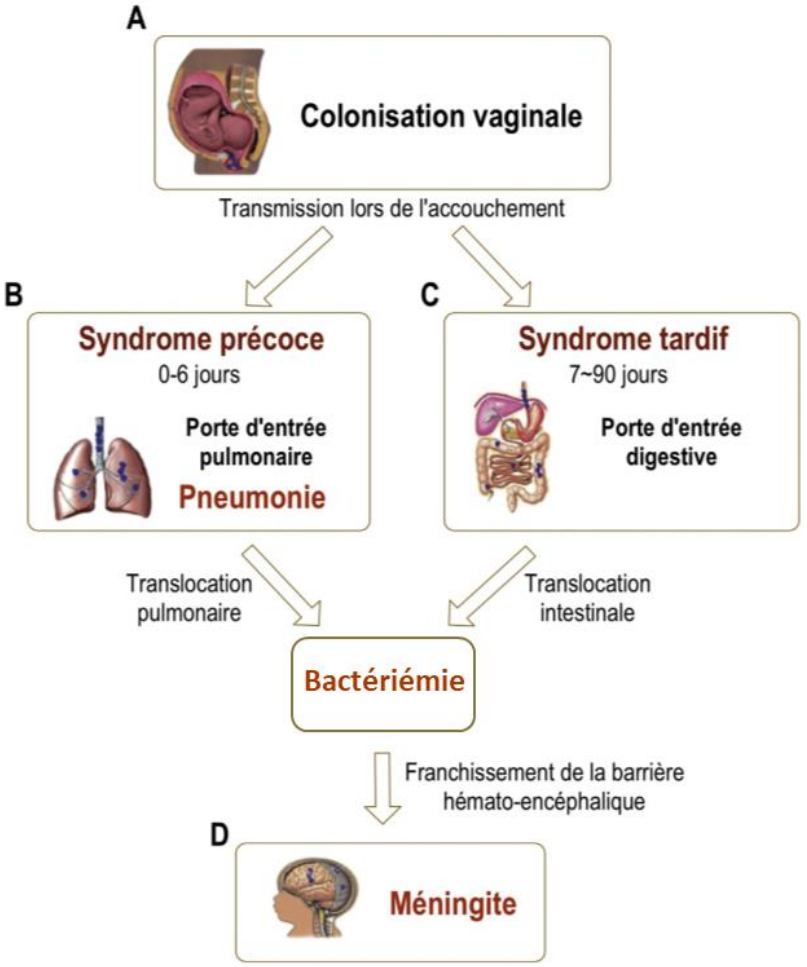


Figure 1 : Physiopathologie des infections néonatales à *S. agalactiae* (d'après Six *et al.*, 2014).
A. La bactérie est principalement transmise de la mère à l'enfant lors de l'accouchement. Celui-ci se contamine par inhalation et/ou ingestion des sécrétions vaginales et/ou de liquide amniotique contaminés.
B. Après inhalation, la bactérie se retrouve au niveau des poumons, entraînant l'apparition d'une pneumonie, c'est le syndrome précoce. Il peut être accompagné d'une bactériémie en cas de translocation de la bactérie à travers l'épithélium pulmonaire.
C. Le syndrome tardif se présente souvent sous la forme d'une bactériémie, probablement due à la translocation digestive de la bactérie suite à une colonisation intestinale.
D. Ces manifestations peuvent s'accompagner d'une méningite en cas de franchissement de la barrière hémato-encéphalique par la bactérie.

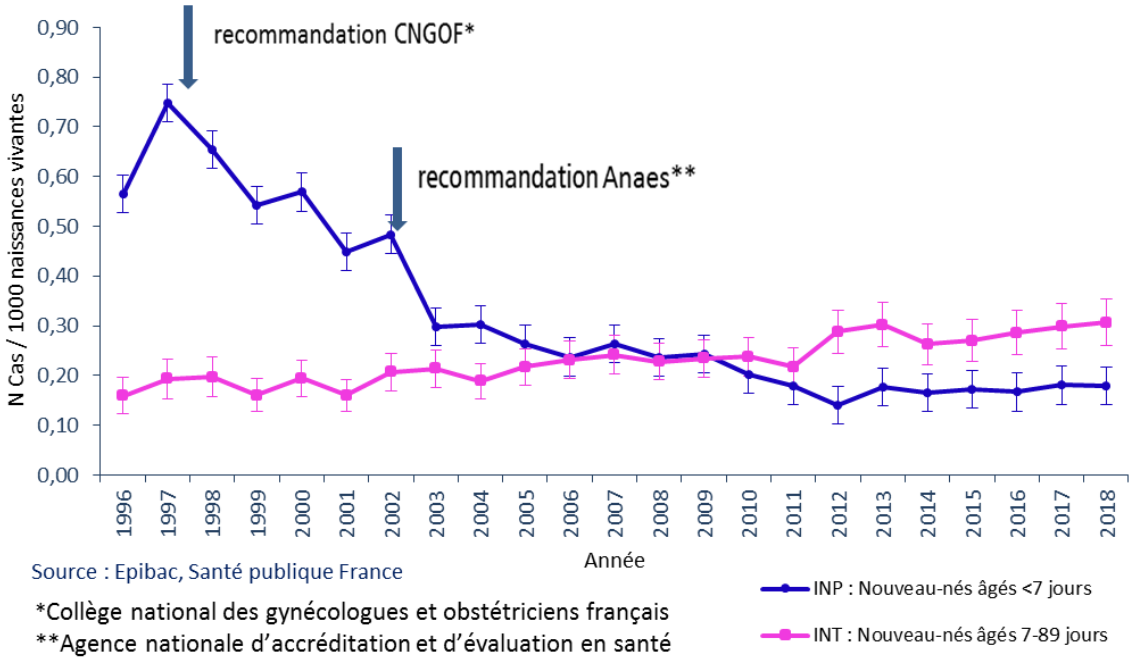
✓ Le syndrome tardif

Les infections néonatales tardives surviennent, quant à elles, entre le 7ème et le 89ème jour de vie, avec un pic de fréquence entre la 4ème et la 8ème semaine (Poyart *et al.*, 2008). Les deux manifestations cliniques les plus observées sont des bactériémies isolées (53 % des cas) ou, plus fréquemment, associées à des méningites (43 % des cas ; Madrid *et al.*, 2017). La porte d'entrée est cependant moins évidente que pour les infections précoces et la physiopathologie de ces infections tardives reste encore incertaine. Plusieurs mécanismes ont ainsi été évoqués, comme une transmission horizontale, de la

mère à l'enfant, lors de l'allaitement, des soins de maternage, ou une transmission entre nouveau-nés (Bingen *et al*., 1992 ; Filleron *et al.*, 2014 ; Morinis *et al*., 2011). Actuellement, l'hypothèse la plus probable semble être une translocation intestinale de la bactérie, suite à un portage digestif prolongé depuis la naissance. Une fois dans la circulation sanguine, *S. agalactiae* pourrait atteindre les méninges, après franchissement de la barrière hémato-encéphalique (Figure 1 ; Six *et al*., 2014 ; Tazi *et al*., 2010 ; Tazi *et al*., 2019). Ces méningites peuvent entraîner secondairement des séquelles neurologiques importantes chez le nouveau-né (Libster *et al*., 2012).

✓ Stratégies de prévention des infections

La morbi-mortalité importante due aux infections néonatales à *S. agalactiae* a entraîné la mise en place de différentes mesures de prévention. Les premières recommandations ont été émises par le Center for Disease Control and prevention aux Etats-Unis dès 1996. Puis en France, l'ANAES a mis en place en 2001 le dépistage systématique du portage vaginal de *S. agalactiae* entre la 34ème et la 38ème semaine d'aménorrhée et l'administration d'une antibioprophylaxie *per-partum* en cas de positivité du dépistage. L'objectif de cette antibioprophylaxie est de diminuer l'inoculum bactérien dans les voies génitales maternelles, afin de diminuer le risque de transmission à l'enfant au moment de l'accouchement. Cette dernière a permis une réelle diminution de l'incidence des infections précoces (Figure 2 ; Six *et al*., 2014). Les cas encore observés de nos jours semblent plutôt dus à la non réalisation du dépistage, ou à un dépistage faussement négatif, étant donné que la colonisation peut être intermittente ou transitoire (Hansen *et al.*, 2004), conduisant à l'absence d'antibioprophylaxie en *per-partum*, alors que les mères sont positives à *S. agalactiae* au moment de l'accouchement. Par contre, l'incidence des infections tardives est restée stable au cours du temps (Figure 2). Les moyens mis en place semblent donc sans effet sur ces infections, ce qui impose le développement de nouvelles approches de prévention. Actuellement des essais de développement de vaccins sont en cours, notamment dirigés contre les antigènes capsulaires (Song *et al*., 2018).



<u>Figure 2 :</u> Incidence des infections invasives néonatales précoces et tardives à Streptocoque du groupe B, France métropolitaine 1996-2018.
La courbe bleue représente les infections néonatales précoces (INP) et la courbe rose, les infections néonatales tardives (INT). Des flèches pointent les dates d'émission des recommandations du CNGOF et de l'ANAES, concernant la mise en place du dépistage systématique et de l'antibioprophylaxie *per partum* (Source : Santé Publique France, réseau Epibac).

1.1.3. Infections chez l'adulte

Depuis quelques années, il s'est produit un changement radical dans l'épidémiologie des infections invasives à *S. agalactiae*. En effet, alors que l'incidence des infections néonatales a connu une forte diminution et semble désormais se stabiliser, l'incidence totale des infections invasives est en augmentation, due au quasi doublement de l'incidence de ces infections chez les adultes de plus de 60 ans (Ballard *et al.*, 2016). Des données similaires sont retrouvées en France, où l'incidence des infections à *S. agalactiae* a largement dépassé celle des infections à *Streptococcus pyogenes* (Source : Santé Publique France, réseau Epibac).

✓ Chez la femme enceinte

A côté de son implication dans les infections néonatales, *S. agalactiae* est également un pathogène reconnu de la femme enceinte. Ainsi, il peut être incriminé dans de simples infections urinaires, ou dans des chorioamniotites, pouvant conduire à un accouchement prématuré, voire à une mort fœtale *in utero*. Il est également responsable d'endométrites dans la période du *post-partum* (Hall *et al.*, 2017 ; Krohn, Hillier, et Baker, 1999 ; Phares *et al.*, 2008).

✓ Chez l'adulte (hors grossesse)

S. agalactiae apparaît, depuis les années 2000 et en dehors de tout contexte de grossesse, comme un pathogène émergent chez l'enfant et l'adulte, notamment chez les personnes âgées et/ou immunodéprimées (Figure 3).

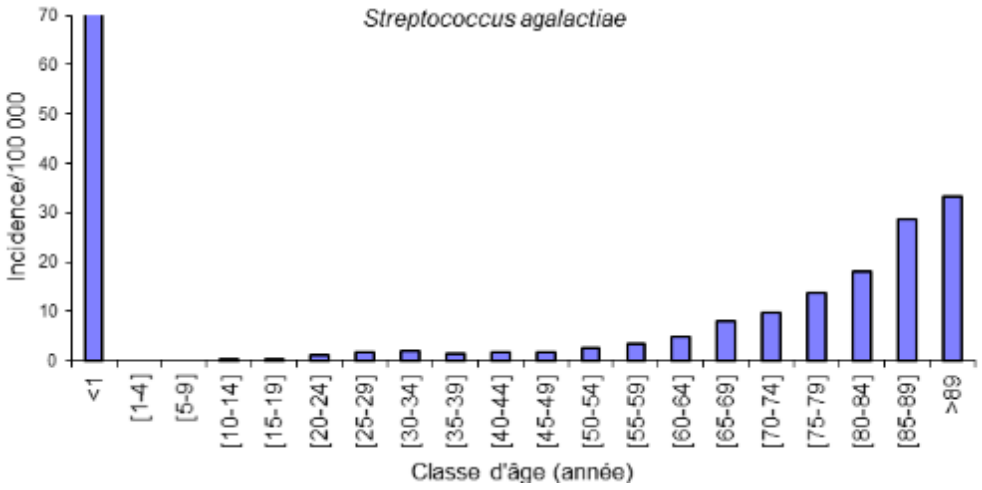


Figure 3 : Incidence pour 100 000 habitants des infections invasives (méningites et bactériémies) à *S. agalactiae* par groupe d'âge, Epibac, France métropolitaine, 2018.

Les différentes manifestations cliniques sont, entre autres, des bactériémies isolées, sans localisation secondaire, des infections de la peau et des tissus mous, des infections ostéo-articulaires, des pneumonies ou des infections urinaires (Figure 4 ; Phares *et al*., 2008 ; Skoff *et al.*, 2009).

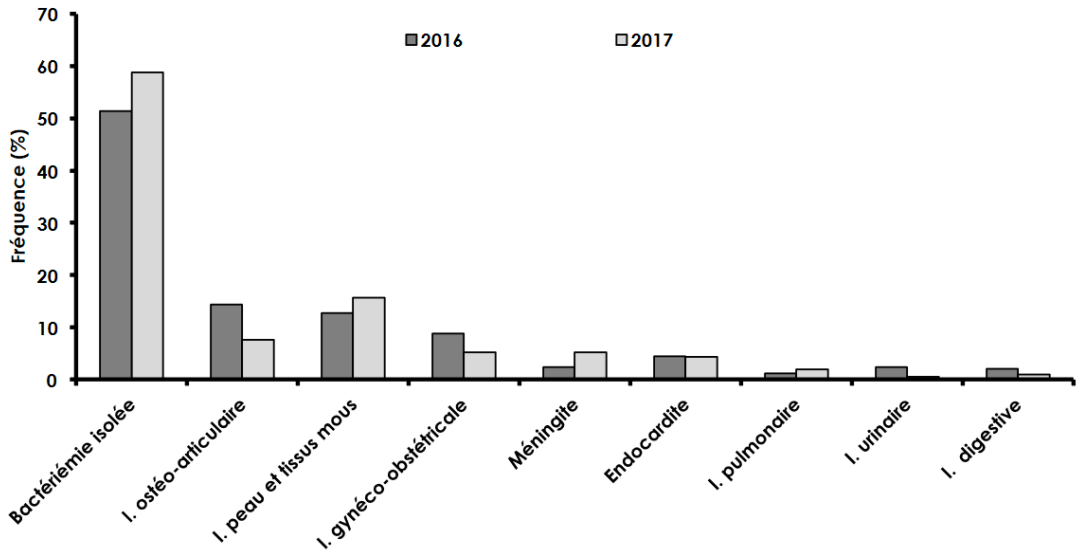


Figure 4 : Comparaison de la fréquence des différentes manifestations cliniques des infections invasives de l'adulte à SGB entre 2016 et 2017 (Rapport d'activité du CNR des Streptocoques, 2017).

✓ Facteurs de risques identifiés

Les infections à *S. agalactiae* touchent principalement des adultes de plus de 60 ans, avec une incidence qui augmente avec l'âge (Figure 3). Dans la plupart des cas, des pathologies sous-jacentes ont été rapportées. La comorbidité la plus fréquemment retrouvée est le diabète (Ballard *et al*., 2016 ; Skoff *et al*., 2009). Chez les patients pour lesquels les données étaient disponibles, le diabète était présent dans 25 % des cas, une anomalie cardiaque dans 13 % des cas et un cancer dans 13% des cas également.

1.1.4. Virulence de S. agalactiae

✓ Facteurs de virulence

En ce qui concerne le processus infectieux, la première étape consiste en l'adhérence de la bactérie aux cellules épithéliales afin d'y établir une colonisation puis de franchir ces barrières physiologiques. Ceci implique de nombreuses protéines de surface, telles que des protéines de liaison au fibrinogène (Fbs, Srr), des protéines de liaison à la laminine (Lmb), des adhésines (BibA, HvgA), ou encore des pili (Armistead *et al*., 2019). La β-hémolysine/cytolysine (β-h/c) figure également parmi les facteurs de virulence de *S. agalactiae* bien caractérisés. Par son activité cytolytique, elle favorise le franchissement des barrières épithéliales et endothéliales par la bactérie (Doran, Liu, et Nizet, 2003). Elle est également à l'origine de la β-hémolyse observée autour des colonies de *S. agalactiae* sur des géloses renfermant du sang de mouton ou de cheval.

Les nouveau-nés sont particulièrement susceptibles aux infections à *S. agalactiae*, du fait de l'immaturité de leur système immunitaire notamment. Il en est de même pour les personnes souffrant d'immunodépression. De plus, la bactérie possède de multiples facteurs qui lui permettent de se multiplier tout en échappant au système immunitaire, tels que la capsule polysaccharidique, la C5a peptidase (ScpB), ou encore la superoxide dismutase SodA (Doran et Nizet, 2004 ; Maisey, Doran, et Nizet, 2008). En plus de favoriser l'échappement à la phagocytose par les macrophages, la capsule participe également à la colonisation et à l'invasion des voies génitales (Noble *et al*., 2021).

✓ Le système à deux composants CovRS

De nombreux facteurs environnementaux participent à la transition entre l'état commensal et l'état pathogène de *S. agalactiae*, par la régulation de l'expression de différents gènes. Les bactéries ont développé des systèmes pour détecter et répondre à ces signaux, parmi lesquels, les systèmes à deux composants (TCSs ; Hoch, 2000). Ces systèmes sont composés d'un capteur histidine kinase, possédant un domaine de reconnaissance des stimuli et d'un régulateur de réponse. La liaison du signal au capteur entraîne son changement de conformation et sa phosphorylation sur un résidu histidine.

Le groupement phosphate est ensuite transféré sur un résidu aspartate du régulateur de réponse, déclenchant son activation. Celui-ci va alors pouvoir exercer ses fonctions de régulation, notamment au niveau de la transcription des gènes, après liaison à l'ADN.

Il existe une vingtaine de TCSs chez *S. agalactiae*, dont le système CovRS, qui est le mieux caractérisé (Lamy *et al.*, 2004). Ce système a été initialement décrit chez *S. pyogenes*, puis des systèmes orthologues ont été identifiés chez plusieurs espèces de streptocoques. La phosphorylation de CovR par l'histidine kinase CovS permet la dimérisation de CovR ce qui augmente sa capacité de liaison aux promoteurs de gènes, sur lesquels il exerce généralement un effet répresseur. CovRS a ainsi été impliqué dans la régulation de plus de 100 gènes, dont de nombreux facteurs de virulence, ce qui lui a valu d'être qualifié de régulateur majeur de la virulence pour l'ensemble des espèces dans lesquelles il a été caractérisé. Ainsi chez *S. agalactiae*, CovRS régule l'expression de plusieurs gènes précédemment identifiés comme des facteurs majeurs de la virulence tels que la β-h/c, codée par l'opéron *cyl*, ScpB et plusieurs adhésines, dont la protéine BibA/HvgA. Même si la plupart de ces voies de régulation sont conservées au sein de l'espèce, certaines apparaissent également spécifiques en fonction des souches (Jiang *et al*., 2008 ; Mazzuoli, 2020).

Selon les études, l'inactivation du système, dans sa totalité ou spécifiquement de CovR ou de CovS, entraîne une hypo ou une hypervirulence des souches de *S. agalactiae*. Ceci suggère une régulation très fine de la virulence en fonction de l'étape du processus infectieux (Jiang *et al*., 2008 ; Lamy *et al*., 2004). Par ailleurs, il a été précédemment démontré que le résidu aspartate situé en position 53 était essentiel à la fonctionnalité de CovR. En effet la substitution de ce résidu par une alanine ($CovR^{D53A}$) ne permet plus la phosphorylation de CovR, altérant sa fonction, comme en témoigne les phénotypes obtenus avec ce mutant, fréquemment similaires à ceux du mutant de délétion (Firon *et al*., 2013 ; Lin *et al*., 2009). Toutefois, d'autres études ont montré que ce mutant non phosphorylé a conservé, au moins en partie, sa capacité de liaison à l'ADN (Gusa *et al*., 2006). Ainsi l'importance de la phosphorylation de CovR pour son activité semble dépendante de plusieurs facteurs, parmi lesquels l'architecture des promoteurs, l'affinité pour les sites de liaison ou encore les mécanismes de liaison à l'ADN.

1.2. Epidémiologie et diversité génétique de *S. agalactiae*

En étudiant les infections à *S. agalactiae*, il a été constaté que toutes les souches n'étaient pas impliquées de la même façon dans les différentes pathologies. Ainsi, le développement de méthodes de typage, reposant sur la variabilité génétique des souches, a permis une étude épidémiologique des infections à *S. agalactiae*.

1.2.1. Sérotypage capsulaire

La capsule est un facteur de virulence majeur, empêchant notamment la phagocytose. Ainsi, la très grande majorité des souches impliquées en pathologie sont capsulées. Elle se compose d'unités polysaccharidiques répétées, codées par l'opéron *cps*. L'assemblage de ces unités est variable en fonction des souches, ce qui a permis de définir 10 sérotypes, numérotés Ia, Ib, puis de II à IX. La détermination du sérotype peut se faire par méthode conventionnelle à l'aide d'antisérum, largement remplacée aujourd'hui par les méthodes moléculaires, amplifiant des régions variables de l'opéron *cps* (Imperi *et al.*, 2010 ; Poyart *et al*., 2007). Les rares souches non typables présentent généralement des mutations dans les gènes codant la capsule.

La distribution des sérotypes varie selon les situations cliniques et ainsi, les sérotypes Ia, Ib, II, III et V, sont les plus retrouvés chez l'homme (>90 %), que ce soit dans les souches de colonisation (Russell *et al*., 2017), ou dans les souches d'infections à *S. agalactiae* (Madrid *et al*., 2017). Parmi ceux-ci, le sérotype capsulaire III apparaît fortement associé aux infections néonatales, puisqu'il est retrouvé dans environ la moitié des cas dans le monde (Figure 5 ; Edmond *et al.*, 2012). Il s'agit du sérotype prépondérant dans les infections précoces (62 %) et dans la majorité des infections tardives (83 %) et notamment, dans les cas de méningites associées (80 % et 87 % respectivement ; Poyart *et al*., 2008). Dans les infections de l'adulte, en dehors d'un contexte de grossesse, les sérotypes les plus fréquemment retrouvés sont les sérotypes Ia, III et V (Phares *et al*., 2008 ; Skoff *et al*., 2009).

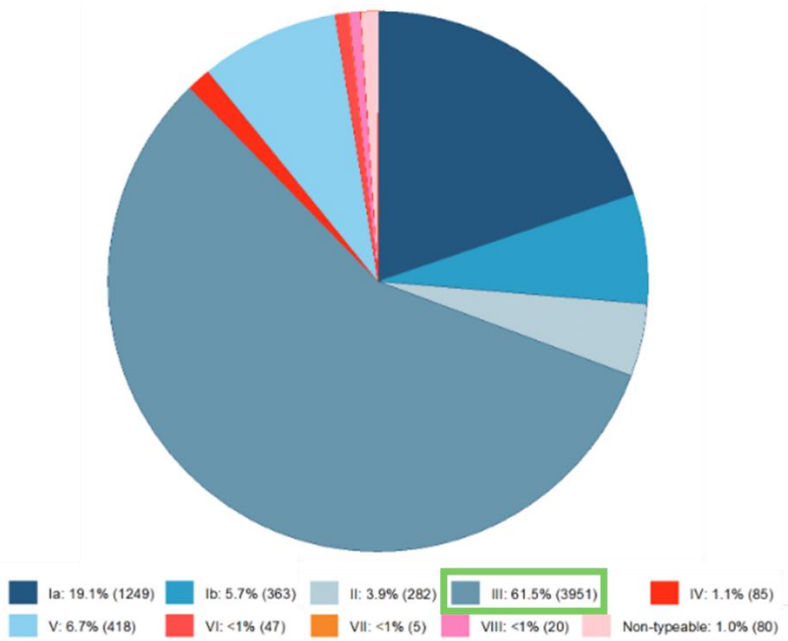


Figure 5 : Distribution mondiale des sérotypes dans les infections invasives néonatales à *S. agalactiae* (d'après Madrid *et al*., 2017).

1.2.2. Multi Locus Sequence Typing

Développée plus récemment, la technique du Multi Locus Sequence Typing (MLST) constitue actuellement la méthode de référence pour la caractérisation des souches de *S. agalactiae* (Jones *et al.*, 2003). Elle repose sur le séquençage de fragments d'environ 500 pb de sept gènes de ménage codant des enzymes du métabolisme (*adhP*, *pheS*, *atr*, *glnA*, *sdhA*, *glcK* et *tkt*). En fonction des variations nucléotidiques observées, un numéro d'allèle unique est attribué à chacun des sept gènes étudiés. La combinaison des sept numéros définit un profil allélique ou Sequence Type (ST). Les STs peuvent ensuite être regroupés, par analyse phylogénétique, en complexes clonaux (CCs ; Feil *et al.*, 2004). Un CC rassemble l'ensemble des STs ayant en commun, au minimum, 6 allèles sur 7. Ceci permet d'établir des relations phylogénétiques entre les isolats. Le CC est identifié par le même numéro que celui de son ST fondateur, c'est-à-dire le ST le plus représenté dans le groupe. Ainsi, le ST fondateur peut varier au cours du temps en fonction de l'enrichissement de la base de données. Si le ST est seul dans son groupe, il constitue un singleton. La recherche ou le dépôt des numéros d'allèle, conduisant à la détermination du ST, se fait en ligne. Le site internet, grâce à une base de données internationale, recense plus de 1600 STs à ce jour (https://pubmlst.org/bigsdb?db=pubmlst_sagalactiae_seqdef). Il semble exister une certaine corrélation entre ces lignées phylogénétiques et les sérotypes capsulaires. Toutefois, même si un sérotype est majoritaire dans un CC, des souches appartenant à d'autres sérotypes peuvent faire partie du même CC (Da Cunha *et al*., 2014).

Ces complexes clonaux sont caractérisés par des différences de tropisme d'hôtes et de pouvoir pathogène, certains semblant plus associés au portage et d'autres, aux infections invasives. De même, des différences ont pu être observées dans les sérotypes capsulaires qui les composent (Jones *et al.*, 2003 ; van der Mee-Marquet *et al.*, 2008). Cinq complexes clonaux prédominent en pathologie humaine à travers le monde. Ce sont les CC1, CC10, CC17, CC19 et CC23 (Figure 6 ; Da Cunha *et al.*, 2014).

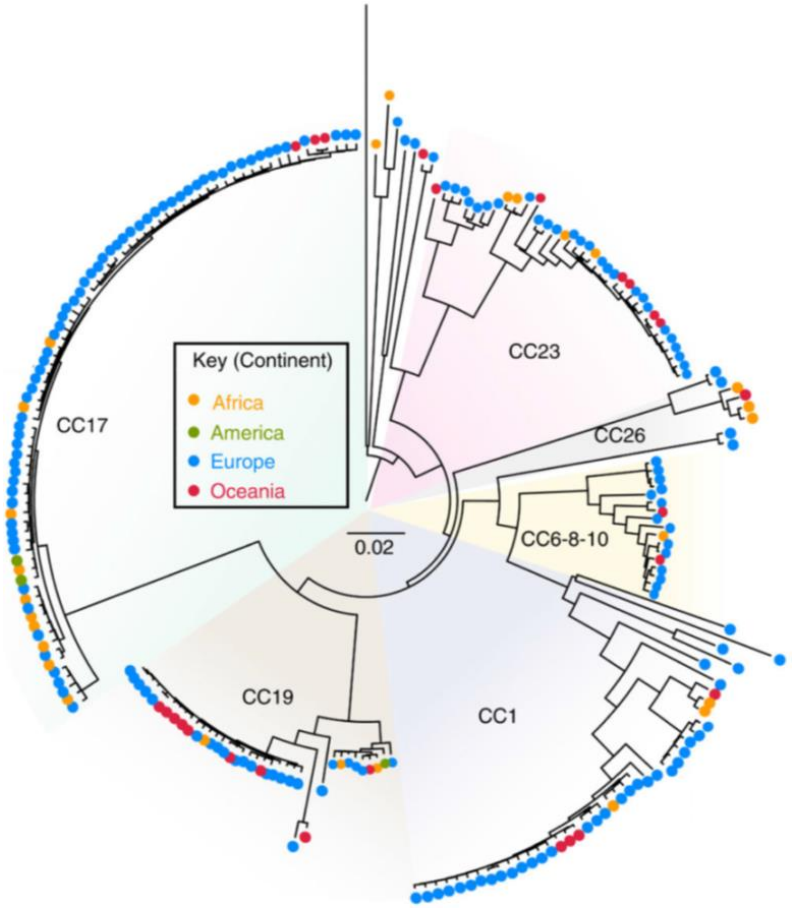


Figure 6 : Répartition phylogénétique mondiale des souches de *S. agalactiae* humaines (d'après Da Cunha *et al*., 2014).
L'analyse phylogénétique de 229 souches a montré qu'elles appartenaient à cinq groupes bien distincts, correspondants aux cinq CCs majeurs définis par MLST.

Les CC1 et CC10, très proches phylogénétiquement, regroupent des souches de sérotypes et de ST variés, retrouvées à la fois chez l'Homme et chez l'animal. Ce sont des souches de colonisation, qui peuvent également être impliquées dans des infections chez l'adulte. Le CC19 se compose de souches strictement humaines, de sérotypes III ou V, qui sont majoritairement des souches de colonisation (Jones *et al*., 2003). Cependant, certaines souches peuvent être responsables d'infections invasives aussi bien chez l'adulte que chez le nouveau-né (Manning *et al*., 2009). Le CC23 rassemble des souches de sérotypes Ia et III, retrouvées en portage ainsi que dans les infections de l'adulte et du nouveau-né. Enfin, par cette approche moléculaire, un clone hypervirulent, de sérotype capsulaire III, a pu être identifié. Il s'agit du CC17, constitué très majoritairement de souches de ST17.

1.2.3. Le clone hypervirulent CC17

Les souches appartenant au ST17, pouvant être regroupées en un complexe clonal portant le même numéro (CC17) car il s'agit du ST fondateur, sont significativement associées aux infections

néonatales invasives, ce qui leur a valu d'être qualifiées d'hypervirulentes. Ce clone serait apparu assez récemment, dans les années 50, du fait de l'utilisation massive de la tétracycline, antibiotique auquel ces souches, entre autres, sont résistantes (Da Cunha *et al*., 2014). Le clone ST17 est particulièrement impliqué dans les infections néonatales tardives et dans plus de 80 % des cas de méningites, qu'elles interviennent au cours d'infections précoces (EOD) ou tardives (LOD ; Tableau I). Il est également incriminé dans plus de la moitié des cas de bactériémies néonatales, alors qu'il est peu représenté dans les infections chez l'adulte (Jones *et al*., 2003 ; Manning *et al.*, 2009 ; Poyart *et al*., 2008 ; Tazi *et al.*, 2010).

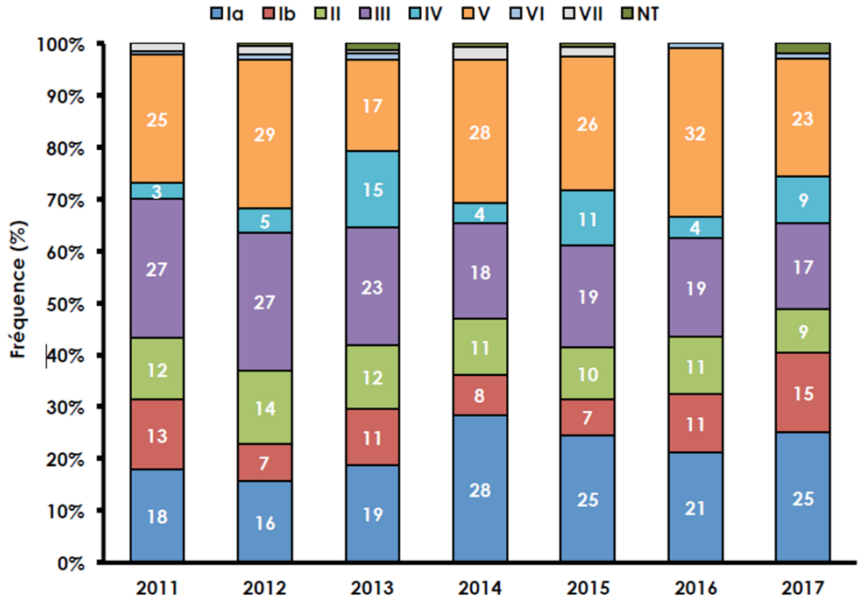
Tableau I : Répartition des sérotypes de 651 souches de *S. agalactiae*, responsables d'infections invasives chez le nouveau-né et l'adulte, isolées en France entre 2006 et 2009 (Tazi *et al*., 2010). L'appartenance au ST17 est indiquée dans la colonne de droite.

Type of infection	Serotype no.							ST-17 no.
	Ia	Ib	II	III	IV	V	Total	
	%							%
Neonatal meningitis	12 (8.7)	4 (2.9)	0	119 (86.2)	0	3 (2.1)	138 (100)	113 (81.9)
EOD (≤6 d)	6 (20.7)	0	0	23 (79.3)	0	0	29 (100)	23 (79.3)
LOD (≥7–89 d)	6 (5.5)	4 (3.7)	0	96 (88.1)	0	3 (2.7)	109 (100)	90 (82.6)
Neonatal bacteremia	34 (20.5)	5 (3)	6 (3.6)	101 (60.8)	3 (1.8)	17 (10.2)	166 (100)	89 (53.6)
EOD (≤6 d)	27 (29)	3 (3.2)	5 (5.4)	41 (44.1)	2 (2.2)	15 (16.1)	93 (100)	32 (34.4)
LOD (≥7–89 d)	7 (9.6)	2 (2.7)	1 (1.4)	60 (82.2)	1 (1.4)	2 (2.7)	73 (100)	57 (78.1)
Adult bacteremia	71 (21.4)	36 (10.3)	37 (11.2)	85 (25.7)	17 (5.1)	85 (25.7)	331 (100)	37 (11.2)
Adult meningitis	4 (25)	2 (12.5)	1 (6.25)	8 (50)	0	1 (6.25)	16 (100)	5 (31.3)

L'hypervirulence du clone ST17 serait en partie liée à sa capacité à traverser les barrières physiologiques. En effet, la comparaison génomique de souches de *S. agalactiae* a permis d'identifier deux protéines de surface spécifiques de ce clone, HvgA et Srr2 (Brochet *et al*., 2006). HvgA est un variant de la protéine BibA, présente chez l'ensemble des souches de *S. agalactiae*. Cette protéine, régulée par le système CovRS, présente une affinité particulière pour la barrière intestinale et la barrière hémato-encéphalique, en lien avec la physiopathologie des infections néonatales tardives (Tazi *et al*., 2010). Srr2 est une protéine de surface spécifique des souches hypervirulentes ST17, les souches appartenant aux autres lignées phylogénétiques exprimant Srr1. La protéine Srr2 est fortement exprimée à la surface bactérienne, en lien avec sa capacité de liaison au fibrinogène. Elle a aussi été impliquée dans l'augmentation de la survie intracellulaire dans les macrophages et dans l'aide au franchissement des barrières physiologiques (Six *et al*., 2015). En plus de ces facteurs de virulence spécifiques, d'autres facteurs apparaissent différemment régulés et exprimés chez les souches de ST17, pouvant contribuer ainsi à la virulence particulière de ces souches (Safadi *et al*., 2011).

1.2.4. Variations épidémiologiques

Les études épidémiologiques réalisées à travers le monde, montrent des répartitions des sérotypes comparables. En effet, les sérotypes Ia, Ib, II, III et V représentent la grande majorité des sérotypes retrouvés dans les infections humaines, pour les différentes zones géographiques (Europe, Amérique, Afrique, Pacifique…). De plus, peu d'évolutions ont été constatées dans cette répartition depuis les 30 dernières années (Le Doare et Heath, 2013). Des constats similaires sont applicables aux souches de *S. agalactiae* isolées en France. Ainsi, les sérotypes capsulaires Ia, Ib, II, III et V ont été retrouvés dans 91% des souches isolées d'infections chez l'adulte, entre 2007 et 2019 (Vuillemin *et al.*, 2021). La proportion de chacun de ces sérotypes, par année, est relativement stable depuis 2011 (Figure 7). Sur cette même période, 96,3% des souches appartenaient aux CC1, CC10, CC17, CC19 et CC23. L'épidémiologie des infections invasives néonatales est relativement stable également, avec une sur-représentation des souches de sérotype III (74%) et du CC17 (66%), parmi les souches isolées entre 2007 et 2019 en France (Plainvert *et al*., 2020).



<u>Figure 7 :</u> Fréquence des sérotypes capsulaires des souches de SGB responsables d'infections invasives chez l'adulte en fonction de l'année (Rapport d'activité du CNR des Streptocoques, 2017).

Récemment, une étude épidémiologique a été réalisée au CHRU de Tours et portait sur des souches de colonisation vaginale et les souches isolées d'infections materno-fœtales, entre 2000 et 2018 (Le Gallou, 2020). Les conclusions de cette étude étaient superposables à l'épidémiologie nationale et mondiale. Ainsi, les sérotypes Ia, III et V étaient les plus fréquemment retrouvés, avec une prédominance du sérotype III dans les infections invasives (56%). La majorité des souches appartenaient aux 5 CCs majeurs (CC1, CC10, CC17, CC19 et CC23 ; Figure 8), avec une association significative entre CC17 et infections invasives néonatales (45,2%).

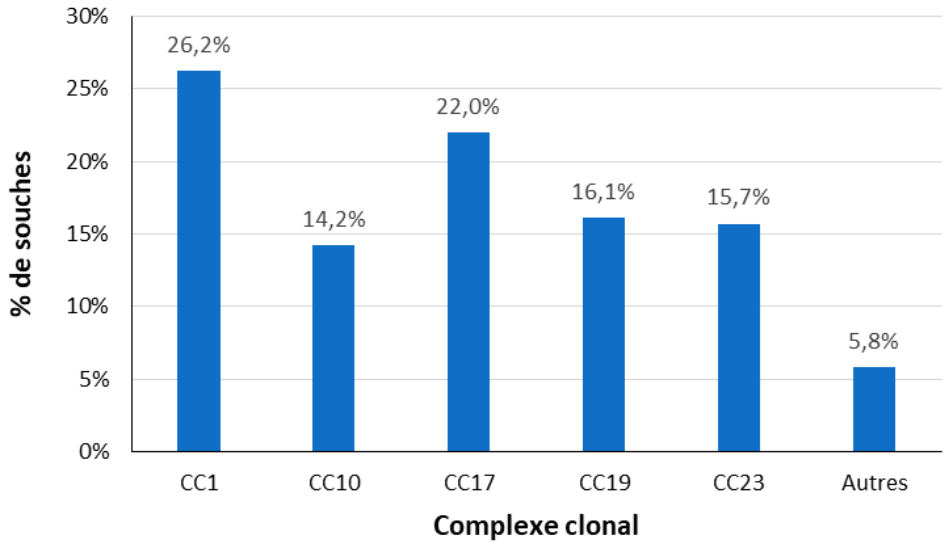


Figure 8 : Répartition au sein des CCs des souches de *S. agalactiae* isolées au CHRU de Tours entre 2000 et 2018 (d'après Le Gallou, 2020).

1.2.5. Autres méthodes de typage

Les méthodes de Whole Genome Sequencing (WGS), permettant le séquençage de génomes entiers, ont révolutionné l'étude des populations bactériennes. Par la comparaison de multiples souches, il est possible de reconstituer l'historique d'une population, ou encore de déterminer la source d'une épidémie. Ainsi, l'analyse des génomes entièrement séquencés de souches de *S. agalactiae* a montré une organisation en mosaïque, c'est-à-dire une partie conservée, le « core genome », présente chez toutes les souches de l'espèce, entrecoupée par de multiples îlots génomiques retrouvés uniquement chez une partie des souches de l'espèce, le « dispensable genome » (Brochet *et al.*, 2006 ; Tettelin *et al.*, 2005 ; Tettelin *et al.*, 2002). Le « core genome » représente approximativement 80 % du génome total de *S. agalactiae*. Le « dispensable genome » correspond fréquemment à des séquences prophagiques ou issues d'autres éléments génétiques mobiles (EGMs), acquises généralement par transfert horizontal de gènes. Malgré les progrès accomplis sur les outils d'analyse, le grand nombre de données générées par ces techniques de séquençage haut débit nécessite une certaine expertise pour leur interprétation.

C'est pourquoi de nouvelles méthodes, plus discriminantes que le MLST et plus faciles à mettre en œuvre que le WGS pourraient présenter un intérêt pour l'étude des populations bactériennes. Récemment, l'utilisation du polymorphisme des spacers au sein des systèmes CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR associated genes) a été proposée comme méthode alternative de typage des souches de *S. agalactiae*. En effet, il a été montré une très bonne

corrélation entre la composition en spacers des systèmes et les lignées phylogénétiques définies par MLST (Figure 9 ; Lier *et al*., 2015 ; Lopez-Sanchez *et al*., 2012).

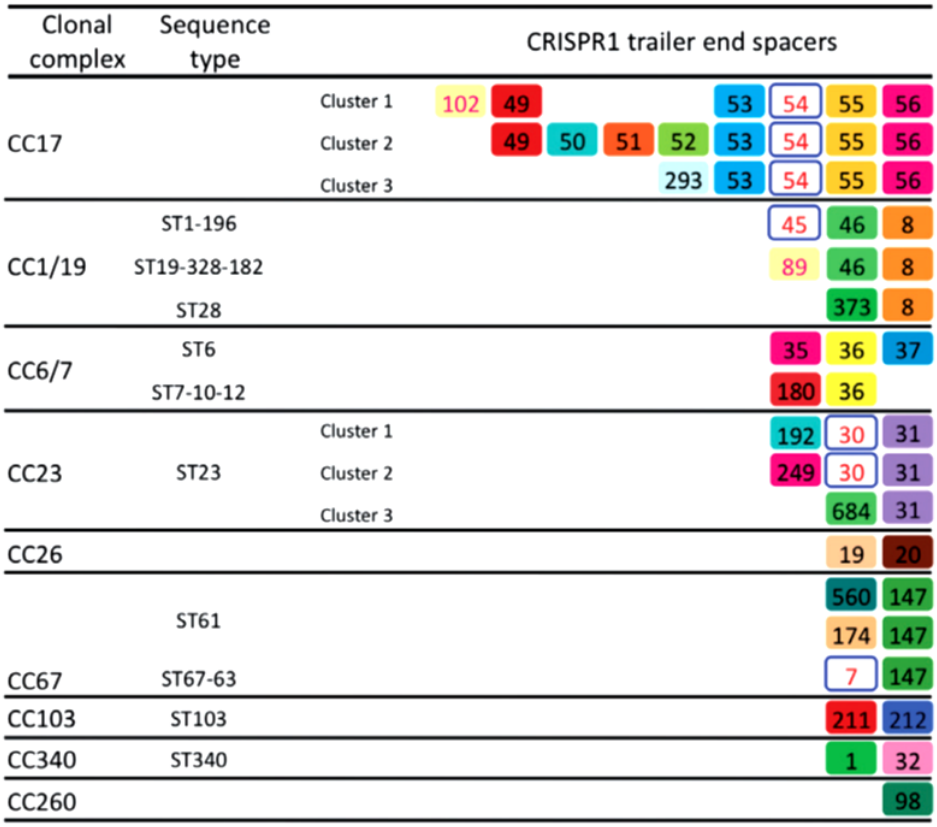


Figure 9 : Corrélation entre la composition en spacers du locus CRISPR1 et le CC chez *S. agalactiae* (Lopez-Sanchez *et al*., 2012).
Les spacers ancestraux du locus CRISPR1 apparaissant spécifiques des CCs définis par MLST, voire de certains STs ou permettent de définir des sous-groupes au sein d'un même CC.

2. Les systèmes CRISPR-Cas

Les systèmes CRISPR-Cas ont été découverts à la fin du XX^{ème} siècle chez *E. coli* et leur rôle de système immunitaire adaptatif des bactéries, contre les bactériophages, a été élucidé vingt ans plus tard. En effet, bien que potentiellement vecteurs de facteurs de virulence, participant ainsi à l'adaptation des pathogènes à leurs hôtes, les bactériophages constituent aussi une menace constante pour la survie des bactéries. C'est pour cela que ces dernières ont développé, au cours de leur évolution, divers mécanismes de défense. Parmi ces mécanismes, plusieurs ciblent des étapes du cycle d'infection phagique (Figure 10). Ainsi, les bactéries peuvent tout d'abord, inhiber l'adsorption du phage à la surface bactérienne, en modifiant le récepteur spécifique du phage, en synthétisant une matrice extracellulaire qui va masquer le récepteur du phage ou encore, en synthétisant des inhibiteurs compétitifs du site de fixation du phage. Les systèmes d'exclusion de la superinfection (Sie) quant à eux, codent des protéines qui bloquent l'entrée de l'ADN phagique dans la cellule bactérienne. Il existe également des systèmes qui, en détournant le cycle infectieux, conduisent à une infection abortive (Abi). Ces systèmes entraînent généralement la mort de la bactérie avant le relargage de nouveaux virions, protégeant ainsi les cellules environnantes (Labrie, Samson, et Moineau, 2010).

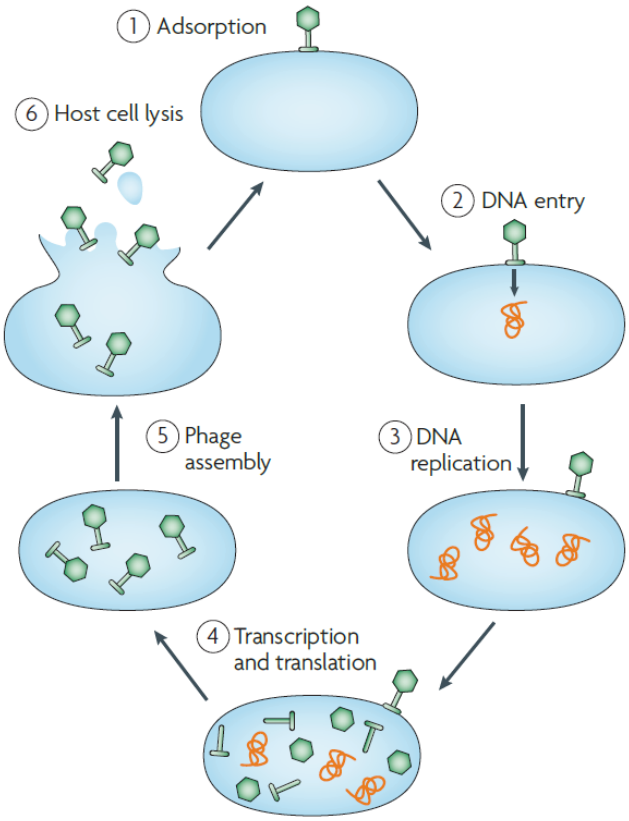


Figure 10 : Les différentes étapes du cycle d'infection phagique (Labrie, Samson, et Moineau, 2010). Les bactéries ont développé au cours de leur évolution des mécanismes de protection variés pouvant bloquer les différentes étapes du cycle de réplication des bactériophages.

D'autres mécanismes vont permettre la dégradation des acides nucléiques étrangers, ne se limitant pas uniquement à la défense anti-virale. C'est le cas des systèmes de Restriction-Modification, présents chez la quasi-totalité des bactéries, mais aussi des systèmes CRISPR-Cas, présents chez une grande majorité d'*Archaea* et environ la moitié des Eubactéries (Grissa, Vergnaud, et Pourcel, 2007).

2.1. Historique

La découverte de ces systèmes remonte à 1987, où l'équipe d'Ishino a détecté des « répétitions » à proximité du gène codant l'isoenzyme phosphastase alcaline chez *E. coli*. Les auteurs décrivent une alternance de courtes séquences répétées de 29 pb, intercalées de courtes séquences variables de 32pb. De plus, les séquences répétées sont partiellement palindromiques et suspectées de jouer un rôle dans la stabilisation des ARN messagers (Ishino *et al.*, 1987). Dans les années 2000, le même type de structure, alors appelées Short Regularly Spaced Repeats (SRSRs), sont décrites chez de nombreuses *Archaea* et divers groupes phylogénétiques de bactéries, apparaissant ainsi comme des systèmes largement répandus chez les procaryotes (Mojica *et al*., 2000). C'est en 2002 que l'acronyme CRISPR, pour Clustered Regularly Interspaced Short Palindromic Repeats est adopté et que les gènes *cas,* associés à ces séquences répétées, sont identifiés (Jansen *et al.*, 2002). Ces systèmes sont retrouvés exclusivement chez les procaryotes, mais leur rôle biologique n'a pas encore été élucidé. Rapidement, un lien est établi entre les séquences variables, les spacers, qui séparent les séquences répétées et les acides nucléiques étrangers. En effet, plusieurs équipes remarquent une homologie de séquence entre les spacers et divers EGMs étrangers, suggérant alors que ces systèmes pourraient conférer une protection immunitaire aux bactéries contre des éléments spécifiquement ciblés par les spacers (Bolotin *et al*., 2005 ; Mojica *et al.*, 2005 ; Pourcel, Salvignol, et Vergnaud, 2005). L'année suivante, il est suspecté que le fonctionnement de ces systèmes soit basé sur un mécanisme d'ARN interférence (RNAi ; Makarova *et al.*, 2006). C'est en 2007, 20 ans après leur découverte, que leur rôle comme système d'immunité anti-virale acquise a été confirmé expérimentalement chez *Streptococcus thermophilus* (Barrangou *et al.*, 2007). Dès lors, de nombreuses équipes ont travaillé à leur caractérisation, à la description de leur structure, de leur fonctionnement et à la découverte de nouveaux systèmes. La compréhension de leurs mécanismes d'action a conduit à leur utilisation dans des domaines variés, parmi lesquels leur utilisation pour l'édition des génomes a explosé mondialement (Figure 11), jusqu'à l'attribution du Prix Nobel de Chimie à Emmanuelle Charpentier et Jennifer Doudna en octobre 2020.

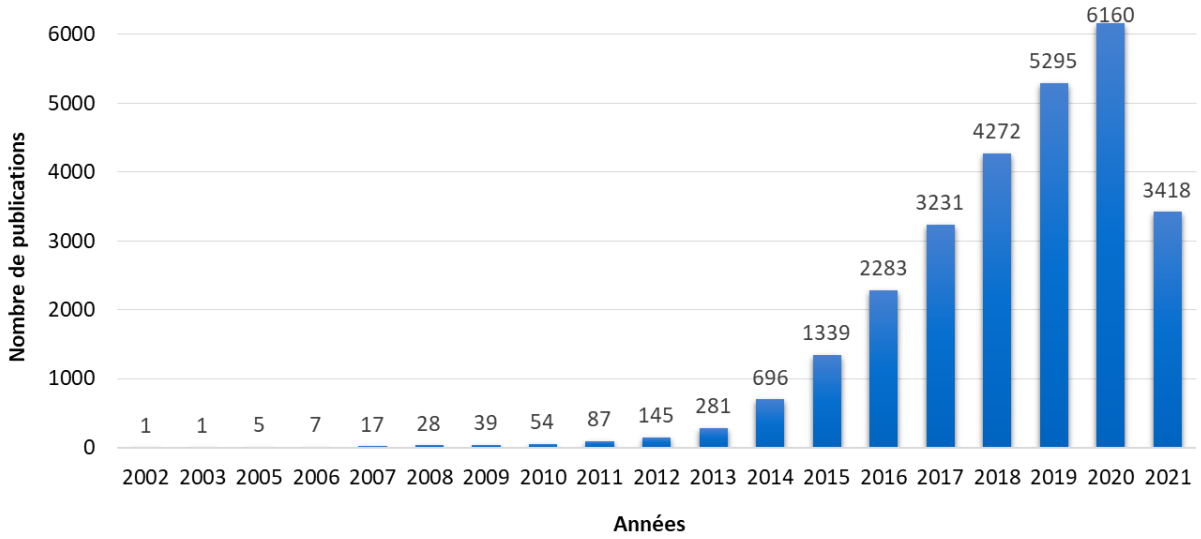


Figure 11 : Nombre de publications indexées dans Pubmed, citant le mot CRISPR, entre 2002 et le 10 juin 2021.

2.2. Organisation des systèmes CRISPR-Cas

Généralement, les bactéries ne renferment qu'un locus CRISPR, mais certaines peuvent en posséder plusieurs (Jansen *et al*., 2002). Les systèmes CRISPR-Cas bien que très divers, présentent une organisation générale similaire (Figure 12). Deux éléments majeurs composent ces systèmes : d'une part, le locus CRISPR, flanqué en amont par une région leader et en aval par une région trailer et d'autre part, les gènes *cas*, organisés en opéron et situés à proximité du locus CRISPR (Cas pour CRISPR-associated).

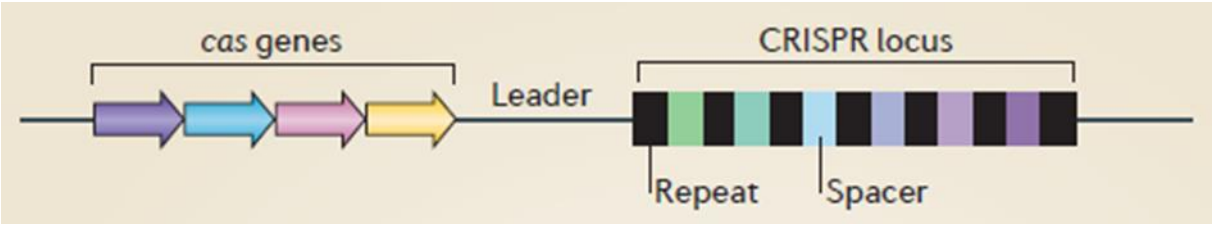


Figure 12 : Organisation générale d'un système CRISPR-Cas (d'après van der Oost *et al*., 2014).
Un système CRISPR-Cas se compose de deux grandes régions. Le locus CRISPR, composé d'une alternance de repeats (rectangles noirs) et de spacers (rectangles colorés). L'opéron *cas*, situé à proximité, est constitué d'un nombre de gènes variables, représentés par des flèches colorées. Ces deux régions sont séparées par le leader.

En dehors de ces éléments majeurs, les systèmes CRISPR-Cas apparaissent hautement variables, probablement en lien avec la nécessaire adaptation rapide du système, en réponse à la pression de sélection imposée par les EGMs rencontrés dans des environnements divers.

2.2.1. Le locus CRISPR

Les loci CRISPR correspondent à une région composée d'une alternance de courtes séquences répétées, les Direct Repeats (DRs), séparées par des séquences variables, les spacers (Figure 12).

✓ Les Direct Repeats (DRs)

Il existe une large diversité des séquences répétées, en fonction des espèces bactériennes étudiées. Ainsi, leur nombre et leur taille varient d'un système à l'autre. Cependant, au sein d'un même locus, la taille des DRs apparaît strictement conservée. Leur taille se situe généralement entre 28 et 37 pb, pouvant varier de 23 à 55 pb. En général, les loci CRISPR contiennent moins de 50 unités repeat-spacer (Barrangou et Marraffini, 2014). De même, la séquence des DRs est très conservée au sein d'un système, mis à part la DR terminale (DRT) qui est souvent tronquée ou dégénérée. Les DRs sont partiellement palindromiques, ce qui leur confère une structure secondaire stable et leur permet d'établir des interactions spécifiques avec les protéines Cas. En effet, la classification des systèmes, lorsqu'elle est basée sur les direct repeats, est fortement corrélée à la classification basée sur les gènes *cas*, suggérant des interactions fortes entre les deux (Al-Attar *et al*., 2011 ; Kunin, Sorek, et Hugenholtz, 2007).

✓ Les spacers

Les spacers (ou espaceurs) sont également uniformes en taille pour un locus donné. Par contre, leur taille est variable d'un système à l'autre, de 21 jusqu'à 72 pb, mais généralement elle est comprise entre 32 et 38 pb. A l'inverse des DRs, leurs séquences sont très variables au sein d'un locus. Ceci s'explique par leur origine généralement extra-chromosomique. En effet, pour la plupart de ces spacers, une homologie est retrouvée entre la séquence des spacers et différents EGMs (phages, plasmides). Ce sont eux qui déterminent ainsi la spécificité du système. En général, ces spacers sont retrouvés en une seule copie au sein des loci CRISPR, mais des duplications sont possibles. De même, il a été remarqué que certains spacers semblaient avoir une origine chromosomique, une homologie de séquence existant entre le spacer et le chromosome hébergeant le système CRISPR-Cas.

2.2.2. Le leader

Directement en amont du premier DR du locus CRISPR se situe le leader (Figure 12). C'est une région non codante, riche en Adénine et en Thymine. Sa taille, de plusieurs centaines de paires de base, et sa séquence, sont variables en fonction des espèces bactériennes. En revanche, elles apparaissent très conservées pour un système CRISPR-Cas donné (Jansen *et al*., 2002). Le leader est un élément primordial des systèmes CRISPR-Cas puisqu'il renferme le promoteur de transcription du locus

CRISPR (Pougach *et al.*, 2010 ; Pul *et al.*, 2010) et il joue un rôle crucial dans l'incorporation de nouveaux spacers (Yosef, Goren, et Qimron, 2012).

2.2.3. *Les gènes* cas

Les gènes *cas* sont généralement situés à proximité d'un locus CRISPR. En effet la présence des gènes *cas* est systématiquement associée à celle d'un locus CRISPR, ce qui leur a valu leur nom de CRISPR associated (Jansen *et al.*, 2002). Organisés en opéron, ils codent les protéines Cas, porteuses de fonctions enzymatiques très diversifiées, nécessaires au fonctionnement du système. L'architecture de cet opéron diffère d'un système à l'autre, ce qui en fait des systèmes complexes et très variables. Néanmoins, les fonctions associées à ces gènes apparaissent conservées entre les systèmes. De même, les protéines Cas1 et Cas2 sont qualifiées d'universelles car elles sont présentes dans la quasi-totalité des systèmes. Ceci s'explique par leur rôle dans l'acquisition de nouveaux spacers, qui est une étape cruciale pour l'immunisation de la bactérie (Yosef, Goren, et Qimron, 2012).

2.3. Classification des systèmes CRISPR-Cas

La classification des systèmes CRISPR-Cas repose sur la nature et l'organisation des gènes *cas* présents dans le système. Avec les progrès sur le séquençage des génomes bactériens et l'augmentation des découvertes sur ces systèmes, la classification est passée de 3 types de systèmes et une dizaine de sous-types en 2011 à 6 types de systèmes et 33 sous-types en 2020 (Makarova *et al.*, 2011 ; Makarova *et al.*, 2020).

Ainsi, les systèmes CRISPR-Cas ont d'abord été répartis en deux classes. Les systèmes de classe 1 font intervenir un complexe multiprotéique à l'étape d'interférence, dont la composition en gènes *cas* varie en fonction des types et des sous-types. Pour les systèmes de classe 2, une seule protéine multidomaine cumule toutes les fonctions du complexe effecteur (Figure 13).

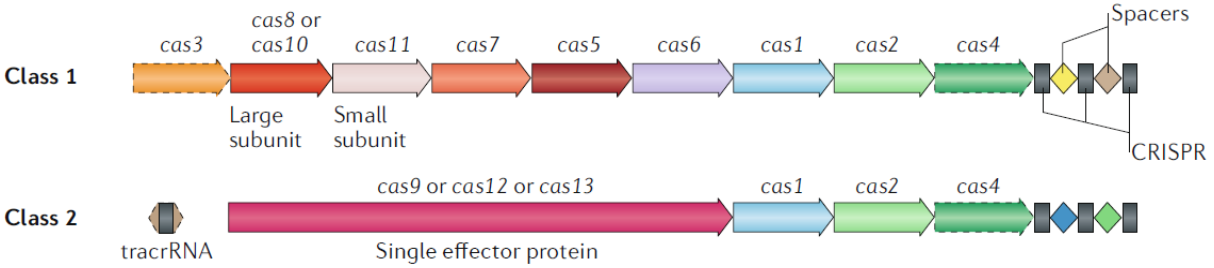
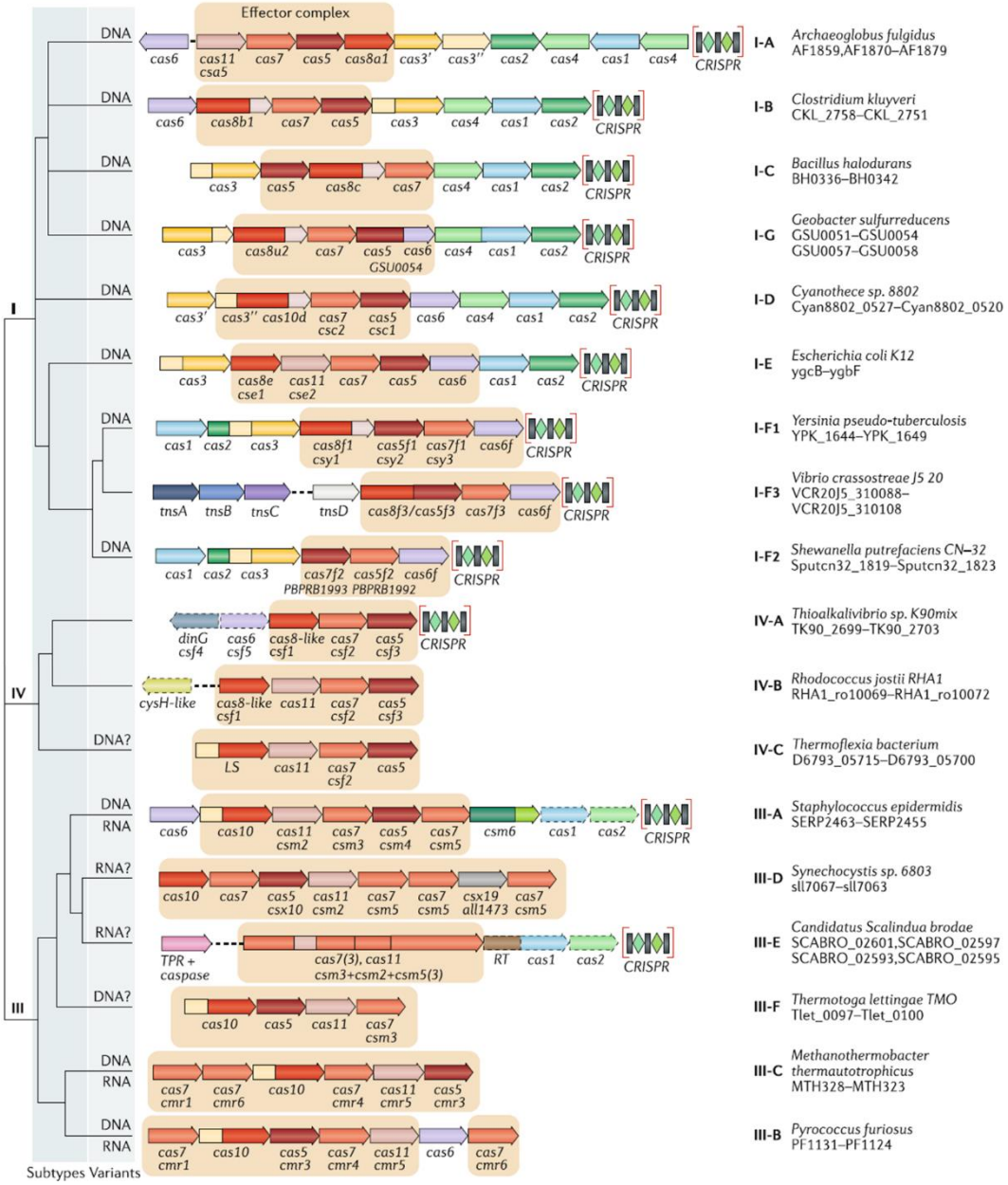


Figure 13 : Organisation des deux classes de systèmes CRISPR-Cas (Makarova *et al.*, 2020).
Le complexe effecteur des systèmes de classe 1 se compose d'une multitude de protéines Cas. A l'inverse, les systèmes de classe 2 possèdent une seule protéine multidomaine, dont les fonctions sont analogues à toutes celles des protéines Cas des systèmes de classe 1.

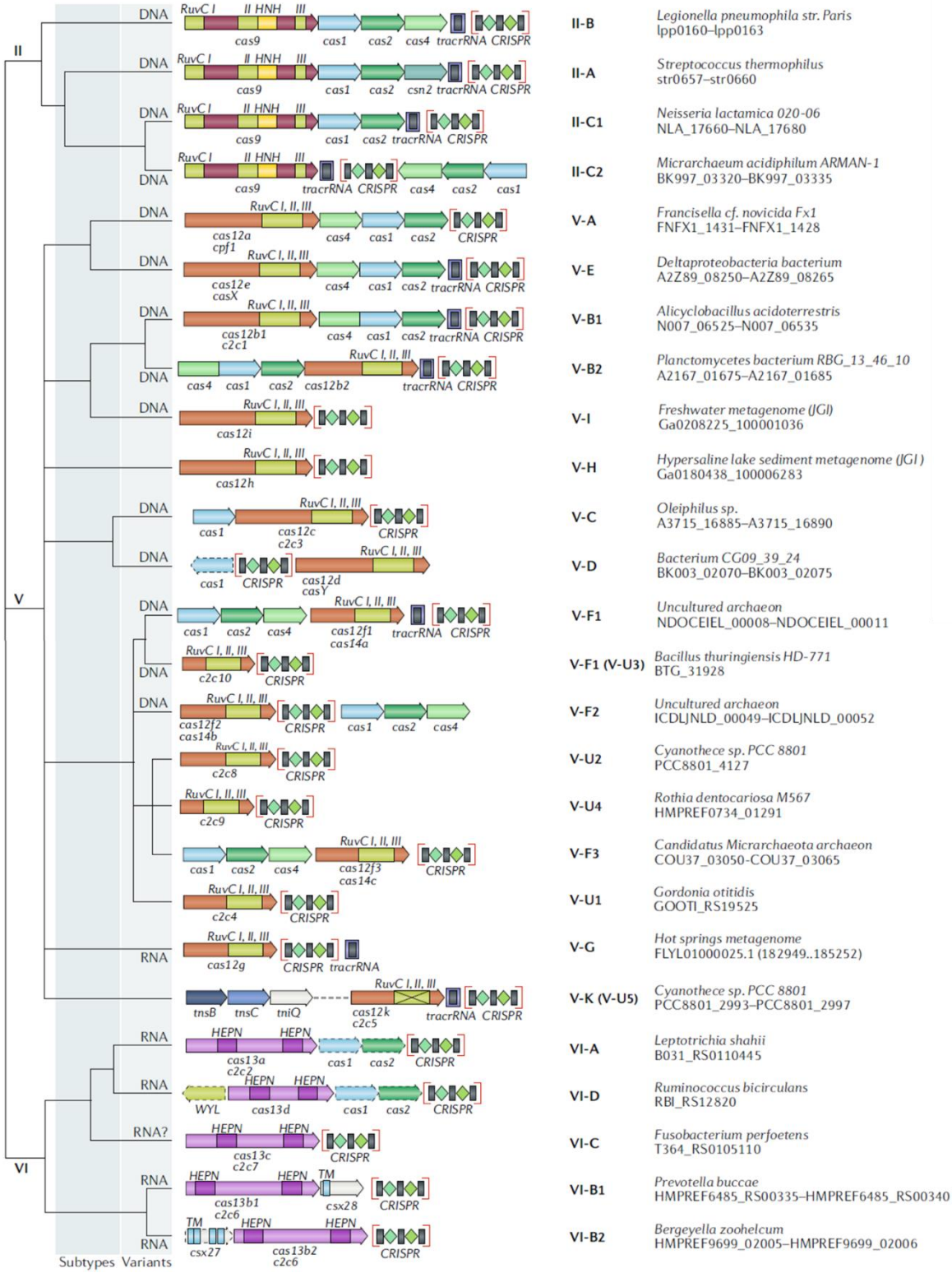
En plus des protéines universelles Cas1 et Cas2, les différents types de systèmes au sein d'une même classe sont définis par une protéine signature Cas. Les sous-types se différencient ensuite par les autres gènes *cas* et par leur organisation.

La classe 1 renferme 3 types de systèmes. Les systèmes de type I se caractérisent par la protéine signature Cas3 et sont divisés en 7 sous-types, caractérisés par des compositions et une organisation en gènes *cas* différentes. Les systèmes de type IV sont divisés en 3 sous-types et caractérisés par le gène *csf1*. La protéine Cas10 est caractéristique des systèmes de type III, subdivisés en 6 sous-types (Figure 14).



<u>Figure 14 :</u> Classification des systèmes de classe 1 (Makarova *et al.*, 2020).
Représentation de l'architecture de l'opéron *cas* des systèmes CRISPR-Cas de classe 1. Les gènes encadrés en orange sont ceux entrant dans la composition du complexe effecteur.

La classe 2 se compose des systèmes de type II, de type V et de type VI. La différence majeure avec les systèmes de classe 1 est le nombre de gènes *cas* qui les composent. En effet ceux-ci sont les plus simples car ils ne sont composés que de 3 ou 4 gènes généralement (Figure 15).



<u>Figure 15 :</u> Classification des systèmes CRISPR-Cas de classe 2 (Makarova *et al*., 2020).
Représentation de l'architecture de l'opéron *cas* des systèmes CRISPR-Cas de classe 2.

Les systèmes de type II sont caractérisés par la protéine effectrice multidomaine Cas9. Les 3 sous-types sont définis par le reste des gènes *cas* constituant l'opéron et par leur organisation. Ainsi, le type II-A se compose de l'enchaînement *cas9*, *cas1*, *cas2*, *csn2* alors que le type II-B contient *cas9*, *cas1*, *cas2* et *cas4* et que le type III-C ne contient que les 3 premiers gènes. Les systèmes de type V et de type VI, de découverte plus récente, sont caractérisés par les protéines Cas12 et Cas13 respectivement. Cette dernière classification a mis en évidence une expansion drastique des systèmes de classe 2 et notamment une multiplication des sous-types et variants des systèmes de type V (10 sous-types) et de type VI (4 sous-types). Ceci s'explique en partie par les nombreuses recherches visant à découvrir de nouveaux outils potentiels d'édition des génomes. Bien que tous caractérisés par la présence d'une seule protéine effectrice, Cas9, Cas12 et Cas13 apparaissent très différentes les unes des autres.

2.4. Mode d'action des systèmes CRISPR-Cas

Différentes études ont permis de suggérer que les systèmes CRISPR-Cas jouaient un rôle dans la défense des bactéries contre les EGMs, lorsque des homologies entre les séquences des spacers et celles de phages ou de plasmides ont été découvertes, jusqu'à la démonstration de l'immunité anti-virale en 2007 (Barrangou *et al.*, 2007). L'immunité s'acquiert selon un processus en 3 étapes : adaptation, expression et interférence, la dernière étape conduisant à la destruction de l'élément étranger par une ou plusieurs protéines Cas à activité nucléasique (Figure 16 ; Al-Attar *et al.*, 2011 ; Horvath et Barrangou, 2010). Ces trois étapes sont conservées entre les systèmes. Cependant, les différences de composition en gènes *cas* entraînent des différences de mécanisme d'action, les protéines Cas intervenant à chaque étape (Richter, Chang, et Fineran, 2012).

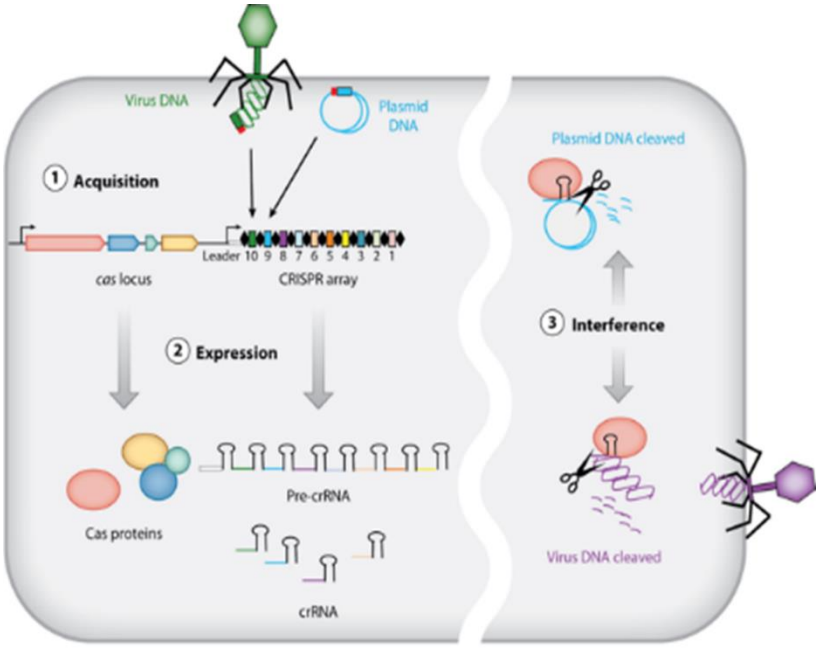


Figure 16 : Récapitulatif du mode d'action des systèmes CRISPR-Cas (The Doudna Lab).

2.4.1. Adaptation

La première étape, appelée adaptation ou encore « immunisation », correspond à la rencontre avec un élément étranger, suivie de l'excision d'un fragment de cet élément, aboutissant à l'obtention d'un pré-spacer. Ce dernier est ensuite incorporé dans le locus CRISPR, sous la forme d'un nouveau spacer, en parallèle de la duplication d'une DR (Figure 17). Ceci permet au système de garder en mémoire ses différentes rencontres avec des EGMs. Les protéines Cas1 et Cas2, quasi-universelles, jouent un rôle primordial dans cette étape (Yosef, Goren, et Qimron, 2012). L'implication d'autres protéines est variable en fonction des systèmes. Ainsi, dans les systèmes de type II, la protéine Cas9, jusque-là principalement étudiée pour son rôle dans l'étape d'interférence, apparaît également nécessaire pour la phase d'immunisation (Heler *et al.*, 2015 ; Wei, Terns, et Terns, 2015).

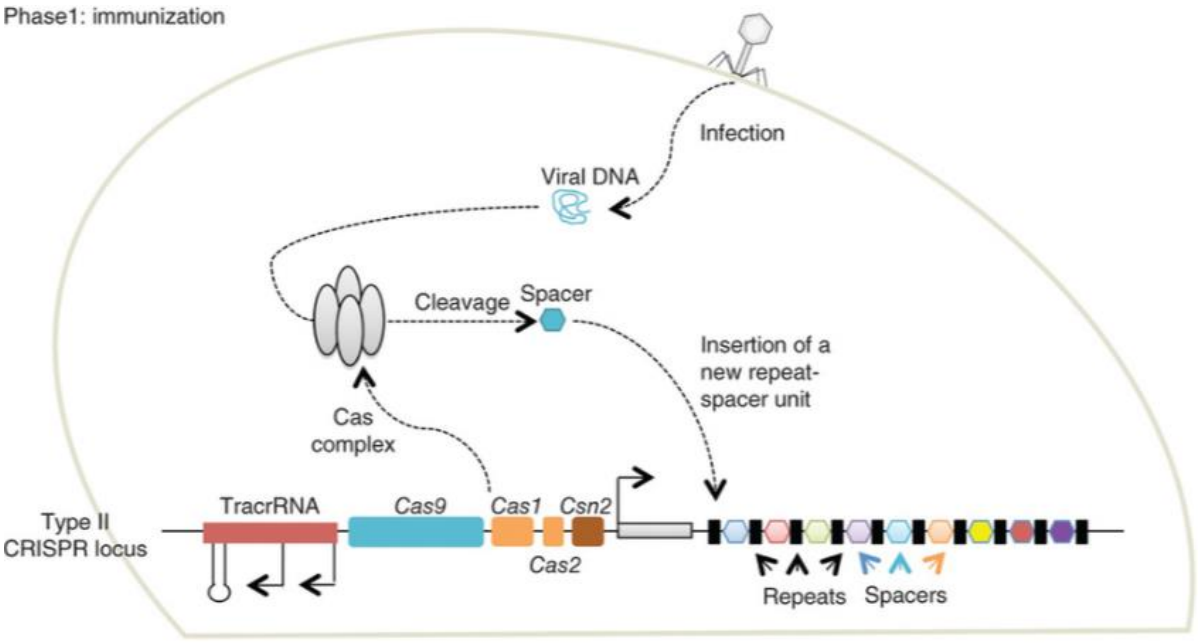


Figure 17 : Description de l'étape d'adaptation (Mali, Esvelt, et Church, 2013).
Après rencontre avec un ADN étranger, la phase d'adaptation correspond à l'excision d'un fragment de cet ADN et à son insertion dans le locus CRISPR, sous la forme d'une nouvelle unité repeat-spacer. Cette intégration se fait de manière polarisée, du côté du leader.

L'intégration de ces nouveaux spacers se produit de manière polarisée, à l'extrémité 5' du locus CRISPR, c'est-à-dire l'extrémité adjacente au leader (Figure 17). Cette intégration polarisée fournit ainsi une vision chronologique des immunisations de la bactérie (Horvath *et al.*, 2008 ; Pourcel, Salvignol, et Vergnaud, 2005). Le leader et plus particulièrement, les nucléotides à proximité immédiate de la première DR du locus CRISPR, sont indispensables à l'intégration de nouveaux spacers (Wei *et al.*, 2015).

2.4.2. Expression

L'étape d'expression se définit comme la transcription des différents éléments du système CRISPR-Cas. Le locus CRISPR est en premier lieu transcrit en un long et unique ARN appelé le pre-crRNA, à partir d'un promoteur situé dans la région leader. Ce pre-crRNA est ensuite clivé au niveau des repeats pour former les ARNs CRISPR matures (crRNAs). Ces crRNAs correspondent chacun à un spacer associé à une partie des DRs adjacentes (Figure 18).

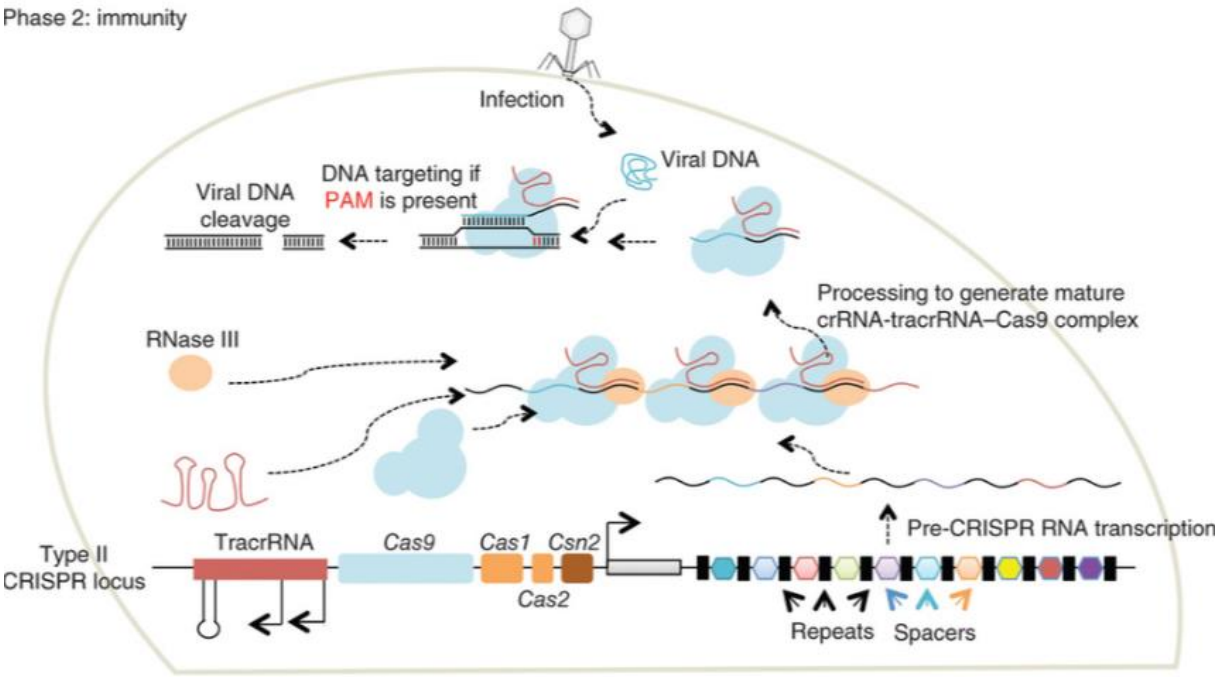


Figure 18 : Représentation des étapes d'expression et d'interférence des systèmes CRISPR-Cas de type II-A (Mali, Esvelt, et Church, 2013).
L'étape d'expression correspond à la transcription du locus CRISPR, à partir du promoteur présent dans le leader, en un long précurseur : le pre-crRNA. Divers acteurs permettent ensuite sa maturation, par clivage, en crRNA. Celui-ci se compose alors d'un spacer et d'un fragment de la DR adjacente (bas de la figure).
Lors d'une nouvelle infection par le même phage, le crRNA, associé aux protéines Cas, va se lier à la séquence étrangère complémentaire de son spacer, après reconnaissance du PAM. Ceci entraîne alors la dégradation de la cible par les protéines Cas. C'est l'interférence (haut de la figure).

Divers acteurs interviennent dans la maturation des crRNAs en fonction des types de systèmes CRISPR-Cas. Par exemple, pour les systèmes de types I et III, le clivage du pre-crRNA est réalisé par une nucléase de la famille des protéines Cas6 (Figure 19). Etonnamment, la maturation des crRNAs dans les systèmes de type II fait intervenir une protéine de l'hôte, la RNAse III bactérienne. Un autre élément important de cette étape est la transcription du tracrRNA (*trans*-activating CRISPR RNA) qui possède une homologie de séquence avec la DR du locus CRISPR (Deltcheva *et al*., 2011). Ainsi, le tracrRNA s'hybride par complémentarité de séquence au pre-crRNA, puis la RNase III endogène, en association

avec Cas9, clive ce duplex, conduisant aux crRNAs matures. Le tracrRNA reste ensuite complexé avec le crRNA afin de le stabiliser (Figure 18 et Figure 19).

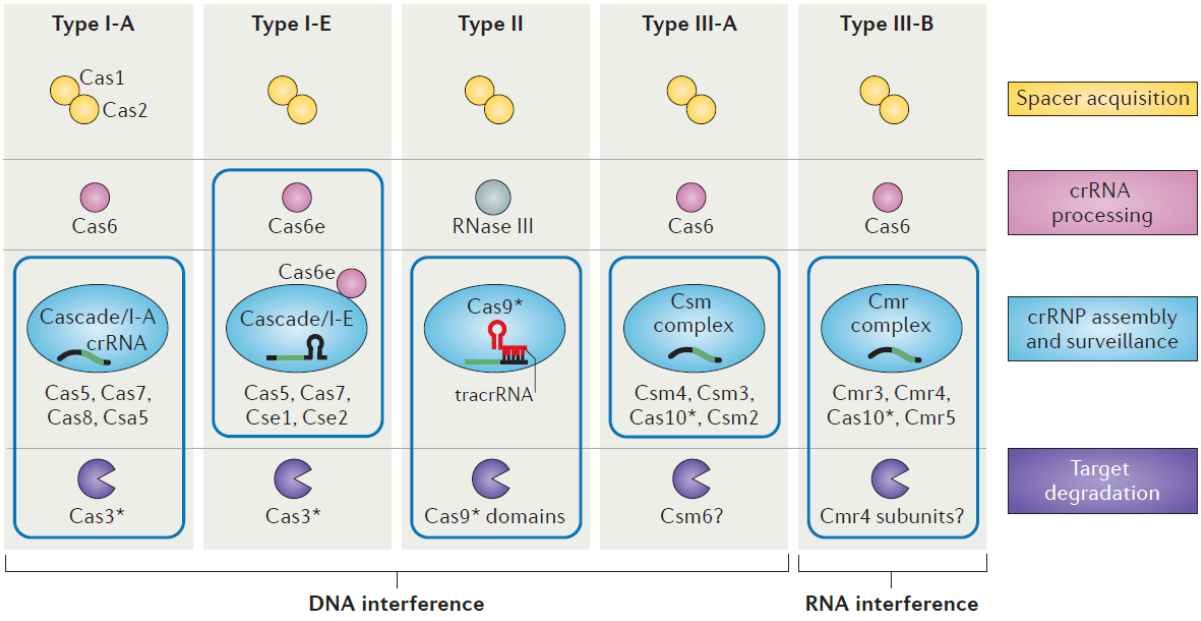


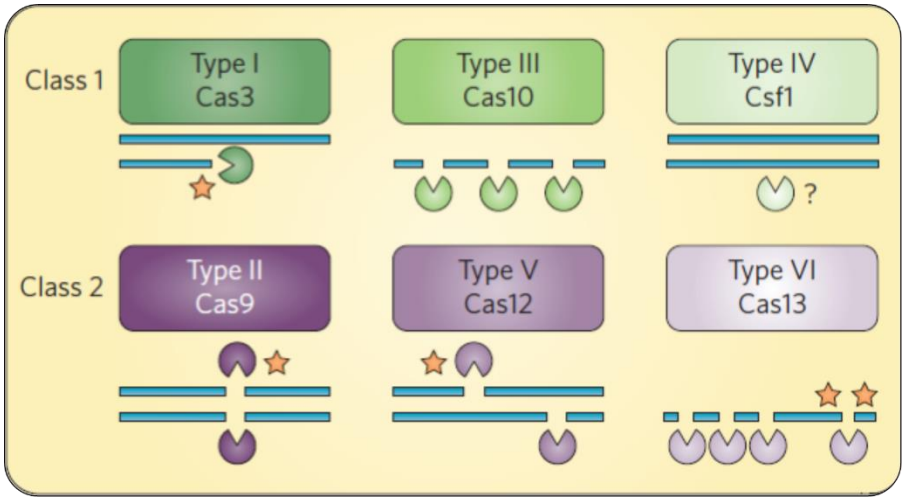
Figure 19 : Implication de différentes protéines Cas en fonction des étapes et en fonction du type de système CRISPR-Cas (van der Oost *et al.*, 2014).

A la fin de l'étape d'expression, les crRNAs matures se retrouvent associés à des protéines Cas, variables en fonction des systèmes, pour former le complexe effecteur (crRNP). Le crRNP permet alors la « surveillance » immunitaire dans le cytoplasme bactérien (Figure 18 et Figure 19). Pour les systèmes de type I, le crRNP, ou complexe « Cascade », est formé de multiples protéines Cas. Il en est de même pour les systèmes de type III (complexes Csm et Cmr). En revanche, pour les systèmes de type II, ce complexe est formé uniquement de la protéine Cas9, associée à deux ARNs, le crRNA et le tracrRNA (Figure 18 et Figure 19).

2.4.3. Interférence

Enfin, l'étape d'interférence ou « immunité », se caractérise par la reconnaissance d'une séquence étrangère, appelée protospacer, qui sera dégradée par une protéine Cas à activité nucléasique (Figure 18). L'action de cette dernière est ciblée spécifiquement grâce aux crRNAs. En effet, le crRNA, s'il porte un spacer présentant une homologie avec le protospacer entrant, va s'hybrider par complémentarité à l'élément étranger. Comme le crRNA est lié aux nucléases Cas au sein du complexe effecteur, on parle d'action guidée (Deltcheva *et al.*, 2011 ; Gasiunas *et al.*, 2012). Les protéines Cas à activité nucléasique et leur mode d'action varient en fonction des types de systèmes. Ainsi, dans les systèmes de type I, l'exonucléase Cas3 clive le brin d'ADN cible puis le dégrade. Dans les systèmes de type III, la

nucléase Cas10 clive les ARNs messagers. Pour les systèmes de type II, la protéine Cas9 entraîne une coupure double brin de l'ADN cible. La protéine Cas12 des systèmes de type V clive également les deux brins d'ADN, mais les sites de coupure sont décalés. Enfin, les systèmes de type VI font appel à la protéine Cas13 pour cliver les ARNs cibles portant les protospacers complémentaires (Figure 20 ; Barrangou et Horvath, 2017).



<u>Figure 20 :</u> Récapitulatif des protéines Cas effectrices et de leur mode d'action en fonction des types de système (d'après Barrangou et Horvath, 2017).
Les protéines Cas à activité nucléasique peuvent cibler l'ADN (présence de deux brins) ou l'ARN (présence d'un seul brin) et engendrent une coupure simple brin ou double brin. Les étoiles jaunes indiquent que le mécanisme de reconnaissance fait intervenir un PAM.

Pour certains systèmes, la reconnaissance de la cible fait intervenir un élément supplémentaire, le Protospacer Adjacent Motif (PAM). Il s'agit le plus souvent d'un di- ou tri-nucléotide, situé juste à l'extrémité du protospacer. Le PAM est très conservé et spécifique au sein d'un système donné. En revanche, sa séquence et sa localisation varient en fonction des types de systèmes CRISPR-Cas (Deveau *et al*., 2008 ; Mojica *et al*., 2009). La présence du PAM est indispensable pour la reconnaissance de la séquence étrangère cible. Une mutation dans le PAM peut être suffisante pour empêcher la reconnaissance de la cible, même en cas de complémentarité parfaite entre le crRNA et le protospacer. Le PAM joue également un rôle essentiel dans l'étape d'adaptation. En effet, les précurseurs des spacers sont sélectionnés sur les séquences étrangères après la reconnaissance du PAM. Ainsi n'importe lequel des brins peut être ciblé. Le PAM déterminera ensuite l'orientation du spacer dans le locus CRISPR (Figure 21 ; Mojica *et al.*, 2009).

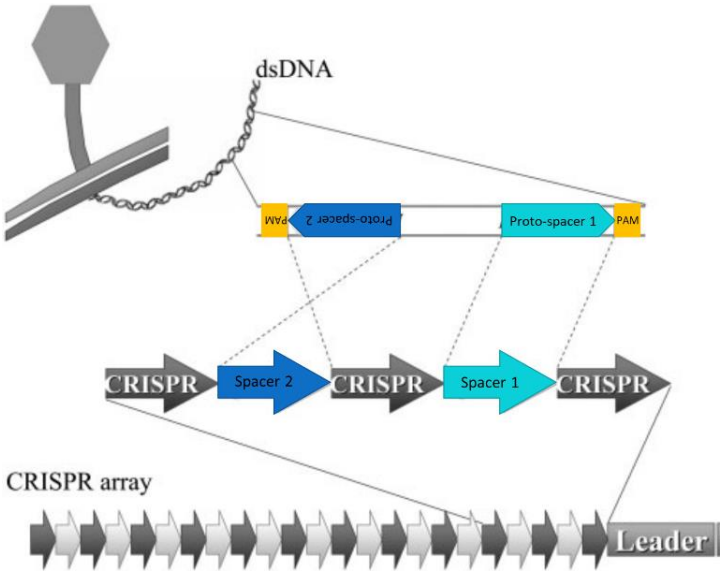


Figure 21 : Le PAM permet la bonne orientation des spacers au sein du locus CRISPR (d'après Mojica *et al*., 2009).
Le protospacer 1 et le protospacer 2 sont situés sur les brins opposés de l'ADN phagique, donc en orientation inverse. Le PAM, en orange, permet l'insertion des spacers qui en dérivent dans le locus CRISPR, dans la même orientation.

2.5. Applications dérivées des systèmes CRISPR-Cas

2.5.1. « Genome editing »

Depuis une dizaine d'années, l'application la plus recherchée et la plus étudiée des systèmes CRISPR-Cas est leur utilisation pour l'édition des génomes (Barrangou et Horvath, 2017 ; Doudna et Charpentier, 2014 ; Horvath et Barrangou, 2013 ; Mali *et al*., 2013). En effet, les systèmes CRISPR-Cas constituent actuellement un sujet de portée scientifique et médicale majeure, qui sort du domaine microbiologique, du fait de l'énorme potentiel de ces systèmes. Cet engouement scientifique est dû à la simplicité relative du fonctionnement de l'étape d'interférence des systèmes CRISPR-Cas de classe 2 et plus particulièrement, des systèmes de type II-A. La base du fonctionnement du système repose sur un crRNA, qui guide la protéine Cas9 effectrice à activité nucléase, pour qu'elle dégrade l'ADN cible complémentaire. Le but de l'outil d'édition des génomes est de reprogrammer le système. En lui fournissant un crRNA complémentaire de la séquence d'intérêt, l'endonucléase Cas9 est utilisée à des fins de modification génétique, en la guidant pour couper l'ADN, dans le but de muter, insérer ou déléter des nucléotides au sein de séquences préalablement ciblées. Le système CRISPR-Cas9 le plus utilisé est celui dérivant de *Streptococcus pyogenes.* En effet, l'action de la nucléase SpCas9 requiert uniquement le crRNA et le tracrRNA, qui peuvent être fusionnés en un ARN chimérique, l'ARN « guide » ou sgRNA (single guide ARN, Figure 22 ; Jinek *et al*., 2012). Depuis, le développement de ces systèmes comme outil de modification du génome et leur potentielle utilisation en thérapie génique font l'objet

de nombreuses publications (Figure 11). Ils ont d'ailleurs été testés avec succès au sein de différentes lignées cellulaires et divers organismes et les possibilités semblent infinies (Broeders *et al*., 2020).

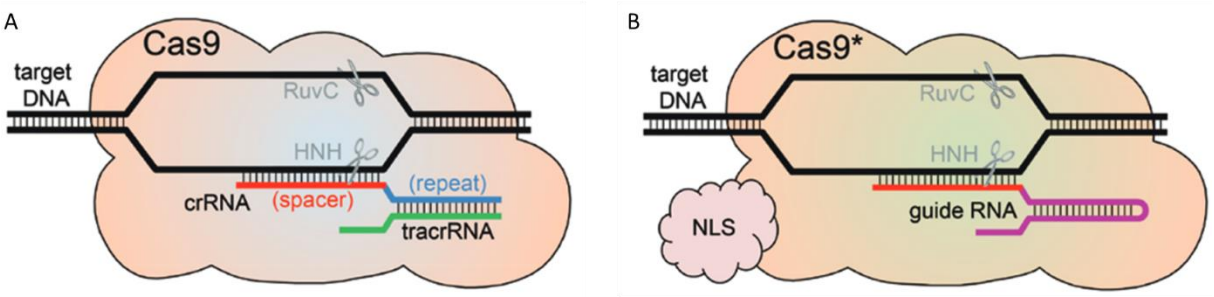
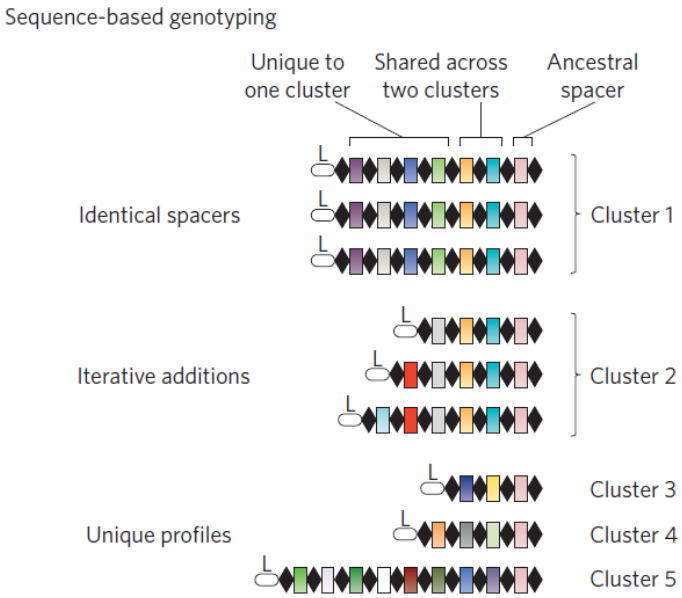


Figure 22 : Principe du « genome editing » dérivé des systèmes CRISPR-Cas de type II-A (Horvath et Barrangou, 2013).
Le schéma A représente le système natif. Le schéma B modélise l'outil d'édition des génomes.

2.5.2. Typage moléculaire

De par leur nature dynamique et surtout l'acquisition polarisée des spacers au sein du locus CRISPR, les systèmes CRISPR-Cas apparaissent également comme des outils intéressants pour la caractérisation des souches au plan moléculaire, ce qui peut être utile en épidémiologie. En effet, l'organisation temporelle des spacers et la composition spécifique en spacers de chaque locus CRISPR en font a priori une méthode performante avec un potentiel discriminant élevé, pour l'analyse des relations phylogénétiques entre les lignées bactériennes. Ainsi, les spacers situés à l'extrémité leader sont les témoins des rencontres récentes de la bactérie, avec des EGMs présents dans son environnement actuel, alors que les spacers proches de la DR terminale reflètent des événements anciens et proviennent d'un ancêtre commun (Figure 23).

Cette méthode de typage a déjà été utilisée avec succès chez diverses bactéries comme *Salmonella enterica,* afin de comparer rapidement les souches isolées au cours d'une épidémie, ou chez *Mycobacterium tuberculosis*. Chez cette dernière, le spoligotyping, ou typage grâce aux systèmes CRISPR-Cas, est utilisé depuis deux décennies, avant même que le rôle de ces systèmes chez les bactéries n'ait été élucidé (Shariat et Dudley, 2014). Bien qu'a priori intéressante pour l'étude des populations, l'analyse des loci CRISPR ne peut s'appliquer à toutes les bactéries. En effet, seulement la moitié des bactéries possède au moins un système CRISPR-Cas. De plus, lorsqu'il est présent, celui-ci doit être dynamique et présenter une variabilité suffisante afin de distinguer les isolats.



<u>Figure 23 :</u> Principe de l'utilisation du polymorphisme des loci CRISPR comme méthode de typage moléculaire (d'après Barrangou et Horvath, 2017).

2.5.3. Les systèmes CRISPR-Cas comme antibactériens

Le principe de cette application est de reprogrammer le système CRISPR-Cas afin qu'il cible le génome bactérien lui-même, conduisant à la mort de la cellule, alors qu'à l'origine, le but d'un système CRISPR-Cas est de protéger la bactérie des EGMs. Etant donnée la spécificité de séquence avec laquelle agissent les systèmes CRISPR-Cas, leur avantage ici serait de pouvoir cibler une bactérie en particulier au sein d'une communauté microbienne, protégeant ainsi les bactéries non pathogènes, à l'inverse d'un antibiotique à large spectre par exemple. Par ailleurs, ce système pourrait également permettre de cibler spécifiquement les gènes de virulence ou les gènes de résistance aux antibiotiques d'une bactérie donnée, qui sont souvent associés à des EGMs. Ce système repose sur l'utilisation d'un vecteur dérivé d'un phage (phagemide), afin de délivrer le crRNA ciblant la séquence souhaitée, voire d'y associer des protéines Cas, notamment pour les bactéries qui ne posséderaient pas de systèmes CRISPR-Cas dans leur génome (Figure 24 ; Bikard et Barrangou, 2017 ; Bikard *et al*., 2014 ; Fage, Lemire, et Moineau, 2021).

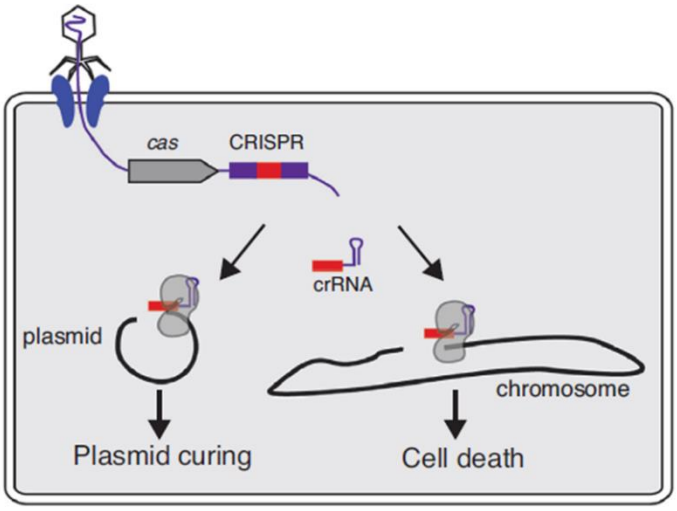


Figure 24 : Principe de l'utilisation des systèmes CRISPR-Cas comme antimicrobiens (d'après Bikard et Barrangou, 2017).
Après hybridation du crRNA, délivré par le phagemide, à la séquence cible sur un plasmide ou sur le chromosome bactérien, le clivage par la nucléase Cas conduit à la perte du plasmide ou à la mort cellulaire.

2.6. CRISPR-Cas et virulence bactérienne

2.6.1. Limitation des transferts horizontaux de gènes par les systèmes CRISPR-Cas

Plusieurs études ont démontré une corrélation inverse entre la présence d'un système CRISPR-Cas et le potentiel de virulence des bactéries. En effet, ces dernières acquièrent régulièrement des facteurs de virulence grâce à des transferts horizontaux de gènes, or le rôle des systèmes CRISPR-Cas est d'empêcher l'entrée de ces EGMs dans la bactérie. Ainsi, la capacité des systèmes CRISPR-Cas à inhiber le transfert de gènes de virulence ou de résistance aux antibiotiques, pourrait être un frein dans le développement du pouvoir pathogène des bactéries (Hatoum-Aslan et Marraffini, 2014).

Parmi les différents exemples de liens entre la virulence liée aux EGMs et les systèmes CRISPR-Cas, il a pu être observé une corrélation négative entre le nombre de spacers et le potentiel de pathogénicité des souches d'*E. coli* productrices de Shiga toxines (STEC), responsables d'épidémies de gastro-entérites et impliquées dans le syndrome hémolytique et urémique (Toro *et al*., 2014). Il a été montré que les souches à haut potentiel épidémique possèdent moins de spacers, or ces souches sont connues pour renfermer de nombreux prophages et autres EGMs dans leur génome. Il a également été mis en évidence une association entre le nombre de spacers et la présence de certains gènes de virulence. Par exemple, les souches de STEC porteuses du gène *eae*, codant l'intimine, ou des 2 gènes *stx*, codant les shigatoxines, possèdent significativement moins de spacers que les souches ne possédant pas ces gènes (Figure 25). De même, la présence d'un système CRISPR-Cas et le nombre de

spacers, lorsqu'un système est présent, semblent inversement associés à l'uropathogénicité des souches d'*E. coli* (Dang *et al.*, 2013). En effet, l'absence de système CRISPR-Cas rend les *E. coli* uropathogènes plus sensibles aux infections par les phages ou les plasmides, leur permettant ainsi de s'adapter plus rapidement à leur environnement, par exemple par l'acquisition de gènes de résistance aux antibiotiques.

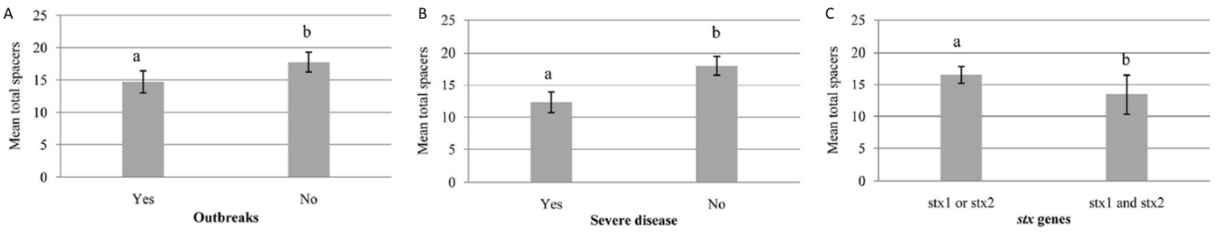


Figure 25 : Association entre le nombre de spacers et la capacité des souches de STEC à engendrer des épidémies (A) ou des pathologies sévères (B) ou avec la présence des gènes *stx* (C) (d'après Toro *et al.*, 2014).
Les lettres a et b indiquent les différences significatives (p<0,05).

Chez les entérocoques multi-résistants, l'étude de la co-distribution des loci CRISPR et des gènes de résistance aux antibiotiques a mis en évidence une corrélation inverse entre la présence d'un locus CRISPR et la présence de gènes de résistance aux antibiotiques (Palmer et Gilmore, 2010). Or, les entérocoques sont également très fréquemment porteurs d'EGMs, que ce soient des plasmides, des prophages ou des îlots de pathogénicité, qui peuvent constituer jusqu'à 25 % de leur génome, ce qui ne favorise pas la cohabitation avec les systèmes CRISPR-Cas. Par ailleurs, les souches possédant un locus CRISPR ont été, pour la plupart, isolées avant l'ère de l'antibiothérapie moderne. Ainsi la perte du locus CRISPR chez les entérocoques pourrait être une conséquence de l'utilisation des antibiotiques à large spectre. Des constats similaires ont été observés chez d'autres espèces bactériennes, telle que *Klebsiella pneumoniae*, d'importance majeure en pathologie humaine et dont la résistance aux antibiotiques peut être problématique (Mackow *et al*., 2019). Plus récemment, une étude portant sur 11 espèces bactériennes impliquées en pathologie humaine, a montré que lorsque la souche possède un système CRISPR-Cas, elle présente, en contrepartie, moins de gènes de résistances aux antibiotiques. En effet la plupart de ces gènes sont présents sur des EGMs (plasmides conjugatifs, intégrons, éléments intégratifs conjugatifs (ICEs),…), permettant un transfert simplifié entre les espèces bactériennes (Pursey *et al*., 2021).

Chez *Streptococcus pyogenes*, il a été montré que les souches qui ne possèdent pas de locus CRISPR possèdent, en revanche, plus de prophages (Figure 26 ; Nozawa *et al*., 2011). De même, il a été trouvé une corrélation inverse significative entre le nombre de spacers au sein du locus, lorsqu'il est présent, et le nombre de prophages. Or, comme d'autres bactéries, *S. pyogenes* a acquis plusieurs toxines et de

nombreux facteurs de virulence par l'intégration de prophages (Beres et Musser, 2007). Ainsi le système CRISPR-Cas, en régulant l'insertion des prophages, peut influencer la virulence des bactéries.

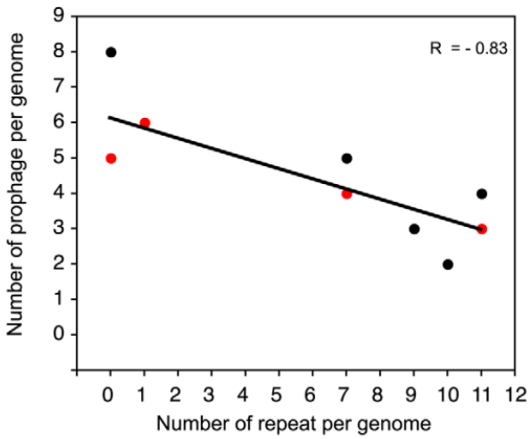


Figure 26 : Corrélation inverse entre le nombre de prophages et le nombre de spacers par génome chez *S. pyogenes* (d'après Nozawa *et al*., 2011).

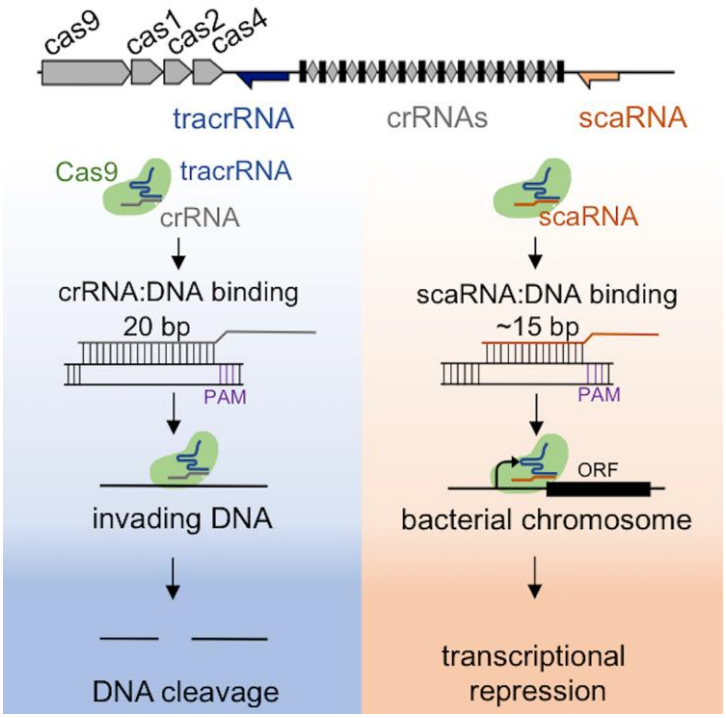
Une hypothèse similaire a été proposée pour un autre pathogène humain majeur, *Staphylococcus aureus*. La pathogénicité des souches de cette espèce repose beaucoup sur les EGMs, tels que les prophages et les plasmides, qui en plus d'être des vecteurs de facteurs de virulence, sont aussi des vecteurs de gènes de résistance aux antibiotiques. Ainsi la quasi-absence de loci CRISPR dans les souches de *S. aureus* pourrait être une conséquence de l'essentialité des EGMs pour la pathogénicité de cette espèce. Ceci pourrait aussi expliquer l'absence de système CRISPR-Cas chez *Streptococcus pneumoniae*. Cette bactérie naturellement compétente utilise énormément le transfert horizontal de gènes pour acquérir des gènes de virulence, ce qui contre-sélectionne les systèmes CRISPR-Cas (Hatoum-Aslan et Marraffini, 2014). Plus récemment, il a été montré que les systèmes CRISPR-Cas ont également été sélectivement inactivés au cours de l'évolution de *Bacillus cereus*, lui permettant une meilleure adaptation à divers environnements (Zheng *et al*., 2020).

2.6.2. Régulation de gènes bactériens par les systèmes CRISPR-Cas

A côté de leur rôle dans la défense immunitaire, les systèmes CRISPR-Cas participent également à l'évolution génétique bactérienne en intervenant dans les réarrangements, les recombinaisons et la régulation de l'expression de gènes, pour certaines espèces bactériennes (Bikard et Marraffini, 2013 ; Sampson et Weiss, 2014). En effet, l'identification de spacers pouvant auto-cibler le chromosome a été, dans un premier temps, plutôt surprenante car la présence de ces spacers devrait entraîner une auto-immunité et conduire ainsi à la mort de la cellule. Or, le fait que ces spacers aient été régulièrement retrouvés et chez plusieurs microorganismes, a rapidement fait suspecter d'autres rôles

pour ces spacers, comme la régulation de gènes (Heussler et O'Toole, 2016 ; Stern *et al*., 2010 ; Westra, Buckling, et Fineran, 2014).

Plusieurs études ont ainsi montré que les systèmes CRISPR-Cas pouvaient directement réguler l'expression des gènes et notamment, les gènes de virulence, contribuant ainsi à améliorer la pathogénicité de la bactérie (Louwen *et al*., 2014). Chez *Francisella novicida,* cette régulation fait intervenir un petit ARN associé au locus CRISPR, le scaRNA. Celui-ci va former un duplex avec le tracrRNA, tout comme les crRNAs, afin de guider l'action de Cas9. Cette dernière pourra alors interagir avec l'ADN bactérien et réguler la transcription de gènes. La complémentarité limitée entre le scaRNA et sa cible chromosomique n'entraîne pas le clivage de la cible, à l'inverse de l'effet obtenu après la liaison du crRNA à sa cible étrangère (Figure 27). Chez *F. novicida*, ce système est à l'origine d'une répression de la transcription de lipoprotéines antigéniques, facilitant l'échappement au système immunitaire et favorisant donc la virulence de la bactérie (Ratner *et al*., 2019 ; Sampson *et al*., 2013). Cas9 semble également à l'origine de la régulation des facteurs de virulence d'autres bactéries pathogènes, comme *Neisseria meningitidis* ou *Campylobacter jejuni* (Louwen *et al*., 2013 ; Sampson *et al.*, 2013). L'absence de Cas9 entraîne une diminution de l'adhésion de ces deux bactéries aux cellules épithéliales et une diminution du processus invasif qui s'en suit.



<u>Figure 27 :</u> Les deux complexes ribonucléoprotéiques de *F. novicida* (Ratner *et al.*, 2019).
A gauche, le complexe Cas9-crRNA:tracrRNA se lie par complémentarité à l'ADN étranger, entraînant son clivage et donc la protection de la bactérie. A droite, le complexe Cas9-scaRNA:tracrRNA interagit avec le chromosome bactérien, entraînant une répression de la transcription des gènes endogènes.

Chez *Legionella pneumophila*, la délétion du gène *cas2* entraîne une diminution de la capacité de multiplication de la bactérie chez les amibes, étape nécessaire au cycle infectieux et à la survie de la bactérie dans l'environnement (Gunderson et Cianciotto, 2013). Chez *S. agalactiae*, deux études ont récemment montré que la délétion de *cas9* entraînait une diminution de la virulence. Ainsi, les souches Δ*cas9* apparaissent atténuées, avec des capacités diminuées d'adhésion aux cellules épithéliales et d'invasion des barrières physiologiques. Plutôt qu'une action directe de Cas9, la modulation de l'expression des gènes de virulence chez *S. agalactiae* semble s'effectuer par le biais d'autres régulateurs (Ma *et al*., 2018 ; Spencer *et al*., 2019).

L'implication des systèmes CRISPR-Cas et particulièrement les systèmes de type II, dans la virulence de certaines bactéries et le fait que ces systèmes sont présents chez un grand nombre de bactéries pathogènes et commensales, suggèrent un rôle supplémentaire que celui de défense antivirale, notamment dans la régulation de la pathogénicité des bactéries (Chylinski *et al*., 2014).

2.7. La régulation des systèmes CRISPR-Cas

A l'état basal, en dehors de toute stimulation, les systèmes CRISPR-Cas sont plus ou moins transcrits en fonction des bactéries. Néanmoins leur expression semble nécessairement constitutive, afin de pouvoir offrir une chance permanente à la bactérie de pouvoir se défendre contre les menaces perpétuelles de l'environnement. En effet, l'expression de ces systèmes est importante pour assurer la survie de la bactérie et sa croissance dans les différents milieux, mais elle représente également un coût énergétique non négligeable, d'où l'importance d'une régulation fine (Vale *et al*., 2015).

Certains facteurs environnementaux sont capables de réguler l'expression des systèmes CRISPR-Cas. C'est notamment le cas lorsque la densité cellulaire augmente. En effet, les bactéries peuvent alors se retrouver, de manière plus importante, au contact de virus et autres éléments pathogènes pouvant porter atteinte à leur intégrité. Ce phénomène a été décrit chez *Serratia* sp. ATCC39006 et a montré que lorsque la densité cellulaire augmente, les signaux de quorum sensing augmentent et engendrent une augmentation de la transcription des systèmes CRISPR-Cas (Figure 28). Ainsi la communication, via le quorum sensing, permet aux bactéries de maintenir l'équilibre entre les exigences en mécanismes de défense pour assurer leur protection, en présence d'une haute densité cellulaire et les coûts énergétiques de l'activité des systèmes CRISPR-Cas (Patterson *et al*., 2016). Ce phénomène a également été observé chez *Pseudomonas aeruginosa* (Høyland-Kroghsbo *et al*., 2017).

Chez *P. aeruginosa*, la température peut aussi réguler l'expression des systèmes CRISPR-Cas. En effet, l'étape d'adaptation, qui correspond à l'acquisition de nouveaux spacers au sein du locus CRISPR, est plus efficace à basse température qu'à la température du corps humain (Figure 29). Ceci s'explique

par un ralentissement de la croissance bactérienne à basse température, qui offre un délai supplémentaire à la bactérie pour acquérir un spacer (Høyland-Kroghsbo, Muñoz, et Bassler, 2018).

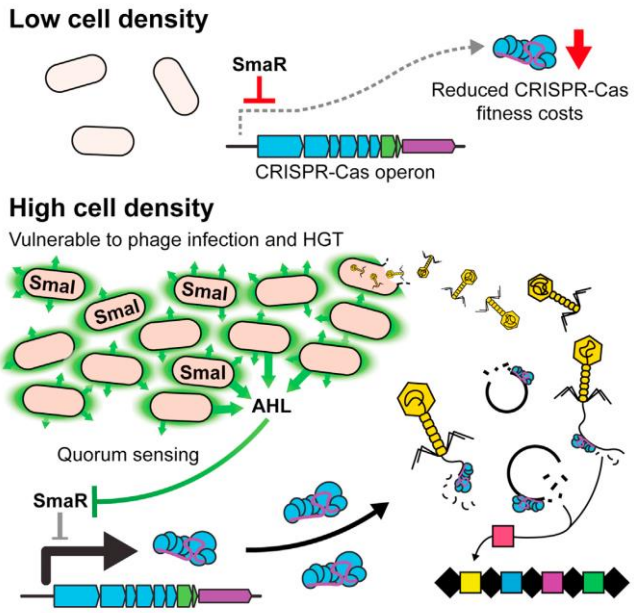


Figure 28 : Le quorum sensing régule l'expression des systèmes CRISPR-Cas de *Serratia* sp (Patterson *et al.*, 2016).
En présence d'une faible densité cellulaire, SmaR réprime l'expression des systèmes CRISPR-Cas. En présence d'une forte densité cellulaire, les bactéries communiquent via le quorum sensing (signaux AHL), ce qui inhibe SmaR et donc lève son effet répresseur sur l'expression des systèmes CRISPR-Cas.

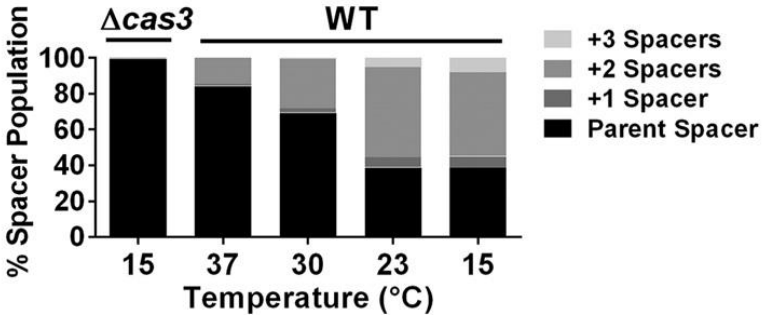
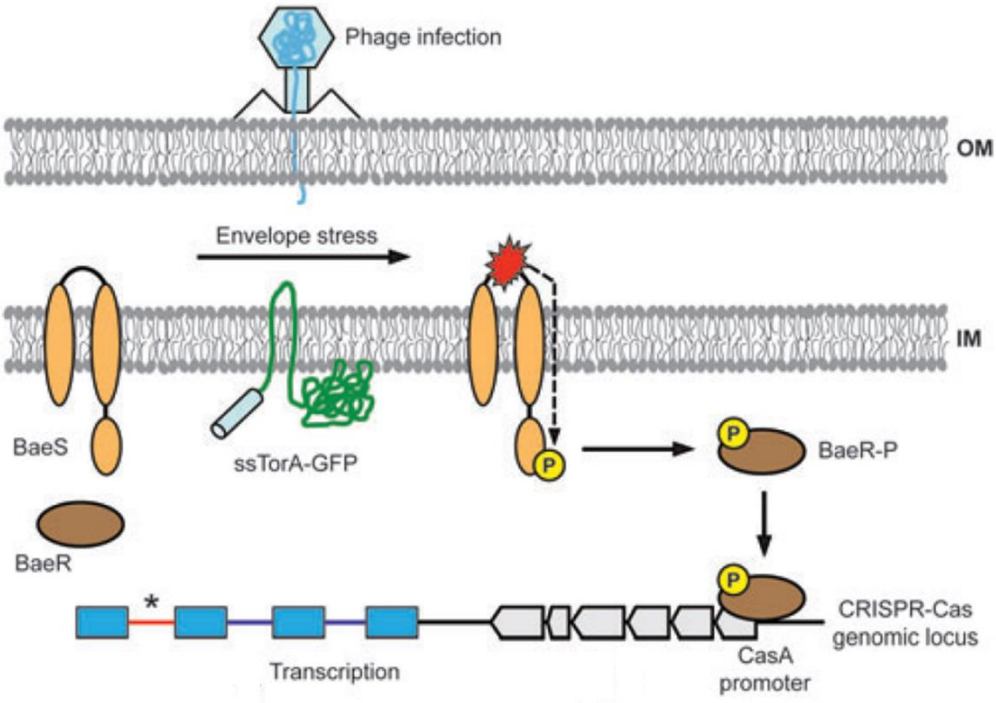


Figure 29 : Etude de l'étape d'adaptation en fonction de la température chez *P. aeruginosa* (d'après (Høyland-Kroghsbo, Muñoz, et Bassler, 2018).
A chaque diminution de température, la proportion de bactéries ayant acquis un ou plusieurs spacers augmente.

La présence d'éléments génétiques étrangers, tels que les bactériophages ou les plasmides, est également capable d'activer les systèmes CRISPR-Cas (He *et al.*, 2017). C'est notamment le cas chez *Streptococcus thermophilus*, pour lequel il a été mis en évidence une transcription constitutive de

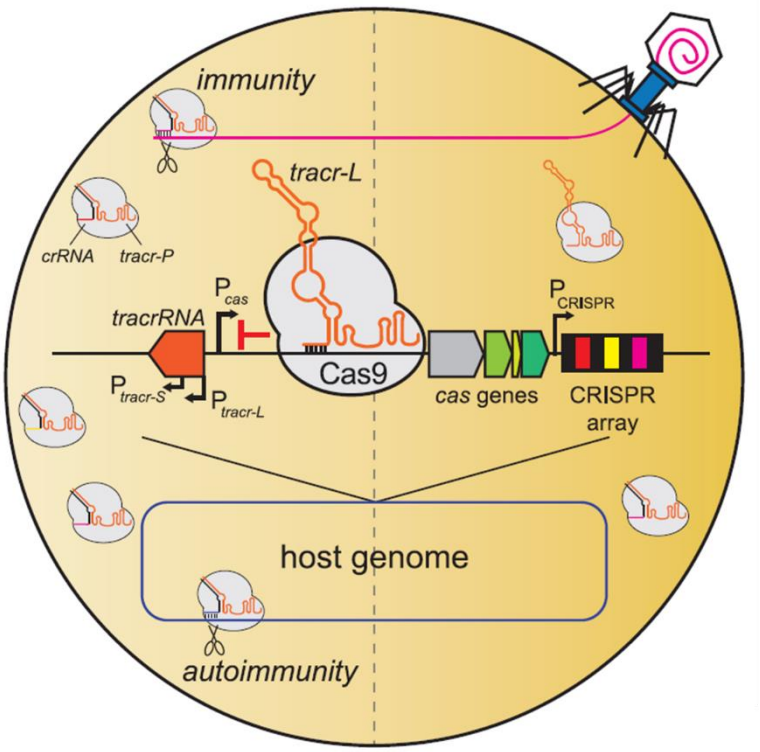
certains éléments composant ces systèmes, ainsi qu'une augmentation importante de cette transcription lors de l'infection de la souche bactérienne par un phage connu de celle-ci (Young *et al*., 2012). L'activation de la transcription se produit probablement en réponse à un stress exercé sur l'enveloppe. En effet, après adsorption du phage à l'enveloppe bactérienne, celui-ci va injecter son ADN dans la bactérie, conduisant à la rupture de l'enveloppe. La bactérie est alors capable de détecter cette effraction, par divers mécanismes, entraînant ensuite l'augmentation de la synthèse des mécanismes de défense, dont les systèmes CRISPR-Cas, afin de se protéger contre les menaces (Ratner, Sampson, et Weiss, 2015). Ainsi, plusieurs régulateurs de réponse au stress semblent pouvoir intervenir dans la régulation de l'expression des systèmes CRISPR-Cas. Le système à deux composant BaeSR est capable d'activer les systèmes CRISPR-Cas en réponse à une perturbation au niveau de l'enveloppe (Figure 30 ; Perez-Rodriguez *et al*., 2011). A l'état basal, H-NS (histone-like nucleoid-structuring protein) inhibe les systèmes CRISPR-Cas de plusieurs espèces bactériennes, telles que *E. coli*, *Salmonella enterica* serovar Typhi et *K. pneumoniae* (Lin *et al*., 2016 ; Medina-Aparicio *et al*., 2011 ; Westra *et al*., 2010). Lors d'un stress au niveau de l'enveloppe, H-NS est inhibée, conduisant à la surexpression des systèmes CRISPR-Cas.



<u>Figure 30 :</u> Représentation schématique de la réponse BaeSR suite à un stress au niveau de l'enveloppe (d'après Perez-Rodriguez *et al*., 2011).
Après détection du signal par BaeS, il phosphoryle BaeR qui peut ensuite activer la transcription du système CRISPR-Cas.

Des changements métaboliques importants, tels que la disponibilité des sucres dans l'environnement, qui peuvent survenir lors de l'infection par un phage, sont également à l'origine du déclenchement de voies de régulation de l'expression des systèmes CRISPR-Cas (Patterson *et al.*, 2015).

Tout récemment, un nouveau mécanisme de régulation de l'expression des systèmes CRISPR-Cas de type II-A a été identifié chez *S. pyogenes* (Figure 31 ; Workman *et al*., 2021). Une forme longue du tracrRNA, *tracr-L*, sert d'ARN guide à la protéine Cas9. La cible de ce complexe effecteur est le promoteur de l'opéron *cas* lui-même, auquel il se fixe sans entraîner de clivage de l'ADN, la taille de la région d'homologie n'étant pas suffisante. Ainsi Cas9 régule sa propre transcription. En régulant l'expression de la protéine Cas9, *tracr-L* régule aussi indirectement la quantité de tracrRNA et de crRNAs, apparaissant ainsi comme un régulateur de l'ensemble du système CRISPR-Cas. Ce mécanisme est suspecté de jouer un rôle dans la prévention de l'auto-immunité, en régulant la quantité de protéine Cas9 présente dans la cellule, tout en la maintenant à un niveau suffisant pour la surveillance des menaces virales. Ce mécanisme semble également exister chez certaines souches de *S. agalactiae* (Workman *et al*., 2021).



<u>Figure 31 :</u> Le *tracr-L* guide Cas9 pour autoréguler l'expression du système CRISPR-Cas (Workman *et al*., 2021).
La régulation de la quantité de protéine Cas9 dans la cellule par le *tracr-L* permet de maintenir l'équilibre entre l'immunité et l'auto-immunité.

3. Le système CRISPR-Cas de type II-A de *S. agalactiae*

Chez *S. agalactiae*, deux loci CRISPR ont été décrits. Le système CRISPR1-Cas, classé parmi les systèmes de type II-A, apparaît ubiquitaire et dynamique, comme en témoigne la diversité du contenu en spacers. Le second, CRISPR2-Cas, est un système de type I-C, présent chez seulement 20% des souches. De plus, lorsqu'il est présent, il apparaît souvent incomplet, suggérant une faible voire une absence d'activité (Lopez-Sanchez *et al*., 2012).

3.1. Organisation

3.1.1. Organisation générale des systèmes de type II-A

Les systèmes de type II-A sont les mieux caractérisés à ce jour et s'organisent en deux grandes régions, l'opéron *cas* et le locus CRISPR (Figure 32). Ils font partie des systèmes CRISPR-Cas les plus simples car ils se composent de seulement 4 gènes *cas*. Un élément supplémentaire caractérise ces systèmes, le tracrRNA pour *trans*-activating CRISPR RNA. Sa position peut être variable, généralement en amont de *cas9* ou entre *cas9* et *cas1* (Chylinski, Le Rhun, et Charpentier, 2013).

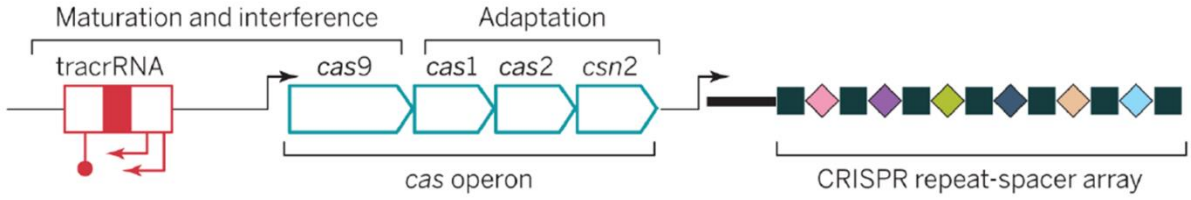


Figure 32 : Organisation générale des systèmes de type II-A (d'après Doudna et Charpentier, 2014). Le système CRISPR-Cas de type II-A se compose de l'opéron *cas* avec l'enchaînement des gènes *cas9*, *cas1*, *cas2* et *csn2* ainsi que du locus CRISPR, constitué d'une alternance de DRs (rectangles noirs) et de spacers (losanges colorés). Le tracrRNA est généralement situé en amont et en antisens de l'opéron *cas.* La séquence leader, en amont du locus CRISPR, renferme le promoteur de transcription du locus. Les différents promoteurs du système sont représentés par des flèches (noires et rouges). L'orientation de la flèche indique le sens de transcription.

3.1.2. Organisation du système CRISPR1-Cas de S. agalactiae

Le système CRISPR1-Cas de *S. agalactiae* présente l'organisation typique des systèmes de type II-A (Figure 33). Au sein du locus CRISPR1, les DRs sont toujours d'une taille de 36 pb (Lopez-Sanchez *et al*., 2012). De plus, leur séquence est très conservée, exceptée pour la DR terminale. Partiellement dégénérée, surtout à l'extrémité 3', la DRT permet d'identifier la fin du locus. D'autre part, l'analyse de 126 souches de *S. agalactiae* a permis d'identifier 4 DRTs différentes, dont la distribution est corrélée avec le typage par MLST (Lier *et al.*, 2015). Ainsi les souches appartenant aux CC17, CC23 et

ST388 possèdent une DRT spécifique, alors que les souches appartenant aux CC1, CC10 et CC19 possèdent la même DRT.

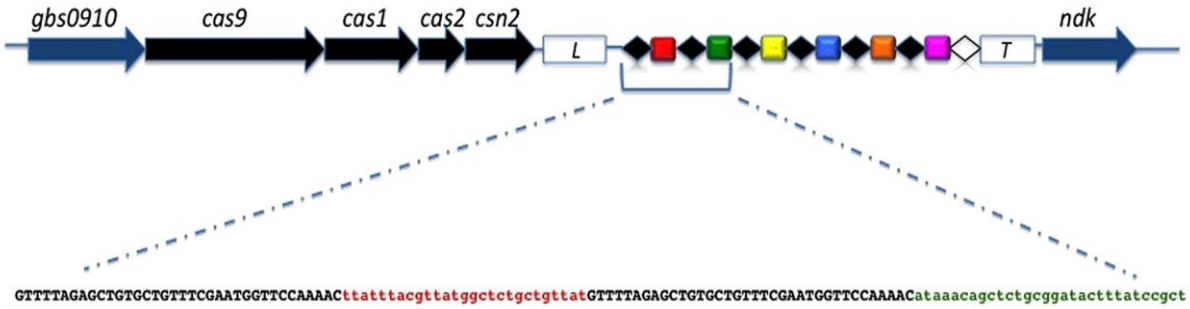


Figure 33 : Organisation du système CRISPR1-Cas de *S. agalactiae* (Lier *et al*., 2015).
Les gènes *cas* sont représentés par des flèches noires. Les gènes en amont et en aval du système CRISPR1-Cas sont représentés par des flèches bleues. La séquence leader (*L*), localisée entre l'opéron *cas* et le locus CRISPR, apparaît en blanc. Les DRs sont représentées par des losanges noirs, sauf la DR terminale, qui est représentée par un losange blanc. Les spacers sont représentés par les rectangles de couleur.
La séquence en dessous montre les deux premières unités repeat-spacer, avec les DR en caractères noirs et les spacers en caractères colorés.

Cette même analyse a permis d'identifier plus de 1700 spacers chez les 126 souches. Intercalés entre les DRs, les spacers sont d'une taille homogène de 30 pb mais leurs séquences sont très variables (Figure 33). Parmi les 1714 spacers, environ 30% étaient uniques, ce qui témoigne de la grande variabilité de ce locus chez l'espèce *S. agalactiae*. Le nombre de spacers par locus est très variable en fonction des isolats, le minimum étant d'un seul spacer, mais pouvant contenir jusqu'à une trentaine de spacers dans le même locus. Avec une moyenne de 13,6 spacers par locus, 93% des souches avait une composition unique en spacers, ce qui témoigne d'un haut pouvoir discriminant et de la possible utilisation du système CRISPR1-Cas comme outil de typage des souches. D'ailleurs, un lien entre la composition en spacers et notamment les spacers ancestraux, et les résultats de la caractérisation des souches par MLST a été démontré (Figure 9 ; Lier *et al*., 2015 ; Lopez-Sanchez *et al*., 2012). Afin d'identifier les éléments ciblés par ces spacers, une homologie de séquence a été recherchée dans les bases de données internationales. Ainsi, il a été montré qu'environ la moitié des spacers présentaient une homologie avec au moins un des génomes de *S. agalactiae* entièrement séquencés. Les éléments ciblés sont principalement des ICEs et des prophages. Lorsque les spacers ne ciblent pas le génome de *S. agalactiae*, ils présentent des homologies avec les EGMs insérés chez les espèces proches du genre *Streptococcus* et principalement, *S. pyogenes*. D'autre part, bien que plus rare, une homologie de séquence a également été retrouvée entre certains spacers et des gènes du « core genome » (Lopez-Sanchez *et al.*, 2012).

Chez *S. agalactiae*, le leader est une séquence d'une longueur de 103 pb, très conservée et spécifique. En effet, aucune similarité de séquence n'a été retrouvée avec une autre espèce bactérienne. Comme observé précédemment, c'est une région riche en Adénine et en Thymine (63 %).

En amont, l'opéron *cas* est constitué par les gènes *cas9*, *cas1*, *cas2* et *csn2*, soit l'organisation caractéristique des systèmes de type II-A. Identifié chez *S. agalactiae* NEM316, le tracrRNA se situe dans la région intergénique entre le gène *gbs0910* et l'opéron *cas* (Lopez-Sanchez *et al.*, 2012 ; Rosinski-Chupin *et al*., 2015). Codé en *trans*, il est capable de s'hybrider avec les crRNAs car il possède une homologie de séquence avec la DR du locus CRISPR et son rôle est primordial dans la stabilité de ces ARNs.

3.2. Fonctionnalité

Chez *S. agalactiae*, l'étude du système CRISPR1-Cas a montré son omniprésence chez les souches étudiées, une forte conservation des DRs, ainsi qu'une grande diversité des spacers intégrés, témoignant de son activité et de son dynamisme (Lier *et al*., 2015 ; Lopez-Sanchez *et al*., 2012). Cette deuxième étude a d'ailleurs permis de démontrer la fonctionnalité du système immunitaire *in vitro*. Ils ont ainsi montré la très forte diminution de l'efficacité de transfert par conjugaison du transposon Tn*GBS*2, dans une souche contenant un spacer dirigé contre cet EGM dans son locus CRISPR1.

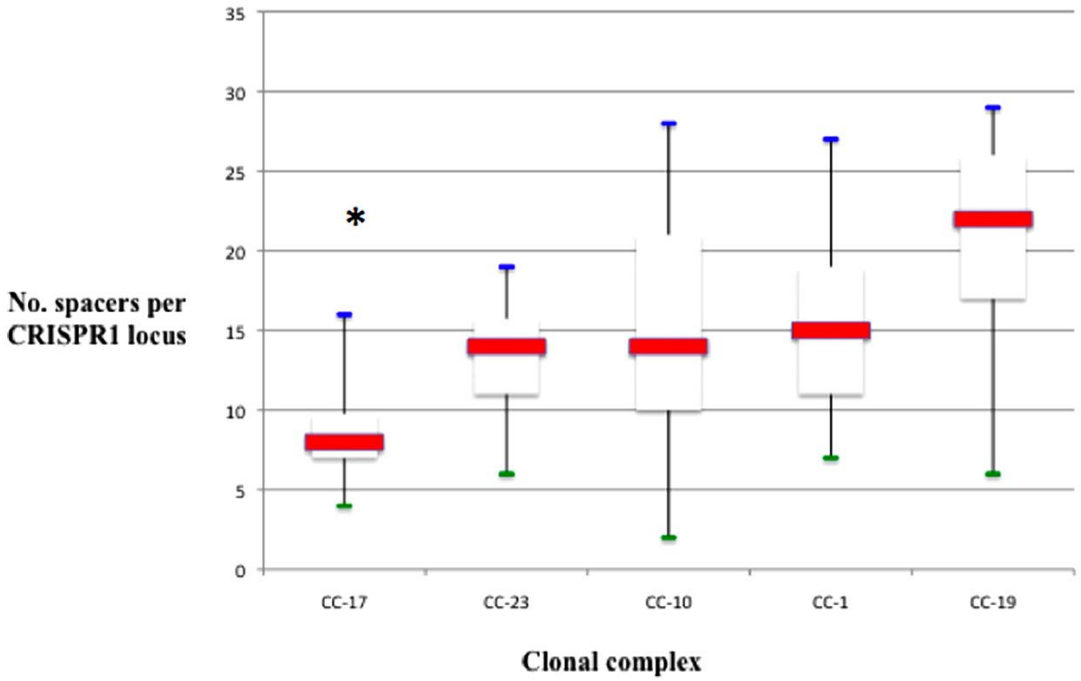
La majorité des spacers présents dans le locus CRISPR1 de *S. agalactiae* sont issus de séquences d'EGMs, tels que des phages ou des plasmides, fréquemment rencontrés dans le genre *Streptococcus*. Plus rarement, certains spacers ciblent les acides nucléiques de *S. agalactiae* lui-même. Cependant, il a été démontré que le système CRISPR1-Cas de *S. agalactiae* est capable de distinguer le soi du non soi. Ceci vient de l'absence de PAM à proximité des protospacers dérivés du « core genome », alors qu'il est quasiment systématiquement présent à l'extrémité 3' des protospacers dérivés d'EGMs. Chez *S. agalactiae*, ce PAM est un dinucléotide GG, en position +2 et +3 (NGG). Ce dinucléotide est indispensable pour le bon fonctionnement du système. Une mutation dans le PAM entraîne une perte de l'activité du système (Lopez-Sanchez *et al*., 2012).

D'un point de vue général, le système CRISPR1-Cas de *S. agalactiae* serait donc un système de défense immunitaire adaptatif mais également, un système de régulation des gènes de la bactérie elle-même, de son « mobilome ». En effet, plus de 40% des spacers ciblent des EGMs présents dans le génome de la bactérie, suggérant que le système régulerait l'intégration de ces EGMs dans le génome, contribuant ainsi à la diversité de la population (Lopez-Sanchez *et al*., 2012).

3.3. Travaux préliminaires

L'étude du locus CRISPR1 a déjà fait l'objet de plusieurs travaux au sein de l'équipe Bactéries et Risque Materno-Fœtal. Ceux-ci ont pu, entre autres, mettre en évidence une grande variabilité de ce locus chez l'espèce *S. agalactiae*, due à l'acquisition, la duplication ou la délétion de spacers. Cette diversité est importante du côté du leader, alors que les spacers proches de la DRT sont plus conservés. Ainsi, il a déjà été montré au laboratoire que la composition en spacers était corrélée aux grands groupes phylogénétiques de cette espèce bactérienne, définis par MLST (Figure 9 ; Lier *et al*., 2015). D'autre part, dans une étude qui visait à étudier la vitesse d'évolutivité du système, il a été montré que le système semblait fonctionnel *in vivo*, puisque certaines souches ont acquis des spacers dérivés de séquences phagiques ou plasmidiques (Beauruelle *et al*., 2017). Ceci a mis en évidence la présence d'interactions phages-bactéries dans leur environnement naturel et leur nécessaire coévolution. Le système CRISPR1-Cas de *S. agalactiae* a également été utilisé pour étudier et comparer la diversité des populations de *S. agalactiae* présentes en portage, chez différentes patientes, ou chez la même patiente (Beauruelle *et al*., 2018). Ces travaux ont également permis de développer une nouvelle méthode de typage, basée sur l'amplification du locus CRISPR1, suivie de restriction enzymatique (CRISPR RFLP). Enfin, l'utilisation du locus CRISPR1 comme outil de typage a été comparée aux méthodes utilisées actuellement et notamment, le MLST, la méthode de référence, afin de déterminer laquelle est la plus discriminante pour différencier les isolats et établir des liens phylogénétiques (Beauruelle *et al*., 2021).

Par ailleurs, lors de ces travaux, il a été observé une diversité moindre du locus CRISPR1 chez les souches hypervirulentes appartenant à la lignée phylogénétique ST17. En effet, la proportion de spacers uniques était significativement plus faible chez ces souches. De plus, celles-ci possèdent significativement moins de spacers en moyenne au sein de leur locus CRISPR1, que les souches des autres lignées (Figure 34 ; Lier *et al*., 2015). Ce nombre limité de spacers pourrait être expliqué par un système CRISPR1-Cas moins ou non fonctionnel, par rapport à celui des souches appartenant aux autres lignées phylogénétiques. Ceci pourrait être lié soit à des différences au niveau de la composition en gènes *cas* associés au locus CRISPR1, soit à des différences dans leurs niveaux d'expression. Si une anomalie de fonctionnement de ce système était retrouvée, cela pourrait constituer, au moins en partie, un élément de compréhension supplémentaire de la virulence particulière des souches de ST17, au sein de l'espèce *S. agalactiae*.



<u>Figure 34 :</u> Nombre de spacers par locus en fonction du complexe clonal (Lier *et al.*, 2015).
Les rectangles rouges représentent la médiane. Les lignes bleues et vertes représentent, respectivement, les valeurs maximales et minimales. Les extrémités basses et hautes des rectangles blancs représentent les premiers et troisièmes quartiles (*p<0,001).

Objectifs

Compte tenu des éléments cités précédemment, l'analyse et la compréhension du fonctionnement des systèmes CRISPR-Cas, dans une espèce bactérienne comme *S. agalactiae*, pathogène majeur dans la période néonatale mais dorénavant, également chez l'adulte, semblent indispensables afin d'appréhender leurs rôles chez cette bactérie.

Au vu des résultats des travaux effectués auparavant sur le locus CRISPR1 de *S. agalactiae*, qui ont montré des particularités pour les souches hypervirulentes de ST17 et des données de la littérature, sur les liens entre virulence et systèmes CRISPR-Cas, l'objectif principal de ce travail, présenté dans la première partie, était d'étudier spécifiquement la capacité de fonctionnement du système CRISPR1-Cas chez les souches de ST17, en la comparant à celle de souches appartenant à d'autres groupes phylogénétiques. Dans un premier temps, nous nous sommes intéressés à la recherche de différences de structure dans les régions promotrices du système CRISPR1-Cas, car une différence d'expression des gènes *cas* pourrait être à l'origine du fonctionnement plus ou moins performant du système. Nous avons notamment utilisé la technique de fusion transcriptionnelle, qui permet de mesurer l'activité d'un promoteur, ainsi que la RT-PCR quantitative pour étudier les ARNs messagers correspondants aux gènes *cas* et mettre en évidence d'éventuelles différences entre les souches, quant à l'expression relative de ces gènes. A partir de ce qui a pu être réalisé pour l'étude d'autres systèmes CRISPR-Cas, nous avons également mis au point des tests de fonctionnalité, permettant d'étudier les capacités d'interférence et d'adaptation du système CRISPR1-Cas.

Dans une seconde partie, notre objectif était d'identifier des conditions dans lesquelles l'expression du système CRISPR1-Cas serait modulée, afin d'identifier les voies de régulation de l'expression de ce système chez *S. agalactiae* et surtout de déterminer si les voies de régulations mises en jeu étaient identiques entre les différentes lignées phylogénétiques. Pour cela, la réponse quantitative de la transcription du gène *cas9* a été étudiée en présence de différents stimuli. Parmi les stimuli sélectionnés, certains ont déjà été impliqués dans la régulation des systèmes CRISPR-Cas. Nous avons également étudié l'expression de la protéine Cas9, préalablement marquée. Les stimuli testés ont été la présence de plasmides pour mimer les « attaques » auxquelles la bactérie peut être soumise, mais également diverses conditions environnementales, telles que des variations de pH et particulièrement un pH acide pour mimer le pH vaginal, une des principales niches écologiques de *S. agalactiae*. Nous avons également étudié la réponse du système suite à l'application d'un stress sur la paroi bactérienne, par l'utilisation d'antibiotiques. Nous avons choisi d'étudier et de suivre la réponse de la protéine Cas9, car c'est un élément essentiel des systèmes CRISPR-Cas de type II-A, tant au niveau de l'intégration de nouveaux spacers, que dans l'étape d'interférence. Ainsi, le suivi de Cas9 au niveau génétique et protéique devrait permettre d'apporter des éléments de compréhension à propos de la régulation du système CRISPR1-Cas chez *S. agalactiae*.

Travail expérimental

Partie 1 : Etude de la fonctionnalité du système CRISPR-Cas de type II-A des souches hypervirulentes de *S. agalactiae*

Dans cette première partie, nous avons cherché à déterminer la fonctionnalité du système CRISPR1-Cas de *S. agalactiae* et surtout à identifier l'existence de différences, dans cette fonctionnalité, entre les lignées phylogénétiques. Nous avons choisi de détailler dans cette partie, la totalité du matériel et méthodes et des résultats obtenus, dont les principaux éléments ont fait l'objet d'une publication en avril 2021 dans le CRISPR Journal. Ainsi, les tableaux et figures légendés en anglais sont ceux présents dans l'article, ceux légendés en français constituent des éléments supplémentaires.

1. Matériel et Méthodes

1.1. Souches étudiées

Pour réaliser la partie expérimentale de ce travail, nous avons sélectionné quatre souches de référence, réparties au sein des lignées phylogénétiques humaines majeures : *S. agalactiae* BM110, NEM316, 2603 V/R et A909, représentant les lignées ST17 (CC17), ST23 (CC23), ST110 (CC19) et ST7 (CC10) respectivement. Ces souches ont été isolées d'infections invasives humaines. Elles ont été caractérisées phénotypiquement et génotypiquement lors de travaux antérieurs (Tableau II ; Glaser *et al.*, 2002 ; Tettelin *et al.*, 2005 ; Tettelin *et al.*, 2002). La composition du locus CRISPR1 est également connue pour chacune de ces souches (Lier *et al.*, 2015) et la séquence complète de leur génome est disponible dans la base de données du NCBI.

Tableau II : Principales caractéristiques des souches de *S. agalactiae* étudiées.
Hémoc. = Hémoculture ; NNé = Nouveau-né ; LCR = Liquide céphalo-rachidien.

Souche	Complexe Clonal	Sequence Type	Sérotype	Origine
BM110	17	17	III	LCR NNé
NEM316	23	23	III	Hémoc. NNé
2603V/R	19	110	V	Humaine
A909	10	7	Ia	Hémoc. NNé

Conservées au laboratoire à -80°C, les souches ont été cultivées en bouillon Todd Hewitt (TH, Becton Dickinson) ou sur géloses TH, ou Trypticase-Soja additionnée de 5% de sang de cheval (géloses TSH, BioMérieux), ou Trypticase-Soja additionnée de 5% de sang de mouton (géloses TSS, BioMérieux), à 37°C en aérobiose. Dix µg/ml d'érythromycine ou 5 µg/ml de chloramphénicol ont été ajoutés au milieu de culture pour maintenir les plasmides dérivés de pG1^{ts}, pTCV-*lac* ou pNZ123.

Des cellules compétentes d'*E. coli* XL1-Blue ont été utilisées pour propager les plasmides, cultivées dans un bouillon LB (MPbio) avec agitation, ou sur géloses LB, supplémentées avec 150 µg/ml d'érythromycine.

1.2. Analyse in silico *: alignement et comparaison nucléotidique*

L'alignement des séquences nucléotidiques des régions promotrices du système CRISPR1-Cas a été réalisé grâce à l'outil BLASTn, disponible sur le site du NCBI. Cet alignement a été effectué sur les génomes de *S. agalactiae* qui étaient disponibles dans la base de données whole-genome shotgun contigs (WGS) en juillet 2020, pour lesquels le groupe phylogénétique avait été précédemment déterminé, et appartenant aux cinq clones les plus prédominants en pathologie humaine. Avec ces paramètres, 331 génomes ont été analysés, parmi lesquels 114 appartenaient au CC1, 60 au CC10, 49 au CC19, 67 au CC23 et 41 au CC17 (Tableau III).

Tableau III : Distribution of the 331 strains on which the alignment of the *cas* promoter region was performed.

Clonal complex	Number of strains	Detailed list of STs
CC1	114	54 ST1 ; 12 ST2 ; 1 ST3 ; 1 ST4 ; 17 ST196 ; 16 ST459 ; 1 ST448 ; 1 ST518 ; 1 ST524 ; 1 ST530 ; 1 ST533 ; 1 ST645 ; 1 ST657 ; 5 new ST ; 1 unknown
CC10	60	3 ST6 ; 8 ST7 ; 15 ST8 ; 2 ST9 ; 8 ST10 ; 22 ST12 ; 1 ST538 ; 1 new ST
CC19	49	30 ST19 ; 7 ST28 ; 3 ST110 ; 1 ST175 ; 1 ST182 ; 1 ST286 ; 1 ST328 ; 1 ST408 ; 1 ST447 ; 1 ST520 ; 2 new ST
CC23	67	48 ST23 ; 6 ST24 ; 2 ST25 ; 9 ST88 ; 1 ST444 ; 1 new ST
CC17	41	41 ST17

1.3. Extraction de l'ADN génomique bactérien

La méthode utilisée pour extraire l'ADN bactérien est une extraction de type phénol/chloroforme, qui repose sur la solubilité différentielle des acides nucléiques et des autres composants bactériens (protéines, lipides) entre deux phases non miscibles (Green et Sambrook, 2017).

Une colonie a été ensemencée dans 10ml de bouillon TH et incubée une nuit à 37°C. Après centrifugation à 7000 rpm pendant 8 minutes à 4°C, le culot a été repris dans 500 µl de tampon de lyse (NaCl 1M, Tris-HCl 10 mM, EDTA 5 mM [pH 8], triton X-100 au 1/200ème). L'ensemble a été transvasé dans un tube contenant 0,4g de billes de verre (Sigma-Aldrich) afin de lyser mécaniquement les bactéries, grâce à l'appareil Fast-prep FP120, en réalisant 3 cycles de 30 secondes à une vitesse de 6 mètres/seconde. Après centrifugation de 10 min à 13000 rpm, le surnageant a été récupéré et additionné de 740 µl d'un mélange Phénol [pH 8] (Sigma-Aldrich) / Chloroforme / Alcool isoamylique (4,898/4,898/0,204 v/v) préparé extemporanément. Une nouvelle centrifugation de 15 minutes à 13000 rpm a été réalisée et la phase aqueuse, contenant l'ADN génomique (ADNg), a été collectée. Après adjonction d'1 ml d'éthanol 100% et 50 µl de NaCl 1 M, la solution a été placée une nuit à -20°C pour précipiter l'ADN. Le lendemain, les tubes ont été centrifugés 15 minutes à 13000 rpm à 4°C et l'éthanol a été éliminé. Ensuite, le culot renfermant l'ADNg a été repris dans 60 µl d'eau Milli-Q stérile, préalablement chauffée à 37°C. Enfin, la concentration en ADN des échantillons a été déterminée par mesure de l'absorbance à 260 nm au NanoDrop lite (Thermo Scientific) et ceux-ci ont été dilués afin d'obtenir une

concentration d'ADN à 50ng/µl. Ce dosage a permis également d'apprécier la pureté des ADNs grâce à la mesure du rapport DO_{260nm}/DO_{280nm}, ce dernier étant supérieur à 1,8 en l'absence de contamination protéique.

1.4. Réactions d'amplification

Les PCR analytiques de routine, comme les PCR de vérification sur colonies, ont été réalisées à l'aide de la One *Taq®* DNA polymerase (New England Biolabs). Pour ce protocole, après introduction d'une colonie dans un microtube à PCR contenant 16,85 µl d'eau Milli-Q, 8,15 µl du mélange réactionnel suivant ont été ajoutés : 1 µl d'un mélange de dNTPs à 10mM, 1 µl des amorces sens et anti-sens à 10mM, 0,15 µl de One *Taq®* DNA polymerase et 5 µl de One *Taq* Standard Reaction Buffer 5X. La PCR a été effectuée selon un protocole comportant une dénaturation initiale de 5 minutes à 94°C, suivie de 30 cycles de 10 secondes à 94°C, 20 secondes à 50°C et 30 à 180 secondes à 68°C, puis une élongation finale de 5 minutes à 68°C, le temps d'élongation étant fonction de la taille du fragment à amplifier.

Les PCR pour les réactions de clonage ou de séquençage ont quant à elles été effectuées avec la Q5® High-Fidelity DNA polymerase (New England Biolabs). L'amplification des différents fragments a été réalisée grâce à un mélange réactionnel qui contenait 2 µl d'un mélange de dNTPs à 10mM, 0,5 µl de Q5® High-Fidelity DNA Polymerase, 1X de Q5® Standard Reaction Buffer, 2,5 µl d'amorce sens et anti-sens et 50ng d'ADNg, dans un volume réactionnel final de 50 µl. Le programme d'amplification comportait une étape de dénaturation initiale de 30 secondes à 98°C, suivie de 30 cycles d'amplification constitués de 10 secondes de dénaturation à 98°C, 20 secondes d'hybridation à une température dépendante des amorces (entre 50°C et 60°C) et 30 à 480 secondes d'élongation à 72°C, puis une élongation finale de 5 à 10 minutes à 72°C. Le temps d'élongation a été adapté à la taille du fragment à amplifier.

Les réactions d'amplification ont été réalisées dans un thermocycleur 2720 Thermal Cycler (Applied Biosystems). La taille des fragments amplifiés a été systématiquement vérifiée par électrophorèse en gel d'agarose 1 % ou 2 %. Les produits amplifiés ont ensuite été purifiés à l'aide du kit NucleoSpin® Gel and PCR clean-up (Macherey Nagel), selon les recommandations fournies par le fabricant, avant d'être dosés au NanoDrop lite.

1.5. Séquençage

Pour les réactions de séquençage, le mélange réactionnel de 10 µl contenait 1 µl de BigDye™ Terminator Mix v.3.1 (Applied Biosystems), 2 µl de tampon de séquençage 5X, 1 µl d'amorce sens ou anti-sens et 50 ng de produit de PCR purifié. Le programme de cette réaction était constitué d'une étape de dénaturation d'1 minute à 96°C, suivie de 30 cycles composés de 20 secondes à 96°C, 15 secondes à 50°C et 4 minutes à 60°C. Les produits ainsi obtenus ont ensuite été purifiés sur colonnes NucleoSEQ® (Macherey-Nagel), selon les recommandations du fournisseur, avant d'être déposés sur une plaque 96 puits pour être analysés par le séquenceur ABI PRISM® (Applied Biosystems).

1.6. Préparation de bactéries compétentes et transformation

Les souches électrocompétentes d'*E. coli* et de *S. agalactiae* ont été produites selon les protocoles précédemment décrits (Dower, Miller, et Ragsdale, 1988 ; Ricci *et al*., 1994). Ces bactéries électrocompétentes ont ensuite été transformées par électroporation, procédé qui a pour but la perméabilisation transitoire de la membrane cytoplasmique sous l'influence d'un champ électrique, afin de permettre l'entrée de matériel plasmidique. Ceci a été réalisé avec l'appareil MicroPulser™ (Bio-Rad), réglé dans les conditions Ec2 (2,5 kV), avec 1 à 2 ug de chaque plasmide. Les transformants ont ensuite été sélectionnés sur des géloses LB-agar (*E. coli*) ou TSH-agar (*S. agalactiae*), en présence des antibiotiques appropriés.

1.7. Extraction de l'ADN plasmidique

Les clones d'*E. coli* contenant les vecteurs d'intérêt ont été cultivés dans 10 ml de bouillon LB additionné d'érythromycine à 150 µg/ml, pendant une nuit à 37°C, sous agitation à 230 rpm. Après centrifugation de cette culture, les plasmides ont été extraits à partir du culot, à l'aide du kit Nucleospin® Plasmid (Macherey-Nagel) selon les recommandations du fabricant, puis dosés au NanoDrop lite.

1.8. Préparation des ARNs totaux de S. agalactiae

Les différentes souches étudiées ont été cultivées en bouillon TH, jusqu'à la phase exponentielle ou la phase stationnaire de croissance, déterminées par la mesure de la DO_{600nm}. Les suspensions bactériennes ont été centrifugées pendant 10 minutes à 7000 rpm à 4°C et les culots ont été congelés à -80°C.

1.8.1. Extraction des ARNs totaux

Une fois encore, l'extraction réalisée est une extraction de type phénol/chloroforme (Lamy *et al*., 2004). Les culots bactériens ont été repris dans 400 µl de solution de resuspension (glucose 20% ; Tris 25 mM pH 7,6 ; EDTA 10 mM) et 60 µl d'EDTA 0,5M. La suspension obtenue a ensuite été transférée dans des tubes contenant 0,4 g de billes de verre et 500 µl de phénol acide (pH 4,5 ; Sigma-Aldrich). Deux passages successifs de 30 secondes à une vitesse de 6 mètres/seconde, au FastPrep, ont permis la lyse mécanique des bactéries. Les tubes ont ensuite été centrifugés 5 minutes à 13000 rpm à 4°C et la phase aqueuse supérieure a été mélangée à 500 µl de Trizol (Sigma-Aldrich), entraînant la dénaturation des protéines et de l'ADN. Après 5 minutes d'incubation à température ambiante, 100 µl d'un mélange chloroforme/alcool isoamylique (24/1, v/v) ont été ajoutés et les tubes ont été laissés 3 minutes à température ambiante avant d'être centrifugés 5 minutes à 13000 rpm à 4°C. La phase aqueuse contenant les ARNs a été récupérée et à nouveau mélangée à 200 µl de chloroforme/alcool isoamylique. Les tubes ont été incubés 5 minutes à température ambiante et centrifugés 5 minutes à 13000 rpm à 4°C. 500 µl d'éthanol 100% ont été ajoutés à la phase aqueuse collectée, puis les tubes ont été placés 2 heures à -80°C, afin de précipiter les ARNs. Après 30 minutes de centrifugation à 13000 rpm à 4°C, le culot a été lavé avec 1 ml d'éthanol à 70%. Une nouvelle centrifugation de 5 minutes à 13000 rpm à 4°C a été réalisée et l'éthanol résiduel a été éliminé. Une fois séchés, les culots ont été repris dans 100 µl d'eau Milli-Q pré-incubée à 37°C. Les ARNs extraits ont été dosés au NanoDrop lite.

1.8.2. Elimination des traces d'ADN génomique résiduel

Afin d'éliminer les traces d'ADNg résiduel, les ARNs extraits ont été traités à la Turbo-DNase™ à l'aide du kit TURBO DNA-*free*™ (Invitrogen™). Pour cela, 10 µg d'ARN ont été traités par 1 µl de Turbo-DNase™ à 2U/µl et 5 µl de tampon (10X Turbo-DNase™ Buffer), dans un volume final de 50 µl. L'ensemble a été incubé 60 minutes à 37°C, puis inactivé par l'addition de 10 µl de réactif d'inactivation. Après 5 minutes à température ambiante, le réactif d'inactivation a été éliminé par 2 minutes de centrifugation à 10 000 rpm. Après un nouveau dosage au NanoDrop lite, la qualité des ARNs traités a été vérifiée par électrophorèse en gel d'agarose à 2%. L'intégrité des ARNs extraits se traduit par la présence de trois bandes correspondant aux ARN ribosomaux 23S, 16S et 5S.

L'absence d'ADNg contaminant a ensuite été contrôlée par une réaction de PCR, cherchant à amplifier le gène de ménage *gyrB*. Pour chaque échantillon, le mélange réactionnel contenait 0,5 µl d'un mélange de dNTPs à 10mM, 0,125 µl de One *Taq*® DNA polymerase, 5 µl de One *Taq* Standard Reaction Buffer 5X, 0,5 µl de chaque amorce (Tableau IV), ainsi que 50 ng d'ARN, pour un volume final de 25 µl. Le programme d'amplification comportait une étape de dénaturation de 30 secondes à 94°C, suivie de 40 cycles d'amplification de 20 secondes à 94°C, 20 secondes à 58°C et 20 secondes à 68°C, ainsi qu'une étape finale d'élongation de 5 minutes à 68°C. L'absence de bande correspondant à une contamination par de l'ADNg a été vérifiée par électrophorèse en gel d'agarose 2%.

Tableau IV : Amorces utilisées pour la PCR de vérification de l'absence de contamination des ARNs par de l'ADNg.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Gène cible	Séquence (5' - 3')	Taille de l'amplicon (pb)
OLM_021 **OLM_022**	F R	*gyrB*	GCTCACATCAGAACTTTACTTTTAACTC TTTAATCTCACTTCCTACTTTGACACC	123

1.9. 5'-RACE PCR

La technique de RACE PCR (Rapid Amplification of cDNA Ends) a été utilisée pour rechercher le site d'initiation de la transcription de l'opéron *cas* chez les souches appartenant au ST17. Ceci a été réalisé à l'aide du kit 5'/3' RACE PCR (2nd generation, Roche Applied Science) et de trois différentes amorces, dessinées spécifiquement (Tableau V). Après croissance des bactéries jusqu'à $DO_{600nm} \approx 0,5$, les ARNs totaux ont été extraits selon le protocole décrit précédemment. Les ADNc ont d'abord été synthétisés à l'aide de l'amorce OPL_028 et de la reverse transcriptase fournie dans le kit. Après purification des ADNc avec le kit NucleoSpin® Gel and PCR clean up, une queue polyA a été ajoutée à l'extrémité 3' des ADNc. Ces ADNc marqués ont ensuite été amplifiés par PCR à l'aide de l'amorce OPL_029 et l'amorce s'hybridant avec la queue polyA (Oligo d(T)-Anchor), fournie dans le kit. Enfin, une PCR nichée a été réalisée grâce aux amorces PCR Anchor et OPL_031. Après purification des produits de PCR, leur séquençage a permis de déterminer un +1 de transcription.

Tableau V : Amorces utilisées pour la réaction de 5'-RACE PCR.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Séquence (5' - 3')[a]	Position
OPL_028	R	GAAGCAGTATTTCCACCATC	+179 pb après l'ATG de *cas9*
Oligo d(T)-Anchor	F	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTV V = A / C / G	
OPL_029	R	GAAGCAGTATTTCCACCATC	+108 pb après l'ATG de *cas9*
PCR Anchor	F	GACCACGCGTATCGATGTCGAC	
OPL_031	R	ATGTCGAGACCTATTGAATATG	+32 pb après l'ATG de *cas9*

1.10. Fusion transcriptionnelle

L'activité du promoteur de l'opéron *cas* a été mesurée par fusion transcriptionnelle, c'est à dire en insérant son promoteur en amont d'un gène rapporteur, lui-même dépourvu de promoteur. Le gène choisi, *lacZ*, code la β-galactosidase. C'est un gène rapporteur très utilisé car la présence de cette enzyme est facilement détectable par l'hydrolyse d'un substrat chromogénique. Il est porté par le plasmide pTCV-*lac*, qui possède également un gène de résistance à l'érythromycine, *ermB*, permettant de sélectionner les clones ayant intégré le vecteur (Figure 35 ; Poyart et Trieu-Cuot, 1997).

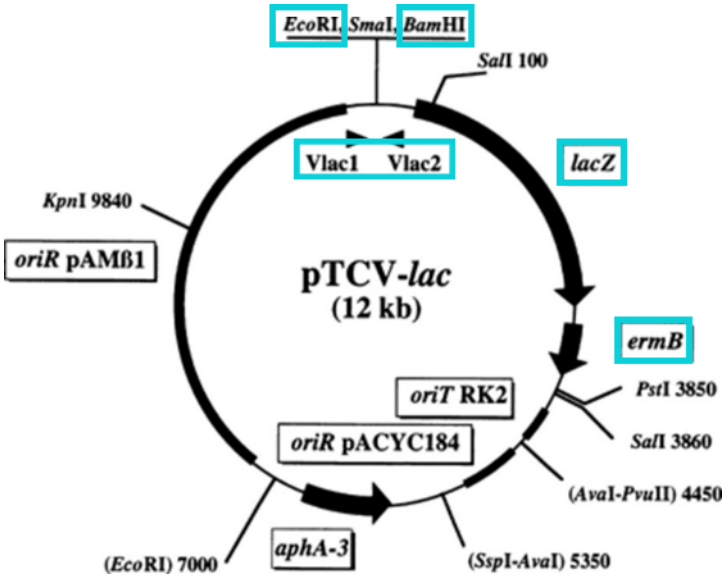


Figure 35 : Organisation du plasmide pTCV-*lac* (d'après Poyart et Trieu-Cuot, 1997).
Les composants d'intérêt ont été encadrés en bleu : *lacZ* : gène codant la β-galactosidase ; *ermB* : gène de résistance à l'érythromycine ; *Eco*RI : site de restriction pour l'enzyme *Eco*RI ; *Bam*HI : site de restriction pour l'enzyme *Bam*HI ; Vlac1 et Vlac2 : positions des amorces permettant l'amplification de l'insert.
Les sites de restriction permettent le clonage d'un fragment en amont du gène rapporteur *lacZ*, dépourvu de promoteur. Le gène de résistance à l'érythromycine (*ermB*) permet la sélection des clones possédant le plasmide.

1.10.1. Préparation de l'insert

Plusieurs couples d'amorces ont été dessinés afin de créer différents inserts contenant soit un seul promoteur, celui de l'opéron *cas* (P_{cas}) ou celui de *srn036* (P_{srn036}), soit les deux. De plus, ces amorces devaient posséder le site de restriction de l'enzyme *Eco*RI pour les amorces sens, ou le site de restriction de l'enzyme *Bam*HI pour les amorces anti-sens, afin de pouvoir insérer ultérieurement le fragment amplifié dans pTCV-*lac* (Tableau VI). Les régions d'intérêt ont été amplifiées par PCR, selon le protocole exposé dans le paragraphe 1.5. Après migration en gel d'agarose 1%, les fragments obtenus ont été purifiés à l'aide du kit NucleoSpin® Gel and PCR clean up et dosés au Nanodrop lite.

Tableau VI : Amorces utilisées pour la construction des fusions transcriptionnelles.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Séquence (5' - 3')[a]	Cible
OPL_016	F	CAAGCCGAATTCTAGCGCTTGCATTTTGTTTG	Région intergénique de *S. agalactiae* NEM316/2603VR
OPL_017	F	GCTCACGAATTCAAGCTGAATACGGACTGGATT	
OPL_018	R	TGGCTTGGATCCAACTTTTCTCCTTTTATTTAA	
OPL_019	F	CAAGCTGAATTCTAGCGCTTGCATTTTGTTTG	Région intergénique de *S. agalactiae* BM110
OPL_020	F	GCTCACGAATTCAAGCTAAATACGGACTGGATT	
OPL_021	R	TGGTTTGGATCCAATTTATCTCCTTTTATTTAA	
Vlac1	F	GTTGAATAACACTTATTCCTATC	pTCV-*lac* : amplification de l'insert
Vlac2	R	CTTCCACAGTAGTTCACCACC	

[a] Les sites de restriction enzymatiques apparaissent soulignés.

1.10.2. Construction des vecteurs

Une double digestion des produits de PCR obtenus précédemment a été réalisée, grâce aux deux sites de restriction présents au niveau des amorces. Le mélange contenait 1 µl de chaque enzyme, *Eco*RI-HF® et *Bam*HI-HF®, à 20000 U/ml (New England Biolabs), 4 µl de tampon CutSmart™ 10X (New England Biolabs) et 30 µl de produit de PCR, pour un volume total de 40 µl. En parallèle, le plasmide pTCV-*lac* a également été digéré par les deux mêmes enzymes. Le tout a été incubé pendant 2h à 37°C, puis purifié avec le kit NucleoSpin® Gel and PCR clean up et dosé au NanoDrop.

La ligature a ensuite été effectuée dans un ratio 3 pour 1, soit une quantité 3 fois plus importante de produit de digestion purifié que de plasmide digéré, en présence d'1 µl de T4 DNA ligase (New England Biolabs). Après

une nuit d'incubation à 16°C, la ligase a été inactivée par chauffage pendant 10 minutes à 65°C. Enfin la solution a été dialysée sur une membrane de 0.025 µm (Merck Millipore) contre de l'eau Milli-Q.

1.10.3. Transformation chez *E. coli*

Dix µl du produit de ligature ont été additionnés à 50 µl d'*E.coli* XL1-Blue compétentes et la totalité a été transférée dans une cuve d'électroporation. Après quelques minutes de contact dans la glace, un champ électrique a été appliqué aux parois de la cuve, permettant une perméabilisation transitoire des membranes bactériennes. Après l'ajout de 800 µl de bouillon LB, la suspension a été incubée 2 heures à 37°C sous agitation à 230 rpm. Après ce temps d'incubation, les bactéries ont été ensemencées sur gélose LB additionnée d'érythromycine à 150 µg/ml et de X-Gal à 60 µg/ml, pour sélectionner les clones possédant le vecteur d'intérêt. Après 24 heures d'incubation à 37°C, la présence du vecteur dans les clones a été vérifiée par PCR sur colonie, grâce aux amorces Vlac1 et Vlac2, pouvant s'hybrider avec le vecteur plasmidique au niveau des bornes d'intégration de l'insert (Figure 35). La taille des produits amplifiés obtenus a été vérifiée par électrophorèse en gel d'agarose 1%.

1.10.4. Séquençage des constructions

Après culture des clones contenant les vecteurs recombinants dans 10 ml de bouillon LB additionnés d'érythromycine à 150 µg/ml, les plasmides ont été extraits avec le kit Nucleospin® Plasmid. Le séquençage des vecteurs a été réalisé selon le même protocole que celui décrit précédemment.

1.10.5. Transformation chez *S. agalactiae*

Un volume X de plasmide, calculé pour que la quantité finale en vecteur soit de 2 µg, a ensuite été mélangé à 70 µl de *S. agalactiae* NEM316 compétentes. L'électroporation a été réalisée de façon similaire à celle décrite chez *E. coli*. Après l'ajout de 800 µl de TH, le tout a été incubé 2 heures à 37°C sans agitation. Puis les bactéries ont été ensemencées sur gélose TSH additionnée d'érythromycine à 10 µg/ml et de X-Gal à 60 µg/ml et incubées à 37°C pendant 24 heures. La mise en évidence de clones contenant le fragment d'intérêt a été réalisée par PCR, directement effectuée sur les colonies, tout comme chez *E. coli*. De plus, pTCV-*lac*, sans insertion en amont de *lacZ*, a également été électroporé chez NEM316, afin de servir de témoin négatif.

1.10.6. Dosage de l'activité β-galactosidase

✓ Extraction des protéines

Les clones possédant le vecteur pTCV-*lac* ont été cultivés dans 20 ml de bouillon TH additionnés d'érythromycine à 10 µg/ml. Dix ml de culture ont été centrifugés à DO_{600nm} ≈ 0,5, 4 minutes à 7 000 rpm. Les culots ont été repris dans 500 µl d'une solution tampon (tampon Z [Na_2HPO_4 60mM, NaH_2PO_4 40mM, KCl 10mM, $MgSO_4$ 1mM, pH7] / Dithiothréitol 1M ; 9,95/0,05 v/v). Cette suspension a ensuite été transvasée dans un tube contenant 0,4g de billes de verre. Deux passages de 40 secondes à une vitesse de 6 mètres/seconde ont été réalisés au FastPrep, en laissant refroidir les tubes dans la glace entre chaque passage. Après 5 minutes de

centrifugation à 6000 rpm, 500 µl de la phase supérieure ont été récupérés, auxquels 0,5 µl de DNase (New England Biolabs) ont été ajoutés. Après 1 heure d'incubation à 37°C, les protéines ont été dosées par la méthode de Bradford, dosage colorimétrique basé sur le changement de couleur du Bleu de Coomassie après liaison aux protéines. Vingt µl d'extrait protéique ont été ajoutés à 980 µl de réactif de Bradford dilué au 1/5 (Bio-Rad). Après 5 minutes d'incubation à l'obscurité, la DO_{595nm} a été mesurée au spectrophotomètre.

✓ Dosage de l'activité β-galactosidase

Les dosages de l'activité β-galactosidase, exprimés en unités arbitraires, ont ensuite été réalisés comme décrit précédemment (Miller, 1972 ; Patron *et al*., 2017). Les tubes témoins, contenant 800 µl de tampon et les tubes tests, contenant 600 µl de tampon et 200 µl d'extrait protéique, ont été préparés avant de démarrer la réaction par l'ajout de 200 µl de 2-orthonitrophényl-β-D-galactoside (ONPG), substrat chromogénique de la β-galactosidase. Les tubes ont été incubés à 28°C. Dès l'apparition d'une coloration jaune, la réaction a été arrêtée par l'addition de 500 µl de Na_2CO_3. La DO_{420nm} a été déterminée au spectrophotomètre, en prenant pour référence un tube témoin arrêté en même temps.

1.11. PCR quantitative en temps réel

1.11.1. Extraction des ARNs totaux

Les souches étudiées ont été cultivées dans 40 ml de bouillon TH. Pour chaque souche, 10 ml de suspension bactérienne ont été prélevés en milieu de phase exponentielle et en phase stationnaire de croissance. Ces phases ont été déterminées préalablement, par mesure de la DO_{600nm} au spectrophotomètre. Les aliquots ont été centrifugés pendant 10 minutes à 7000 rpm à 4°C et les culots ont été congelés à -80°C. L'extraction a ensuite été réalisée comme décrit précédemment.

1.11.2. Synthèse des ADN complémentaires

Après vérification de l'absence de contamination par de l'ADNg, les ARNs ont ensuite été rétro-transcrits en ADN complémentaire (ADNc) en mélangeant 1 µg d'ARN à 1 µl d'iScript™ Reverse Transcriptase (Bio-Rad) et 4µl d'iScript™ Reaction Mix 5X, dans un volume final de 20µl. Les étapes de la réaction étaient : 5 minutes à 25°C, 20 minutes à 46°C et 5 minutes à 85°C. Les ADNc ainsi obtenus ont été dilués afin d'obtenir une concentration finale de 10 ng/µl.

1.11.3. PCR quantitative

Pour déterminer le niveau d'expression du gène *cas9*, une mesure des transcrits de ce gène a été réalisée par qPCR, au cours de deux phases de croissance, chez des souches de *S. agalactiae* appartenant à différentes lignées phylogénétiques. La technique décrite ici repose sur l'utilisation du SYBR Green, marqueur qui s'intercale dans l'ADN double brin, émettant alors une fluorescence proportionnelle à la quantité d'amplicons formés. Le gène *recA*, qui est un gène de ménage codant la recombinase A, a été choisi comme gène de référence pour

normaliser les résultats (Florindo *et al.*, 2012). Par ailleurs, un couple d'amorces permettant l'amplification de *cas9*, chez l'ensemble des souches testées, a été dessiné (Tableau VII).

Tableau VII : Amorces utilisées pour la réaction de PCR quantitative en temps réel.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Gène cible	Séquence (5' - 3')	Taille de l'amplicon (pb)
OLM_319 **OLM_320**	F R	*recA*	ATTCAGGCGCAGTTGATTTAGTT TCAATCTCAGCACGAGGAACA	67
OPL_012 **OPL_013**	F R	*cas9*	CCATATTCAATAGGYCTYGACATC Y=T/C CCATCAAAAAGCAGAGCACCTAT	155

L'efficacité d'amplification de chaque couple d'amorces a été calculée sur une gamme d'ADN génomique (20 ; 2 ; 0,2 et 0,02 ng/µl). Pour cela, 7,5 µl de LightCycler® 480 SYBR Green I Master 2X (Roche) et 0,5 µl d'amorces sens et anti-sens à 10 µM (sauf *cas9*-F à 20 µM) ont été ajoutés à 5 µl de chacune des concentrations d'ADNg testées, pour un volume final de 15 µl. Le protocole de RT-qPCR, réalisé grâce au thermocycleur LightCycler® 480 Instrument II (Roche), comprenait une première étape de dénaturation à 95°C pendant 5 minutes, suivie de 40 cycles de 10 secondes à 95°C, 20 secondes à 58°C et 20 secondes à 72°C. A la fin de cette réaction d'amplification, une courbe de dissociation a été réalisée, par augmentation progressive de la température jusqu'à 95°C. Pour chaque couple d'amorces, une droite standard est établie, représentant les valeurs de Ct obtenues expérimentalement en fonction du log de la concentration en ADNg. La pente de cette droite donne alors l'efficacité des amorces, par la formule Eff = $10^{-1/\text{pente}}$. L'efficacité doit être comprise entre 1,8 et 2, ce qui correspond environ au doublement du nombre de copies d'amplicon à chaque cycle. Par ailleurs, la spécificité de la réaction est vérifiée grâce au calcul de la dérivée première de la courbe de dissociation, où un unique pic doit être visualisé. Il s'agit de la température de fusion (Tm), qui est caractéristique de l'amplicon et qui correspond à la température pour laquelle 50 % de l'ADN double brin est dénaturé.

Les réactions de RT-qPCR pour les échantillons testés ont été réalisées en duplicata, à partir de 20 ng d'ADNc. Le reste du mélange réactionnel et le protocole de RT-qPCR étaient identiques à ceux utilisés précédemment pour le calcul de l'efficacité des amorces. Après calcul de la moyenne des valeurs de Ct pour chaque échantillon, l'expression relative des gènes a été obtenue grâce à la méthode du $2^{\Delta\Delta Ct}$:

$$\Delta Ct = Ct\ (cible) - Ct\ (GR)\ ;$$

$$\Delta\Delta Ct = ((\Delta Ct\ (CR) - \Delta Ct\ (CT))\ ;$$

$$FC = Eff^{\Delta\Delta Ct}$$

Dans un premier temps, les résultats du gène testé (cible) sont normalisés par rapport au gène de référence (GR). Puis le calcul du ΔΔCt permet la normalisation de la condition testée (CT) par rapport à la condition choisie comme référence (CR). Enfin le Fold Change (FC) donne l'expression relative du gène testé, en tenant compte de l'efficacité des amorces (Eff). Chaque détermination a été réalisée avec au moins trois échantillons d'ADNc provenant d'extractions d'ARN biologiquement indépendantes. Les comparaisons statistiques des moyennes ont ensuite été effectuées à l'aide du logiciel Prism (GraphPad).

1.12. Mutagénèse chez S. agalactiae

Afin d'étudier plus précisément l'impact des différences mises en évidence dans la région promotrice de l'opéron *cas* entre les souches des différentes lignées phylogénétiques, mais également le rôle de certains éléments du système CRISPR1-Cas, plusieurs mutants ont été construits. Pour cela nous avons utilisé la méthode de double recombinaison homologue et le vecteur thermosensible pG+host1 (pG1^{ts}, Figure 36).

L'échange allélique entre la copie mutée, portée par le plasmide et la copie sauvage, portée par le chromosome, va s'effectuer grâce à des variations de température (Biswas *et al.*, 1993). Ainsi, la bactérie transformée par pG1^{ts} est d'abord incubée à une température de 30°C, température optimale de réplication du plasmide chez les bactéries à Gram positif. Les clones possédant le vecteur sont ensuite sélectionnés sur gélose TSH additionnée de 10 µg/mL d'érythromycine. Puis, une suspension en bouillon TH + érythromycine à 10 µg/ml est réalisée à partir d'une colonie et incubée plusieurs heures à 30°C, jusqu'à l'apparition d'un trouble. La suspension est alors placée à 38°C, température ne permettant plus la réplication du vecteur. Ce choc thermique favorise la sélection des bactéries ayant intégré le vecteur dans leur chromosome. Les clones ainsi obtenus, sélectionnés sur gélose TSH additionnée de 10 µg/mL d'érythromycine, sont à nouveau cultivés en bouillon TH, à 30°C et sans antibiotique. A partir de cette suspension, plusieurs dilutions au 1/1000ème vont être effectuées, matin et soir, dans 10 ml de bouillon TH. Ceci doit permettre le deuxième événement de recombinaison, aboutissant à la perte du plasmide et à l'excision de la séquence du gène à muter. A partir du 3ème ou 4ème passage, les clones sont repiqués sur gélose TH et TH + érythromycine à 10 µg/ml en parallèle, afin d'identifier les clones ayant perdu le plasmide (sensibles à l'érythromycine). La présence de la délétion ou des mutations désirées a été systématiquement vérifiée par PCR, puis séquençage de l'ADN, selon les protocoles précédemment décrits.

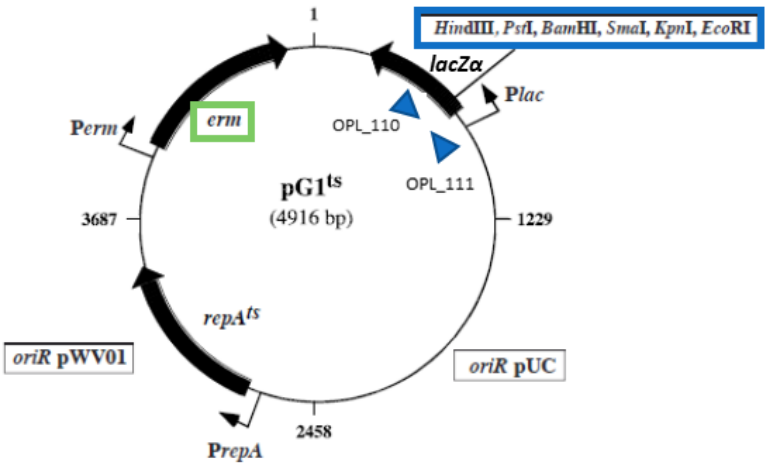


Figure 36 : Organisation du plasmide pG1^{ts}.
Le site de clonage multiple présente les sites de restriction pour les enzymes *Hin*dIII, *Pst*I, *Bam*HI, *Sma*I, *Kpn*I et *Eco*RI (encadré bleu). Les amorces OPL_110 et OPL_111, positionnées de part et d'autre du site de clonage, permettent l'amplification des inserts. Le gène de résistance à l'érythromycine (*erm*, encadré vert) permet la sélection des clones possédant le plasmide.

1.12.1. Construction du mutant BM110Δ*cas9*

Afin de confirmer le rôle de la protéine Cas9 dans les différentes étapes de fonctionnement du système CRISPR1-Cas des souches de ST17, le mutant *S. agalactiae* BM110Δ*cas9* a été construit. Il s'agit d'un mutant non polaire de la souche BM110, délété de la séquence codante de *cas9*, par échange allélique. Il s'agit d'une délétion propre, sans insertion de cassette de résistance. Pour cela des régions d'environ 500 pb, en amont et en aval de *cas9*, ont été amplifiées par PCR à l'aide des amorces OPL_060/OPL_061 et OPL_062/OPL_063 respectivement (Tableau VIII). Ces deux régions ont ensuite été assemblées par PCR overlap à l'aide des amorces externes OPL_060 et OPL_063. Cet assemblage peut avoir lieu grâce à une homologie de séquence apportée par les amorces internes OPL_061 et OPL_062. Le fragment ainsi obtenu a été digéré par les enzymes de restriction *Eco*RI-HF® et *Kpn*I-HF®, avant d'être ligaturé dans le site de clonage de pG1^{ts}, comme décrit dans le paragraphe 1.10.2. Le plasmide recombinant a été électroporé chez *E. coli* afin de l'amplifier et de le purifier, avant de pouvoir l'électroporer chez *S. agalactiae* BM110. Après double recombinaison homologue, la délétion de *cas9* a été confirmée par PCR, suivie du séquençage de l'ADN.

Pour complémenter le mutant de délétion BM110Δ*cas9*, la séquence codante de *cas9* a été amplifiée par PCR à l'aide des amorces OPL_114 et OPL_115 (Tableau VIII), avant d'être insérée dans le vecteur pG1^{ts}, comme expliqué ci-dessus. De même, la complémentation de *cas9* a été vérifiée par PCR et séquençage de l'ADN.

1.12.2. Construction des mutants BM110ΔP$_{srn036}$ et NEM316ΔP$_{srn036}$

Afin d'étudier le rôle du promoteur de *srn036* (P$_{srn036}$) dans la fonctionnalité du système CRISPR1-Cas de *S. agalactiae*, les mutants BM110ΔP$_{srn036}$ et NEM316ΔP$_{srn036}$ ont été construits par mutagénèse dirigée. Ainsi, les amorces OPL_067 et OPL_068 ont été conçues avec des mutations ponctuelles pour remplacer la boîte -10 du promoteur par un site de restriction *Eag*I. Le nucléotide du site d'initiation de la transcription a également été remplacé (A→G, Tableau VIII). Les amorces OPL_066/OPL_067 et OPL_068/OPL_069 ont été utilisées pour amplifier deux fragments d'ADN se chevauchant et contenant les mutations souhaitées. Puis, une PCR supplémentaire avec les oligonucléotides externes OPL_066 et OPL_069 a permis de fusionner les deux fragments. Le fragment amplifié résultant a été inséré dans le site de clonage du vecteur pG1^{ts}. La présence des mutations désirées dans le promoteur de *srn036* a été confirmée par séquençage de l'ADN.

Les compléments ont été réalisés par amplification de la séquence sauvage du promoteur de *srn036* à l'aide des amorces OPL_066 et OPL_069.

1.12.3. Construction du mutant NEM316ΔP$_{cas}$

Le but de cette construction était de recréer le promoteur de la souche de ST17 dans la région intergénique de la souche de ST23 (NEM316). Cette construction a été réalisée selon le même procédé que pour la construction du mutant BM110Δ*cas9*. Ainsi, des fragments d'environ 900 pb en amont et en aval de la zone à échanger ont été amplifiés avec les couples d'amorces OPL_051/OPL_002 et OPL_003/OPL_053 (Tableau VIII). Ces deux régions ont ensuite été fusionnées par PCR overlap à l'aide des amorces externes OPL_051 et OPL_053.

Après digestion du fragment obtenu et clonage dans pG1^{ts}, le plasmide recombinant a été transformé chez *E. coli*, puis chez *S. agalactiae* NEM316. Les mutants ainsi obtenus ont été confirmés par PCR et séquençage de l'ADN.

Tableau VIII : Amorces utilisées pour la construction des différents mutants de *S. agalactiae*.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Fragment amplifié	Mutant	Séquence (5' – 3')[a, b, c]
OPL_060	F	Box amont	BM110Δ*cas9*	CCTATTGAATTCGTAGTATGTGCTTTTCTGTTGG
OPL_061	R			CTAACTTACCTGGTTTGTTCATAATTTATCTCC
OPL_062	F	Box aval		GAACAAACCAGGTAAGTTAGGAGAAGATTGAT
OPL_063	R			CCGTAGGGTACCCCAGCATTTATTGGATTATCATG
OPL_114	F	*cas9*	BM110Δ*cas9::cas9*	CCTATTCCCGGGGTAGTATGTGCTTTTCTGTTGG
OPL_115	R			CCGTAGCTGCAGCCAGCATTTATTGGATTATCATG
OPL_066	F	Box amont	BM110ΔP$_{srn036}$ BM110ΔP$_{srn036::Psrn036}$ NEM316ΔP$_{srn036}$ NEM316ΔP$_{srn036::Psrn036}$	TCATCCGAATTCCTCTGATTTAAATATGCCTAGC
OPL_067	R			CATTTCTCA**C**ATTTGT**CGGCCG**AAAAAATAG
OPL_068	F	Box aval		CTATTTTTT**CGGCCG**ACAAAT**G**TGAGAAATG
OPL_069	R			GTATAACCCGGGCGAGCAGTTCGCTTCAAGC
OPL_051	F	Box amont	NEM316ΔP$_{cas}$	TAGCAAGAATTCATGAATGGACATTTAATGAAGG
OPL_002	R			CATTCTAAAAACATAAGTTTTAAAGCTTTGCT GATTTTTATTTTTTAC
OPL_003	F	Box aval		GCTTTAAAACTTATGTTTTTAGAATGGTG TTTTAAATAAAAG
OPL_053	R			ATATAGGGGCCCCGCTACTGAGAATAAGTCTG
OPL_054	F	Box amont	BM110:: IG$_{NEM316ΔPcas}$	TAGCAAGGGCCCATGAATGGACATTTAATGAAGG
OPL_117	R			CATATAATAATTTTGTTGGAATTATTC
OPL_118	F	IG $_{NEM316ΔPcas}$		CAAAATTATTATATGTTATTTTAGTG
OPL_119	R			GTTTGTTCATAACTTTTCTCCTTTTA TTTAAAACAC
OPL_120	F	Box aval		GAGAAAAGTTATGAACAAACCATATTCAATAGG
OPL_055	R			CTACTGACTGCAGGTCAGCAAATTCATCACCAATC
OPL_051	F	Box amont	NEM316ΔP$_{cas}$:: IG$_{BM110}$	TAGCAAGAATTCATGAATGGACATTTAATGAAGG
OPL_121	R			CATATAATAATTTTGTTAGAACTATTC
OPL_122	F	IG$_{BM110}$		CAAAATTATTATATGTTATTTTAATGTA
OPL_123	R			GCTTATTCATAATTTATCTCCTTTTA TTTAAAATATC
OPL_124	F	Box aval		GAGATAAATTATGAATAAGCCATATTCAATAGG
OPL_053	R			ATATAGGGGCCCCGCTACTGAGAATAAGTCTG
OPL_110	F	pG1^{ts} : vérification de l'insert		CGCCAGGGTTTTCCCAGTCACGAC
OPL_111	R			AGCGGATAACAATTTCACACAGGA

[a] Les sites de restriction enzymatique apparaissent soulignés.
[b] Les régions d'homologie entre les fragments apparaissent en couleur.
[c] Les bases mutées apparaissent en gras.

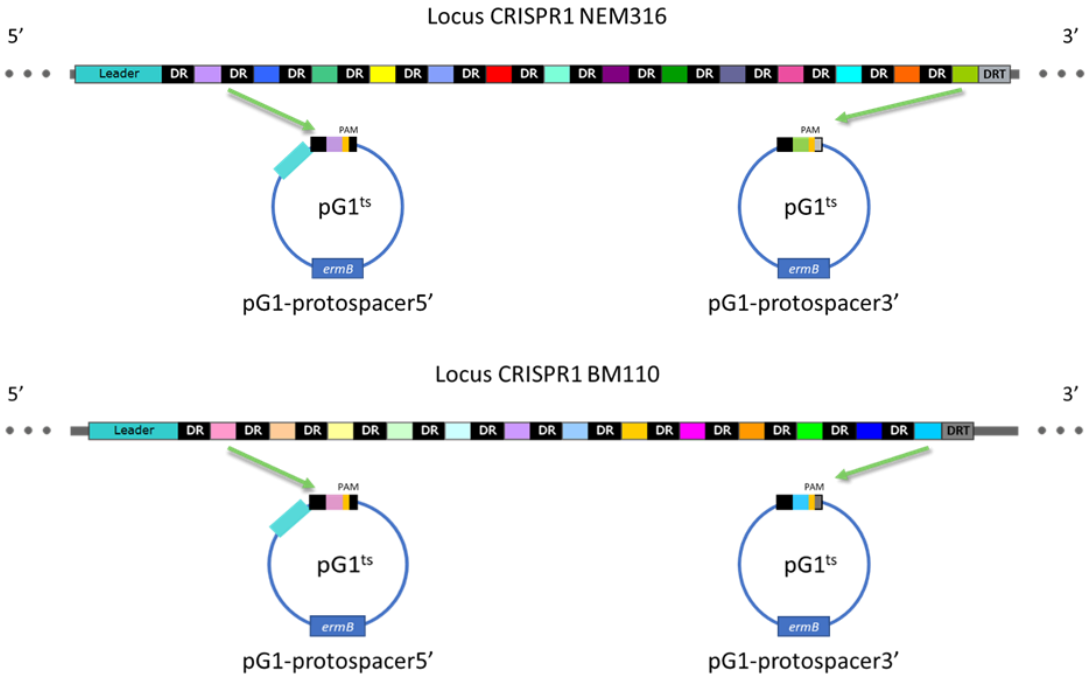
1.12.4. Construction des mutants NEM316ΔP$_{cas}$::IG$_{BM110}$ et BM110::IG$_{NEM316\Delta Pcas}$

Afin d'étudier leurs activités promotrices respectives, nous avons interverti les régions allant du tracrRNA à *cas9* entre les souches BM110 et NEM316ΔP$_{cas}$. Pour ces constructions, il a fallu amplifier 3 fragments possédant des extrémités homologues, afin de pouvoir les assembler deux par deux. Ainsi, les couples d'amorces OPL_054/OPL_117 et OPL_051/OPL_121 ont été utilisés afin d'amplifier un fragment d'environ 870 pb en amont de la région à intervertir, chez BM110 et NEM316 respectivement. Les couples d'amorces OPL_120/OPL_055 et OPL_124/OPL_053 ont quant à eux permis d'amplifier un fragment d'environ 900 pb en aval de la région à intervertir. Enfin, le couple OPL_118/OPL_119 a été utilisé pour amplifier la région intergénique de NEM316ΔP$_{cas}$ (IG$_{NEM316\Delta Pcas}$) et le couple OPL_122/OPL_123, pour amplifier la région intergénique de BM110 (IG$_{BM110}$). Les amorces OPL_118 et OPL_122 possédaient une homologie de séquence avec les amorces OPL_117 et OPL_121 respectivement, permettant aux deux fragments de s'accoler. Il en était de même pour les amorces OPL_119 et OPL_123 avec les amorces OPL_120 et OPL_124 (Tableau VIII). Ainsi, après amplification et purification des trois fragments nécessaires pour chacune des constructions, une PCR overlap, utilisant les amorces externes OPL_O54/OPL_055 et OPL_051/OPL_053, a permis la synthèse des fragments composites. Puis les étapes de digestion enzymatique et de ligature avec pG1^{ts} ont été réalisées comme décrites précédemment. Les plasmides recombinants ont ensuite été transformés chez *E. coli*, puis purifiés et séquencés, avant d'être transformés chez *S. agalactiae* NEM316ΔP$_{cas}$ ou BM110, en fonction de la mutation désirée. Les deux constructions obtenues, NEM316ΔP$_{cas}$::IG$_{BM110}$ et BM110::IG$_{NEM316\Delta Pcas}$ ont été vérifiées par séquençage.

1.13. Expériences d'interférence

1.13.1. Construction des vecteurs

Afin de mimer une attaque par un EGM étranger que la bactérie soit en capacité de reconnaître, des vecteurs porteurs de protospacer ont été construits. Pour cela, le dernier spacer inséré dans le locus CRISPR1 de chacune des souches, c'est-à-dire le plus proche du leader, a été inséré sur le plasmide pG1^{ts} (Figure 37). Nous avons également réalisé des constructions similaires en insérant le spacer le plus ancien, c'est-à-dire celui de l'extrémité 3' du locus CRISPR1. Des amorces permettant l'amplification de ces spacers ont donc été dessinées, auxquelles il a fallu ajouter le PAM, qui est un motif indispensable pour la reconnaissance de l'ADN étranger. Les sites de restriction des enzymes *Eco*RI et *Bam*HI ont également été apportés par les amorces (Tableau IX).



<u>Figure 37 :</u> Représentation schématique de la construction des plasmides vecteurs de protospacer à partir du locus CRISPR1 des souches *S. agalactiae* NEM316 (ST23) et BM110 (ST17).
Pour la construction pG1-protospacer5', le fragment inséré sur le plasmide comprenait le leader, la première DR, le premier spacer et une partie de la deuxième DR. Pour la construction pG1-protospacer3', le fragment inséré comprenait une partie de l'avant-dernière DR, le dernier spacer et une partie de la DR terminale. Pour chacune des constructions, le PAM (en jaune) a été ajouté grâce à l'amorce anti-sens.
Le gène de résistance à l'érythromycine *ermB* permet la sélection des clones possédant le plasmide.

L'amplification du spacer a été réalisée avec la Q5® High Fidelity DNA polymerase, comme décrit précédemment. Après l'amplification de l'insert, les mêmes étapes de digestion et de ligature que celles réalisées pour la construction des vecteurs utilisés en fusion transcriptionnelle ont été suivies. L'électroporation des constructions chez *E. coli* XL1-Blue s'est également déroulée de façon similaire. Après 24 heures d'incubation à 37°C, la présence du vecteur dans les clones a été vérifiée par PCR sur colonie, grâce aux amorces OPL_110 et OPL_111. Une préculture de ces clones a ensuite été réalisée dans 10 ml de bouillon LB additionnés d'érythromycine à 150 µg/ml, pendant une nuit à 37°C sous agitation à 230 rpm, afin de pouvoir extraire les plasmides à l'aide du kit Nucleospin® Plasmid. Les constructions ont ensuite été vérifiées par séquençage.

1.13.2. Transformation chez *S. agalactiae* et dénombrement

Après avoir réalisé des bactéries compétentes des souches NEM316 et BM110 sauvages, ainsi que de leurs souches mutantes respectives, les différentes constructions ont été électroporées selon les mêmes modalités que décrites précédemment, au paragraphe 1.10.5. Le vecteur vide pG1^{ts} a été électroporé dans chacune des souches afin de servir de témoin de transformation. Les plasmides porteurs des protospacer de NEM316 ont été électroporés chez la souche NEM316 sauvage et ses mutants. De même, les plasmides porteurs des protospacer de BM110 ont été électroporés chez les souches BM110 sauvage et mutées. Afin de standardiser le test, la même quantité de vecteur a systématiquement été électroporée, à savoir 1500 ng de plasmide. La présence du gène

ermB sur ce dernier a permis la sélection des transformants par l'ajout d'érythromycine dans le milieu de culture (Figure 37). Après 2h d'incubation à 30°C, 100 µl de chaque suspension ont été ensemencés sur gélose TSH additionnée d'érythromycine à 10 µg/ml. Après 48 h d'incubation à 30°C, le nombre de colonies ayant cultivé sur chacune des géloses a été compté.

Tableau IX : Amorces utilisées pour amplifier les spacers d'intérêt avant leur insertion sur le vecteur pG1^{ts}.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Séquence 5' – 3' [a, b]	Cible	Souche
OPL_042	F	TATGTTGAATTCAAGGATTATTTCTTGATAGGC	Spacer 5'	NEM316 et BM110
OPL_035	R	CGAAAGGATCCAGCTCTAA**CCT**CATACATGTA		NEM316
OPL_038	R	CGAAAGGATCCAGCTCTAA**CCT**CGGCGACTGA		BM110
OPL_183	F	GCTGTGGAATTCCGAATGGTTCCAAAACGATGG	Spacer 3'	NEM316
OPL_184	R	CTAGCAGGATCCCAGCACAGCTTTAA**CCT**ATCAGTAC		
OPL_185	F	CTGTGCGAATTCGAATGGTTCCAAAACATTTGAAC	Spacer 3'	BM110
OPL_186	R	CTAGCAGGATCCCCGCACAGCTCTAA**CCT**ACATATC		
OPL_110	F	CGCCAGGGTTTTCCCAGTCACGAC	pG1^{ts} : vérification de l'insert	
OPL_111	R	AGCGGATAACAATTTCACACAGGA		

[a] Les sites de restriction enzymatiques apparaissent soulignés.
[b] Le PAM apparaît en rouge.

1.14. Acquisition de spacers

Enfin, pour étudier la fonctionnalité de l'étape d'adaptation du système CRISPR1-Cas de *S. agalactiae*, nous avons réalisé des « challenges » avec des plasmides. Ces tests de fonctionnalité s'appuient plus généralement sur des expériences de « bactériophages challenges », mais à ce jour, nous n'avons pas mis en évidence de bactériophages lytiques chez *S. agalactiae*, condition indispensable pour réaliser ces expériences. Cependant, des équipes ont démontré que la pression exercée par les plasmides sur le métabolisme bactérien, même si elle est inférieure à celle exercée par les bactériophages, pouvait être suffisante pour entraîner l'acquisition de nouveaux spacers par les bactéries ciblées (Garneau *et al*., 2010).

Le principe de ces tests repose sur l'électroporation d'un plasmide, porteur d'une cassette de résistance à un antibiotique, dans une bactérie, sauvage ou mutante. Ainsi, différents plasmides ont été testés, tels que pTCV-*lac*, pG1^{ts} et pNZ123, les deux premiers possédant une cassette de résistance à l'érythromycine (*ermB*) et le dernier, une cassette de résistance au chloramphénicol (*cat*). Après avoir électroporé l'un de ces plasmides dans les différentes bactéries cibles, les transformants ont été sélectionnés grâce à la présence d'antibiotique dans le

milieu de culture, avant d'être confirmés par PCR. Un transformant a ensuite été mis en culture dans 10 ml de bouillon TH, sans antibiotique. Après une journée de croissance, 100 µl de cette culture ont servi à inoculer 10 ml d'un nouveau bouillon TH, incubé une nuit. Le lendemain matin, 100 µl de cette culture ont été ensemencés dans 10 ml de bouillon TH. Ceci a été répété deux fois par jour pendant 5 jours, pour un total de 9 inoculations (Figure 38A).

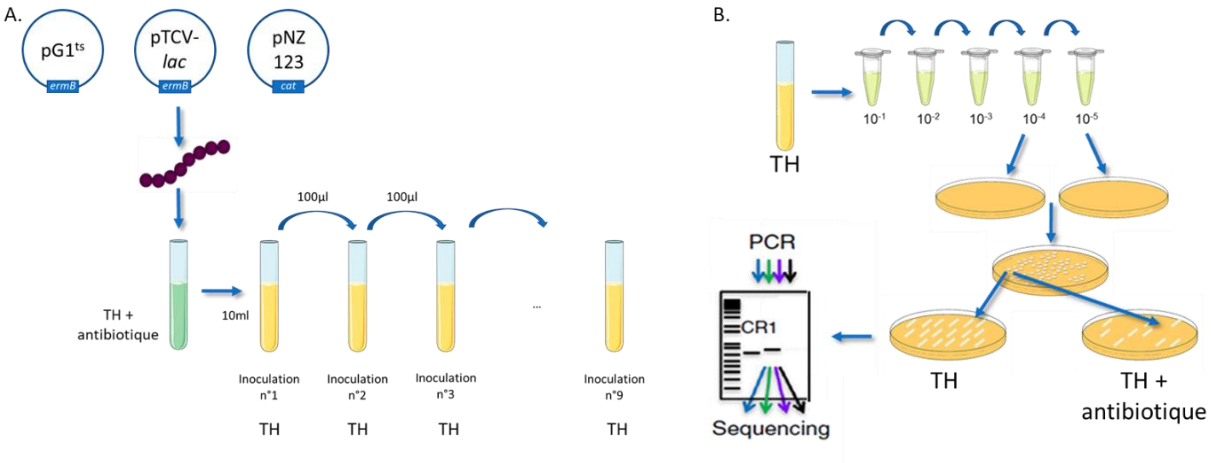


Figure 38 : Principe des expériences d'acquisition (d'après Garneau *et al.*, 2010 ; Hynes *et al.*, 2017).
A. Après électroporation d'un plasmide porteur d'une cassette de résistance chez *S. agalactiae*, un transformant, obtenu en bouillon TH avec antibiotique, a été mis en culture en bouillon TH sans antibiotique (inoculation n°1). Puis 100 µl de cette culture ont été utilisés pour ensemencer 10 ml de TH, ceci deux fois par jour pendant 5 jours.
B. 100 colonies de chaque passage ont ensuite été repiquées sur gélose TH et TH + antibiotique pour rechercher les clones ayant perdu le plasmide. Le locus CRISPR1 de ces clones sensibles a été amplifié par PCR et tous les clones présentant une modification de taille de leur locus CRISPR1 ont été séquencés.

Les dilutions 10^{-4} et 10^{-5} de chacun des passages ont été ensemencées sur gélose TH. Puis, 100 colonies de chaque passage ont été repiquées sur gélose TH et TH + antibiotique en parallèle (érythromycine 10 µg/ml ou chloramphénicol 5 µg/ml en fonction du plasmide électroporé). Les bactéries ayant perdu le plasmide, que ce soit par l'acquisition d'un spacer ou par un autre mécanisme, apparaissaient ainsi sensibles à l'antibiotique et ne cultivaient donc que sur la gélose sans antibiotique. Le locus CRISPR1 de tous les clones sensibles a ensuite été amplifié par PCR, selon les protocoles utilisés auparavant (Figure 38B). Pour cela, plusieurs amorces ont été dessinées, positionnées sur le locus CRISPR1 de façon à permettre de bien différencier la taille du locus sauvage et la taille du locus possédant une unité repeat-spacer supplémentaire (+ 66 pb). Ainsi, l'amorce sens, OPL_100, était positionnée sur le gène *csn2* et les amorces anti-sens sur des spacers spécifiques de chaque souche (OPL_102 sur le 8ème spacer du locus de BM110, OPL_105 sur le 6ème spacer du locus de NEM316 et OPL_107 sur le 7ème spacer du locus d'A909 ; Tableau X). Les modifications de la taille du locus CRISPR1 ont été visualisées par migration des produits de PCR en gel d'agarose 2%. Enfin, le locus CRISPR1 de l'ensemble des clones pour lesquels une modification a été mise en évidence, a été séquencé afin de confirmer la présence d'une nouvelle unité repeat-spacer.

<u>Tableau X :</u> Amorces permettant l'amplification du locus CRISPR1 des différentes souches de *S. agalactiae* d'intérêt.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Séquence 5' – 3'	Position	Souche
OPL_100	F	GAAGACTCTATGATTTACCGC	*csn2*	*S. agalactiae*
OPL_102	R	GGCCGAACAAGTTACACCAG	8ème spacer	BM110
OPL_105	R	CCTTGAGGCGGTATCTCTC	6ème spacer	NEM316
OPL_107	R	GCATAGATTGCTGATTTGGAC	7ème spacer	A909

Un deuxième test, dérivé du précédent mais se déroulant uniquement en milieu liquide a également été utilisé (McKenzie *et al.*, 2019). Les premières étapes étaient identiques à celles décrites précédemment. Après électroporation du plasmide dans la bactérie cible, un transformant a été cultivé dans du TH sans antibiotique pendant 5 jours, selon les mêmes procédés. L'amplification du locus CRISPR1 a ensuite été réalisée directement sur chaque suspension, en utilisant 2 µl de suspension comme matrice d'ADN pour la PCR. L'expérience s'est donc déroulée sans criblage des clones sensibles sur milieu gélosé. Ce test a été utilisé pour étudier l'étape d'acquisition chez les différentes souches mutantes de *S. agalactiae* construites.

2. Résultats

2.1. Etude des régions promotrices du système CRISPR1-Cas de S. agalactiae

Dans un premier temps, nous nous sommes intéressés à l'étude des régions promotrices présentes dans le système CRISPR1-Cas, car si une différence dans la transcription des gènes *cas* était mise en évidence, ceci pourrait expliquer, au moins en partie, des différences dans la fonctionnalité du système entre les lignées phylogénétiques. Habituellement, deux promoteurs au moins sont nécessaires pour la transcription d'un système CRISPR-Cas, le promoteur du locus CRISPR, permettant la transcription puis la formation des crRNAs et le promoteur de l'opéron *cas*, permettant l'expression des protéines Cas (Figure 32).

2.1.1. Démarche d'identification et de localisation des promoteurs

Afin de déterminer la position putative des promoteurs d'intérêt du système CRISPR1-Cas de *S. agalactiae*, nous avons tout d'abord utilisé le logiciel BPROM, qui permet la prédiction des promoteurs bactériens de type σ70 (Solovyev et Salamov, 2011). Cette analyse nous a permis d'identifier un promoteur dans la région du leader, conservé dans l'ensemble des 4 souches de référence de *S. agalactiae*. En ce qui concerne l'opéron *cas*, bien qu'un promoteur ait pu être identifié à l'aide de BPROM, sa position semblait moins évidente et non identique entre les souches. L'analyse par séquençage ARN du génome de la souche *S. agalactiae* NEM316 (ST23) a permis de préciser la position des deux promoteurs présents dans le système CRISPR1-Cas (Rosinski-Chupin *et al.*, 2015). Ainsi les sites d'initiation de la transcription (TSS pour Transcription Start Site) et les régions -10 et -35 de chacun des promoteurs ont pu être déterminés (Figure 39A). Le séquençage ARN a permis d'identifier également deux autres éléments dans la région en amont de l'opéron *cas*. Le premier, le tracrRNA, se situe sur le brin opposé de l'opéron *cas* et possède un promoteur principal et un promoteur alternatif (non représenté). Le second, un petit ARN nommé *srn036*, se situe partiellement en anti-sens du tracrRNA et possède son propre promoteur (Figure 39B).

2.1.2. Comparaison nucléotidique des séquences promotrices

Pour étudier la conservation des régions promotrices entre les lignées phylogénétiques, une comparaison nucléotidique *in silico* a été réalisée. L'outil Blastn et la base de données WGS, nous ont permis de réaliser cette analyse sur un total de 331 génomes de *S. agalactiae*, pour lesquels les appartenances à un ST et un CC étaient disponibles. Nous n'avons pris en compte que les souches appartenant aux 5 CC majeurs humains. Ainsi, le CC1 était représenté par 114 souches, le CC10 par 60 souches, le CC19 par 49 souches, le CC23 par 67 souches et le CC17 par 41 souches (Tableau III). Chacun

des CC était représenté par un ensemble de ST variés, dans le but de refléter au mieux la diversité au sein de l'espèce.

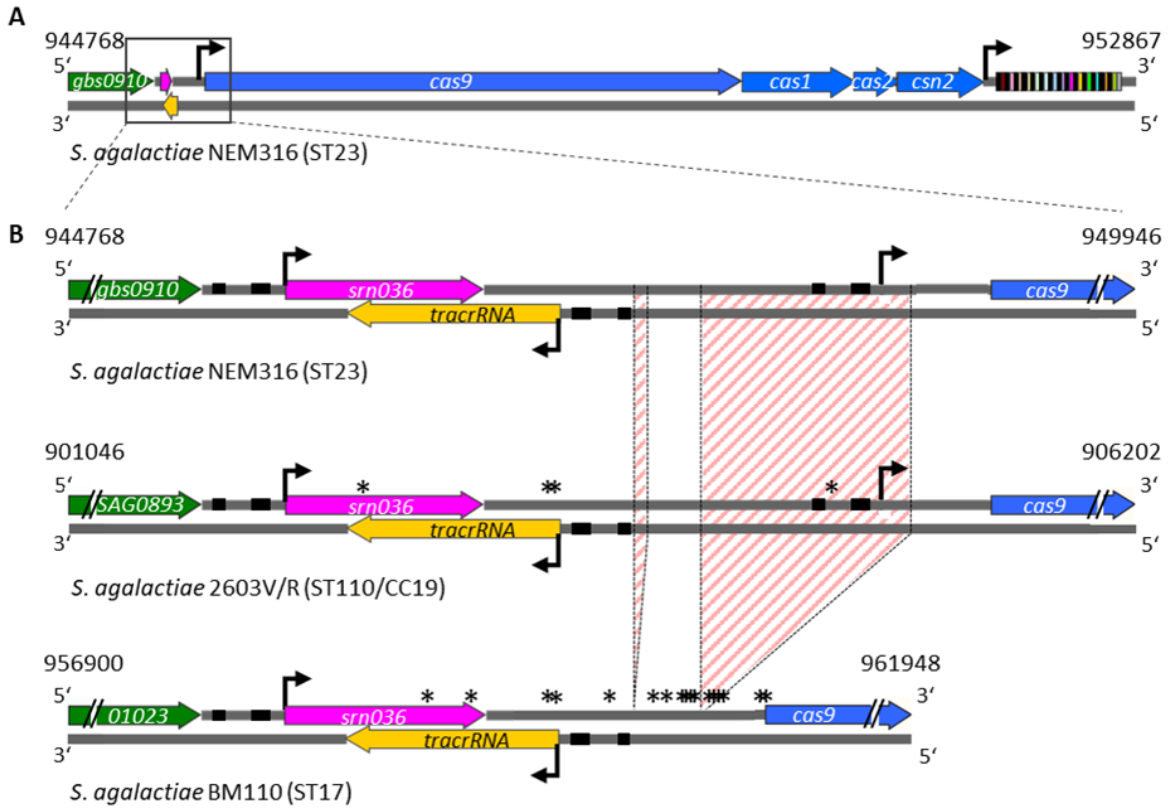


Figure 39 : Deletion of the regular *cas* promoter in the ST17 lineage.
A. Organisation of the CRISPR1-Cas system of *S. agalactiae* NEM316. *cas* genes are shown as blue arrows, the gene upstream the CRISPR1-Cas system is in green, the antisense tracrRNA in yellow and *srn036* in pink. The leader sequence is located downstream the *cas* genes and upstream the CRISPR locus. CRISPR1 array contains 14 repeat-spacer units, represented as black and colored rectangles respectively. The TSSs of the *cas* gene promoter and the CRISPR1 locus promoter are represented with a right black arrow.
B. Close up of the intergenic region upstream of the *cas* operon of three strains representative of three major lineages. Promoter elements have been mapped with RNA-Seq. They are represented in black: the black arrows represent the TSSs, and the black boxes represent the predicted -10 and -35 boxes. Black stars indicate the SNPs observed in strains belonging to CC1, CC10, CC19, and ST17 in comparison to CC23 strains. Red striped boxes indicate the two deleted regions in ST17 strains.

✓ Etude du promoteur du locus CRISPR1

Lors de notre comparaison, nous avons aligné la totalité de la séquence leader, contenant le promoteur du locus CRISPR1. Nous avons constaté que cette séquence apparaît très conservée au sein de l'espèce *S. agalactiae*. Quelques SNPs ont toutefois été observés (Tableau XI), majoritairement à l'extrémité 3' du leader, dont certains avaient été précédemment décrits (Lier *et al.*, 2015). Cependant aucun de ces SNPs n'était localisé dans la séquence promotrice précédemment identifiée. De plus,

aucun ne semble spécifique d'une lignée phylogénétique et donc ne semble pouvoir être à l'origine de différences de fonctionnalité entre les lignées.

Tableau XI : Résumé des différents SNPs observés dans la séquence du leader, leur fréquence relative et leur distribution au sein des CCs.
Les SNPs apparaissent en rouge dans la séquence. Les trois premières séquences représentent l'extrémité 3' du leader, alors que la dernière montre l'extrémité 5'.

Séquence Leader (sens 5' - 3')	Fréquence (%)	Complexe clonal (nombre de souches)
…AAAATTTTCTACGA-	7,5%	CC1 (1) CC10 (22) CC19 (2)
…AAAATTTTCTACGAGG	1,2%	CC1 (1) CC10 (2) CC19 (1)
…AAAATTTTCTATGAG	0,3%	CC1 (1)
TAATATTAGTAAAAGCACAGTAATAACAAGGAATCATCGAAAG…	0,3%	CC17 (1)

✓ Etude du promoteur de l'opéron *cas*

Nous avons ensuite effectué une comparaison des séquences nucléotidiques sur la région promotrice de l'opéron *cas*. Pour cela, un alignement de la totalité de la région intergénique a été réalisé, entre le gène en amont de l'opéron *cas* (*gbs0910*) et *cas9*, premier gène de l'opéron, soit une séquence de 388 pb chez *S. agalactiae* NEM316 (Figure 39B).

La comparaison a d'abord été effectuée sur les souches appartenant au même CC, mais à des STs différents. Ceci n'a pas permis de mettre en évidence de différence dans la séquence de la région alignée entre les souches, exception faite des souches appartenant aux ST28 et ST520, au sein du CC19. Ces dernières présentent un SNP, juste en amont de l'ATG de *cas9* (non représenté). Il apparaît donc que cette séquence est très conservée au sein des CCs, étudiés individuellement.

Puis, nous avons comparé la séquence de cette même région intergénique sur l'ensemble des 331 souches, en prenant la séquence de NEM316 comme référence. Ceci a permis de répartir les souches en deux grands groupes distincts, caractérisés par des séquences différentes de cette région promotrice. D'une part, les souches appartenant aux CC1, CC10, CC19 et CC23 forment un groupe homogène, dont la séquence promotrice est très proche. Quatre SNPs ont été observés et retrouvés chez l'ensemble des souches de CC1, CC10 et CC19 par rapport aux souches de CC23 (Figure 39B et Figure 40). Dans l'autre groupe figurent les souches de ST17, chez lesquelles 16 SNPs ont été identifiés, une délétion de 7 pb et surtout une délétion de 103 pb, portant la région a une longueur de 279 pb

(Figure 39B et Figure 40). Cette dernière délétion entraîne, chez les souches appartenant au ST17, la suppression du promoteur de l'opéron *cas*, identifié chez la souche NEM316.

```
                ************************************************************
NEM316          TAGCGCTTGCATTTTGTTTGCTATTTTTTTATAATACAAATATGAGAAATGTCG
2603V/R         TAGCGCTTGCATTTTGTTTGCTATTTTTTTATAATACAAATATGAGAAATGTCG
BM110           TAGCGCTTGCATTTTGTTTGCTATTTTTTTATAATACAAATATGAGAAATGTCG

                ****************** ***********************************
NEM316          GAATTATAACAATGACATAAAAAAAGCACCGAATCGGTGCTCACTTTTTCAAGC
2603V/R         GAATTATAACAATGACAT-AAAAAAGCACCGAATCGGTGCTCACTTTTTCAAGC
BM110           GAATTATAACAATGACATAAAAAAAGCACCGAATCGGTGCTCACTTTTTCAAGC

                * ***************  ***********************************
NEM316          TGAATACGGACTGGATT-AAAAATCACTGCCTTATTTTAACGTGCTGTGCTGTT
2603V/R         TGAATACGGACTGGATT-AAAAATCACTGCCTTATTTTAACGTGCTGTGCTGTT
BM110           TAAATACGGACTGGATTAAAAAATCACTGCCTTATTTTAACGTGCTGTGCTGTT


                ******* *** ************************ *************
NEM316          TCGAATAGTTCTAACAAAATTATTATATGTTATTTTAGTGTAATAATCAAGTAA
2603V/R         TCGAATAATTCCAACAAAATTATTATATGTTATTTTAGTGTAATAATCAAGTAA
BM110           TCGAATAATTCCAACAAAATTATTATATGTTATTTTAATGTAATAATCAAG---

                    ******* ****** *******  **
NEM316          AAAATAAAAATCAGCAAAGCTTTAAAACTTAAAACATACTTTAGGATATTTTAA
2603V/R         AAAATAAAAATCAGCAAAGCTTTAAAACTTAAAACATACTTTAGGATATTTTAA
BM110           ----TAAAAATTAGCAAAACTTTAAAGTTTG----------------------


NEM316          CAAATAACAGATGAAAGATAAAGATAATGCTTTTAAAACTCAAAAAAATAATAT
2603V/R         CAAATAACAGATGAAAGATAAAGATAATGCTTTTAAAATTCAAAAAAATAATAT
BM110           ------------------------------------------------------

                                          ******* * ** * ************
NEM316          ATAATTAAAGTATGTTTTAAAGGTGTTGTTTTTAGAATGGTGTTTTAAATAAAA
2603V/R         ATAATTAAAGTATGTTTTAAAGGTGTTGTTTTTAGAATGGTGTTTTAAATAAAA
BM110           -------------------------CGTTTTTAAGTGATATTTTAAATAAAA

                ***** ** **
NEM316          GGAGAAAAGTT
2603V/R         GGAGAAAAGTT
BM110           GGAGATAAATT
```

Figure 40 : Nucleotidic sequence of the intergenic region upstream *cas9* of the three *S. agalactiae* strains NEM316 (ST23/CC23), 2603V/R (ST110/CC19) and BM110 (ST17/CC17).
Single nucleotide polymorphisms and deletions are highlighted in red. *cas* gene promoter elements (the transcription start site, and the predicted -10 and -35 boxes) are underlined.

Dans la perspective de relier cette délétion du promoteur *cas* avec un défaut potentiel de fonctionnalité du système CRISPR1-Cas des souches de ST17, qui pourrait également expliquer l'observation précédente d'un plus faible nombre de spacers dans cette lignée (Lier *et al*., 2015), nous

avons commencé par vérifier que les souches de ST17 étudiées par Lier *et al.*, appartenant à la collection du laboratoire, possédaient cette même région promotrice. Pour cela, un couple d'amorces a été spécialement dessiné afin d'amplifier et de séquencer la région intergénique (Tableau XII). Après analyse des échantillons sur le séquenceur ABI PRISM®, les séquences ont été vérifiées à l'aide du logiciel ApE (A Plasmid Editor), puis alignées avec le logiciel MEGA (Tamura *et al.*, 2013). Les résultats obtenus étaient en parfait accord avec ceux de l'analyse *in silico* réalisée sur les souches de la base du NCBI. Ainsi, les 30 souches de ST17 de la collection du laboratoire présentaient les 16 SNPs et les délétions de 7 et 103 pb.

Tableau XII : Amorces utilisées pour le séquençage de la région intergénique.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Gène cible	Séquence (5' - 3')	Taille de l'amplicon (pb)
OPL_014	F	*BM110_01023*	TTTTTCCTCCATTGTTTCATCA	500
OPL_015	R	*cas9*	CTCATCTTCTTAGCAGGTAC	

2.1.3. Recherche d'un promoteur alternatif de l'opéron *cas*

Le promoteur principal de l'opéron *cas*, identifié chez NEM316 (ST23), étant délété chez les souches de ST17, nous avons cherché l'existence éventuelle d'un autre promoteur, plus ou moins spécifique des souches de cette lignée. Pour cela, nous avons utilisé la technique de 5'-RACE PCR. Celle-ci nous a permis de déterminer la position d'un potentiel site d'initiation de la transcription, conservé dans l'ensemble des lignées phylogénétiques mais dont les boîtes -10 (TTATTTTAA) et -35 (GGACTGGAT) n'avaient pas une séquence typique (Figure 41).

```
5' TAGCGCTTGCATTTTGTTTGCTATTTTTTTATAATACAAATATGAGAAATGTCGGAATTATAACA

   ATGACATAAAAAAAGCACCGAATCGGTGCTCACTTTTTCAAGCTAAATACGGACTGGATTAAAAA
                                                          -35
   ATCACTGCCTTATTTTAACGTGCTGTGCTGTTTCGAATAATTCCAACAAAATTATTATATGTTAT
             -10            TSS
   TTTAATGTAATAATCAAGTAAAAATTAGCAAAACTTTAAAGTTTGCGTTTTTAAAGTGATATTTT

   AAATAAAAGGAGATAAATT 3'
            RBS
```

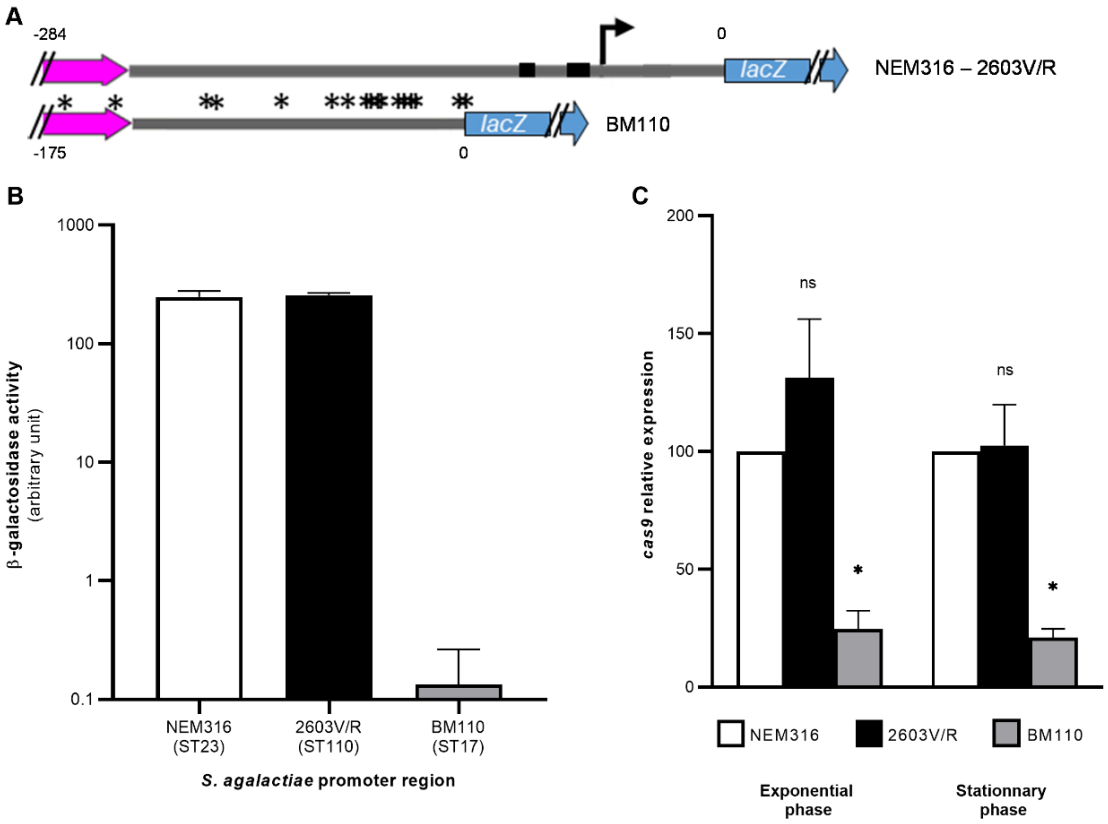
Figure 41 : Identification d'un promoteur alternatif par 5'-RACE PCR chez *S. agalactiae* BM110.
La RACE PCR a permis de déterminer la position de l'éventuel site d'initiation de la transcription (TSS). Les boîtes -10 et -35 ont été localisées en fonction des séquences consensus et de leur position par rapport au TSS. Le RBS apparaît en gras.

2.2. Etude de l'activité promotrice par fusion transcriptionnelle

La comparaison nucléotidique de la région promotrice de l'opéron *cas* a montré une importante délétion chez les souches appartenant au ST17, entraînant la suppression du promoteur *cas* identifié précédemment chez NEM316 (ST23). Afin d'étudier l'impact de cette délétion et les activités respectives des différents promoteurs, des expériences de fusion transcriptionnelle ont été réalisées. Ainsi, la région promotrice de NEM316 (ST23), a été comparée à celle de BM110 (ST17), représentante du ST chez lequel la délétion a été identifiée et dont le système CRISPR1-Cas est composé d'un nombre plus faible de spacers (Lier *et al.*, 2015). De plus, la région promotrice de la souche 2603V/R, (ST110/CC19), a été ajoutée à l'expérience. En effet, d'après les travaux antérieurs, ce sont les souches de ce CC qui possèdent le plus grand nombre de spacers. Il pourrait donc exister une différence d'activité avec les souches des autres lignées, d'autant plus que des SNPs ont été observés dans cette région promotrice par rapport aux souches de CC23.

Des couples d'amorces ont donc été dessinés afin d'amplifier les différents fragments à insérer dans pTCV-*lac*. Ces amorces ont été positionnées de façon à exclure le promoteur de *srn036* du fragment inséré dans le vecteur, afin de s'affranchir de son influence sur les résultats du dosage de l'activité β-galactosidase (Figure 42A). Par contre, le fragment créé contenait bien le promoteur putatif de l'opéron *cas* identifié par RACE-PCR. Les trois constructions ont été électroporées dans le même fond génétique, *S. agalactiae* NEM316. Le vecteur sans insert a également été électroporé afin d'évaluer le bruit de fond et de le retrancher systématiquement à chacune des activités mesurées.

Lorsque la région promotrice de NEM316 (ST23) a été insérée en amont de *lacZ*, l'activité spécifique de la β-galactosidase, sur 3 dosages, s'élevait en moyenne à 250 Unités arbitraires. Pour la région promotrice de 2603V/R (ST110), dans les mêmes conditions, celle-ci était en moyenne de 260 Unités arbitraires (Figure 42B). Ainsi, aucune différence significative n'a pu être mise en évidence entre l'activité des promoteurs de ces deux souches. Les quatre SNPs observés dans la région promotrice des souches appartenant au CC19, ne semblent pas influer sur l'activité du promoteur de l'opéron *cas* et ne peuvent donc pas expliquer les différences existantes au niveau du nombre de spacers composant le locus CRISPR1, chez les souches appartenant à ces deux lignées.



<u>Figure 42 :</u> The large deletion in ST17 strains reduces *cas* promoter activity and *cas9* transcription.
A. Activity of the *cas* promoter was determined by transcriptional fusion between the promoter region of three representative *S. agalactiae* strains and the reporter gene *lacZ*. To reflect *cas* promoter activity only, *srn036* promoter was excluded from the transcriptional fusion. Numbers above and below the diagrams indicate the relative position from the start of *cas9*.
B. Following bacterial growth to mid-exponential phase, β-galactosidase activity was measured. Results are presented as means ± standard deviations (SD) from three independent experiments.
C. Transcriptional profiling of *cas9* in *S. agalactiae* 2603V/R (ST110, black bars) and *S. agalactiae* BM110 (ST17, gray bars) was compared to *cas9* transcription in *S. agalactiae* NEM316 (ST23, white bars) during growth in TH medium to mid-exponential phase (OD_{600nm} = 0.5) and stationary phase. Error bars show the mean ± SD (n = 3 biological replicates). Statistical comparisons of means were made using one-way ANOVA and Tukey's multiple comparison tests. *$p < 0.05$.

Par contre, pour la région promotrice de BM110 (ST17), l'activité spécifique de la β-galactosidase, toujours dosée dans les mêmes conditions, s'élevait en moyenne à 1,6 Unité arbitraire (Figure 42B). Cette activité a été considérée comme négligeable en comparaison avec les valeurs obtenues pour le vecteur vide. Ce résultat semble cohérent avec celui obtenu lors de la comparaison nucléotidique. En effet, pour les souches de ST17, une délétion de 103 pb entraîne la suppression du promoteur principal de l'opéron *cas*, identifié chez NEM316 (ST23). Or en l'absence de promoteur actif, *lacZ* ne peut être transcrit et par conséquent, ne peut être traduit en β-galactosidase. Cette absence d'activité laisse donc supposer qu'il n'existe pas de promoteur alternatif, dans la région clonée, chez les souches de ST17.

2.3. Etude de la transcription de cas9 *par qPCR*

Afin de compléter le résultat précédent, une analyse comparative de la transcription de *cas9* en fonction des lignées phylogénétiques a été réalisée par RT-PCR quantitative. L'analyse a été effectuée sur les trois mêmes souches de référence. Par ailleurs, comme des différences de transcription de *cas9* en fonction de la phase de croissance des bactéries sont possibles, à l'instar de ce qui a pu être observé pour de nombreux gènes, la quantification a été réalisée sur deux phases de croissance distinctes, la phase exponentielle et la phase stationnaire. Ces croissances ont été réalisées en triplicata et les mesures de RT-qPCR ont été effectuées en duplicata (Figure 42C).

Après avoir normalisé les résultats obtenus avec *cas9* par rapport à ceux obtenus avec le gène de ménage *recA* chez chacune des souches et pour chaque phase de croissance, le calcul du ΔΔCt a permis la normalisation des résultats obtenus chez les souches BM110 et 2603V/R par rapport à ceux obtenus chez NEM316. Ensuite, le calcul du Fold Change (FC), qui tient compte de l'efficacité des amorces, déterminée au préalable, a donné les résultats en expression relative de *cas9* chez les souches testées, en comparaison à la souche de référence, dont l'expression du gène étudié a été normalisée à 100 (Figure 42C). Tout d'abord, nous n'avons constaté aucune différence significative de transcription de *cas9* entre la phase exponentielle et la phase stationnaire de croissance, quel que soit le CC auquel les souches sont rattachées. Par ailleurs, en ce qui concerne l'étude en fonction des lignées phylogénétiques, la transcription de *cas9* chez la souche 2603 V/R (ST110/CC19) était similaire à celle de la souche NEM316 (ST23/CC23). Ceci est en accord avec les résultats obtenus précédemment. En effet, ces deux souches possèdent des régions promotrices très proches et aucune différence d'activité du promoteur n'avait été retrouvée. En revanche, chez la souche BM110 (ST17/CC17), malgré la délétion de 103pb et l'absence de détection d'activité promotrice, *cas9* et probablement l'opéron dans son ensemble, semble bel et bien transcrit. Cependant, dans les conditions de croissance testées, le niveau de transcription de *cas9* est significativement plus faible chez la souche de ST17 par rapport aux souches des autres lignées ($p < 0.05$).

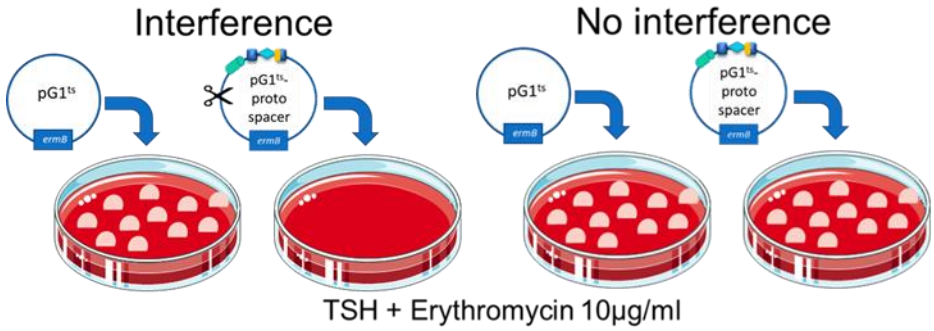
2.4. Démonstration de la fonctionnalité du système

2.4.1. Etude de l'étape d'interférence chez NEM316 et BM110

Pour évaluer la fonctionnalité du système CRISPR1-Cas, un test ciblant l'étape d'interférence a été mis au point. Pour mémoire, l'interférence correspond à la dégradation par la protéine Cas9 d'un acide nucléique étranger déjà rencontré, après hybridation avec un crRNA. En général, ces tests s'appuient sur l'utilisation de bactériophages, mais en l'absence de phage lytique disponible chez *S. agalactiae*, nous nous sommes inspirés de tests déjà décrits chez *S. agalactiae* (Lopez-Sanchez *et al.*, 2012). Ainsi,

nous avons construit des plasmides porteurs d'une séquence complémentaire d'un des spacers du locus CRISPR1 de la souche étudiée. La reconnaissance d'un protospacer nécessite la présence d'un motif spécifique, le PAM (NGG), qui a pu être ajouté à la construction grâce aux amorces.

Les constructions ont ensuite été électroporées chez les souches sauvages, en fonction du protospacer apporté. Le vecteur vide a également été électroporé en parallèle, afin de vérifier la possibilité de transformation chez l'ensemble des souches. Par comparaison du nombre de transformants obtenu avec le plasmide vide et celui obtenu avec le plasmide porteur du protospacer, la fonctionnalité du système a pu être déduite. Lorsque le système est fonctionnel, il doit y avoir reconnaissance du protospacer sur le plasmide et clivage de ce dernier par Cas9, ce qui se traduirait par une efficacité de transformation très faible voire l'absence totale de transformants en culture. En effet, les bactéries qui ne possèdent pas le plasmide ne peuvent pas pousser sur le milieu rendu sélectif par l'ajout d'érythromycine. A l'inverse, dans le cas où le système ne serait pas fonctionnel, le plasmide porteur du protospacer ne serait ni reconnu, ni clivé et le nombre de transformants serait donc similaire à celui obtenu après électroporation du plasmide vide (Figure 43).



<u>Figure 43 :</u> Résultats attendus lors des tests d'interférence.
En cas de reconnaissance du protospacer porté par le plasmide, ce dernier est clivé par la machinerie Cas et aucun transformant n'est obtenu en culture sur milieu sélectif. C'est l'interférence (à gauche). Dans le cas contraire, le plasmide reste intact et un nombre équivalent de transformants est obtenu avec le vecteur vide et le vecteur de protospacer.

Dans un premier temps, nous avons réalisé ces tests avec le spacer le plus récemment acquis par la bactérie, c'est-à-dire celui de l'extrémité 5' du locus CRISPR1, le plus proche du leader et de son promoteur (Figure 37) Ceci nous a permis de confirmer la fonctionnalité du système chez la souche NEM316 et par là même, de valider notre test. En effet, alors que l'électroporation du vecteur vide a permis d'obtenir de nombreux transformants sur milieu sélectif, celle du vecteur de protospacer n'a conduit à aucun transformant, témoignant de la reconnaissance et du clivage du plasmide (Figure 44). Le même résultat a été obtenu avec la souche BM110, aucun transformant n'a pu être dénombré lors de l'électroporation du plasmide porteur du protospacer. Ainsi, l'interférence est toujours possible

chez les souches de ST17, malgré les différences observées dans la région promotrice *cas*. Les résultats concernant la souche BM110Δ*cas9* seront discutés au paragraphe 2.5.

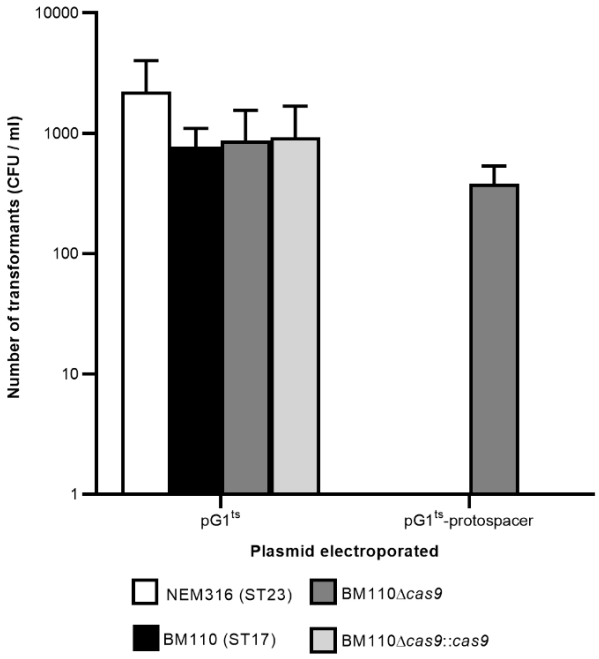


Figure 44 : CRISPR interference in ST17 strain.
An interference assay was performed by cloning the CRISPR1 leader-proximal spacer (and its PAM) of *S. agalactiae* NEM316 (ST23) or BM110 (ST17) into a plasmid (pG1^{ts}-protospacer). This construct was electroporated into *S. agalactiae* wild-type strains as well as in *cas9* deletion derivatives. After electroporation, cells were plated on selective medium and colonies were counted. The experiments were replicated three times. Error bars show the mean ± SD.

Un plasmide n'exerçant pas exactement la même pression sur une bactérie qu'un phage, nous avons également cloné le spacer le plus ancien du locus, c'est-à-dire celui de l'extrémité 3' (Figure 37), sur le vecteur. Le crRNA correspondant serait potentiellement moins exprimé car plus éloigné du promoteur et donc présent en moins grande quantité dans la cellule. Ceci pourrait permettre de mieux mettre en évidence des différences quantitatives de fonctionnement entre les souches, dépendantes de la quantité de protéine Cas9 par exemple. Etonnamment, aucune interférence n'a été détectée lors de ces tests, que ce soit avec la souche NEM316 (ST23) ou la souche BM110 (ST17 ; Figure 45).

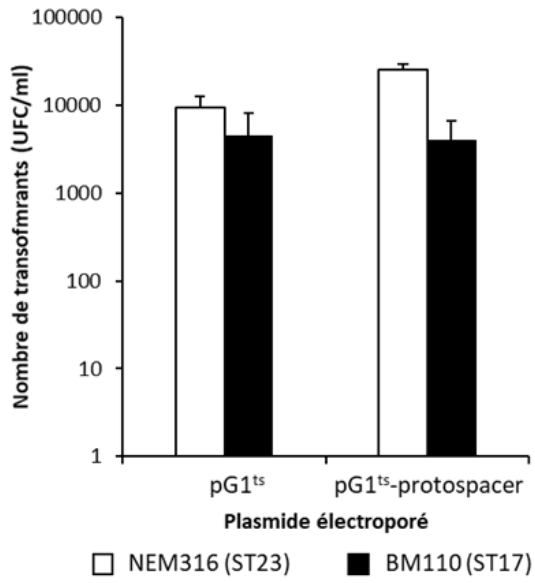


Figure 45 : Perte de la capacité d'interférence face à un protospacer homologue d'un spacer anciennement acquis.
Ces tests d'interférence ont été réalisés selon le même protocole que décrit précédemment, mis à part que le spacer cloné sur pG1^{ts} était celui de l'extrémité terminale 3' du locus CRISPR1 de NEM316 ou BM110 respectivement. Les dénombrements ont été réalisés 3 fois et le graphique présente la moyenne de ces triplicatas.

2.4.2. Etude de l'étape d'adaptation chez BM110 et A909

Par ailleurs, même si le système semble toujours capable de se défendre contre un EGM préalablement rencontré, la capacité d'intégration de nouveaux spacers reste à démontrer. En effet, une étude récente effectuée au sein du laboratoire, n'a montré aucune variation *in vivo* du locus CRISPR1 des souches de ST17 à l'inverse des souches des autres lignées, notamment les souches du ST6 (CC10 ; Beauruelle *et al*., 2017). Cette éventuelle impossibilité d'intégration de nouveaux spacers pourrait constituer un avantage pour ces souches en les rendant plus permissives à l'acquisition d'EGM porteurs de facteurs de virulence. Là encore, les bactériophages étant volontiers utilisés pour tester cette étape du fonctionnement du système CRISPR-Cas, nous avons dû chercher une alternative. Ainsi, nous nous sommes intéressés à l'utilisation de plasmides, comme décrit dans les expériences réalisées par Garneau *et al*. en 2010. Ce travail a d'ailleurs fait l'objet d'une collaboration avec l'équipe du Pr Moineau de l'Université Laval à Québec, au sein de laquelle j'ai pu effectuer un stage de recherche de quasiment 3 mois.

Après avoir électroporé un plasmide porteur d'une cassette de résistance à un antibiotique dans une souche donnée, les transformants vont être cultivés dans un milieu sans antibiotique. Ceci a pour but de ne pas forcer le maintien du plasmide mais au contraire, de favoriser la perte du plasmide, notamment grâce à l'acquisition d'un nouveau spacer au sein du locus CRISPR1, dirigé contre le

plasmide et donc entraînant son clivage par Cas9. Dans un premier temps, nous avons électroporé des plasmides habituellement utilisés chez *S. agalactiae*, pTCV-*lac* et pG1^{ts}, porteurs d'une cassette de résistance à l'érythromycine. Bien qu'un certain nombre de clones sensibles aient pu être mis en évidence avec chacun de ces plasmides et dans les différents fonds génétiques testés, aucune acquisition de spacer n'a pu être détectée. Nous avons alors testé un nouveau plasmide, pNZ123, précédemment utilisé pour démontré l'acquisition de nouveaux spacers chez *S. thermophilus*. Ce test a été réalisé chez *S. agalactiae* BM110 (ST17) et A909 (ST7).

Sur un total de 900 clones testés, 207 et 235 clones sensibles ont été isolés chez BM110 et A909 respectivement. L'extrémité 5' du locus CRISPR1 de l'ensemble de ces clones a ensuite été amplifiée et après migration en gel d'agarose, une modification dans la taille du locus, correspondant le plus souvent à l'addition d'une nouvelle unité repeat-spacer de 66 pb, a été observée pour 92 clones (44%) chez BM110 et 208 clones (88,5%) chez A909 respectivement (Figure 46). La capacité d'acquisition semble donc meilleure chez la souche de ST7 par rapport à celle de ST17.

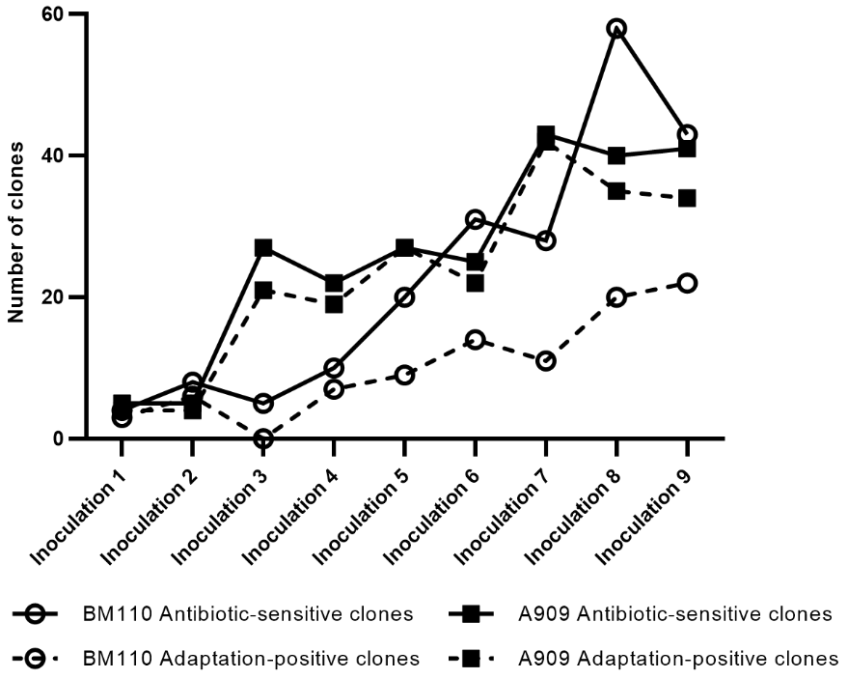
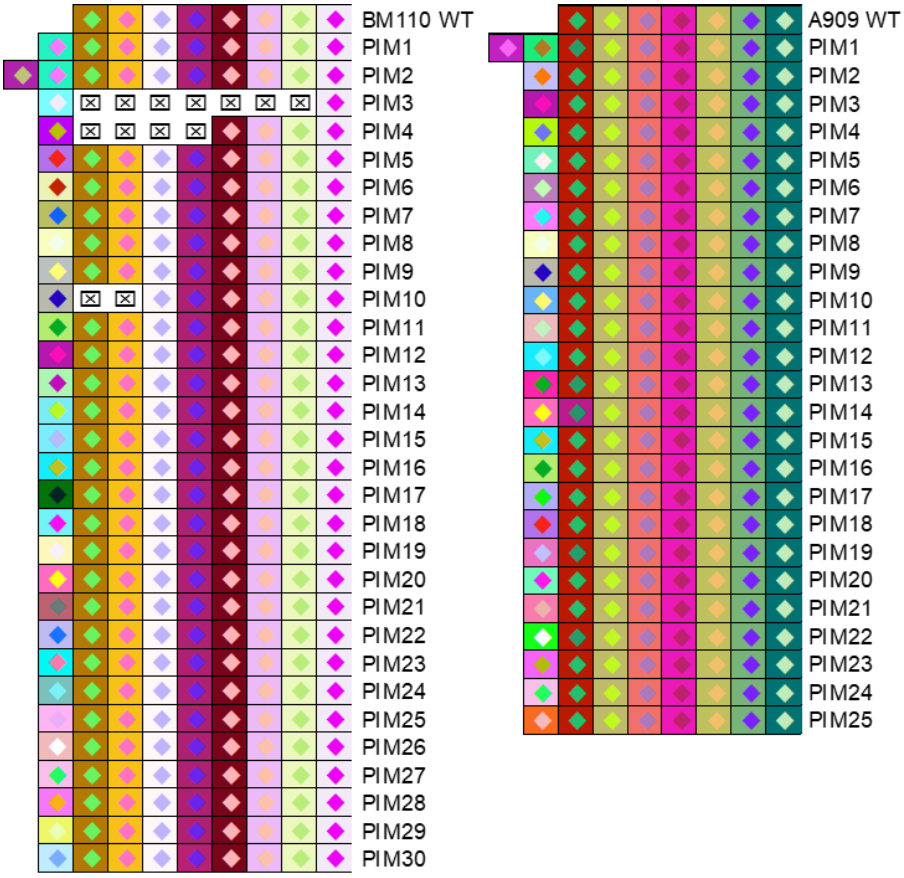


Figure 46 : Adaptation-positive clones over time in response to plasmid challenges.
pNZ123 was electroporated into *S. agalactiae* BM110 (ST17) and A909 (ST7). A confirmed transformant was grown in 10 ml of fresh TH twice a day for five days, for a total of nine inoculations. For each inoculation, 100 colonies were screened for antibiotic sensitivity (solid lines). All the antibiotic-sensitive clones were then tested for new spacer acquisition (adaptation-positive clones, dotted lines).

Le séquençage de ces clones a permis de confirmer l'acquisition d'un nouveau spacer, dont la séquence était parfaitement homologue de celle du plasmide pNZ123. 30 spacers différents ont ainsi été identifiés parmi les 92 clones de BM110 présentant une modification et 26 spacers différents sur les 208 clones d'A909 (Figure 47). La diversité des spacers étaient donc plus importante chez la souche de ST17. Par ailleurs, nous avons remarqué que parmi les clones de BM110 ayant acquis un nouveau spacer, plusieurs ont simultanément perdus entre 2 et 7 spacers plus anciens.



<u>Figure 47 :</u> Représentation schématique des loci CRISPR1 des souches BM110 et A909 après les tests d'acquisition.
Ces représentations ont été réalisées à l'aide d'un outil Excel (P. Horvath), dans lequel les spacers sont convertis en symboles bicolores, en fonction de leur séquence. Les délétions intra-locus (spacers manquants) sont indiquées par des croix.
PIM: Plasmid interfering mutant.

Les protospacers étaient répartis sur l'ensemble du plasmide et systématiquement associés au PAM NGG (Figure 48).

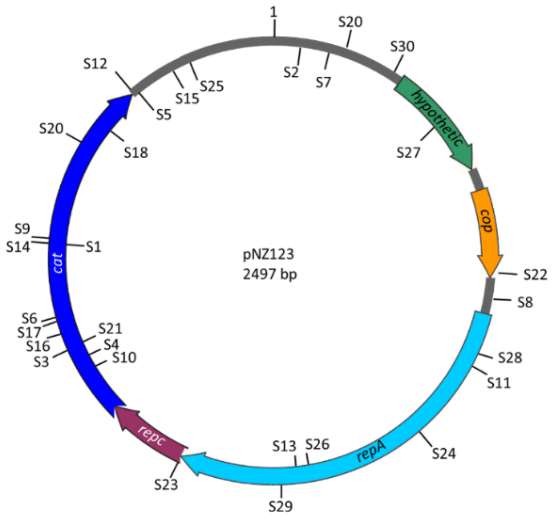


Figure 48 : CRISPR adaptation in ST17 strain.
Schematic representation of 30 pNZ123 proto-spacers, perfectly matching the spacers newly acquired in *S. agalactiae* BM110 CRISPR1 locus. The numbers outside and inside the plasmid map correspond to proto-spacers from the positive and negative strand, respectively.

Enfin, nous avons retransformé un clone ayant acquis un spacer, dont la séquence est homologue de celle de pNZ123, avec pNZ123. Aucun transformant n'a pu être obtenu, confirmant l'efficacité du spacer nouvellement acquis (données non présentées).

2.5. Implication de Cas9 dans la fonctionnalité du système

Alors que nos premiers résultats ont montré une très faible transcription de *cas9*, pouvant suggérer une moindre efficacité du système des souches de ST17, nous avons par la suite démontré qu'il restait pleinement fonctionnel, que ce soit dans sa capacité à se défendre contre un EGM déjà rencontré ou dans sa capacité d'adaptation. C'est pourquoi nous avons construit un mutant délété du gène *cas9* (BM110Δ*cas9*), afin d'étudier son importance dans les événements observés.

Dans un premier temps, nous avons répété les expériences d'interférence avec ce mutant. Ainsi nous avons obtenu un nombre similaire de transformants après électroporation du vecteur vide (pG1^{ts}) ou après électroporation du plasmide porteur du protospacer (Figure 44). Ceci témoigne de la perte de fonctionnalité du système en l'absence de Cas9. Ce résultat a été confirmé avec la souche complémentée BM110Δ*cas9*::*cas9*, pour laquelle la fonctionnalité du système a été restaurée.

Dans un second temps, nous avons souhaité étudier le rôle de Cas9 dans l'étape d'adaptation. Pour cela nous avons utilisé un protocole dérivé de celui précédemment décrit. Cette fois cependant, l'amplification du locus CRISPR1 a été effectuée directement à partir des passages en bouillon TH, sans passer par une étape de détection des clones sensibles sur milieu gélosé. Ceci permet de s'affranchir de l'étape d'interférence et d'étudier uniquement l'étape d'adaptation. L'amplification se faisant sur

l'ensemble de la population présente dans le milieu, l'acquisition d'un nouveau spacer dans le locus par certains clones se traduisait par la présence de deux bandes sur le gel d'électrophorèse. La première bande correspondait à la taille du locus sauvage, amplifiée à partir des clones n'ayant pas acquis de nouveau spacer et l'autre bande correspondait à celle du locus présentant une unité repeat-spacer supplémentaire. C'est ce que nous avons pu observer pour les souches BM110 WT et BM110Δ*cas9*::*cas9* (Figure 49). En revanche aucune acquisition de nouveaux spacers n'a pu être détectée chez la souche BM110Δ*cas9*, sur la totalité des 9 passages effectués lors de cette expérience. Ceci confirme le rôle indispensable de Cas9 dans l'étape d'acquisition également, comme cela avait été évoqué précédemment, bien que beaucoup moins étudié que son implication dans l'étape d'interférence (Heler *et al*., 2015). Cas9 joue ainsi un rôle essentiel à chaque étape du fonctionnement du système CRISPR-Cas, certifiant qu'il s'agit bien de la protéine majeure des systèmes de type II-A.

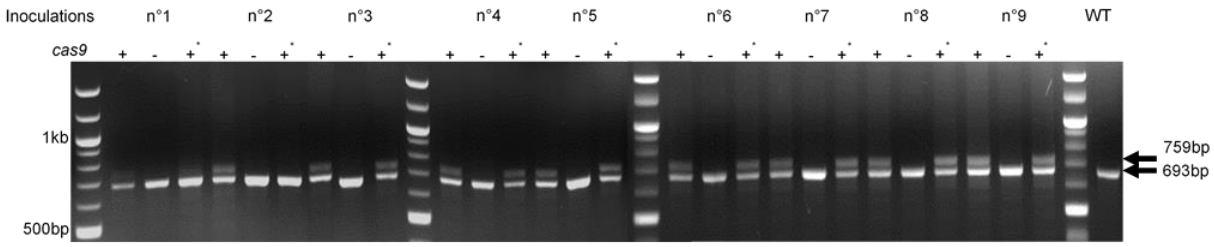


Figure 49 : Cas9 is essential for adaptation.
After amplification of the CRISPR1 locus during the plasmid challenge assay, the acquisition of one new repeat-spacer unit results in a 66-bp increase of the locus size. Gel electrophoresis indicated the coexistence of different-sized CRISPR loci, one corresponding to the unexpanded CRISPR1 array (693 bp), and the other to the population that acquired a new spacer (759 bp), as shown by the two black arrows on the right. For each inoculation, the first lane represents the CRISPR1 locus of *S. agalactiae* BM110 wild-type strain (+), the second lane represents BM110Δ*cas9* (-), and the third lane represents the complemented strain BM110Δ*cas9*::*cas9* (+*). The last lane (WT) corresponds to the native *S. agalactiae* BM110 CRISPR1 array, used as a control.

2.6. Identification du promoteur de l'opéron cas *chez les souches du ST17*

Par la suite, nous avons donc souhaité identifier le promoteur qui pouvait être à l'origine de la transcription des gènes *cas* et donc de la fonctionnalité du système, chez les souches de ST17 spécifiquement. Nous nous sommes naturellement intéressés au promoteur de *srn036*, situé juste en amont de l'opéron *cas* et dans le même sens de transcription.

2.6.1. Détermination de l'activité promotrice de P_{srn036}

Afin de déterminer si le promoteur de *srn036* était capable de transcrire les gènes *cas* situés en aval, nous avons réalisé de nouvelles fusions transcriptionnelles. Nous avons ainsi cloné la totalité de la région intergénique, de P_{srn036} jusqu'à l'ATG de *cas9*, dans pTCV-*lac*, en amont du gène rapporteur *lacZ* (Figure 50A). Les plasmides obtenus ont été électroporés chez *S. agalactiae* NEM316 et les

transformants ont été cultivés jusqu'à DO_{600nm} ≈ 0,5. Après extraction des protéines, l'activité de la β-galactosidase a été déterminée au cours de trois expériences, comme décrit précédemment. Cette activité s'élevait en moyenne à 1200 unités arbitraires pour la région intergénique de NEM316 (ST23) et à 570 unités arbitraires pour la région de BM110 (ST17, Figure 50B). Ceci implique une activité promotrice notable de P_{srn036}, contrairement à ce que nous avions obtenu avec la construction l'excluant (Figure 42A et B). P_{srn036} semble donc capable de permettre la transcription de gènes situés en aval. L'activité promotrice de la souche de ST17 reste toutefois inférieure à celle obtenue avec la souche de ST23, mais cette dernière possède deux promoteurs dans la région clonée, P_{srn036} et P_{cas}, ce qui peut légitimement expliquer ce résultat (Figure 50A).

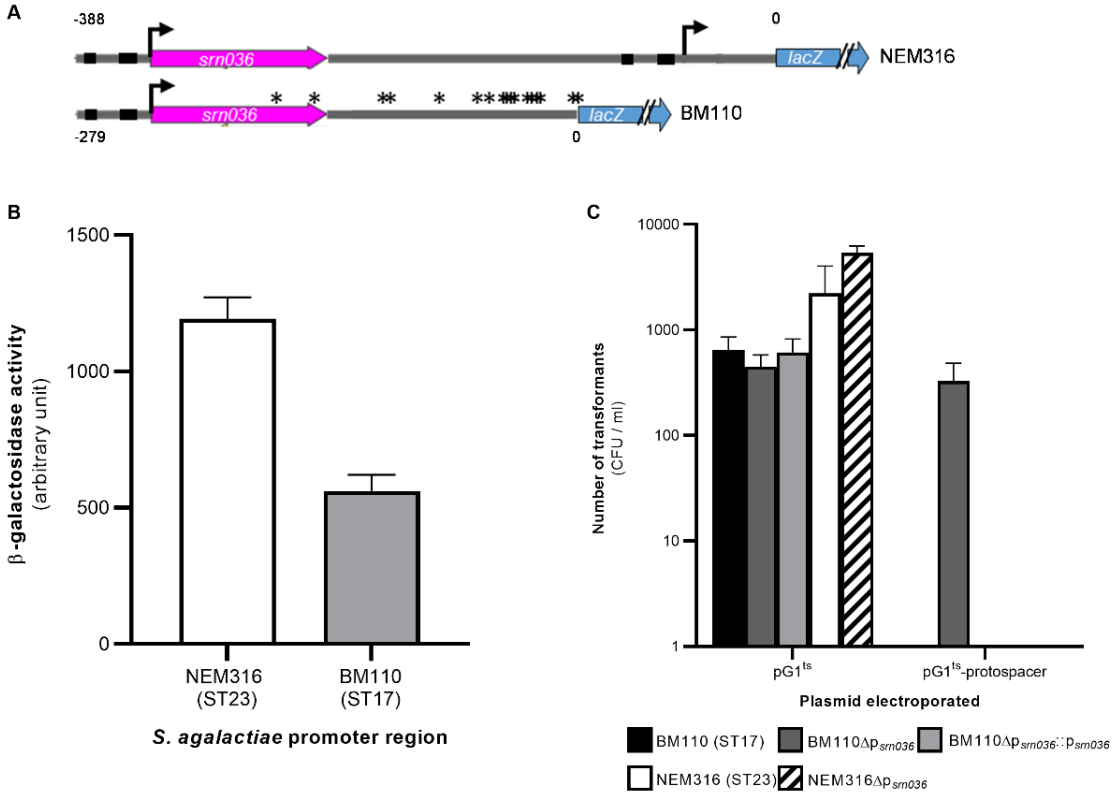


Figure 50 : *srn036* plays a key role in the functionality of the CRISPR1-Cas system in ST17 lineage.
A. *S. agalactiae* BM110 (ST17) and NEM316 (ST23) regions upstream of the *cas* operon, including the *srn036* promoter, were cloned into pTCV-*lacZ*. Numbers above and below the diagrams indicate the relative position from the start of *cas9*.
B. β-galactosidase activity was measured in three independent experiments, as in Figure 42B.
C. CRISPR1-Cas functionality was determined using an interference assay, which was performed as in Figure 44, with an *S. agalactiae* BM110 WT strain (ST17), a derivative with a mutated *srn036* promoter ($BM110\Delta P_{srn036}$), and its complement ($BM110\Delta P_{srn036}::P_{srn036}$). In comparison, the interference assay was also performed with *S. agalactiae* NEM316 (ST23) and its P_{srn036} promoter mutant ($NEM316\Delta P_{srn036}$). The experiments were repeated three times.
B. and **C.** Error bars show the mean ± SD.

2.6.2. Implication de P_{srn036} dans la fonctionnalité du système ST17

Pour étudier précisément le rôle de P_{srn036} dans la transcription des gènes *cas* et donc potentiellement dans la fonctionnalité du système CRISPR1-Cas des souches de ST17, nous avons construit des souches mutées de ce promoteur, où la boîte -10 a été remplacée par un site de restriction d'*Eag*I et la base du TSS a également été modifiée (Figure 51). Ceci doit empêcher la fixation de l'ARN polymérase et la transcription des gènes en aval.

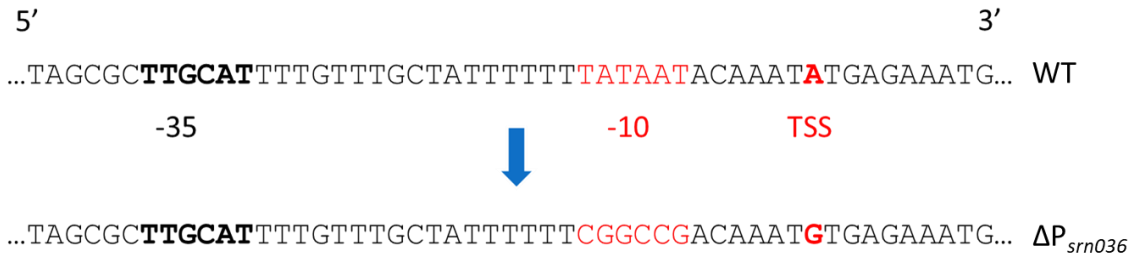


Figure 51 : Représentation de la séquence promotrice de *srn036* chez les souches sauvages (WT) et son pendant chez les souches mutées (ΔP_{srn036}).
La boîte - 10 du promoteur, en rouge a été remplacée par un site de restriction *Eag*I. L'adénine située au site d'initiation de la transcription (TSS), en rouge et en gras, a été remplacée par une guanine. La boîte - 35, en gras, n'a pas été modifiée.

2.6.2.1. Quantification de la transcription de *cas9*

Afin de confirmer l'implication de P_{srn036} dans la transcription des gènes *cas*, nous avons quantifié la transcription de *cas9* par qPCR, selon la méthodologie précédemment décrite, dans les différents fonds génétiques. Ces expériences, réalisées en triplicata, n'ont montré aucune différence significative dans la transcription de *cas9* entre les souches NEM316 WT et NEM316ΔP_{srn036}, semblant indiquer l'absence d'implication de P_{srn036} chez les souches de ST23. En revanche, aucune transcription de *cas9* n'a pu être détectée dans la souche BM110ΔP_{srn036}, semblant confirmer l'essentialité de P_{srn036} pour la transcription des gènes *cas* chez les souches appartenant au ST17.

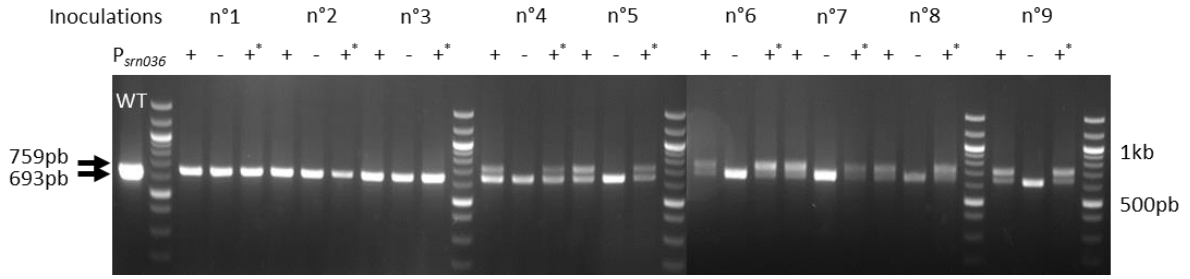
2.6.2.2. Interférence et adaptation

Nous avons ensuite réalisé de nouvelles expériences d'interférence et d'acquisition de spacers chez les mutants ΔP_{srn036}, selon les protocoles précédemment utilisés, afin de déterminer le rôle de P_{srn036} dans la fonctionnalité du système CRISPR1-Cas.

Dans un premier temps, les expériences d'interférence ont montré que le système CRISPR1-Cas de la souche BM110ΔP_{srn036} était devenu non fonctionnel. En effet, nous avons obtenu un nombre de transformants similaire après électroporation du vecteur vide et du plasmide vecteur du protospacer (Figure 50C). Ce résultat a été confirmé grâce à la souche complémentée BM110ΔP_{srn036}::P_{srn036}, pour

laquelle la fonctionnalité du système a été restaurée, puisqu'aucun transformant n'a été obtenu avec le plasmide porteur du protospacer. En ce qui concerne la souche appartenant au ST23, la mutation de P_{srn036} ne semble pas impacter la fonctionnalité du système. En effet, aucun transformant n'a été obtenu en culture après electroporation du plasmide avec le protospacer chez la souche NEM316ΔP_{srn036}.

Dans un second temps, l'étape d'adaptation a été étudiée à l'aide du protocole utilisé pour le mutant BM110Δ*cas9*. Après électroporation des souches BM110 sauvage, BM110ΔP_{srn036} et BM110ΔP_{srn036}::P_{srn036} avec pNZ123, les transformants ont été cultivés en milieu TH sans antibiotique, repiqués matin et soir dans 10 ml de TH, pendant 5 jours. L'amplification du locus CRISPR1 a été réalisée directement sur les suspensions et n'a pas permis de détecter d'acquisition de nouveaux spacers chez la souche mutée pour P_{srn036} (Figure 52). A l'inverse, l'acquisition de nouveaux spacers a pu être visualisée avec la souche sauvage, comme précédemment démontré, mais aussi avec la souche complémentée BM110ΔP_{srn036}::P_{srn036}. La complémentation permettant la restauration du phénotype sauvage, ceci confirme que l'absence d'acquisition de nouveaux spacers est due à la mutation de P_{srn036}.



<u>Figure 52 :</u> Migration électrophorétique des produits de PCR obtenus après amplification du locus CRISPR1 après « challenge » avec pNZ123.
Le premier puits (WT) correspond au locus CRISPR1 natif de *S. agalactiae* BM110, utilisé comme témoin. Sa taille est de 693 pb. L'acquisition d'une nouvelle unité repeat-spacer (+66 pb) se traduit par la présence d'une deuxième bande à 759 pb. Pour chaque inoculation, le premier puits représente la souche sauvage *S. agalactiae* BM110 (+), le deuxième puits représente BM110ΔP_{srn036} (-) et le troisième puits représente la souche complémentée BM110ΔP_{srn036}::P_{srn036} (+*).

Ainsi, le promoteur de *srn036* apparaît essentiel pour la fonctionnalité du système CRISPR1-Cas des souches de ST17, que ce soit au niveau de l'étape d'interférence ou de l'étape d'adaptation, en permettant notamment la transcription des gènes *cas*. En revanche, il ne semble pas jouer de rôle majeur pour les souches appartenant au ST23, qui possèdent par ailleurs un promoteur *cas* pleinement fonctionnel.

2.7. Spécificités du système CRISPR1-Cas des souches hypervirulentes

Afin de déterminer si l'implication de P_{srn036} dans la fonctionnalité du système CRISPR1-Cas est spécifique des souches appartenant à la lignée phylogénétique ST17, nous avons construit un mutant de la souche NEM316 (ST23) délété du promoteur *cas*, NEM316ΔP_{cas}. Celui-ci a été construit en délétant les 103 pb absentes naturellement chez les souches du ST17, conservant uniquement P_{srn036} en amont de l'opéron *cas*. Ainsi, NEM316ΔP_{cas} mime la région promotrice ST17, mais dans un fond génétique de ST23 (Figure 53).

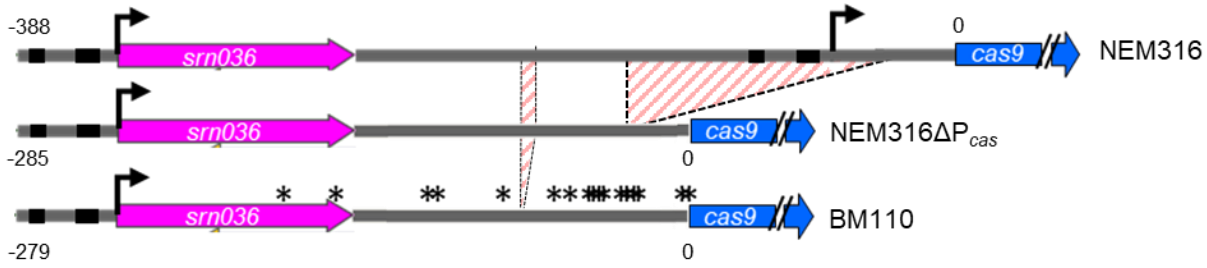


Figure 53 : Représentation graphique du mutant NEM316ΔP_{cas}.
Les régions en amont de l'opéron *cas* de *S. agalactiae* NEM316 (ST23), NEM316ΔP_{cas} et BM110 (ST17), ont été représentées. Les astérisques montrent la position des SNPs chez les souches de ST17. Les parties hachurées indiquent les délétions. Les nombres au-dessus et en dessous des schémas indiquent la position relative depuis le début de *cas9*.

Par la suite nous avons réalisé de nouveaux tests d'interférence avec les souches NEM316 sauvage et NEM316ΔP_{cas}. Pour rappel, un système pleinement fonctionnel se traduit par l'absence de transformant en culture après électroporation du plasmide avec le protospacer (Figure 43), comme nous l'observons pour la souche sauvage (Figure 54). A l'inverse, chez la souche NEM316ΔP_{cas} le système apparaît défectueux car nous avons obtenu un nombre de transformants non négligeable après électroporation du plasmide vecteur du protospacer. Le promoteur P_{cas} semble donc essentiel au bon fonctionnement du système CRISPR1-Cas chez les souches appartenant au ST23. Ce résultat apparaît étonnant compte tenu des résultats obtenus précédemment avec la souche BM110. En effet, bien que délétée de P_{cas}, la souche NEM316ΔP_{cas} possède toujours P_{srn036} en amont de l'opéron *cas*, qui apparait totalement suffisant pour la fonctionnalité du système chez BM110 (ST17). Il semble donc exister des spécificités de fonctionnement entre les lignées phylogénétiques.

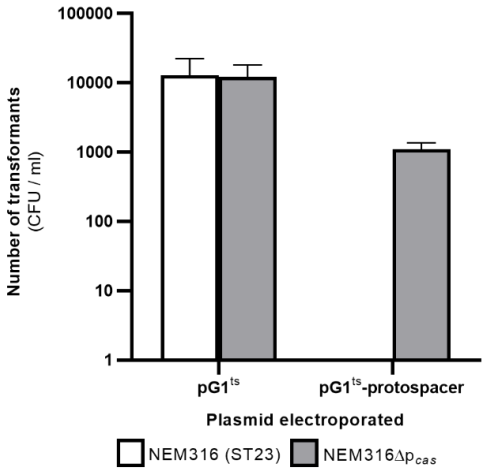
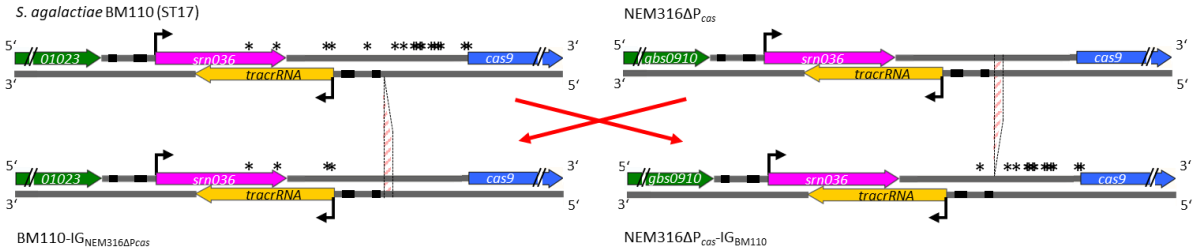


Figure 54 : P_{cas} is essential for CRISPR interference in ST23.
CRISPR1-Cas functionality was determined using an interference assay, performed as previously described, with an *S. agalactiae* NEM316 WT strain (ST23), and a derivative with a deleted *cas* promoter, mimicking the deletion found in ST17 strains (Figure 39B). The experiments were repeated three times. Error bars show the mean ± SD.

2.8. Etude des SNPs et du « terminateur de transcription » de srn036

En plus de la délétion de 103 pb supprimant le P_{cas}, de nombreux SNPs sont présents dans la région intergénique en amont de l'opéron *cas* chez les souches appartenant à la lignée ST17 (Figure 39 et Figure 40). Nous avons alors émis l'hypothèse qu'un ou plusieurs de ces SNPs puissent se situer dans le terminateur de transcription de *srn036*, permettant la poursuite de la transcription jusqu'à *cas9* chez les souches de ST17, par rapport à celles des autres lignées. Ceci pourrait en partie expliquer le constat observé précédemment avec la souche NEM316ΔP_{cas} dont le système CRISPR1-Cas apparaît défectueux. Cependant, le séquençage ARN effectué chez la souche NEM316 n'a pas permis d'identifier avec précision la position du terminateur (Rosinski-Chupin *et al.*, 2015). Nous avons alors construit de nouveaux mutants, en permutant les régions intergéniques allant du tracrRNA à l'ATG de *cas9*, entre les souches BM110 et NEM316ΔP_{cas} (Figure 55). Ces mutants nous ont ainsi permis d'étudier plusieurs SNPs en même temps, afin de vérifier ou non notre hypothèse. Chez BM110::$IG_{NEM316\Delta Pcas}$, 4 SNPs persistaient, situés en amont de la région permutée et chez NEM316ΔP_{cas}::IG_{BM110}, 12 SNPs ont été apportés ainsi qu'une délétion de 6 pb.



<u>Figure 55 :</u> Représentation schématique des régions intergéniques en amont de *cas9* des mutants BM110::IG$_{NEM316\Delta P cas}$ et NEM316ΔP$_{cas}$::IG$_{BM110}$.
Ces mutants ont été construits par échange des régions situées entre le tracrRNA et *cas9* entre *S. agalactiae* BM110 et NEM316ΔP$_{cas}$. Les différents SNPs présents dans ces régions sont représentés par des astérisques. La partie hachurée représente une délétion (BM110) ou une addition (NEM316ΔP$_{cas}$) de 6 pb.

Une fois les mutants vérifiés par séquençage, nous avons effectué de nouveaux tests d'interférence afin d'étudier l'impact de ces mutations sur la fonctionnalité du système CRISPR1-Cas. La première constatation que nous avons pu établir est que les SNPs et la délétion de 6 pb apportés par la région de BM110 ont permis de rétablir la fonctionnalité du système de la souche NEM316ΔP$_{cas}$::IG$_{BM110}$ (Figure 56). En effet, aucun transformant n'a été obtenu après électroporation de pG1^{ts}-protospacer avec cette souche, contrairement aux résultats obtenus chez la souche NEM316ΔP$_{cas}$. Ceci pourrait donc valider notre hypothèse qu'un ou plusieurs des SNPs présents chez les souches du ST17 permettent la transcription depuis P$_{srn036}$ jusqu'à *cas9*. Etonnamment, lorsque nous avons inséré la région intergénique de NEM316ΔP$_{cas}$ dans BM110, nous n'avons obtenu aucun transformant avec le plasmide vecteur du protospacer. Ainsi le système de BM110::IG$_{NEM316\Delta P cas}$ est resté parfaitement fonctionnel, malgré l'absence des SNPs et de la délétion de 6 pb. Ceci suggère l'intervention probable d'un mécanisme régulateur supplémentaire chez les souches appartenant au ST17, pouvant conduire à une surexpression de P$_{srn036}$ par exemple.

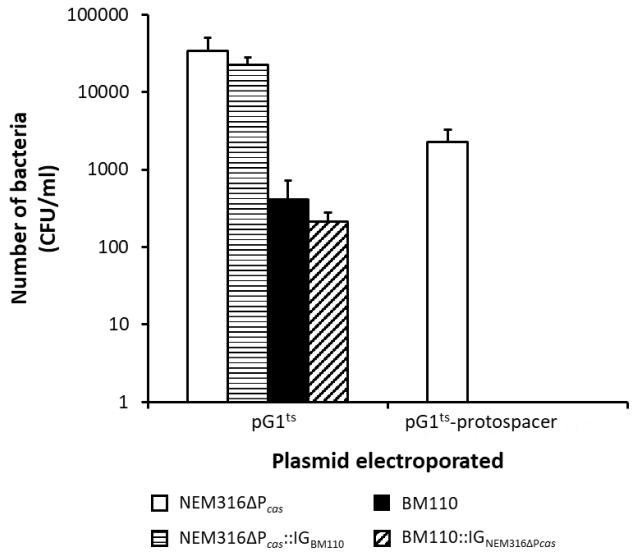


Figure 56 : Détermination de la fonctionnalité du système CRISPR1-Cas de NEM316ΔP*cas*::IG$_{BM110}$ et BM110::IG$_{NEM316ΔPcas}$.
Les tests d'interférence ont été réalisés selon les mêmes protocoles que décrits précédemment. Les dénombrements ont été réalisés 3 fois et le graphique présente la moyenne de ces triplicatas.

2.9. Valorisation

Les principaux résultats présentés dans cette première partie ont été publiés dans le CRISPR Journal en avril 2021.

PUBLICATION

Functional study of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* hypervirulent strains

Adeline Pastuszka, Clémence Beauruelle, Emilie Camiade, Geneviève M. Rousseau, Sylvain Moineau, Laurent Mereghetti, Philippe Horvath, and Philippe Lanotte

CRISPR J. 2021 ; 4:233–42

PMID : 33876956

DOI: 10.1089/crispr.2020.0145

The CRISPR Journal
Volume 4, Number 2, 2021
© Mary Ann Liebert, Inc.
DOI: 10.1089/crispr.2020.0145

The CRISPR Journal

RESEARCH ARTICLE

Functional Study of the Type II-A CRISPR-Cas System of *Streptococcus agalactiae* Hypervirulent Strains

Adeline Pastuszka,[1,2,†] Clémence Beauruelle,[3,4] Emilie Camiade,[1] Geneviève M. Rousseau,[5] Sylvain Moineau,[5,6,‡] Laurent Mereghetti,[1,2] Philippe Horvath,[7] and Philippe Lanotte[1,2,*]

Abstract

Nearly all strains of *Streptococcus agalactiae*, the leading cause of invasive infections in neonates, encode a type II-A clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system. Interestingly, *S. agalactiae* strains belonging to the hypervirulent Sequence Type 17 (ST17) contain significantly fewer spacers in their CRISPR locus than other lineages, which could be the result of a less functional CRISPR-Cas system. Here, we revealed one large deletion in the ST17 *cas* promoter region and we evaluated its impact on the transcription of *cas* genes as well as the functionalities of the CRISPR-Cas system. We demonstrated that Cas9 interference is functional and that the CRISPR-Cas system of ST17 strains can still acquire new spacers, despite the absence of a regular *cas* promoter. We demonstrated that a promoter sequence upstream of *srn036,* a small RNA partially overlapping the antisense tracrRNA, is responsible for the ST17 CRISPR-Cas adaptation and interference activities.

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

Introduction

Streptococcus agalactiae was initially described as an animal pathogen that causes bovine mastitis.[1] Since the 1960s, *S. agalactiae* has emerged in human pathology and has quickly become one of the main causes of maternal and neonatal infections.[2–6] This bacterium is also an emerging pathogen in immunocompromised nonpregnant adults and elderly people, causing a broad spectrum of invasive infections.[7,8] However, in almost 30% of healthy people, *S. agalactiae* resides in the commensal microbiota of the lower digestive and genitourinary tracts.[9]

Research into *S. agalactiae* populations has highlighted that the infection risk associated with different strains is varied. As such, diverse typing methods have been applied to distinguish *S. agalactiae* isolates. Multilocus sequence typing (MLST), the current reference method, is based on the sequencing of seven housekeeping genes.[10] Closely related STs cluster into well-defined clonal complexes (CCs), characterized by differences in host tropism and pathogenicity.[10] The strains that have been isolated in human infections belong mostly to CC1, CC10, CC17, CC19, and CC23.[11] Among those, CC17 is a highly homogenous group, mainly composed of Sequence Type 17 (ST17) isolates, which are overrepresented in invasive neonatal infections and are thus considered hypervirulent.[12]

This diversity among *S. agalactiae* strains can be partly explained by genomic rearrangements that occur during the evolution and adaptation of the species.[13,14] The analysis of *S. agalactiae* sequenced genomes has revealed that the core genome represents ~80% of the gene content. Several genes within this core genome are interspaced with multiple genomic islands (GIs) that are specific to some strains.[15] These GIs generally contain sequences derived from phages or other mobile genetic elements (MGEs) that were acquired by horizontal transfer. MGEs frequently carry virulence factors that play a role in the adaptation of pathogens to their hosts.[16] However, they are also a threat to bacterial survival, forcing bacteria to develop various defense mechanisms to maintain the balance

[1]*ISP, Université de Tours, INRAE, Tours, France;* [2]*Service de Bactériologie-Virologie-Hygiène Hospitalière, CHRU de Tours, Tours, France;* [3]*Département de Bactériologie-Virologie, Hygiène Hospitalière et Parasitologie-Mycologie, Centre Hospitalier Régional Universitaire (CHRU) de Brest, Brest, France;* [4]*Univ Brest, Inserm, EFS, UMR 1078, GGB, Brest, France;* [5]*Département de Biochimie, de Microbiologie, et de Bio-informatique, Faculté des Sciences et de Génie, Groupe de Recherche en Écologie Buccale, Faculté de Médecine Dentaire, Université Laval, Québec City, Canada;* [6]*Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Québec City, Canada; and* [7]*DuPont Nutrition & Biosciences, Dangé-Saint-Romain, France.*
[†]ORCID ID (https://orcid.org/0000-0002-2807-4699).
[‡]ORCID ID (https://orcid.org/0000-0002-2832-5101).

*Address correspondence to: *Philippe Lanotte, Service de Bactériologie-Virologie Hospitalière, CHRU de Tours, 2 Boulevard Tonnellé, F-37044 Tours Cedex, France,* E-mail: philippe.lanotte@univ-tours.fr

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

between beneficial and adverse effects of these MGEs.[17] Recent research has revealed a defense role for the clustered regularly interspaced short palindromic repeats (CRISPR) and their associated proteins (Cas) systems.[18,19]

Two CRISPR-Cas systems have been identified in *S. agalactiae*. According to the latest classification,[20] one of the identified systems (CRISPR1) belongs to type II-A, and is ubiquitous, highly polymorphic, and dynamic. However the second system (CRISPR2), belonging to type I-C, is rare and often incomplete, with seemingly little to no activity.[21] Type II-A systems are among the simplest ones as they contain only four *cas* genes, namely *cas9*, *cas1*, *cas2*, and *csn2*. They are also characterized by an additional element, the tracrRNA. This small RNA, generally transcribed in opposite orientation to the *cas* operon, is partially complementary to the repeat sequence and is essential for the expression and interference steps of CRISPR-Cas immunity.[22] Previous studies have revealed a strong correlation between the spacer content of the CRISPR1 locus and *S. agalactiae* phylogenetic lineages, determined using the MLST method.[21,23] Examination of the *S. agalactiae* CRISPR1-Cas system has also highlighted differences between lineages. For example, strains belonging to the CC17 group contain significantly fewer spacers per locus than other CCs.[23]

Our objective was to determine whether the differences observed in spacer content between *S. agalactiae* lineages could result from differences in the functionality of their CRISPR1-Cas systems. In this respect, we first investigated these differences through *in silico* analysis; we then tested the adaptation and interference stages, which are characteristics of the system's functionality, on representative strains of various lineages.

Materials and Methods

S. agalactiae BM110, NEM316, 2603V/R, and A909 were selected as representative strains of the ST17 (CC17), ST23 (CC23), ST110 (CC19), and ST7 (CC10) lineages, respectively. These strains were isolated from humans with invasive infections and are well characterized, including genome sequencing.[15,24,25]

Nucleotide sequence alignment of promoter regions was achieved using the NCBI BLASTn web server with the default parameter values. This alignment was performed on *S. agalactiae* genomes that were available in the whole-genome sequence database (as of July 2020), and comprised only the most predominant clones in human pathology.

Plasmid pTCV-*lacZ*[26] was used to construct transcriptional fusions between the *lacZ* reporter gene and upstream regions of the *cas* operon. Regions of interest were amplified using Q5® High-Fidelity DNA polymerase (Supplementary Table S1; New England Biolabs) before cloning into the vector. All inserts were confirmed by DNA sequencing and the plasmids obtained were electroporated into *S. agalactiae* NEM316. A confirmed transformant was cultivated in a selective Todd Hewitt (TH) broth until OD_{600nm} reached 0.5. Assays of β-galactosidase were then performed as previously described.[27,28]

To study *cas9* transcription, *S. agalactiae* strains were grown in triplicate to midlog and stationary phases, and total RNA was extracted with a phenol/TRIzol-based purification method as previously described.[29] Quantitative reverse transcription PCR (RT-qPCR) was performed using the LightCycler® 480 SYBR Green I mix, with the associated instrument and software (Roche). Fold changes in transcript abundance were calculated using $\Delta\Delta$Ct, where *cas9* transcript levels were normalized to those of the *S. agalactiae* housekeeping gene *recA*.[30]

Construction of the *cas9* deletion mutant (BM110Δ*cas9*) was performed as previously described with a thermosensitive shuttle vector pG+host1 ($pG1^{ts}$).[28,31] The primers used to construct this mutant are described in Supplementary Table S1. To complement this mutant, the entire coding sequence of *cas9* was PCR-amplified and inserted into $pG1^{ts}$. After allelic exchange, the deletion or the complementation of *cas9* was confirmed using PCR and DNA sequencing. Site-specific mutagenesis was performed to obtain the *srn036* promoter mutations. Primers were designed with point mutations to replace the −10 box with an *Eag*I restriction site. The nucleotide at the transcription start site was also replaced (Supplementary Table S1). After insertion into the cloning site of the $pG1^{ts}$ vector, the presence of mutations in the *srn036* promoter and the absence of secondary mutations were confirmed by DNA sequencing.

The interference assay was based on cloning the last spacer at the 5′ end of the CRISPR1 locus of *S. agalactiae* NEM316 (ST23) or BM110 (ST17) into a plasmid that carries an erythromycin resistance gene. The consensus PAM (5′-NGG-3′)[21] was added in the reverse primer (Supplementary Table S1), as a PAM is essential for the recognition of a protospacer by the CRISPR-Cas machinery.[32] If this targeted region is recognized, the corresponding plasmid should be cleaved upon entry into the cell, and no colony should be obtained after plating on TSH-agar containing 10 μg/mL of erythromycin.

The spacer acquisition assay was performed as previously described.[33] Briefly, pNZ123, a plasmid that carries a chloramphenicol resistance gene, was electroporated into *S. agalactiae* BM110 (ST17) and A909 (ST7). The ST7 (CC10) strain was used as a control because *in vivo* spacer acquisition was previously demonstrated in this lineage.[34] A confirmed transformant was grown in 10 mL of

TH without antibiotics. A 10-mL volume of fresh TH medium was inoculated with 100 μL of the previous culture twice a day for 5 days, for a total of nine inoculations. For each culture, 100 colonies were screened for chloramphenicol sensitivity, appearing following plasmid loss. The CRISPR1 locus of all the antibiotic-sensitive clones was PCR-amplified and sequenced. BM110Δ*cas9* and its complement were also transformed with pNZ123. A confirmed transformant of each BM110 derivative was grown in 10 mL of fresh TH twice a day for 5 days, as described above. The CRISPR1 locus was amplified directly, without screening for antibiotic-sensitive clones, using 2 μL of each culture as a sample for PCR analysis.[35]

Results

Deletion of the regular *cas* promoter in the *S. agalactiae* ST17 lineage

We first focused on the promoter regions of the CRISPR1-Cas system, as we hypothesized that differences in *cas* gene transcription could explain differences in the functionality of the systems between *S. agalactiae* lineages. There are two promoters in the CRISPR1-Cas system: the CRISPR1 locus promoter and the *cas* gene promoter (Fig. 1A). The CRISPR1 locus promoter is located in the leader region, immediately upstream of the repeat-spacer region, and is highly conserved among *S. agalactiae* lineages (data not shown). The *cas* gene promoter was previously identified in *S. agalactiae* NEM316 through a

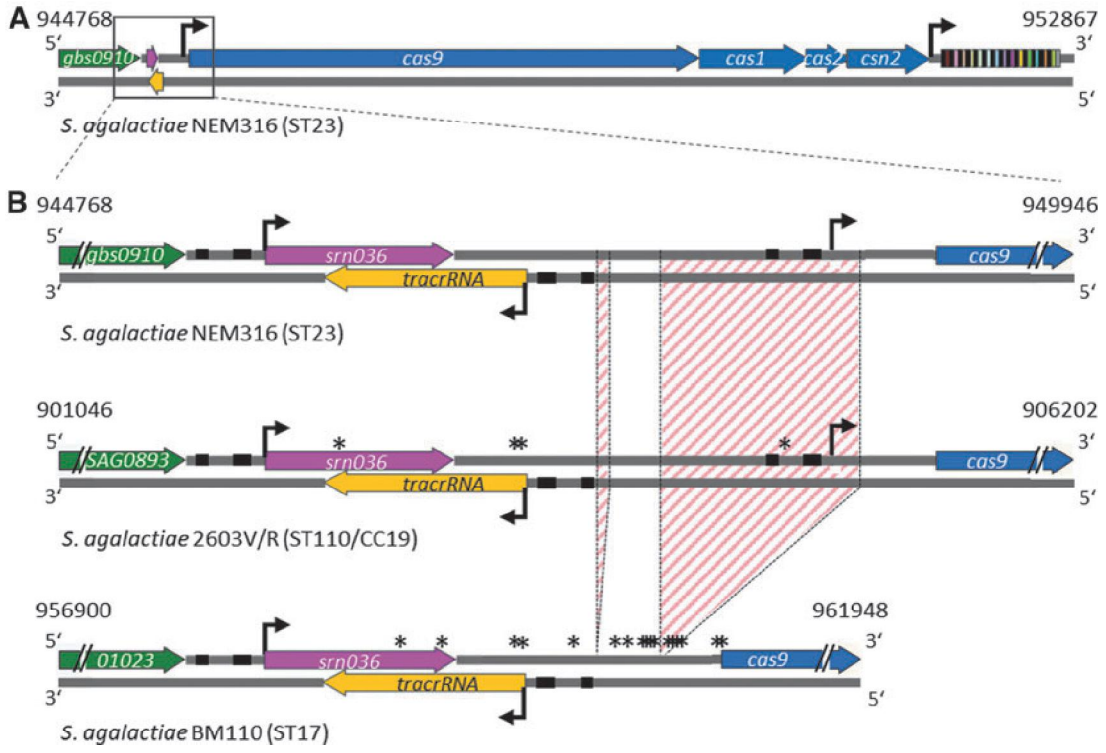


FIG. 1. Deletion of the regular *cas* promoter in the ST17 lineage. **(A)** Organization of the CRISPR1-Cas system of *Streptococcus agalactiae* NEM316. *cas* genes are shown as *blue arrows*, the gene upstream the CRISPR1-Cas system is in *green*, the antisense tracrRNA in *yellow*, and *srn036* in *pink*. The leader sequence is located downstream the *cas* genes and upstream the CRISPR locus. CRISPR1 array contains 14 repeat-spacer units, represented as *black* and *colored rectangles*, respectively. The TSSs of the *cas* gene promoter and the CRISPR1 locus promoter are represented with a *right black arrow*. **(B)** Close up of the intergenic region upstream of the *cas* operon of three strains representative of three major lineages. Promoter elements have been mapped with RNA-Seq. They are represented in *black*: the *black arrows* represent the TSSs, and the *black boxes* represent the predicted −10 and −35 boxes. *Black stars* indicate the SNPs observed in strains belonging to CC1, CC10, CC19, and ST17 in comparison with CC23 strains. *Red striped boxes* indicate the two deleted regions in ST17 strains. CCs, clonal complexes; CRISPR, clustered regularly interspaced short palindromic repeats; SNPs, single-nucleotide polymorphisms; ST17, Sequence Type 17; TSSs, transcription start sites.

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

combination of differential RNA sequencing (RNA-Seq, Fig. 1B).[36] RNA-Seq analyses identified two other elements in the intergenic region upstream of the *cas* operon: the tracrRNA and its promoter in reverse orientation, and another small RNA that possesses its own promoter, *srn036* (Fig. 1B).[36] To search for differences in this region between *S. agalactiae* strains, we performed a nucleotide sequence alignment, using the BLASTn web server. This alignment was performed on a total of 331 sequences that belong to the main CC isolated from human infections (Supplementary Table S2). When the nucleotide sequence alignment of the promoter region was carried out on strains belonging to the same CC but different STs, no difference was observed, with the exception of strains belonging to ST28 and ST520 in CC19. These two STs presented the same single-nucleotide polymorphism (SNP, data not shown). Taken altogether, our analysis demonstrates that nucleotides are highly conserved in this region within the same lineage.

Sequence alignment was also performed between CCs using *S. agalactiae* NEM316 as a template, from which isolates were classified into two groups. In one group, strains belonging to CC1, CC10, CC19, and CC23 shared very similar promoter regions. Four SNPs differentiated CC1, CC10, and CC19 from CC23: a one base-pair deletion and three substitutions (Fig. 1B and Supplementary Fig. S1). In the second group, composed of ST17 strains, this promoter region was very different. We identified 1 insertion, 16 substitutions, and 2 deletions of 7- and 103-bp long (Fig. 1B and Supplementary Fig. S1). The large deletion is located in the previously identified *cas* promoter, which is consequently no longer present in ST17 strains.

Impact of the deletions on *cas* promoter activity and *cas9* transcription

To study the impact of the observed nucleotide differences, and especially those affecting the *cas* promoter region, transcriptional fusion assays were performed to determine the activity of the *cas* promoter according to the phylogenetic lineage (Fig. 2A). β-galactosidase activity was calculated to be $\sim$250 and 260 arbitrary units when *S. agalactiae* NEM316 and 2603V/R promoter regions were cloned into pTCV-*lacZ*, respectively (Fig. 2B). These results were expected as these two strains possess very similar *cas* promoter sequences. Conversely, the β-galactosidase activity was negligible when expressed from the *S. agalactiae* BM110 promoter, suggesting the absence of an active promoter in the cloned region, which is likely due to the 103-bp deletion upstream of the *cas* operon in ST17 strains.

In the case of absence of an active upstream promoter, the *cas* genes would not be transcribed in these ST17 strains. To test this hypothesis, we performed an RT-qPCR to study the transcription of *cas9*, according to the phylogenetic lineage of the strain and the growth phase (Fig. 2C). First, we demonstrated that *cas9* transcription is similar between exponential and stationary growth phases in all lineages. As expected, *cas9* was transcribed in ST110/CC19 and ST23/CC23 strains and there was no difference observed in *cas9* transcription between both groups. Interestingly, *cas9* was still transcribed in the ST17/CC17 strain, but significantly less, by four- to fivefold, in the growth conditions tested.

Interference is still occurring in ST17 strains

To test if the CRISPR1-Cas system is still functional in the ST17 strains, we performed an interference assay. When pG1^{ts} (the empty vector) was electroporated into *S. agalactiae* NEM316, a substantial number of colonies were obtained on selective medium. However, no transformant was obtained with the plasmid carrying the protospacer (pG1^{ts}-protospacer), indicating a functional interference by the CRISPR1-Cas system (Fig. 3). To our surprise, similar results were observed with the ST17 strain BM110 (Fig. 3).

To address the discrepancy between *cas9* transcription levels and the interference activity of the ST17 strain, we also performed the interference assay with a *cas9* deletion mutant. A comparable number of transformants were obtained after electroporation of the plasmid carrying the protospacer and the empty vector, confirming the role of the CRISPR1-Cas system in the above results (Fig. 3). Of note, the nonfunctional CRISPR1-Cas system of *S. agalactiae* BM110Δ*cas9* became functional again when it was complemented with the wild-type *cas9* (BM110Δ*cas9*::*cas9*), validating the results with the deleted *cas9* strain (Fig. 3). Taken altogether, Cas9 plays an essential role in the interference step of the CRISPR1-Cas system found in ST17 strains, despite having a significantly different *cas* promoter region.

Spacer acquisition in response to plasmid challenges

After confirming the interference activity of the CRISPR1-Cas system in various *S. agalactiae* strains, we examined the functionality of the adaptation step, which is the acquisition of new spacers from MGE, by way of a plasmid challenge.[33] Out of 900 screened clones, a total of 207 antibiotic-sensitive clones (plasmid lost) were isolated for *S. agalactiae* BM110 (ST17). Of the 207 sensitive clones analyzed, the CRISPR1 array was modified in 92 of them (44%). All the clones presenting a CRISPR1 modification were sequenced and analyses showed the addition

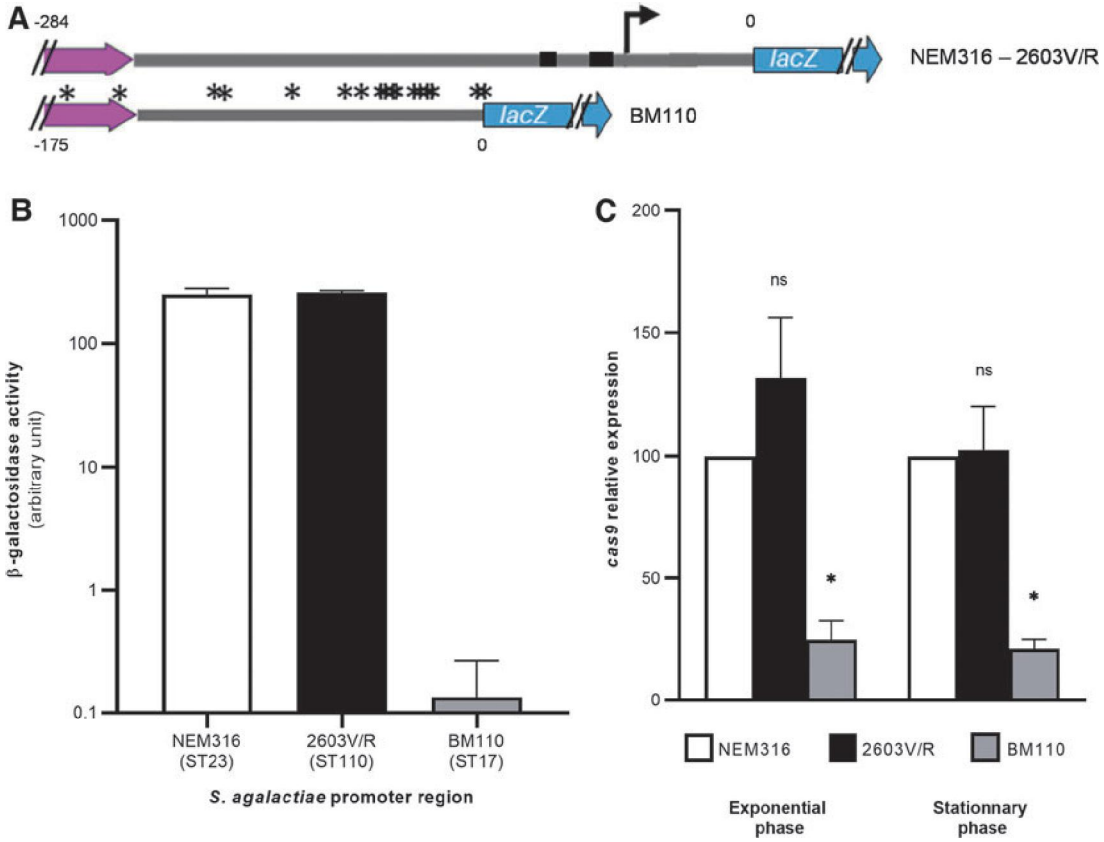


FIG. 2. The large deletion in ST17 strains reduces *cas* promoter activity and *cas9* transcription. **(A)** Activity of the *cas* promoter was determined by transcriptional fusion between the promoter region of three representative *S. agalactiae* strains and the reporter gene *lacZ*. To reflect *cas* promoter activity only, *srn036* promoter was excluded from the transcriptional fusion. Numbers above and below the diagrams indicate the relative position from the start of *cas9*. *Black stars* indicate the SNPs observed in strains belonging to ST17 in comparison with CC23 strains. **(B)** Following bacterial growth to midexponential phase, β-galactosidase activity was measured. Results are presented as mean ± SD from three independent experiments. **(C)** Transcriptional profiling of *cas9* in *S. agalactiae* 2603V/R (ST110, *black bars*) and *S. agalactiae* BM110 (ST17, *gray bars*) was compared with *cas9* transcription in *S. agalactiae* NEM316 (ST23, *white bars*) during growth in TH medium to midexponential phase ($OD_{600nm} = 0.5$) and stationary phase. Error bars show the mean ± SD ($n = 3$ biological replicates). Statistical comparisons of means were made using one-way ANOVA and Tukey's multiple comparison tests. *$p < 0.05$. SD, standard deviation; TH, Todd Hewitt.

at the 5′ end of the CRISPR1 array of a new 66-bp repeat-spacer unit, which corresponded to a 36-bp repeat and a 30-bp new spacer. The acquisition of one new repeat-spacer unit could be detected from the first inoculation. Most clones acquired a single spacer from the pNZ123 plasmid, but one clone acquired two spacers. Interestingly, 10 clones acquired 1 new spacer but also lost 2 to 7 spacers from the wild-type array. Sequence analysis of the CRISPR1 locus identified 30 different spacers among the 92 acquisition events (Fig. 4). All these new spacers matched perfectly with the sequence of pNZ123. Each protospacer was associated with the consensus PAM NGG (Supplementary Table S3). As a control, we performed the same plasmid challenge with *S. agalactiae* A909 (ST7/CC10). Of the 235 antibiotic-sensitive clones isolated, the CRISPR1 array was modified in 208 of them (88.5%, Supplementary Fig. S2). Adaptation efficiency thus appears stronger in the ST7 strain, compared with ST17. However, the sequencing analysis of the adaptation-positive clones revealed only 26 new different spacers among the 208 acquisition events. Spacer diversity is therefore greater for the ST17 lineage.

It was previously shown that Cas9 is also involved in the adaptation step in other bacterial systems.[37] Therefore, we tested whether this is also the case in *S. agalactiae*. New spacer acquisition assays were performed with

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

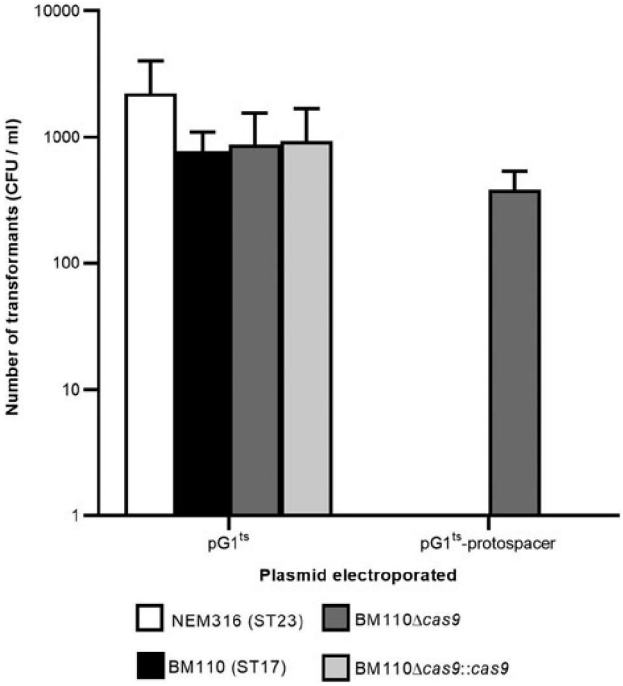


FIG. 3. CRISPR interference in ST17 strain. An interference assay was performed by cloning the CRISPR1 leader-proximal spacer (and its PAM) of *S. agalactiae* NEM316 (ST23) or BM110 (ST17) into a plasmid (pG1^{ts}-protospacer). This construct was electroporated into *S. agalactiae* wild-type strains as well as in *cas9* deletion derivatives. After electroporation, cells were plated on selective medium and colonies were counted. The experiments were replicated three times. Error bars show the mean ± SD.

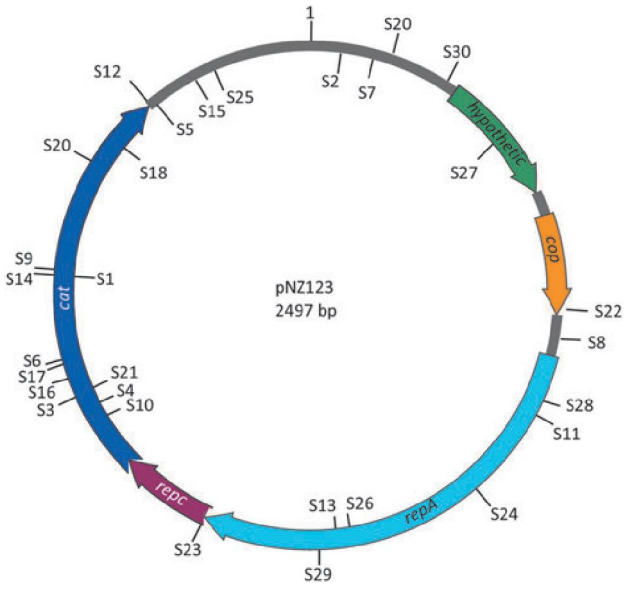


FIG. 4. CRISPR adaptation in ST17 strain. Schematic representation of 30 pNZ123 protospacers, perfectly matching the spacers newly acquired in *S. agalactiae* BM110 CRISPR1 locus. The numbers outside and inside the plasmid map correspond to protospacers from the positive and negative strands, respectively.

S. agalactiae BM110 wild-type strain, BM110Δ*cas9*, and its complement. Inoculations were carried out as previously described over 5 days, but CRISPR1 locus amplification was performed directly, using each inoculation as a template, without plating or screening for antibiotic-sensitive clones.[35] As expected, new repeat-spacer unit acquisitions occurred for the wild-type and the complemented strains. However, we were unable to detect new spacer acquisition in the strain (BM110Δ*cas9*) in which *cas9* was deleted, confirming its essential role in spacer acquisition (Supplementary Fig. S3). Taken altogether, the ST17 CRISPR1-Cas system remains capable of *in vitro* adaptation, that is, the acquisition of new spacers, despite the absence of the regular *cas9* promoter.

Implication of the *srn036* promoter in ST17 CRISPR1-Cas functionality

Our data strongly suggest the presence of another promoter from which *cas* genes are transcribed in the ST17 lineage. We focused on the promoter of *srn036*, which is located upstream of the *cas* genes and is in the same transcriptional orientation (Fig. 1B). To test this hypothesis, we performed additional *lacZ* transcriptional fusion experiments, including the *srn036* promoter in the region cloned upstream of the reporter gene (Fig. 5A). With these new constructs, we observed significant promoter activity resulting in about 570 arbitrary units in the ST17 strain (Fig. 5B). However, this reported β-galactosidase activity was much lower than the one obtained with the ST23 strain, which resulted in about 1200 arbitrary units. It is worth noting that the ST23 strain possesses two promoters in the region studied (*srn036* and *cas* promoters). These results suggest that the *srn036* promoter may enable the transcription of the *cas* genes located downstream.

We also performed another set of plasmid interference assays. A similar number of transformants were obtained with the empty vector as with the plasmid carrying the protospacer with a strain that was mutated in the *srn036* promoter (Fig. 5C). Thus, our results showed that the CRISPR1-Cas system of *S. agalactiae* BM110Δp$_{srn036}$ was nonfunctional compared with the wild-type and complemented strains. These observations were corroborated with an RT-qPCR performed on the p$_{srn036}$ promoter mutants as we were unable to detect *cas9* transcription in BM110Δp$_{srn036}$. Moreover, the *srn036* promoter was

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

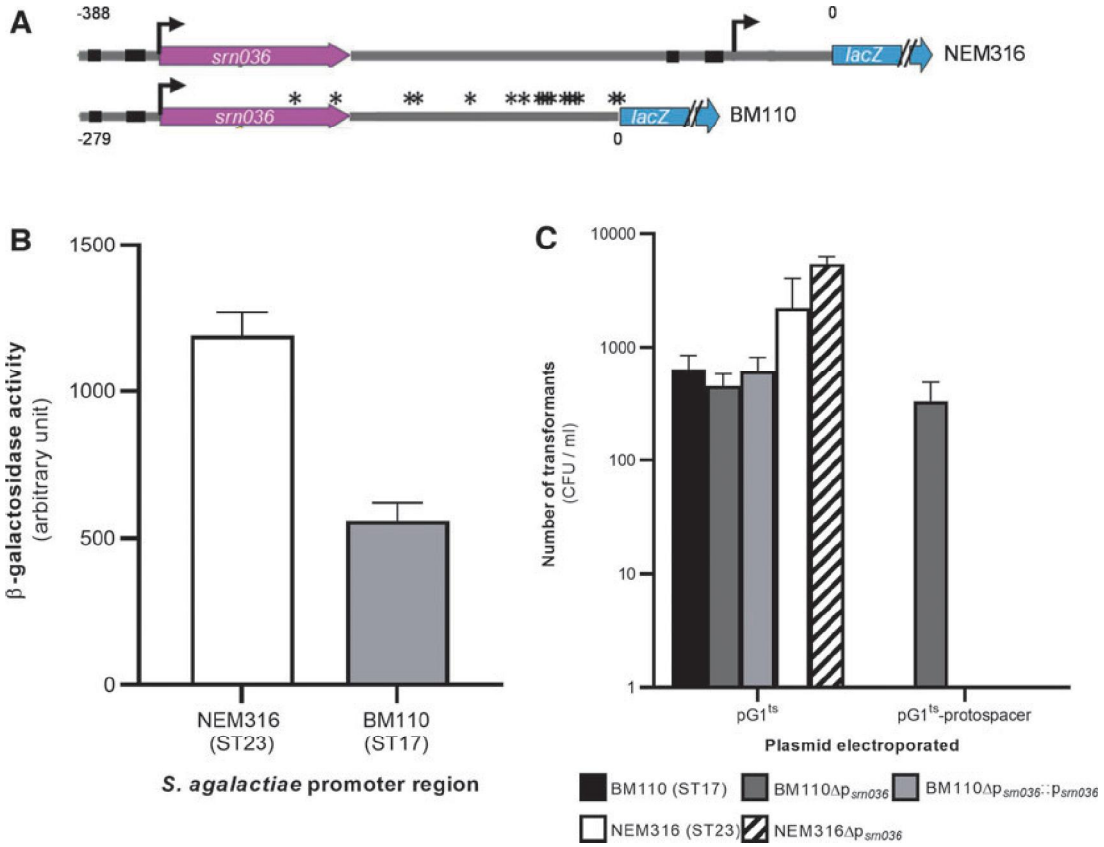


FIG. 5. *srn036* plays a key role in the functionality of the CRISPR1-Cas system in ST17 lineage. **(A)** *S. agalactiae* BM110 (ST17) and NEM316 (ST23) regions upstream of the *cas* operon, including the *srn036* promoter, were cloned into pTCV-*lacZ*. Numbers above and below the diagrams indicate the relative position from the start of *cas9*. *Black stars* indicate the SNPs observed in strains belonging to ST17 in comparison with CC23 strains. **(B)** β-galactosidase activity was measured in three independent experiments, as in Figure 2B. **(C)** CRISPR1-Cas functionality was determined using an interference assay, which was performed as in Figure 3, with an *S. agalactiae* BM110 WT strain (ST17), a derivative with a mutated *srn036* promoter (BM110Δp_{srn036}), and its complement (BM110Δp_{srn036}::p_{srn036}). In comparison, the interference assay was also performed with *S. agalactiae* NEM316 (ST23) and its p_{srn036} promoter mutant (NEM316Δp_{srn036}). The experiments were repeated three times. **(B)** and **(C)** Error bars show the mean ± SD.

also mutated in strain NEM316 (NEM316Δp_{srn036}) and the CRISPR1-Cas system remained functional (Fig. 5C). Overall, our data demonstrate that the *srn036* promoter is essential in the ST17 strain, but not in the ST23 strain, in which the *cas* promoter is responsible for the functionality of the CRISPR1-Cas system.

Discussion

Within a bacterial species, the functionality of CRISPR-Cas systems can have a dramatic impact on strains' specificities. Indeed, CRISPR-Cas systems can limit genetic exchanges,[38] and also, they can play a role in bacterial virulence,[39,40] in responses to bacterial stress or in gene regulation,[41] as well as protect against phage infection.[18] A nucleotide sequence alignment of the region upstream of the *cas* operon, performed on a total of 331 *S. agalactiae* strains, revealed similarities but also several differences that were linked to phylogenetic lineages. Previous studies have shown that the most ancestral spacers and the 3′ terminal repeat of the CRISPR1 array correlate with the phylogenetic lineage that was determined using the MLST method.[21,23] Our nucleotide sequence alignment of the *cas* promoter region of the *S. agalactiae* CRISPR1-Cas system also showed polymorphisms that are specific to some phylogenetic groups within the species.

Of the two deletions that were highlighted in ST17 strains, one is located within the *cas* promoter, and is potentially responsible for differences in *cas* gene transcription. To evaluate the impact of this deletion, transcriptional

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

fusion experiments were performed. The β-galactosidase activity results confirm that the promoter identified by RNA-Seq in *S. agalactiae* NEM316[36] is functional and enables the transcription of the genes located downstream. They also suggest that the 103-bp deletion observed in ST17 strains removes the active promoter and perhaps leads to no *cas* gene transcription. Thus, *cas9* transcription was investigated by performing quantitative reverse transcription PCR (RT-qPCR). Surprisingly, transcription was detected in all strains, regardless of lineage. Cas9 seems to be continually expressed, which is consistent with results from previous research[42]; indeed, Cas proteins must be continuously expressed to offer an adaptive immune response in case of foreign aggression. However, *cas9* is significantly under-transcribed in ST17 strains, in our growth conditions. While these data do not exclude the existence of a regulatory mechanism that may increase transcription when cells are exposed to a foreign element, our experiments did not trigger the CRISPR1-Cas system. A recent study discovered a new regulatory mechanism involving a long-form tracrRNA, capable of autoregulating CRISPR-Cas expression and especially *cas* gene transcription, to avoid autoimmunity.[43] The *tracr-L* binds to *cas* promoter, which dampens transcription. While this regulatory mechanism may occur in the ST23 strain, the *tracr-L* and its target are deleted in the ST17 strain.

As our transcription results suggested that the CRISPR1-Cas system of ST17 strains may not be properly functioning, we performed an interference assay. First, we established the functionality of the system and the assay with the ST23 strain. Then, we demonstrated that this step of the CRISPR1-Cas system was still effective in the ST17 strain, despite a different *cas* promoter region. However, in a mutated ST23 strain, mimicking the *cas* promoter deletion found in ST17 strains, the interference step became nonfunctional (Supplementary Fig. S4), suggesting that the ST17 CRISPR1-Cas system functions in a specific way that is different from the ST23 system.

CRISPR array diversity is driven by the acquisition of spacers at the 5′ leader-end. It has been well established that *S. agalactiae* CRISPR1-Cas systems are highly diverse and therefore appear to be dynamic,[21,23] but *in vitro* spacer acquisition has never been reported for this bacterial species. Moreover, previous work did not report any *in vivo* variation of the ST17 CRISPR1 locus, whereas leader-end spacer acquisition has been reported for other lineages, in a study that followed vaginal carriage of *S. agalactiae* in women.[34] Our work provides another model to study spacer acquisition *in vitro,* for which very few are currently available.[44] We demonstrated spacer acquisition through an *in vitro* plasmid challenge, with the ST17 strain. Spacer acquisition was detected from the first inoculation, suggesting that the system is very reactive. We also noticed that the acquisition of a new spacer could be accompanied sometimes by the deletion of older spacers. Although this has been observed in ∼10% of the ST17 clones presenting an expansion of their CRISPR1 array, this spacer loss has not been observed in strains belonging to other lineages (data not shown). This could perhaps explain the low number of spacers in ST17 CRISPR1 loci, but it remains to be confirmed through experimentation on a larger number of strains. On the contrary, the lower transcription of *cas* genes in the ST17 lineage likely results in a lower adaptation efficiency, which may explain the shorter CRISPR array.

Previous observations suggested that the CRISPR1-Cas system of ST17 strains was less active, but we have demonstrated here that it is still functional, both for interference and adaptation. We focused on the 103-bp deletion because it removed the previously identified *cas* promoter. However, there were also many SNPs in this region compared with other lineages. An analysis of the transcriptional landscape of *S. agalactiae* NEM316 identified a small RNA, *srn036*, in the immediate vicinity, which possesses its own promoter.[36] Unlike the transcription start site (TSS) and the −10 and −35 boxes, the *srn036* terminator location has not been clearly identified through differential RNA-Seq, but it has been described as leaky, suggesting that transcription could continue toward the *cas* genes. Moreover, SNPs observed in this area could lead to the complete absence of transcriptional termination. Our results, obtained from the transcriptional fusion experiments including the *srn036* promoter and from the interference assays with BM110Δp_{srn036}, are consistent with this hypothesis. *srn036* seems to play a key role in the functionality of the CRISPR1-Cas system of ST17 strains, particularly its promoter, which seems able to promote the transcription of *cas* genes in these strains. In addition, its location, partially overlapping the tracrRNA on the other DNA strand, is most likely critical, as the tracrRNA is an essential component of some class 2 CRISPR-Cas systems.[22,37,45] This location, characteristic of *cis*-encoded antisense RNAs, indicates that *srn036* probably targets tracrRNA and might thus interfere with its biological functions, as previously suggested.[36] *srn036* could also regulate the interactions between tracrRNA, crRNA, and Cas9 during the expression or the interference steps.

Conclusion

Significant nucleotide differences have been highlighted in the region upstream of the *cas* operon in the ST17 CRISPR1-Cas system compared with other *S. agalactiae* lineages. These differences do not prevent the functionality of the system. Our results suggest that other elements

such as *srn036* likely play an important regulatory role, emphasizing the specificity of these hypervirulent strains within the *S. agalactiae* species.

Authors' Contributions

A.P., P.L., G.R., S.M., and P.H. designed the research. A.P. and C.B. performed the experiments. A.P., G.R., and E.C. analyzed the data. A.P., L.M., and P.L. wrote the article with help from all authors. All coauthors have reviewed and approved of the article before submission.

Acknowledgments

We thank Vanessa Rong for technical support. We thank Amanda Toperoff and Michi Waygood for editorial assistance. We thank the Fédération de recherche en Infectiologie, the Centre-Val de Loire region, and the SSBCV doctoral school (Tours University) for funding. S.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada (Discovery program). S.M. holds a T1 Canada Research Chair in Bacteriophages.

Author Confirmation Statement

The data presented in this article have not been published or submitted for publication elsewhere.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

No funding was received for this article.

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3

References

1. Nocard E, Mollereau R. Contagious mastitis in dairy cows. *Ann Inst Pasteur.* 1887;1:109–126.
2. Baker CJ, Barrett FF, Gordon RC, et al. Suppurative meningitis due to streptococci of lancefield group B: A study of 33 infants. *J Pediatr.* 1973;82:724–729. DOI: 10.1016/S0022-3476(73)80606-7.
3. Barton LL, Feigin RD, Lins R. Group B beta hemolytic streptococcal meningitis in infants. *J Pediatr.* 1973;82:719–723. DOI: 10.1016/S0022-3476(73)80605-5.
4. Franciosi RA, Knostman JD, Zimmerman RA. Group B streptococcal neonatal and infant infections. *J Pediatr.* 1973;82:707–718. DOI: 10.1016/S0022-3476(73)80604-3.
5. McCracken GH. Group B streptococci: The new challenge in neonatal infections. *J Pediatr.* 1973;82:703–706. DOI: 10.1016/S0022-3476(73)80603-1.
6. Edmond KM, Kortsalioudaki C, Scott S, et al. Group B streptococcal disease in infants aged younger than 3 months: Systematic review and meta-analysis. *Lancet.* 2012;379:547–556. DOI: 10.1016/S0140-6736(11)61651-6.
7. Phares CR, Lynfield R, Farley MM, et al. Epidemiology of invasive group B streptococcal disease in the United States, 1999–2005. *JAMA.* 2008;299:2056–2065. DOI: 10.1001/jama.299.17.2056.
8. Skoff TH, Farley MM, Petit S, et al. Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990–2007. *Clin Infect Dis.* 2009;49:85–92. DOI: 10.1086/599369.
9. van der Mee-Marquet N, Fourny L, Arnault L, et al. Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites. *J Clin Microbiol.* 2008;46:2906–2911. DOI: 10.1128/JCM.00421-08.
10. Jones N, Bohnsack JF, Takahashi S, et al. Multilocus sequence typing system for group B *Streptococcus*. *J Clin Microbiol.* 2003;41:2530–2536. DOI: 10.1128/jcm.41.6.2530-2536.2003.
11. Da Cunha V, Davies MR, Douarre P-E, et al. *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nat Commun.* 2014;5:4544. DOI: 10.1038/ncomms5544.
12. Tazi A, Disson O, Bellais S, et al. The surface protein HvgA mediates group B *Streptococcus* hypervirulence and meningeal tropism in neonates. *J Exp Med.* 2010;207:2313–2322. DOI: 10.1084/jem.20092594.
13. Héry-Arnaud G, Bruant G, Lanotte P, et al. Acquisition of insertion sequences and the GBSi1 intron by *Streptococcus agalactiae* isolates correlates with the evolution of the species. *J Bacteriol.* 2005;187:6248–6252. DOI: 10.1128/JB.187.17.6248-6252.2005.
14. Fléchard M, Gilot P. Physiological impact of transposable elements encoding DDE transposases in the environmental adaptation of *Streptococcus agalactiae*. *Microbiol Read Engl.* 2014;160:1298–1315. DOI: 10.1099/mic.0.077628-0.
15. Tettelin H, Masignani V, Cieslewicz MJ, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." *Proc Natl Acad Sci U S A.* 2005;102:13950–13955. DOI: 10.1073/pnas.0506758102.
16. Brüssow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev.* 2004;68:560–602. DOI: 10.1128/MMBR.68.3.560-602.2004.
17. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nat Rev Microbiol.* 2010;8:317–327. DOI: 10.1038/nrmicro2315.
18. Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007;315:1709–1712. DOI: 10.1126/science.1138140.
19. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* 2010;327:167–170. DOI: 10.1126/science.1179555.
20. Makarova KS, Wolf YI, Iranzo J, et al. Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nat Rev Microbiol.* 2020;18:67–83. DOI: 10.1038/s41579-019-0299-x.
21. Lopez-Sanchez M-J, Sauvage E, Da Cunha V, et al. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol Microbiol.* 2012;85:1057–1071. DOI: 10.1111/j.1365-2958.2012.08172.x.
22. Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature.* 2011;471:602–607. DOI: 10.1038/nature09886.
23. Lier C, Baticle E, Horvath P, et al. Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. *Front Genet.* 2015;6:214. DOI: 10.3389/fgene.2015.00214.
24. Tettelin H, Masignani V, Cieslewicz MJ, et al. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A.* 2002;99:12391–12396. DOI: 10.1073/pnas.182380799.
25. Glaser P, Rusniok C, Buchrieser C, et al. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol Microbiol.* 2002;45:1499–1513. DOI: 10.1046/j.1365-2958.2002.03126.x.
26. Poyart C, Trieu-Cuot P. A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to beta-galactosidase in Gram-positive bacteria. *FEMS Microbiol Lett.* 1997;156:193–198. DOI: 10.1111/j.1574-6968.1997.tb12726.x.

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

27. Miller JH. Experiments in molecular genetics. New York, NY: Cold Spring Harbor Laboratory. 1972.
28. Patron K, Gilot P, Rong V, et al. Inductors and regulatory properties of the genomic island-associated *fru2* metabolic operon of *Streptococcus agalactiae*. *Mol Microbiol.* 2017;103:678–697. DOI: 10.1111/mmi.13581.
29. Lamy M-C, Zouine M, Fert J, et al. CovS/CovR of group B *Streptococcus*: A two-component global regulatory system involved in virulence. *Mol Microbiol.* 2004;54:1250–1268. DOI: 10.1111/j.1365-2958.2004.04365.x.
30. Florindo C, Ferreira R, Borges V, et al. Selection of reference genes for real-time expression studies in *Streptococcus agalactiae*. *J Microbiol Methods.* 2012;90:220–227. DOI: 10.1016/j.mimet.2012.05.011.
31. Biswas I, Gruss A, Ehrlich SD, et al. High-efficiency gene inactivation and replacement system for Gram-positive bacteria. *J Bacteriol.* 1993;175:3628–3635. DOI: 10.1128/jb.175.11.3628-3635.1993.
32. Mojica FJM, Díez-Villaseñor C, García-Martínez J, et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology.* 2009;155:733–740. DOI: 10.1099/mic.0.023960-0.
33. Garneau JE, Dupuis M-È, Villion M, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature.* 2010;468:67–71. DOI: 10.1038/nature09523.
34. Beauruelle C, Pastuszka A, Horvath P, et al. CRISPR: A useful genetic feature to follow vaginal carriage of group B *Streptococcus*. *Front Microbiol.* 2017;8:1981. DOI: 10.3389/fmicb.2017.01981.
35. McKenzie RE, Almendros C, Vink JNA, et al. Using CAPTURE to detect spacer acquisition in native CRISPR arrays. *Nat Protoc.* 2019;14:976–990. DOI: 10.1038/s41596-018-0123-5.
36. Rosinski-Chupin I, Sauvage E, Sismeiro O, et al. Single nucleotide resolution RNA-seq uncovers new regulatory mechanisms in the opportunistic pathogen *Streptococcus agalactiae*. *BMC Genomics.* 2015;16:419. DOI: 10.1186/s12864-015-1583-4.
37. Heler R, Samai P, Modell JW, et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature.* 2015;519:199–202. DOI: 10.1038/nature14245.
38. Lindenstrauss AG, Pavlovic M, Bringmann A, et al. Comparison of genotypic and phenotypic cluster analyses of virulence determinants and possible role of CRISPR elements toward their incidence in *Enterococcus faecalis* and *Enterococcus faecium*. *Syst Appl Microbiol.* 2011;34:553–560. DOI: 10.1016/j.syapm.2011.05.002.
39. Louwen R, Horst-Kreft D, de Boer AG, et al. A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré syndrome. *Eur J Clin Microbiol Infect Dis.* 2013;32:207–226. DOI: 10.1007/s10096-012-1733-4.
40. Louwen R, Staals RHJ, Endtz HP, et al. The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol Mol Biol Rev.* 2014;78:74–88. DOI: 10.1128/MMBR.00039-13.
41. Sampson TR, Weiss DS. CRISPR-Cas systems: New players in gene regulation and bacterial physiology. *Front Cell Infect Microbiol.* 2014;4:37. DOI: 10.3389/fcimb.2014.00037.
42. Young JC, Dill BD, Pan C, et al. Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS One.* 2012;7:e38077. DOI: 10.1371/journal.pone.0038077.
43. Workman RE, Pammi T, Nguyen BTK, et al. A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression. *Cell.* 2021;184:675–688.e19. DOI: 10.1016/j.cell.2020.12.017.
44. Mosterd C, Rousseau GM, Moineau S. A short overview of the CRISPR-Cas adaptation stage. *Can J Microbiol.* 2020:1–12. DOI: 10.1139/cjm-2020-0212.
45. Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337:816–821. DOI: 10.1126/science.1225829.

Downloaded by 87.231.95.228 from www.liebertpub.com at 04/20/21. For personal use only.

Partie 2 : Etude de la régulation de l'expression du système CRISPR1-Cas

Dans cette deuxième partie, l'objectif était d'identifier les voies de régulations mises en jeu par *S. agalactiae* pour réguler l'expression de son système CRISPR1-Cas. Ainsi dans un premier temps, nous avons étudié le rôle du système à deux composants CovRS dans cette régulation, puis nous avons étudié la réponse du système CRISPR1-Cas suite à l'application de différents stimuli.

1. Matériel et Méthodes

Plusieurs des méthodes utilisées dans cette partie ont été préalablement décrites dans la partie 1, notamment toutes les techniques générales d'analyse comme les réactions d'amplification, la transformation, la fusion transcriptionnelle ou la PCR quantitative en temps réel, mais également les méthodes qui concernent l'étude de la fonctionnalité du système CRISPR1-Cas, par analyse de l'étape d'interférence ou de l'étape d'adaptation.

1.1. Souches étudiées

1.1.1. Souches de référence

Les travaux présentés dans cette deuxième partie ont été effectués à l'aide de trois souches de référence : S*. agalactiae* BM110, qui appartient à la lignée d'intérêt, le ST17 ; A909, qui appartient au ST7, proche phylogénétiquement du ST17 et NEM316, une souche de ST23, qui en est plus éloignée. Leurs caractéristiques sont présentées dans le Tableau II.

1.1.2. Mutants du système CovRS

Pour étudier le rôle éventuel du système à deux composants CovRS dans la régulation de l'expression du système CRISPR1-Cas de *S. agalactiae*, nous avons utilisé différents mutants de ce système, mis à notre disposition par le Docteur Arnaud Firon, de l'équipe Biologie des Bactéries Pathogènes à Gram-Positif de l'Institut Pasteur. D'une part, nous avions le mutant de délétion du gène *covR* (Δ*covR*), le régulateur cytosolique, ainsi qu'un mutant de substitution de CovR ($CovR^{D53A}$), où l'aspartate en position 53 a été substitué par une alanine, altérant la fonction de CovR en empêchant sa phosphorylation. D'autre part, nous avions également un mutant de substitution de CovS, l'histidine kinase du système. Ce dernier, $CovS^{T282A}$, est un mutant où CovS a été inactivé pour son activité phosphatase, entraînant une activation constitutive de CovR, ce qui équivaut, en quelque sorte, à une surexpression de CovR. Ces mutants étaient disponibles dans les fonds génétiques *S. agalactiae* NEM316 (ST23) et BM110 (ST17).

1.2. Etude de la fonctionnalité du système CRISPR1-Cas

Afin de déterminer l'impact du système CovRS sur la régulation de l'expression du système CRISPR1-Cas de *S. agalactiae*, nous avons reproduit les tests de fonctionnalité précédemment décrits. Ainsi, nous avons effectué

des tests d'interférence et d'acquisition de spacers en comparant les résultats obtenus avec les 3 mutants (CovR^{D53A}, Δ*covR* et CovS^{T282A}) à ceux obtenus avec leur souche sauvage respective (NEM316 ou BM110).

Pour les tests d'interférence, nous avons utilisé les deux vecteurs de protospacer construits, celui portant le protospacer homologue du spacer le plus récent, situé à l'extrémité 5' du locus CRISPR1 et celui portant le protospacer homologue du spacer le plus ancien, situé à l'extrémité 3' du locus (Figure 37). Les deux vecteurs ont été électroporés dans les souches BM110 et NEM316 sauvages et leurs mutants respectifs. Le vecteur vide a été systématiquement électroporé en parallèle afin de servir de témoin de transformation. Les expériences ont été réalisées en triplicata.

Les tests d'acquisition de nouveaux spacers ont été réalisés comme pour les mutants BM110Δ*cas9* ou BM110ΔP$_{srn036}$, c'est-à-dire avec l'amplification du locus CRISPR1 directement sur les passages après « challenge » avec le plasmide pNZ123, ceci afin de s'affranchir de l'étape d'interférence et de n'étudier que l'étape d'adaptation.

1.3. Détermination de l'activité promotrice de P$_{srn036}$ dans le mutant BM110ΔcovR

1.3.1. Transformation chez *S. agalactiae*

Afin d'étudier le rôle répresseur éventuel de CovR sur P$_{srn036}$, le vecteur contenant ce promoteur en amont de *lacZ*, précédemment décrit et utilisé dans la partie 1 (Figure 50A), a été électroporé dans la souche BM110 sauvage et dans la souche BM110Δ*covR*. Le plasmide pTCV-*lac* vide a également été electroporé dans ces deux fonds génétiques afin de servir de témoin négatif. Les différents transformants ont tous été confirmés par PCR sur colonie à l'aide des amorces Vlac1 et Vlac2 (Tableau VI).

1.3.2. Dosage de l'activité β-galactosidase

Chacun des transformants a ensuite été cultivé en bouillon TH additionné d'érythromycine à 10 µg/ml jusqu'à une DO$_{600nm}$ ≈ 0,5. Après centrifugation 4 min à 7000 rpm, les protéines ont été extraites à partir des culots obtenus, selon le protocole précédemment utilisé. Puis, la quantité de protéines extraites a été déterminée par la méthode de Bradford. Enfin, l'activité spécifique de la β-galactosidase dans chacun des fonds génétiques a été calculée à l'aide du protocole précédemment décrit.

1.3.3. PCR quantitative

Dans un second temps, nous avons réalisé de nouvelles réactions de qPCR afin de déterminer le niveau de transcription du gène *cas9* dans les mutants du système CovRS et de corroborer les résultats obtenus avec les dosages de l'activité β-galactosidase. Ces réactions ont été effectuées en suivant le protocole décrit dans le paragraphe 1.11 de la première partie de ce travail. Par ailleurs, nous avons également dessiné un nouveau couple d'amorces, permettant la quantification de *srn036* (Tableau XIII).

Tableau XIII : Amorces utilisées pour quantifier la transcription de *srn036* par PCR quantitative en temps réel.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Gène cible	Séquence (5' - 3')	Taille de l'amplicon (pb)
OPL_134	F	*srn036*	ATGAGAAATGTCGGAATTATAACA	97 (*S. agalactiae* NEM136) 98 (*S. agalactiae* BM110)
OPL_095	R		GGCAGTGATTTTTAATCCAGTC	
OPL_096	R		GGCAGTGATTTTTTAATCCAGTC	

1.4. Identification de conditions régulant l'expression du système CRISPR1-Cas

Afin d'identifier des conditions dans lesquelles l'activité du système CRISPR1-Cas pourrait être régulée, nous avons quantifié la transcription du gène *cas9* après l'application de divers stimuli. Nous avons choisi d'étudier la réponse du système suite à un stress acide, en présence d'un EGM et en présence d'antibiotiques.

1.4.1. Application d'un stress acide

Les trois souches de référence de *S. agalactiae*, BM110, A909 et NEM316, ont été cultivées en bouillon TH à 37°C, jusqu'à une DO_{600nm} ≈ 0,5. Lorsque la mi-phase exponentielle de croissance a été atteinte, un aliquot de 5 ml de culture, correspondant au T0, a été prélevé. La culture a alors été répartie en deux tubes de 25 ml. Ces tubes ont été centrifugés 5 minutes à 7000 rpm à 4°C, puis un des culots a été remis en suspension en bouillon TH (pH 6,8) et le deuxième a été remis en suspension en bouillon TH auquel X µl d'HCl 10M ont été ajoutés, afin d'amener le pH de la culture à un pH voisin de celui du vagin (pH ≈ 4). De nouveaux aliquots de 5ml ont alors été prélevés après 5 (T5) et 10 (T10) minutes d'incubation à 37°C. Les aliquots ont été centrifugés 5 minutes à 7000 rpm à 4°C pour ensuite extraire les ARNs après congélation des culots à -80°C.

1.4.2. Effet de la présence d'un élément génétique mobile

Les plasmides ont été utilisés pour imiter la rencontre avec un EGM étranger. Les trois souches ont ainsi été cultivées dans 10 ml de bouillon TH, jusqu'à DO_{600nm} ≈ 0,5. Puis, 10 ml d'un mélange glycine/saccharose ont été ajoutés à la suspension afin d'affaiblir les membranes bactériennes et de rendre les bactéries « compétentes ». Après 1h d'incubation à 37°C sous agitation à 50 rpm, les tubes ont été centrifugés 10 minutes à 7000 rpm à 4°C. Les culots ont ensuite été lavés 3 fois avec du tampon d'électroporation, avant d'être remis en suspension dans 120 µl d'un mélange tampon d'électroporation/glycérol 10%. Les bactéries ainsi préparées ont été transférées dans une cuve d'électroporation et 1500 ng d'un des plasmides ont été ajoutés. L'application d'un champ électrique aux parois de la cuve a alors permis la perméabilisation transitoire de la membrane cytoplasmique et l'entrée du plasmide dans la bactérie. Enfin, la bactérie transformée a été remise en suspension dans 5 ml de bouillon TH préchauffé à 30°C. Après 15 minutes d'incubation à 30°C, la culture a été centrifugée 5 minutes à 7000 rpm à 4°C. Les culots bactériens ainsi obtenus ont été congelés à -80°C afin d'extraire les ARNs.

Comme plusieurs étapes sont nécessaires *in vitro* pour introduire un plasmide dans la bactérie, des extractions d'ARN correspondant à ces différentes étapes ont été réalisées, afin de pouvoir objectiver le rôle potentiel de chacune d'entre elles sur le système CRISPR1-Cas. La condition de référence correspondait à la culture en bouillon TH jusqu'à la mi-phase exponentielle. La deuxième extraction a été réalisée après l'ajout du

mélange glycine/saccharose. La troisième extraction a été réalisée après l'application du champ électrique. La quatrième extraction a été réalisée après l'ajout du plasmide vide pG1^{ts}, utilisé pour mimer un EGM jamais rencontré par la souche. Enfin la dernière extraction correspondait à l'ajout du plasmide pG1^{ts}-protospacer, portant un protospacer spécifique à chacune des souches testées, imitant un EGM déjà rencontré. Chaque extraction a été réalisée en triplicata.

1.4.3. Croissance en présence d'antibiotiques

Nous avons sélectionné deux antibiotiques dont la cible cellulaire est la paroi bactérienne, la pénicilline G, qui appartient à la famille des β-lactamines et la vancomycine, qui est un glycopeptide. Tout d'abord, nous avons réalisé des courbes de croissance des 3 souches de *S. agalactiae* en présence de concentrations croissantes de chacun des antibiotiques, afin de déterminer la concentration subinhibitrice qui n'impacterait pas la croissance bactérienne. Ceci a été réalisé dans une microplaque de 96 puits où 150 µl d'antibiotique ont été ajoutés à 150 µl de suspension bactérienne à DO_{600nm} = 0,05. Les gammes de concentrations testées étaient comprises entre 0,004 µg/mL et 0,032 µg/mL pour la pénicilline G et entre 0,125 µg/mL et 1 µg/mL pour la vancomycine. La microplaque a été incubée 24h à 37°C sous agitation dans un spectrophotomètre Eon™ (Biotek™), permettant une mesure de l'absorbance toutes les 20 minutes.

Une fois la concentration subinhibitrice déterminée, les expériences de stress ont pu être réalisées. Ainsi, chacune des 3 souches a été mise en culture en bouillon TH sans antibiotique et en bouillon TH contenant soit 0,008 µg/mL de pénicilline G, soit 0,25 µg/mL de vancomycine. Après incubation à 37°C jusqu'à $DO_{600nm} \approx 0,5$, 10 ml de suspension ont été centrifugés 5 minutes à 7000 rpm à 4°C.

1.5. Etude de la transcription de cas9 *par PCR quantitative en temps réel*

1.5.1. Préparation des ARNs totaux

Afin d'extraire les ARNs totaux à partir des culots bactériens obtenus dans les différentes conditions de stimulation, l'ensemble des étapes a été reproduit selon les protocoles décrits dans la partie 1. Brièvement, l'extraction des ARNs totaux a été réalisée selon une méthode de type phénol/chloroforme. Les ARNs ont ensuite été traités à la DNase afin d'éliminer les traces d'ADNg résiduel. L'efficacité de ce traitement a été vérifiée par PCR, confirmant l'absence d'ADNg contaminant.

1.5.2. qPCR

Les ADNc ont été synthétisés à l'aide du kit iScript™ cDNA synthesis (Bio-Rad) à partir des ARNs traités à la DNase. Chaque ADNc obtenu a été déposé en triplicata dans la plaque de qPCR afin de garantir la reproductibilité des résultats. Puis, le niveau de transcription du gène d'intérêt, ici *cas9*, face aux divers stimuli a été déterminé par qPCR, à l'aide d'un mélange réactionnel prêt à l'emploi contenant du SYBR Green et de l'instrument LightCycler® 480 (Roche).

Les mesures ont été effectuées sur les souches *S. agalactiae* NEM316 (ST23), A909 (ST7), et BM110 (ST17) et chacune des conditions a été réalisée en triplicata. Pour chaque extrait d'ARN, la valeur du Ct obtenue pour le

gène *cas9* a été normalisée à celle obtenue pour le gène de référence *recA* (EGM et antibiotiques, Tableau VII) ou rp*oB* (pH, Tableau XIV). Le ΔΔCt a ensuite été calculé pour normaliser les résultats obtenus dans les différentes conditions testées par rapport à la condition de référence. Puis le Fold Change (FC), qui tient compte de l'efficacité des couples d'amorces, a été calculé pour obtenir l'expression relative du gène *cas9*. La mesure en condition de référence a été réalisée en phase exponentielle de croissance bactérienne car nous avons montré qu'il n'existait pas de différence significative de transcription de *cas9* entre la phase exponentielle et la phase stationnaire de croissance. La transcription de *cas9* en condition de référence a systématiquement été fixée à 100. Les comparaisons statistiques des moyennes ont ensuite été effectuées à l'aide du logiciel Prism (GraphPad).

Tableau XIV : Amorces utilisées pour les réactions de PCR quantitative en temps réel.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Gène cible	Séquence (5' - 3')	Taille de l'amplicon (pb)
OLM_019	F	*rpoB*	TATTGAAAAAGACATTCATAAAAATCCA	137
OLM_020	R		CAAAGAAACGTGCAACTAAAAGACTAC	

1.6. Etude de l'expression de la protéine Cas9

1.6.1. Marquage de la protéine Cas9 avec un 3Xflag

Dans le but d'étudier l'expression de la protéine Cas9, suite à l'application des divers stimuli, nous avons marqué la protéine Cas9 par un 3Xflag en position N-terminale dans chacune de nos 3 souches (Figure 57).

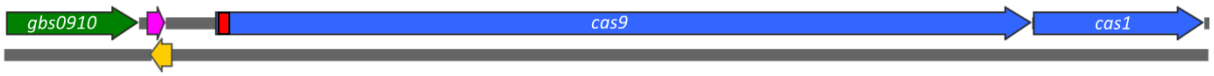


Figure 57 : Schématisation de la position du 3Xflag sur le génome de *S. agalactiae* NEM316.
Le code couleur employé est strictement identique à celui utilisé précédemment. Les gènes *cas* sont en bleus ; le gène en amont est en vert ; *srn036* en rose et le tracrRNA en jaune. Le 3Xflag a été représenté sous la forme d'un rectangle rouge.

1.6.1.1. Construction des inserts

Plusieurs couples d'amorces ont ainsi été dessinés afin d'amplifier les fragments en amont et en aval de l'ATG de *cas9* chez nos différentes souches, ainsi qu'un couple d'amorces permettant l'amplification du 3Xflag (Tableau XV). Celui-ci a été amplifié à partir d'un fragment d'ADN de *Salmonella*, qui nous a été fourni par le Docteur Isabelle Virlogeux-Payant, de l'équipe Signalisation, Portage et Virulence Bactérienne de l'UMR Infectiologie et Santé Publique. Les amorces devaient apporter les régions d'homologie nécessaires pour qu'il y ait assemblage des fragments entre eux, ainsi que les sites de restriction enzymatique, requis pour l'intégration dans le plasmide. De plus, les fragments devaient être d'une taille suffisante pour ensuite permettre la recombinaison avec le chromosome bactérien.

L'amplification des différents fragments a été réalisée à l'aide de la Q5® High-Fidelity DNA polymerase, comme décrit dans le paragraphe 1.4 de la partie 1. Nous avons ensuite procédé à l'assemblage des 3 fragments

entre eux grâce à deux PCR overlap. Tout d'abord, la première PCR overlap a permis d'assembler la région en amont avec le 3Xflag grâce à l'utilisation des amorces externes (OPL_056/OPL_170 pour NEM316 et A909 et OPL_060/OPL_176 pour BM110) et à la complémentarité de séquences apportée par les amorces internes (OPL_168/OPL_169 et OPL_174/OPL_175 ; Tableau XV). Puis, la deuxième PCR overlap a permis d'assembler le fragment ainsi obtenu avec la région en aval. Pour cela, la réaction d'amplification a été réalisée avec les amorces sens des régions en amont et les amorces antisens des régions en aval, la complémentarité de séquences ayant été apportée par les amorces antisens du 3Xflag et les amorces sens de la région en aval (Tableau XV).

Après purification des fragments obtenus, ceux-ci ont été digérés par les enzymes dont les sites de restriction étaient présents sur les amorces externes et au niveau du site de clonage de pG1^{ts}, afin de pouvoir ensuite réaliser la réaction de ligature (protocole décrit dans la partie 1, paragraphe 1.10.2).

Tableau XV : Amorces utilisées pour le marquage de la protéine Cas9 chez les différentes souches de *S. agalactiae*.
F : amorce sens. R : amorce anti-sens.

Amorce	Sens	Fragment amplifié	Mutant	Séquence (5' – 3')[a, b]
OPL_056	F	Box amont	NEM316$_{3Xflag-Cas9}$ A909$_{3Xflag-Cas9}$	CTAACGGAATTCGTAATTTGGATACTTCCTGAAAC
OPL_168	R			CTTTGTAGTCCATAACTTTTCTCCTTTTATTTAAA
OPL_060	F	Box amont	BM110$_{3Xflag-Cas9}$	CCTATTGAATTCGTAGTATGTGCTTTTCTGTTGG
OPL_174	R			CTTTGTAGTCCATAATTTATCTCCTTTTATTTAAAA
OPL_169	F	3Xflag	NEM316$_{3Xflag-Cas9}$ A909$_{3Xflag-Cas9}$	GAAAAGTTATGGACTACAAAGACCATGACGG
OPL_175	F		BM110$_{3Xflag-Cas9}$	ATAAATTATGGACTACAAAGACCATGACGG
OAH_283	R	ADN *Salmonella*		CGCGCTTGCTGCTTGGATGCC
OPL_170	R	3Xflag	NEM316$_{3Xflag-Cas9}$ A909$_{3Xflag-Cas9}$	ATGGCTTATTTTTATCGTCGTCATCTTTGTAG
OPL_176	R		BM110$_{3Xflag-Cas9}$	ATGGTTTGTTTTTATCGTCGTCATCTTTGTAG
OPL_171	F	Box aval	A909$_{3Xflag-Cas9}$	CGACGATAAAAATAAGCCATATTCAATAGGCC
OPL_172	R			TTGTTTGGTACCGTCTGTATTCCTGACATCAAAG
OPL_171	F	Box aval	NEM316$_{3Xflag-Cas9}$	CGACGATAAAAATAAGCCATATTCAATAGGCC
OPL_173	R			GTTTTTGGTACCCGGTATTCCTCACATCAAATC
OPL_177	F	Box aval	BM110$_{3Xflag-Cas9}$	CGACGATAAAAACAAACCATATTCAATAGGTCT
OPL_178	R			CTCACACTGCAGCTATCATCCTCAATTAGGAAATG
OPL_110	F	pG1^{ts} : vérification de l'insert		CGCCAGGGTTTTCCCAGTCACGAC
OPL_111	R			AGCGGATAACAATTTCACACAGGA

[a] Les sites de restriction enzymatique apparaissent soulignés.
[b] Les régions d'homologie entre les fragments apparaissent en couleur.

1.6.1.2. Obtention du marquage au niveau chromosomique

L'ensemble des étapes suivantes a été réalisé selon les modalités décrites précédemment. Ainsi les plasmides recombinants ont d'abord été électroporés chez *E. coli* afin de les amplifier. Ils ont ensuite été purifiés puis séquencés avant d'être électroporés chez *S. agalactiae* NEM316, A909 ou BM110 respectivement. Après

double recombinaison homologue, l'insertion du 3Xflag à la position désirée a été vérifiée par séquençage de l'ADN.

1.6.2. Extraction des protéines

1.6.2.1. Conditions de croissance testées

Les 3 souches marquées, NEM316$_{3Xflag\text{-}Cas9}$, A909$_{3Xflag\text{-}Cas9}$ et BM110$_{3Xflag\text{-}Cas9}$ ont été cultivées dans 10 ml de bouillon TH, jusqu'à DO_{600nm} ≈ 0,5, avant d'être rendues compétentes par l'ajout de 10 ml d'un mélange glycine/saccharose et enfin, d'être transformées par le plasmide pG1^{ts} vide ou porteur d'un protospacer spécifique de la souche. Différents aliquots, correspondant aux 5 conditions décrites dans le paragraphe 1.4.2 de la partie 2, ont été prélevés et centrifugés afin d'en extraire les protéines. Ainsi, la première extraction correspondait à la croissance en bouillon TH jusqu'à la mi-phase exponentielle. La deuxième extraction correspondait à la croissance après l'ajout du mélange glycine/saccharose. La troisième extraction a été réalisée après l'application du champ électrique. La quatrième extraction correspondait à l'ajout du plasmide vide pG1^{ts} et la dernière extraction correspondait à l'ajout du plasmide pG1^{ts}-protospacer.

1.6.2.2. Extraction et dosage des protéines

Les protéines ont été extraites à partir des différents culots selon le même protocole que celui utilisé pour la détermination de l'activité β-galactosidase (Partie 1, paragraphe 1.10.6). Les concentrations des extraits protéiques obtenus ont été mesurées par la méthode de Bradford, après réalisation d'une gamme d'étalonnage à l'aide de dilutions successives de sérum albumine bovine (BSA). L'équation de la droite de corrélation obtenue entre l'absorbance à DO_{595nm} et la concentration en µg/mL, nous a permis de calculer les concentrations protéiques de nos extraits. Les extraits ont ensuite été dilués afin de les amener tous à la même concentration pour réaliser les Western Blots.

1.6.3. Western Blot

1.6.3.1. Préparation des échantillons

Les extraits protéiques ont été repris dans un tampon de charge 5X (tampon Laemmli [Tris-HCl 0,5M pH 6,8, Glycérol 100%, SDS 10%, β-mercaptoéthanol, Bleu de Bromophénol]), chauffés 10 minutes à 95°C, puis placés au congélateur à -20°C. Les échantillons ont été centrifugés puis à nouveau chauffés 5 min à 100°C juste avant la réalisation des Western Blots.

1.6.3.2. Séparation des protéines (SDS-PAGE)

La protéine Cas9 étant d'une taille approximative de 160 kDa, nous avons utilisé des gels prêts à l'emploi, contenant un gradient de concentration de 4 à 15% de polyacrylamide, permettant la séparation des protéines entre 5 et 200 kDa (4-15% Mini-PROTEAN® TGX™ Precast Protein Gels, Bio-Rad). Nous avons réalisé deux migrations en parallèle. Les gels ont ainsi été placés dans une cuve contenant du tampon de migration 1X (Tris, glycine, SDS, pH 8,3). Nous avons ensuite déposé 10 µl de chaque extrait protéique dans les puits, soit environ 7

µg de protéines, ainsi qu'un marqueur de taille. L'application d'un champ électrique a alors permis la migration des protéines vers l'électrode positive et leur séparation en fonction de leur poids moléculaire.

1.6.3.3. Transfert des protéines

Les protéines ont ensuite été transférées sur une membrane de nitrocellulose à l'aide du kit Trans-Blot® Turbo™ RTA Transfer Kit (Bio-Rad) et de l'appareil Trans-Blot® Turbo™ Transfer System associé. Après vérification du transfert de la coloration bleue sur la membrane de nitrocellulose, celle-ci a été saturée avec du TBS (Tris Buffered Saline)-Tween 0,05% contenant 3% de lait écrémé. Après 1h sous agitation, le mélange TBS-lait a été éliminé par retournement et trois lavages successifs de 5 minutes en TBS-Tween 0,05% ont été réalisés.

1.6.3.4. Révélation

La membrane a ensuite été incubée avec l'anticorps primaire, un anticorps monoclonal de souris ANTI-FLAG® M2 (Sigma-Aldrich), dans un tampon TBS-Tween 0,05%, 2% lait, pendant 1h sous agitation. Après trois nouveaux lavages de 5 minutes, la membrane a ensuite été incubée avec l'anticorps secondaire, un anticorps polyclonal de lapin, anti-Ig de souris, couplé à la peroxydase (Dako). Après 2h d'incubation sous agitation, la membrane a à nouveau été lavée trois fois avec un tampon TBS-Tween. La révélation du marquage a été réalisée à l'aide du kit SuperSignal™ West Dura Extended Duration Substrate (Thermo Scientific). La membrane a été mise en contact avec un mélange contenant 1 ml de luminol et 1 ml de peroxyde, pendant 5 minutes à l'obscurité. La quantification de la chimiluminescence et l'acquisition des images ont été réalisées avec le système avancé d'imagerie Fusion FX7 (Vilber Lourmat™).

1.6.3.5. Coloration au Bleu de Neuhoff

En parallèle, nous avons également coloré un des gels au Bleu de Neuhoff, afin d'avoir un aperçu de la quantité et de la qualité des protéines déposées.

2. Résultats

2.1. Régulation du système CRISPR1-Cas par le régulateur majeur de virulence CovR

Récemment, des travaux ont été réalisés sur le système à deux composants CovRS par Maria Vittoria Mazzuoli, pour sa Thèse de doctorat, dans l'équipe Biologie des Bactéries Pathogènes à Gram-Positif de l'Institut Pasteur (Mazzuoli, 2020). Ces travaux, visant à caractériser les voies de régulation de ce système et à déterminer leur spécificité, chez les souches hypervirulentes du ST17, ont montré un site de fixation du régulateur CovR au niveau du promoteur de *srn036*. D'abord identifiée par ChIP-Seq, la fixation de CovR à P_{srn036} a ensuite été validée par retard sur gel. Puis, pour identifier précisément le site de fixation, une expérience de DNase I footprinting a été réalisée (Figure 58). Brièvement, la fixation de CovR à l'ADN protège ce dernier de la digestion par la DNase I. Ceci a permis d'identifier deux sites de liaison de CovR, dont un englobant le TSS et la boîte -10 de P_{srn036}. Ceci suggère fortement une altération de la transcription à partir de P_{srn036} lorsque CovR est fixé à l'ADN. Si CovR était impliqué dans la régulation de l'expression du système CRISPR1-Cas, ceci serait la découverte d'un nouveau mécanisme de régulation non décrit à ce jour.

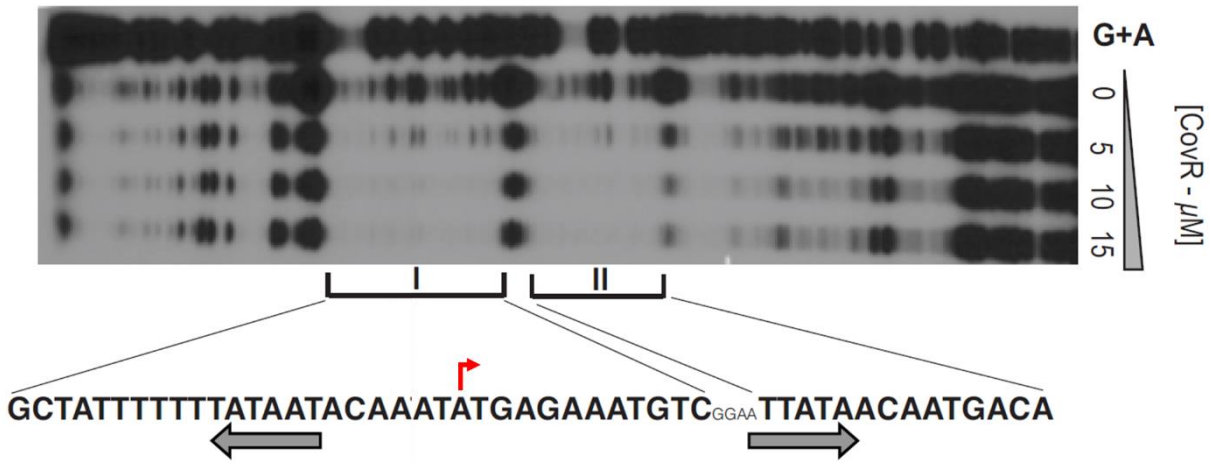


Figure 58 : P_{srn036} CovR binding sites detected by DNase I footprint.
The two DNase I protected boxes are numbered and spaced by four nucleotides. The position of the transcriptional start site is highlighted in red. The two opposite ATTARA sequences are underlined by grey arrows *(Figure aimablement transmise par le Dr Arnaud Firon et issue des travaux de Thèse de Maria Vittoria Mazzuoli)*.

2.1.1. Impact de CovR sur la fonctionnalité du système CRISPR1-Cas

Dans un premier temps, nous avons donc souhaité connaître l'impact de la fixation de CovR sur P_{srn036} dans la fonctionnalité du système CRISPR1-Cas de *S. agalactiae* BM110 (ST17) et NEM316 (ST23), le site de fixation étant présent et conservé dans les deux souches. En effet, nous avons démontré le rôle essentiel de P_{srn036} pour le fonctionnement du système CRISPR1-cas des souches appartenant au

ST17 mais pas pour celui des souches appartenant au ST23. De plus, comme cela a été montré dans le travail sur l'étude des voies de signalisation de CovR, il semble exister des spécificités de régulation en fonction des lignées phylogénétiques. Nous avons donc reproduit des expériences d'interférence et d'acquisition de spacers comme décrites précédemment, avec les différents mutants du système CovRS à notre disposition.

Nous avons ainsi électroporé le plasmide pG1^{ts} vide et le plasmide porteur du protospacer homologue du spacer en 5' du locus CRISPR1 (pG1^{ts}-protospacer5') dans les mutants de substitution de CovR (CovR^{D53A}), dans les mutants de délétion de *covR* (Δ*covR*) et dans les mutants de substitution de CovS (Covs^{T282A}). Nous avons pu constater que l'interférence restait possible dans tous les mutants, qu'ils soient dans un fond génétique ST17 ou ST23, comme pour les souches sauvages, puisque nous n'avons obtenu aucun transformant avec le plasmide vecteur du protospacer (Figure 59).

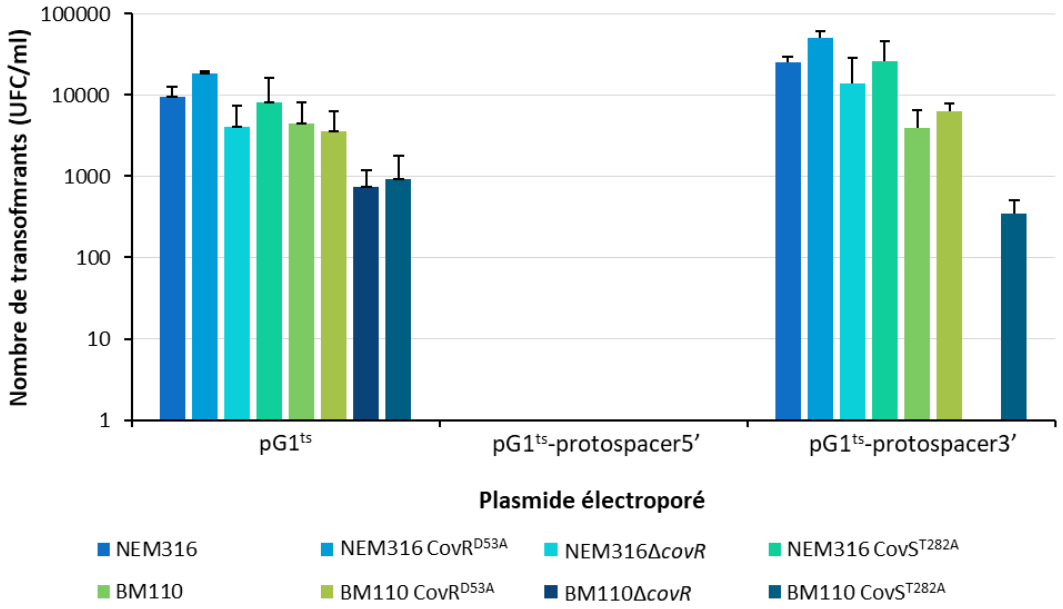
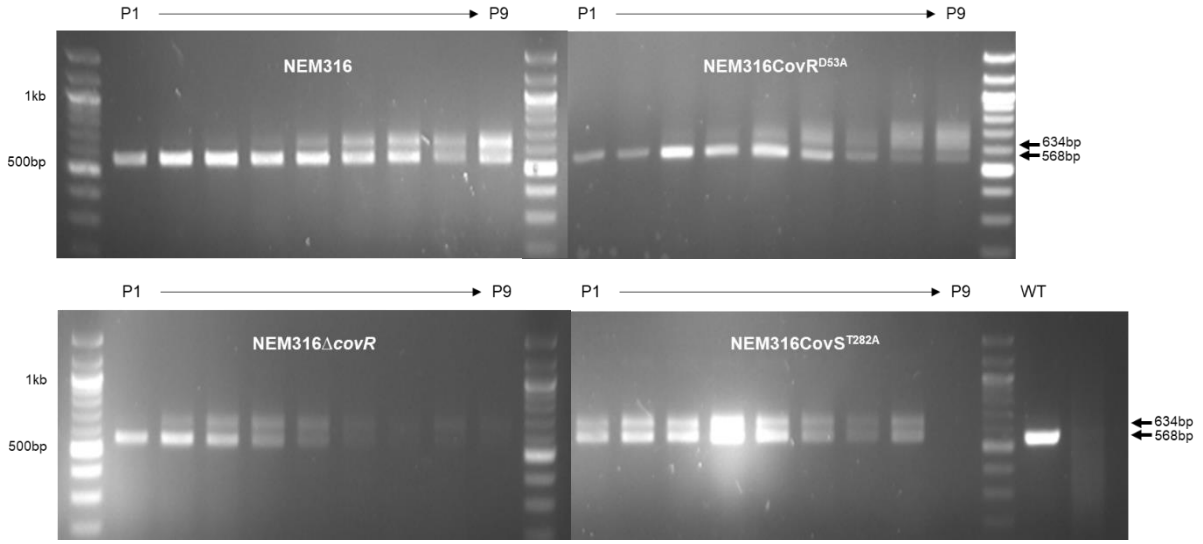


Figure 59 : L'absence de CovR améliore la capacité d'interférence de *S. agalactiae* BM110 (ST17).
La capacité d'interférence a été déterminée par comparaison du nombre de transformants obtenus après électroporation d'un vecteur vide (pG1^{ts}) ou d'un vecteur porteur d'un protospacer (pG1^{ts}-protospacer5' et pG1^{ts}-protospacer3'). Les dénombrements ont été réalisés 3 fois et le graphique présente la moyenne de ces triplicatas.

En revanche, lorsque nous avions réitéré les expériences chez les souches sauvages avec le plasmide portant le protospacer homologue du spacer en 3' du locus CRISPR1, nous avions constaté une perte de fonctionnalité du système, qui ne semblait plus capable de reconnaître ou de cliver le plasmide avec le protospacer (Figure 45). Nous avons obtenu des résultats similaires dans le cas présent, avec un nombre de transformants sensiblement identiques après électroporation du vecteur

vide ou électroporation du plasmide vecteur du protospacer (pG1^{ts}-protospacer3'), sauf pour le mutant BM110Δ*covR* (Figure 59). Pour ce dernier, aucun transformant n'a été obtenu en culture sur milieu sélectif après électroporation du vecteur du protospacer, témoignant d'une reconnaissance et d'un clivage du plasmide par la machinerie enzymatique Cas. Ainsi l'absence de CovR semble améliorer les capacités d'interférence du système CRISPR1-Cas des souches appartenant au ST17.

Nous avons ensuite réalisé de nouveaux « challenges » plasmidiques avec le plasmide pNZ123, afin d'étudier la capacité d'acquisition d'un nouveau spacer par le système chez les différents mutants. Pour rappel, après l'amplification du locus CRISPR1, celle-ci se traduit par la coexistence de deux tailles de locus dans la population totale, l'une correspondant à la population ayant acquis un spacer supplémentaire et l'autre à la population sauvage. L'acquisition d'un nouveau spacer a ainsi pu être détectée dans l'ensemble des souches appartenant au ST23, témoignant d'un système fonctionnel pour la souche sauvage NEM316 et ses dérivées (Figure 60).



<u>Figure 60 :</u> Absence d'impact du système CovRS sur la capacité d'acquisition des souches du ST23.
Cette figure représente la migration électrophorétique des produits de PCR obtenus après amplification du locus CRISPR1, à différents passages, de la souche *S. agalactiae* NEM316 (ST23) et de ses dérivées, mutées pour le système CovRS, après « challenge » avec pNZ123. Le puits WT correspond au locus CRISPR1 natif de NEM316, utilisé comme témoin. Sa taille est de 568 pb. L'acquisition d'une nouvelle unité repeat-spacer (+66 pb) se traduit par la présence d'une deuxième bande à 634 pb. Pour chaque souche le premier puits représente le premier passage (P1) et le dernier puits, le neuvième passage (P9).

En ce qui concerne les souches appartenant au ST17, la capacité d'acquisition de la souche sauvage BM110 ayant déjà été déterminée dans la première partie de ce travail, nous avons repris ici, uniquement les résultats concernant les mutants du système CovRS (Figure 61). Ainsi la souche BM110Δ*covR* est apparue capable d'acquérir de nouveaux spacers, tout comme la souche sauvage, à

la différence du mutant de « surexpression » de CovR (BM110 CovS^{T282A}) et du mutant de CovR non fonctionnel (BM110 CovR^{D53A}). En effet, pour ces deux derniers, aucune acquisition de spacer supplémentaire n'a pu être détectée dans nos expériences. Cette fois encore, l'absence de CovR semble permettre un fonctionnement optimal du système CRISPR1-Cas.

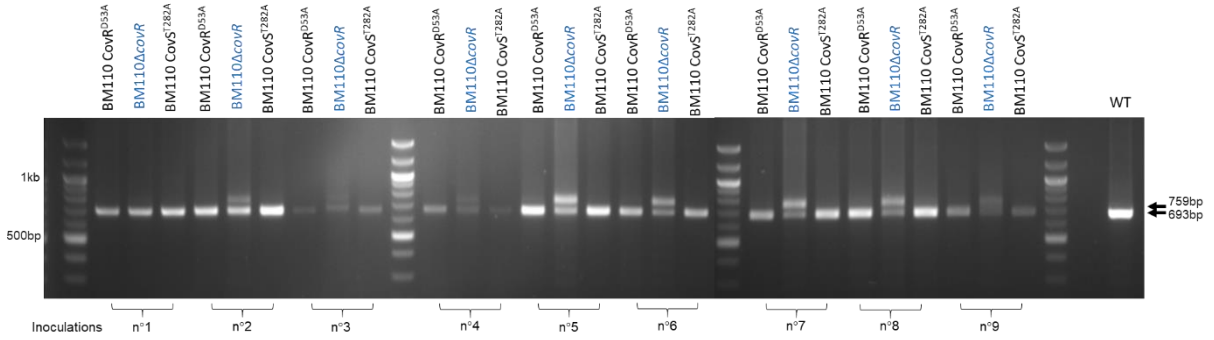


Figure 61 : Restauration de la capacité d'acquisition du système CRISPR1-Cas de *S. agalactiae* BM110 (ST17) en l'absence de CovR.
Le dernier puits (WT) correspond au locus CRISPR1 natif de *S. agalactiae* BM110, utilisé comme témoin. Sa taille est de 693 pb. Après « challenge » avec le plasmide pNZ123, l'acquisition d'une nouvelle unité repeat-spacer (+66 pb) dans le locus CRISPR1, se traduit par la présence d'une deuxième bande à 759 pb. Pour chaque inoculation, le premier puits représente la souche BM110 CovR^{D53A}, le deuxième puits représente BM110Δ*covR* et le troisième puits représente BM110 CovS^{T282A}.

En conclusion, CovR ne semble pas jouer de rôle dans la régulation du système CRISPR1-Cas des souches appartenant au ST23, probablement en lien avec l'absence de rôle de P$_{srn036}$ chez ces souches également. A l'inverse CovR, qui est connu pour avoir un rôle répresseur sur les promoteurs qu'il cible, semble pouvoir réprimer P$_{srn036}$ chez les souches de ST17, puisque la fonctionnalité du système ST17 est améliorée lorsque la répression est levée, en l'absence de CovR.

2.1.2. CovR réprime la transcription des gènes *cas* chez les souches du ST17

Nous avons vérifié notre hypothèse en comparant l'activité promotrice de P$_{srn036}$ en présence et en absence de CovR. En effet, d'après nos résultats précédemment obtenus, P$_{srn036}$ semble être le P$_{2\text{-}cas}$ des souches appartenant au ST17. Ainsi nous avons électroporé le plasmide contenant P$_{srn036}$ en amont du gène rapporteur *lacZ* dans la souche BM110 sauvage et dans la souche délétée de *covR*, puis nous avons déterminé l'activité de la β-galactosidase pour chacune des souches, après avoir retranché l'activité obtenue avec le vecteur vide. Cette activité s'élevait en moyenne à 755 unités arbitraires pour la souche BM110 sauvage (Figure 62A). Par contre, lorsque P$_{srn036}$ a été électroporé chez BM110Δ*covR*, l'activité spécifique de la β-galactosidase était en moyenne de 3000 unités arbitraires, soit une activité de P$_{srn036}$ environ 4 fois plus importante que dans la souche sauvage. Ceci semble donc confirmer l'effet répresseur de CovR sur P$_{srn036}$.

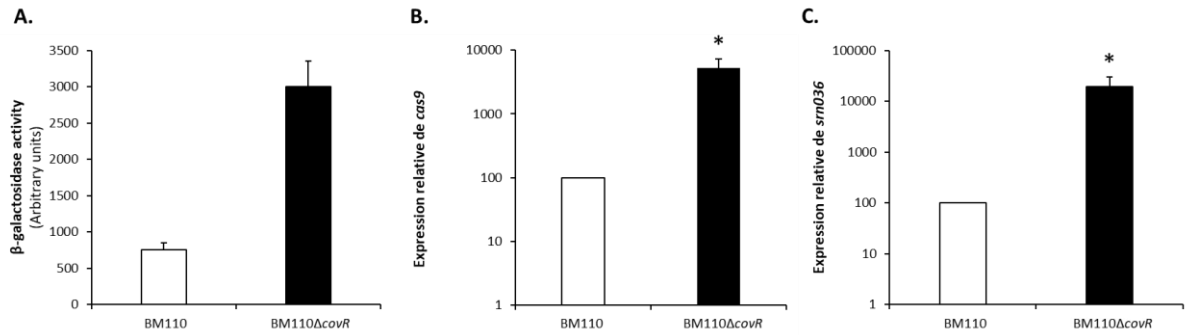


Figure 62 : Effet répresseur de CovR sur P_{srn036}.
A. Détermination de l'activité promotrice de P_{srn036} dans les différents fonds génétiques par mesure de l'activité β-galactosidase.
B. Expression relative de *cas9* chez BM110Δ*covR* par rapport à BM110 (WT).
C. Expression relative de *srn036* chez BM110Δ*covR* par rapport à BM110 (WT).
A. B. et **C.** Chaque graphique présente la moyenne des résultats obtenus au cours de trois expériences.
B. et **C.** La comparaison statistique des moyennes a été réalisée avec un test-t (*p<0,05).

Enfin, P_{srn036} permettant la transcription des gènes *cas* chez les souches appartenant au ST17, nous avons étudié la transcription de *cas9* par PCR quantitative chez le mutant BM110Δ*covR*. *cas9* est apparu largement surexprimé chez la souche BM110Δ*covR* par rapport à la souche sauvage (Figure 62B). Ceci corrobore le résultat précédemment obtenu lors de la détermination de l'activité promotrice de P_{srn036}. Ceci semble également confirmer que P_{srn036} est bien responsable de la transcription de *cas9* chez les souches appartenant au ST17 et que les effets observés lors de la mutation du promoteur sont bien dus à l'absence de promoteur et non à la mutation du site de fixation de CovR. En effet chez le mutant BM110ΔP_{srn036}, *cas9* n'est plus du tout transcrit, alors qu'il est surexprimé chez le mutant BM110Δ*covR*.

Par ailleurs, nous avons également dessiné un couple d'amorces permettant la quantification de *srn036* par qPCR. Comme attendu, nous avons démontré que *srn036* était lui aussi largement surexprimé chez le mutant BM110Δ*covR* (Figure 62C).

Nous pouvons ainsi conclure que la fixation de CovR à P_{srn036} entraîne une répression de la transcription de *cas9* et donc régule l'expression du système CRISPR1-Cas dans son ensemble.

2.2. Etude de la transcription de cas9 *à travers l'application de divers stimuli*

Nous avons montré que le système CRISPR1-Cas n'était que peu ou pas transcrit à l'état basal, mais qu'il restait néanmoins fonctionnel dans les conditions expérimentales testées. L'objectif a donc été de comparer son activité dans cet état basal à son activité après l'application de divers stimuli, afin d'identifier des éléments qui pourraient activer le système et éventuellement d'avoir une meilleure

compréhension de sa régulation. Nous avons ainsi testé l'application d'un stress acide, la présence d'EGMs et la présence d'antibiotiques.

2.2.1. Evaluation de l'impact d'un stress acide

S. agalactiae reste à ce jour la première cause d'infections bactériennes materno-fœtales, notamment dû au fait de sa présence en portage dans le tractus génito-urinaire d'environ un tiers des femmes. De plus il a déjà été démontré que les facteurs environnementaux peuvent intervenir, par des mécanismes indirects, dans la régulation des systèmes CRISPR-Cas. Nous avons donc décidé d'étudier l'effet d'un pH acide, mimant celui du tractus vaginal, sur la transcription du système CRISPR1-Cas.

La transcription du gène *cas9* a ainsi été mesurée sur des extraits obtenus après 5 minutes (t=5) et 10 minutes (t=10) d'incubation à 37°C, suivant l'application du stress acide ou du milieu liquide neutre, représentant la condition de référence. Après avoir normalisé les résultats obtenus en stress acide à ceux obtenus en condition de référence, nous avons calculé l'expression relative de *cas9* pour chacune des souches testées (Figure 63). Aucune différence significative de transcription du gène *cas9* n'a été montrée dans les conditions testées, quel que soit le temps d'application du stress et quelle que soit la souche. Le pH acide ne semble donc pas jouer de rôle, direct ou indirect, sur la transcription des gènes *cas* chez *S. agalactiae*.

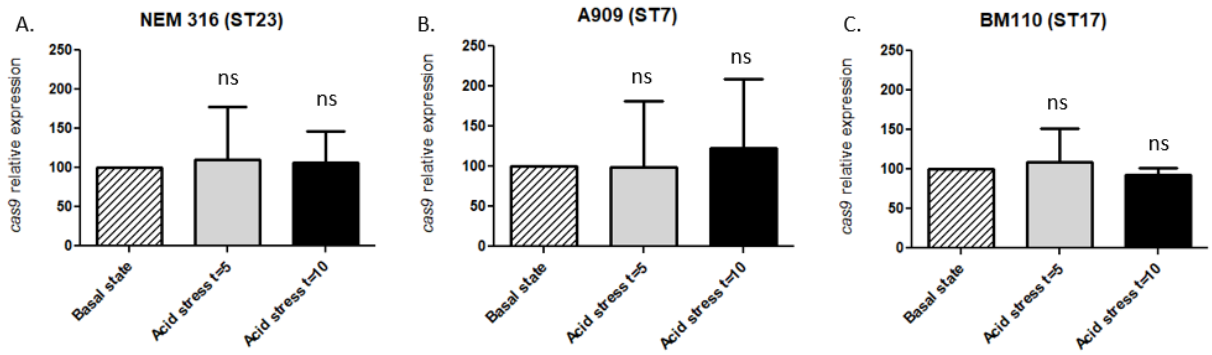


Figure 63 : Expression relative du gène *cas9* après application d'un « stress acide ».
La figure représente l'expression relative du gène *cas9* dans la condition « stress acide » à t=5 minutes (en gris clair) et à t=10 minutes (en noir), comparativement à la condition « pH neutre » de la même souche prise comme condition de référence, chez *S. agalactiae* NEM316 (**A**), A909 (**B**) et BM110 (**C**). Les mesures ont été répétées 3 fois. La comparaison statistique des moyennes a été réalisée avec un test-t (ns : non significatif).

2.2.2. Evaluation de l'impact de la présence d'un EGM

Ces expériences ont été réalisées dans l'objectif de déterminer si la présence d'un élément étranger, connu ou inconnu de la souche, pouvait stimuler le système CRISPR1-Cas et donc augmenter

la transcription du gène *cas9*. Pour cela, la transcription de *cas9* a été mesurée en présence d'un plasmide vide (pG1^{ts}, élément étranger inconnu) et en présence d'un plasmide contenant un protospacer spécifique de la souche testée (pG1^{ts}-protospacer, élément étranger connu). Les résultats ont été comparés à différentes conditions, correspondant aux différentes étapes nécessaires *in vitro* pour introduire un plasmide chez *S. agalactiae*, afin de déterminer le rôle potentiel de chacune d'entre elles (Figure 64).

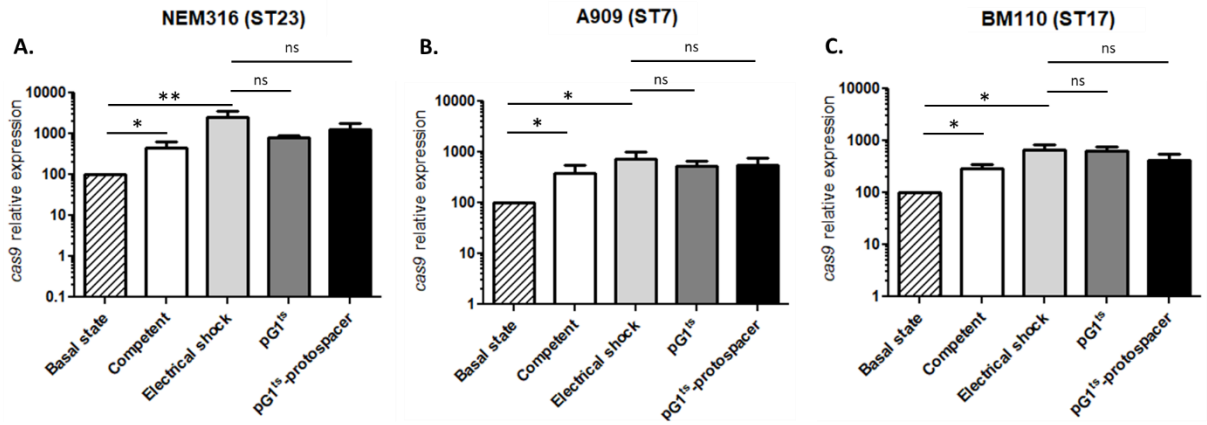


Figure 64 : Evaluation de l'impact de la présence d'un plasmide sur la transcription de *cas9*.
Chaque histogramme représente l'expression relative de *cas9* dans une condition test particulière (compétente, choc électrique, plasmide vide et plasmide avec protospacer) par rapport à la condition de référence, chez *S. agalactiae* NEM316 (**A**), A909 (**B**) et BM110 (**C**). Les mesures ont été répétées 3 fois. La comparaison statistique des moyennes a été réalisée avec un test-t (ns : non significatif ; **p*<0,05 ; ***p*<0,005).

Dans un premier temps, nous avons comparé la présence d'un plasmide à l'application d'un choc électrique, chez la même souche, mais sans ajout de plasmide, celui-ci pouvant potentiellement être responsable d'effets non négligeables sur la transcription des gènes bactériens. Dans ces conditions, nous n'avons pas mis en évidence de différence de transcription du gène *cas9*, que la bactérie soit en contact avec un plasmide vide ou contenant un spacer spécifique, en comparaison à l'application d'un choc électrique. L'ajout d'un élément étranger, qu'il soit déjà connu ou non par la souche, ne semble pas engendrer d'augmentation de la transcription du gène *cas9*, dans les conditions testées.

Nous nous sommes ensuite interrogés sur l'impact potentiel du choc électrique et du traitement glycine/saccharose utilisé pour rendre les bactéries compétentes, étapes indispensables *in vitro* pour la transformation avec un plasmide, mais inexistantes *in vivo*. Pour cela, nous avons comparé la transcription de *cas9* dans ces deux conditions à la condition de référence, qui correspond à la croissance de la bactérie en bouillon TH jusqu'à la mi-phase exponentielle. Nous avons ainsi montré que le traitement contenant le mélange glycine/saccharose impacte légèrement la transcription de *cas9*, qui est environ 4 fois plus transcrit dans cette condition comparativement à la condition de

référence, quelle que soit la souche étudiée. De plus, l'application du choc électrique aux parois bactériennes entraîne une surexpression de *cas9*, environ 25 fois plus importante chez NEM316 et 8 fois plus importante chez A909 et BM110, par rapport à la condition de référence. Ce résultat apparaît difficilement extrapolable, étant donné que le choc électrique appliqué est une condition artificielle.

2.2.3. Evaluation de l'impact d'antibiotiques ciblant la paroi

Comme nous avons mis en évidence qu'une action sur la paroi bactérienne, sous la forme d'un traitement chimique, pouvait modifier la transcription de *cas9*, et comme cela a déjà été démontré par ailleurs, nous avons souhaité tester d'autres stimuli pouvant agir à ce niveau. Nous avons ainsi décidé de comparer la transcription de *cas9* lors de la croissance en bouillon TH (condition de référence) à la croissance en présence d'antibiotiques ciblant la paroi. Nous avons choisi d'étudier la pénicilline G et la vancomycine.

Dans un premier temps nous avons déterminé les concentrations subinhibitrices des deux antibiotiques sur les 3 souches testées, afin de déstabiliser les bactéries, sans impacter leur croissance. Nous avons obtenu des concentrations subinhibitrices de 0,008 µg/mL pour la pénicilline G et de 0,25 µg/mL pour la vancomycine. Nous avons ensuite étudié la transcription de *cas9* après croissance des différentes souches en milieu liquide, avec ou sans antibiotique et incubées à 37°C, jusqu'à atteinte de la phase exponentielle (Figure 65). Ni la présence de pénicilline G, ni la présence de vancomycine, n'ont entraîné de différence significative de transcription de *cas9*, quelle que soit la souche étudiée.

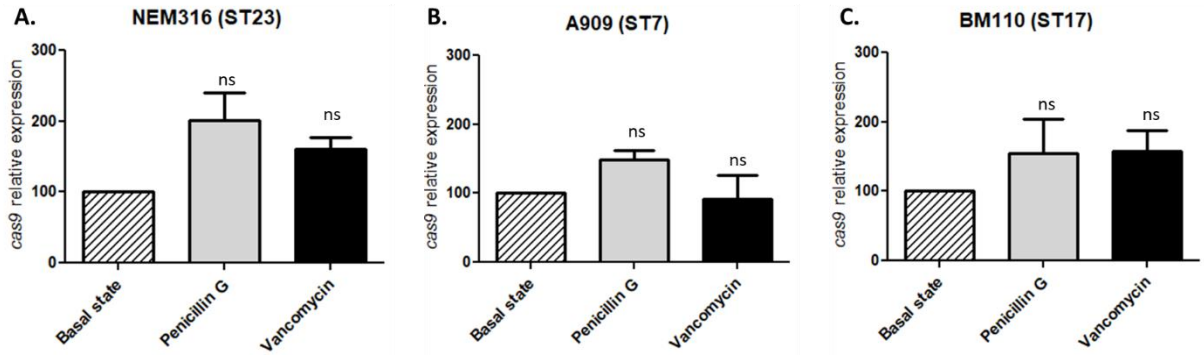
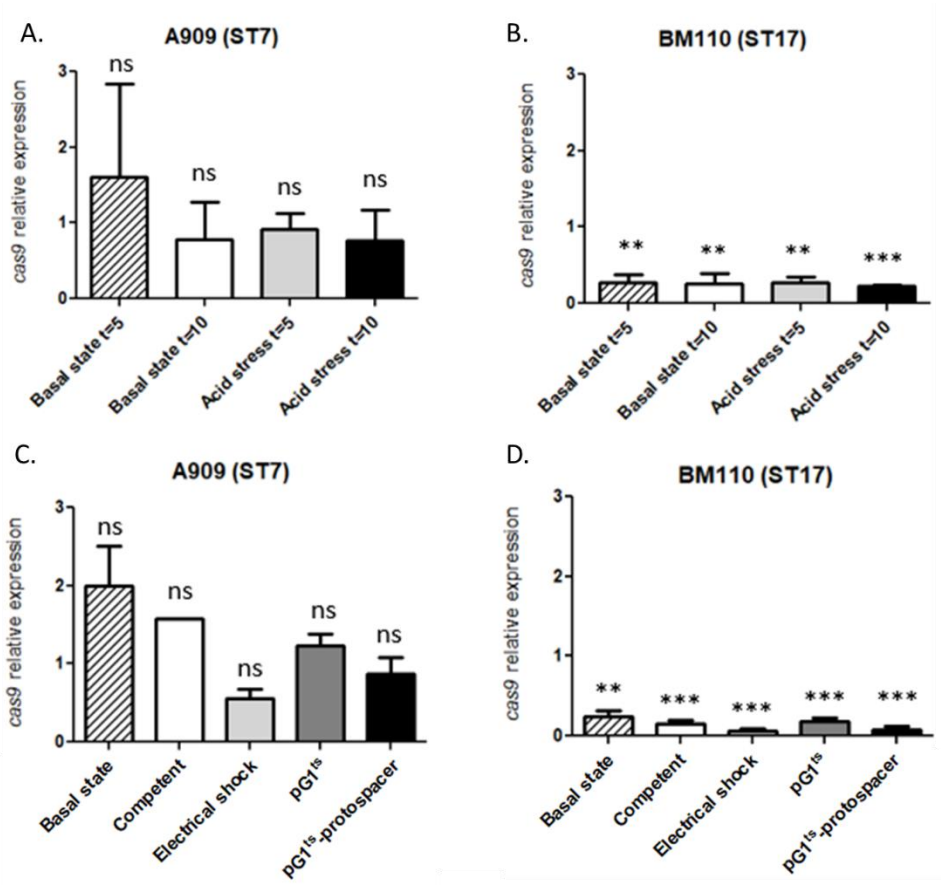


Figure 65 : Effet des antibiotiques sur la transcription de *cas9*.
Les graphiques représentent l'expression relative du gène *cas9* en présence d'antibiotique (pénicilline G ou vancomycine) dans le milieu de culture, comparativement à la condition de croissance de référence, sans antibiotique, chez *S. agalactiae* NEM316 (**A**), A909 (**B**) et BM110 (**C**). Les mesures ont été répétées 3 fois. La comparaison statistique des moyennes a été réalisée avec un test-t (ns : non significatif).

2.2.4. Comparaison entre les souches

Nous avons par la suite comparé l'expression du gène *cas9* entre les souches, dans plusieurs des conditions testées (stress acide et présence d'EGM) pour voir notamment si une différence était retrouvée. Ces résultats n'ont pas montré de différence de transcription significative entre NEM316 (ST23) et A909 (ST7), quelle que soit la condition testée (Figure 66). En revanche, *cas9* est systématiquement moins transcrit chez BM110 (ST17) par rapport à NEM316 (ST23) et son expression est 4 à 20 fois plus faible lors de l'application d'un des stimuli (stress acide, compétente, choc électrique, plasmide vide, plasmide avec protospacer). Ces résultats sont en accord avec les résultats observés dans la première partie de ce travail, qui portaient sur l'étude de la transcription de *cas9* lors de la croissance en bouillon TH et qui avaient mis en évidence une sous-expression de *cas9* chez les souches hypervirulentes de ST17 par rapport aux souches de ST23 et ST110.



<u>Figure 66 :</u> Expression relative du gène *cas9* chez A909 et BM110 par rapport à NEM316.
La figure représente l'expression relative du gène *cas9* après l'application de divers stimuli : stress acide (**A** et **B**), compétente, choc électrique, plasmide vide et plasmide avec protospacer (**C** et **D**) chez *S. agalactiae* A909 (**A** et **C**) et BM110 (**B** et **D**) par rapport à NEM316. Les graphiques montrent les résultats de 3 mesures. La comparaison statistique des moyennes a été réalisée avec un test-t (ns : non significatif ; **p*<0,05 ; ***p*<0,005 ; ****p*<0,0005).

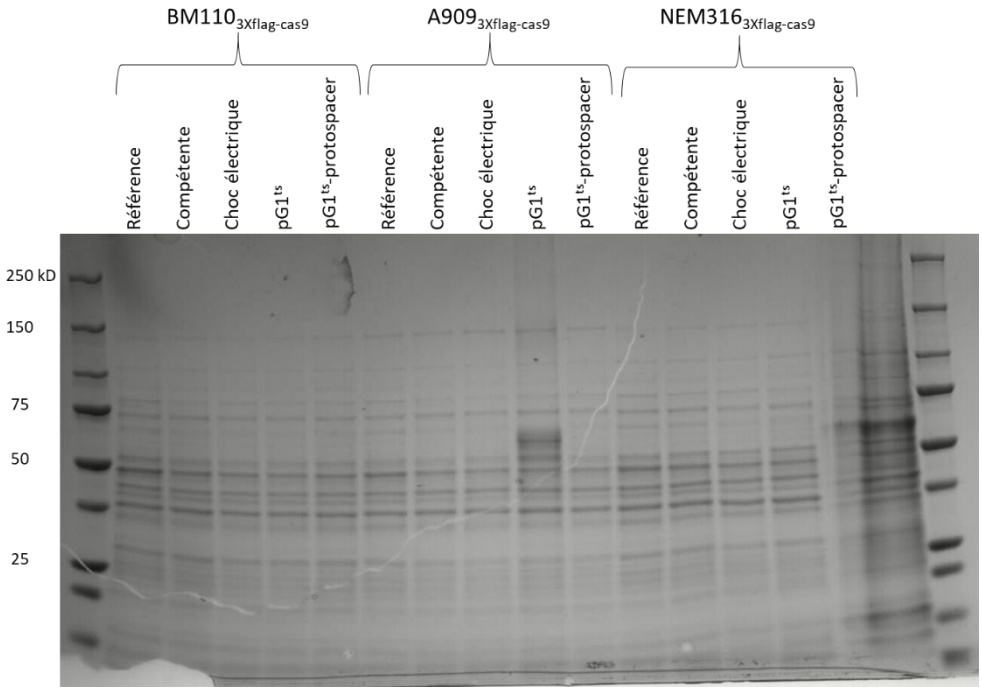
2.3. Etude de l'expression de Cas9

Nous avons également souhaité étudier la production de Cas9 en réponse à l'application de divers stimuli, en marquant la protéine par un 3Xflag, en position N-terminale.

2.3.1. Validation du marquage de Cas9

Après avoir vérifié par séquençage la présence du 3Xflag dans nos trois fonds génétiques et l'absence de décalage du cadre de lecture, nous avons procédé à la croissance des trois mutants dans différentes conditions. Les conditions testées étaient identiques à celles présentées lors de l'étude de la transcription de *cas9* en présence d'un EGM. Ainsi nous avions la condition de référence en bouillon TH et les 4 conditions tests (compétente, choc électrique, pG1^{ts} et pG1^{ts}-protospacer), pour chacune des trois souches. Après extraction des protéines, nous avons utilisé une technique de Western Blot afin de révéler le marquage de Cas9.

En parallèle de la révélation du marquage par un anticorps anti-3Xflag, nous avons également réalisé une coloration au bleu de Neuhoff (Figure 67). Ceci nous a permis d'apprécier la qualité de nos extractions protéiques et de vérifier que nous avions déposé des quantités de protéines similaires, pour chaque souche et pour chaque condition, afin de ne pas introduire de biais pour l'interprétation du Western Blot.



<u>Figure 67 :</u> Migration des extraits protéiques colorée au bleu de Neuhoff.
Pour chaque souche, le premier puits correspondait à la condition de référence, le deuxième puits à la condition compétente, le troisième puits à la condition choc électrique, le quatrième puits à la condition plasmide vide et le dernier puits à la condition plasmide avec protospacer.

Le Western Blot nous a quant à lui permis de valider le marquage de Cas9 et l'absence de marquage non spécifique ou de bruit de fond parasite (Figure 68). En effet nous avons observé une seule bande marquée, dont la taille, aux environs de 150 kDa, était compatible avec la taille attendue de Cas9 (entre 159 et 160 kDa selon les souches). Nous avions également déposé dans le dernier puits de notre gel un extrait protéique de *Salmonella*, dont la protéine PagN était elle aussi marquée par un 3Xflag, afin de servir de témoin pour la technique. Les résultats du Western Blot n'ont pas montré de différence significative dans la quantité de Cas9 en fonction des conditions testées, chez une même souche. Il ne semble pas non plus exister de différence dans la quantité de Cas9 entre les souches, alors que nous l'avons démontré à plusieurs reprises en PCR quantitative. Ces expériences seront à réitérer.

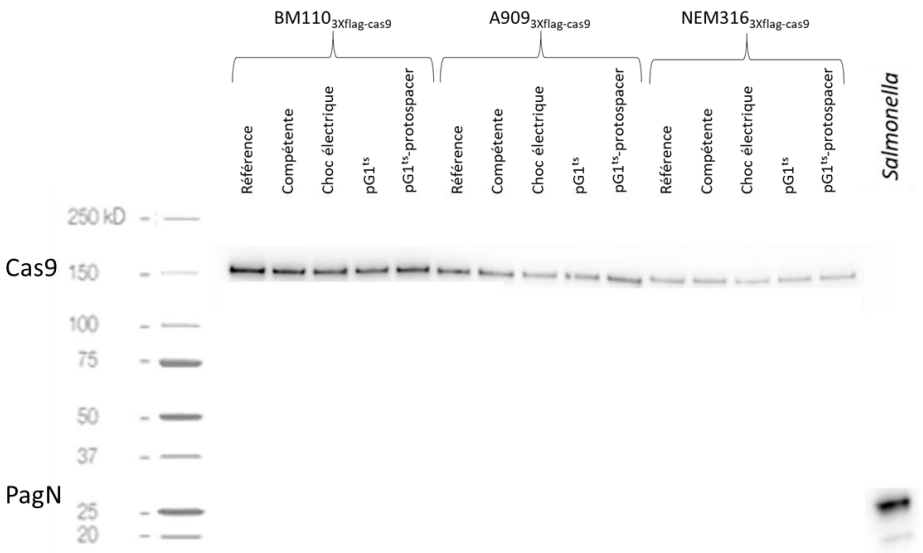
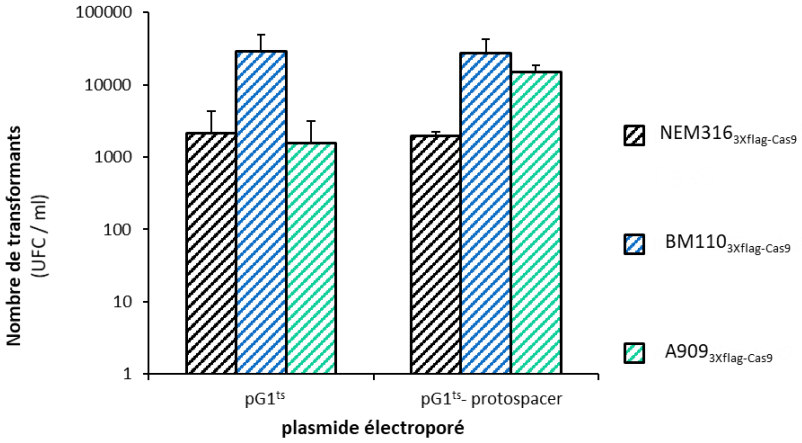


Figure 68 : Evaluation de l'expression de la protéine 3Xflag-Cas9 chez *S. agalactiae*.
Pour chaque souche, le premier puits correspondait à la condition de référence, le deuxième puits à la condition compétente, le troisième puits à la condition choc électrique, le quatrième puits à la condition plasmide vide et le dernier puits à la condition plasmide avec protospacer. Le dernier puits du gel contenait un extrait protéique de *Salmonella*, dont la protéine 3Xflag-PagN, qui a servi de témoin positif de la révélation du marquage.

2.3.2. Tests de fonctionnalité chez les mutants marqués

Par ailleurs, nous avons voulu nous assurer que l'ajout du 3Xflag à l'extrémité N-terminale de Cas9 ne perturbait pas son fonctionnement, étant donné l'importance des nombreux domaines enzymatiques présents sur cette protéine. Pour cela nous avons procédé à des tests d'interférence, selon les mêmes protocoles que ceux utilisés auparavant (Figure 69). Nous avons ainsi remarqué que le système CRISPR1-Cas de l'ensemble des mutants marqués était devenu non fonctionnel, puisque un nombre significatif de transformants a été obtenu en culture après électroporation du plasmide porteur du protospacer. L'ajout du 3Xflag en N-terminal semble donc à l'origine d'un défaut de

fonctionnalité de Cas9, soit par blocage d'un site enzymatique, soit par modification de la conformation de la protéine, soit par encombrement allostérique.



<u>Figure 69 :</u> L'ajout du 3Xflag entraîne une perte de fonctionnalité du système CRISPR1-Cas.
Des expériences d'interférence ont été réalisées avec les souches $NEM316_{3Xflag\text{-}Cas9}$, $A909_{3Xflag\text{-}Cas9}$ et $BM110_{3Xflag\text{-}Cas9}$. Les transformants obtenus après électroporation des plasmides ($pG1^{ts}$ ou $pG1^{ts}$-protospacer) ont été dénombrés 3 fois. La moyenne des triplicatas est représentée sur le graphique.

Discussion

Chez *S. agalactiae*, pathogène néonatal majeur et pathogène émergent chez l'adulte, deux systèmes CRISPR-Cas ont été décrits (Lopez-Sanchez *et al*., 2012), dont le système CRISPR1, de type II-A, soit le type le plus fréquemment retrouvé chez les bactéries commensales et pathogènes de l'Homme (Chylinski e*t al.*, 2014). Ce système se caractérise par la présence de la nucléase Cas9, protéine effectrice majeure, dont le rôle est indispensable pour le fonctionnement du système. Des travaux, portant sur l'étude du système CRISPR1-Cas chez plusieurs souches de *S. agalactiae*, ont montré une très grande diversité des spacers, témoignant de la dynamique de ce système, sans pour autant démontrer la fonctionnalité de l'ensemble des étapes (Lier *et al*., 2015 ; Lopez-Sanchez *et al.*, 2012). De plus, ces travaux ont également mis en évidence un nombre de spacers significativement plus faible pour les souches de CC17, par rapport aux souches classées dans les autres CCs (Lier *et al*., 2015). Or, les souches de ST17, composant très majoritairement le CC17, sont considérées comme un clone hypervirulent chez *S. agalactiae*, du fait de leur prédominance dans les infections néonatales tardives (Poyart *et al*., 2008). Les systèmes CRISPR-Cas peuvent jouer un rôle dans les spécificités des isolats au sein d'une espèce bactérienne, puisqu'ils peuvent limiter les échanges génétiques (Lindenstrauss *et al*., 2011), intervenir dans la virulence (Louwen *et al.,* 2013 ; Louwen *et al.*, 2014), dans la réponse au stress bactérien ou la régulation de gènes (Sampson et Weiss, 2014) et protéger les bactéries contre les bactériophages (Barrangou *et al*., 2007). L'objectif de ce travail était donc d'étudier spécifiquement la fonctionnalité du système CRISPR1-Cas des souches de ST17 par rapport à d'autres lignées phylogénétiques.

Le promoteur du locus CRISPR1, localisé au niveau du leader, étant très conservé entre les souches, nous nous sommes focalisés sur l'étude de l'expression de l'opéron *cas* et en particulier sur la région intergénique située en amont de ce locus. En effet, les gènes *cas* codent des protéines essentielles pour le fonctionnement du système CRISPR-Cas, quelle que soit l'étape où elles sont impliquées (adaptation, expression ou interférence). Par conséquent, l'impact d'une différence dans la composition de cet opéron ou dans l'expression de ces gènes, pourrait être majeur. L'alignement de la région en amont de l'opéron *cas*, sur un total de 331 souches, a permis de retrouver des analogies de séquences, qui peuvent être corrélées à la classification des souches selon leur appartenance aux groupes phylogénétiques définis par MLST. Ainsi, pour les souches appartenant aux CC1 et CC10, qui sont deux CC très proches phylogénétiquement, aucune différence n'a été constatée dans la totalité des 387 pb composant cette région intergénique. De même, un seul SNP a été observé, pour deux STs particuliers, le ST28 et le ST520, au sein du CC19, lui-même très proche des deux CC précédents. Ainsi, l'ensemble du système CRISPR1-Cas a une structure corrélée au MLST, comme cela avait été montré lors de l'analyse détaillée du contenu en spacers du locus CRISPR1 (Lier *et al.*, 2015 ; Lopez-Sanchez *et*

al., 2012). Fait majeur, cet alignement a mis en lumière l'existence de nombreux SNPs et de deux délétions chez les souches appartenant au ST17, réduisant cette région à une longueur de 279 pb. Une des deux délétions, d'une longueur de 103 pb, correspond à la région contenant le promoteur de transcription de l'opéron *cas* identifié chez NEM316 (ST23 ; Rosinski-Chupin *et al.*, 2015), entraînant ainsi sa suppression, alors que c'est une région qui semble plutôt conservée sur l'ensemble des souches des autres lignées phylogénétiques. Ceci pourrait donc éventuellement être à l'origine d'une différence d'expression des gènes *cas*. En effet, en l'absence de promoteur actif, les gènes composant l'opéron *cas* pourraient ne pas être transcrits en ARN messagers, ni traduits en protéines Cas.

Afin d'évaluer l'activité du promoteur de l'opéron *cas,* nous avons réalisé des expériences de fusion transcriptionnelle. Nous avons sélectionné trois souches de référence, représentatives de trois lignées phylogénétiques majeures : NEM316 (ST23 / CC23), 2603V/R (ST110 / CC19) et BM110 (ST17 / CC17). Ces expériences nous ont permis de confirmer que le promoteur identifié chez NEM316 est actif et peut transcrire les gènes *cas*. Par ailleurs, aucune différence significative d'activité n'a été montrée chez la souche 2603V/R (ST110) par rapport à celle du promoteur de NEM316 (ST23), ce qui est en accord avec les résultats précédents, qui ont montré que ces souches ont des séquences promotrices très proches. Par contre, pour la région promotrice de BM110 (ST17), pour laquelle il existe une délétion de 103 pb supprimant le promoteur identifié dans cette région, l'expérience de fusion transcriptionnelle n'a mesuré aucune activité β-galactosidase. En l'absence d'un promoteur actif en amont, *lacZ* ne peut pas être transcrit et par conséquent, ne peut pas être traduit en protéine, comme le prouvent d'ailleurs les résultats obtenus avec le vecteur sans insert. Ainsi, dans les conditions testées, il ne semble pas exister de promoteur actif chez les souches de ST17. Or, si les gènes *cas* ne sont ni transcrits ni traduits, il est difficile d'imaginer que le système CRISPR1-Cas de ces souches soit fonctionnel, étant donné l'importance des protéines Cas à chaque étape du fonctionnement du système.

Afin de préciser ces résultats, l'expression des gènes *cas* a été étudiée par PCR quantitative en temps réel, chez les trois mêmes souches de référence. Ces gènes étant en opéron, seule l'expression de *cas9*, premier gène de l'opéron, a été examinée. Dans un premier temps, nous n'avons pas observé de différence significative de transcription de *cas9* entre la phase exponentielle et la phase stationnaire de croissance, quelle que soit la souche étudiée. Ainsi la densité bactérienne ne semble pas moduler l'expression du système CRISPR1-Cas de *S. agalactiae*, comme cela a pu être démontré dans d'autres espèces ou pour d'autres systèmes (Høyland-Kroghsbo *et al*., 2017 ; Patterson *et al*., 2016). Par ailleurs, aucune différence d'expression de *cas9* n'a pu être mise en évidence entre les souches 2603/R (ST110) et NEM316 (ST23), ce qui est cohérent avec le fait que ces souches ont des séquences promotrices très proches et une activité du promoteur similaire. Par contre, il existe une sous-expression significative

de *cas9* chez la souche BM110 (ST17) par rapport à celle de NEM316 (ST23). Cependant, étant donné la délétion de 103 pb en amont de l'opéron *cas*, entraînant la suppression de P_{cas} et l'absence d'activité β-galactosidase retrouvée lors de l'expérience de fusion transcriptionnelle, le résultat qui semblait le plus probable chez les souches de ST17 était l'absence totale de transcription du gène *cas9*. Or l'analyse des résultats de la RT-qPCR a permis de constater que ce gène était bien transcrit chez ces souches et donc probablement l'opéron dans son ensemble, même s'il semble que ce soit à un niveau très faible. Ainsi, dans nos conditions de culture, c'est-à-dire en l'absence de toute stimulation du système CRISPR1-Cas par un acide nucléique étranger (bactériophage ou plasmide), *cas9* semble être un gène exprimé de façon constitutive, quelle que soit la phase de croissance et quelle que soit la souche étudiée. Ceci est en accord avec des résultats présentés dans d'autres modèles. En effet, beaucoup de protéines Cas sont exprimées de façon constitutive afin d'être toujours disponibles pour jouer leur rôle dans la réponse immunitaire en cas d'infection (Young *et al.*, 2012).

Par la suite, nous avons réalisé une expérience reproduisant une étape d'interférence, destinée à vérifier la fonctionnalité du système en tant que mécanisme de défense anti-EGM. Pour cela, nous avons amplifié le premier spacer du locus CRISPR1 de chacune des souches et nous l'avons cloné sur un plasmide portant une résistance à l'érythromycine, après avoir ajouté le PAM. Le premier constat que nous avons établi est la fonctionnalité du système chez la souche NEM316 (ST23), validant du même coup, la technique employée. En effet, lorsque le plasmide vecteur de protospacer est électroporé chez la souche NEM316 sauvage, aucun transformant n'est obtenu en culture, témoignant du clivage du plasmide. Le même résultat a été obtenu avec la souche sauvage BM110, malgré les différences observées dans la région promotrice *cas*. Les résultats de ces expériences ont été vérifiés chez une autre souche de ST17, COH1, dans l'hypothèse où BM110 présenterait un fonctionnement particulier (données non présentées). Ceci est en accord avec les résultats de qPCR, qui avaient montré la transcription de *cas9* chez les souches de ST17, même si le niveau de transcription était très faible. En effet, la protéine Cas9 est indispensable à cette étape d'interférence. Comme l'ont démontré des expériences antérieures, une bactérie dont le gène *cas9* a été inactivé, est incapable de résister à l'attaque d'un phage, même si son locus CRISPR contient un spacer complémentaire de celui-ci (Barrangou *et al*., 2007). Nous avons d'ailleurs réitéré nos tests de fonctionnalité avec une souche délétée du gène *cas9* (BM110Δ*cas9*) et son complément, qui nous ont permis de confirmer cette essentialité de Cas9 chez *S. agalactiae* également. De plus, si les gènes *cas* semblent exprimés de façon constitutive, des études ont également constaté que la présence d'un EGM pouvait augmenter l'expression des protéines Cas et contribuer ainsi à la stimulation du système (Young *et al*., 2012). Lors de l'expérience de qPCR, réalisée en l'absence de stimulation, nous étions donc dans les conditions d'expression constitutive à bas niveau des gènes *cas*, alors que lors de l'expérience d'interférence, la

présence d'un EGM pourrait stimuler l'expression des protéines Cas, augmentant à leur tour, la fonctionnalité du système.

A l'inverse, lorsque nous avons cloné sur le vecteur le spacer le plus anciennement acquis par la bactérie, c'est-à-dire celui de l'extrémité 3' du locus CRISPR1, le système est apparu non fonctionnel, potentiellement en lien avec une quantité de crRNA trop faible. En effet, certaines études ont montré que les crRNAs les plus abondants dans la cellule correspondaient aux spacers les plus proches du leader (Hale *et al*., 2008). Ceci semble en accord avec le fait que les spacers récemment acquis serviraient à se défendre contre une menace actuelle et doivent donc être facilement disponibles, tandis que les spacers plus anciens pourraient être moins exprimés car dirigés contre des EGMs moins courants dans l'environnement présent. Il est possible également que le nombre de molécules de Cas9 dans la cellule soit insuffisant pour s'associer à tous les crRNAs présents. En effet, *cas9* étant faiblement transcrit, il est probable que l'expression de la protéine correspondante soit faible également et que toutes les copies de la protéine soient déjà occupées par d'autres crRNAs. Ainsi la quantité de Cas9 pourrait être un facteur limitant les capacités d'interférence du système et nous pourrions alors envisager de réitérer nos tests avec un spacer situé vers le milieu du locus, après avoir déterminé la quantité relative de chacun des crRNAs.

Les expériences d'interférence n'ayant pas permis de mettre en évidence de différences entre les lignées phylogénétiques, nous avons poursuivi l'étude de la fonctionnalité du système CRISPR1-Cas par l'étude de l'étape d'adaptation. En effet la diversité des systèmes CRISPR-Cas est en lien direct avec leur capacité d'intégration de nouveaux spacers. Or il a été montré une grande diversité de spacers chez *S. agalactiae* (Lier *et al.*, 2015 ; Lopez-Sanchez *et al*., 2012) mais la démonstration de l'acquisition de nouveaux spacers n'a jamais été réalisée *in vitro*. Les expériences de « challenges » avec des plasmides ont permis de faire cette démonstration, chez la souche de ST17 (BM110) et chez une souche de ST7 (A909). Nous avons remarqué que la capacité d'acquisition de BM110 (ST17) était deux fois plus faible que celle d'A909 (ST7), potentiellement en lien avec la plus faible expression des gènes *cas* et à l'origine du plus faible nombre de spacers constaté chez ces souches. Cependant le nombre de spacers uniques était plus important chez BM110. D'autre part, nous avons constaté que l'acquisition d'un nouveau spacer était parfois accompagnée par la délétion de spacers plus anciens, suggérant que les deux phénomènes pourraient se produire dans le même temps (Deveau *et al*., 2008). Ce phénomène de délétion intra-locus n'a été observé que chez la souche BM110 et, s'il était généralisé à l'ensemble des souches de ST17, pourrait aussi constituer en partie une explication au plus faible nombre de spacers chez ces souches. Comme pour l'étape d'interférence, lorsque Cas9 a été inactivée, cela a conduit à l'impossibilité de détecter de nouveaux événements d'acquisition, confirmant le rôle de Cas9 dans l'étape d'adaptation chez *S. agalactiae*. Ceci avait été décrit

précédemment pour les systèmes de type II-A (Heler *et al.*, 2015 ; Wei, Terns, et Terns, 2015). Autre fait intéressant, l'acquisition de nouveaux spacers n'a pu être observée qu'après « challenge » avec le plasmide pNZ123 et pas avec les autres plasmides testés (pTCV-*lacZ* et pG1^{ts}), ce qui est cohérent avec d'autres études, préconisant l'utilisation d'un plasmide à haut nombre de copies afin d'augmenter la pression sélective sur les bactéries (Hynes, Labrie, et Moineau, 2016).

L'analyse par séquençage ARN, effectuée chez *S. agalactiae* NEM316, a identifié un autre promoteur dans la région intergénique en amont de *cas9*, P$_{srn036}$ (Rosinski-Chupin *et al.,* 2015). Cette analyse a permis d'identifier de façon assez précise la position du site d'initiation de la transcription (TSS) et des boîtes -10 et -35 de P$_{srn036}$. Par contre, la position du terminateur de transcription est plus incertaine et surtout, son efficacité est apparue partielle, ce qui autoriserait la transcription de l'opéron *cas* à partir de P$_{srn036}$ et expliquerait le faible niveau de transcription retrouvé en RT-qPCR pour les souches de ST17. Ceci est cohérent avec les résultats obtenus lors des expériences de fusion transcriptionnelle incluant P$_{srn036}$ en amont de *lacZ*. Celles-ci ont montré l'existence d'une activité β-galactosidase d'une intensité significative avec la souche de ST17, témoignant de la présence d'un promoteur actif. Pour confirmer ces premiers résultats, nous avons construits des souches mutées pour ce promoteur, NEM316ΔP$_{srn036}$ et BM110ΔP$_{srn036}$. Celles-ci nous ont permis de confirmer le rôle essentiel de P$_{srn036}$ pour la fonctionnalité du système CRISPR1-Cas des souches appartenant au ST17. En effet, les capacités d'interférence et d'acquisition ont été abolies chez la souche mutée pour le promoteur, en lien avec l'absence de transcription des gènes *cas*. En revanche, la mutation semble n'avoir eu aucun effet sur la fonctionnalité du système de la souche de ST23. P$_{srn036}$ ne semble rien apporter à ces souches, qui possèdent le promoteur *cas* parfaitement fonctionnel.

Afin de confirmer ce dernier constat, nous avons construit un nouveau mutant, NEM316ΔP$_{cas}$, délété du promoteur P$_{cas}$ mais possédant toujours P$_{srn036}$, à l'instar des souches de ST17. Etonnamment, le système CRISPR1-Cas de ce mutant est apparu défectueux, malgré la présence de P$_{srn036}$, qui semble suffisant pour assurer le fonctionnement du système chez les souches de ST17. Cependant, nous avons tout de même constaté que le nombre de transformants obtenus après électroporation du plasmide avec le protospacer reste plus faible que le nombre de transformants obtenus avec le plasmide vide. Ceci pourrait témoigner d'une activité partielle du système CRISPR1-Cas de la souche NEM316ΔP$_{cas}$, potentiellement due à une faible expression des gènes *cas*, en lien avec l'efficacité partielle du terminateur de transcription de *srn036*. De plus, nous nous sommes focalisés jusqu'à présent sur la délétion de 103 pb qui existe dans la région en amont de l'opéron *cas* chez les souches de ST17, car il s'agit de la différence majeure que nous avons observée, mais lors de l'alignement de cette région, nous avons également montré que ces souches étaient caractérisées par la présence de nombreux SNPs par rapport aux souches des autres lignées. Or ces SNPs pourraient, s'ils étaient localisés au

niveau du terminateur, conduire à une efficacité encore moindre du terminateur voire à l'absence totale de terminaison de transcription chez les souches de ST17. C'est pourquoi nous avons construit le mutant NEM316ΔP$_{cas}$::IG$_{BM110}$, mutant chez lequel ont été apportés 12 des 16 SNPs présents chez BM110, ainsi qu'une délétion de 6pb. Ceux-ci ont permis de restaurer la fonctionnalité du système CRISPR1-Cas de NEM316ΔP$_{cas}$, validant du même coup notre hypothèse selon laquelle les différences observées dans la région intergénique des souches de ST17 pourraient permettre la transcription de P$_{srn036}$ jusqu'aux gènes *cas*. Cependant, lorsque nous avons inséré la région de NEM316ΔP$_{cas}$ dans BM110, le système est resté fonctionnel. Ceci pourrait éventuellement s'expliquer par les SNPs restants en amont de la région interchangée. En effet, après les échanges de régions intergéniques, 4 SNPs étaient toujours présents chez BM110::IG$_{NEM316\Delta Pcas}$ par rapport à NEM316ΔP$_{cas}$, dont les effets pourraient se compenser avec ceux des SNPs présents en aval (Figure 55). D'autre part, l'existence d'un régulateur de transcription, qui agirait spécifiquement chez les souches de ST17, pourrait expliquer les résultats observés. En effet, il semble exister une faible « fuite » de transcription depuis *srn036* chez NEM316ΔP$_{cas}$, qui pourrait devenir suffisante pour permettre un fonctionnement optimal du système, si P$_{srn036}$ était surexprimé. Un rôle cumulé de ces différentes hypothèses pourrait également être envisagé.

Parmi les sources majeures de régulation bactérienne, figurent les systèmes à deux composants (TCS). Ceux-ci sont constitués d'une histidine kinase, qui capte les signaux de l'environnement et d'un régulateur, qui module l'expression des gènes (Hoch, 2000). Parmi la vingtaine de TCSs identifiés chez *S. agalactiae*, le système CovRS, le régulateur majeur de virulence, fait partie des mieux caractérisés à ce jour (Lamy *et al*., 2004). Récemment, la caractérisation du régulon de CovR chez les souches hypervirulentes de ST17 a mis en évidence un site de fixation de CovR au niveau de P$_{srn036}$. L'identification du site de fixation et la fixation de CovR ont été vérifiées par des expériences de DNase I footprinting et de retard sur gel, réalisées par Maria Vittoria Mazzuoli pendant ses travaux de thèse, dans l'équipe Biologie des Bactéries Pathogènes à Gram-Positif de l'Institut Pasteur (Mazzuoli, 2020). Comme nous avons montré l'importance de P$_{srn036}$ pour la fonctionnalité du système CRISPR1-Cas chez les souches de *S. agalactiae* appartenant au ST17, nous avons souhaité connaître le rôle de CovR dans la régulation de l'expression du système. En parallèle, comme le rôle de P$_{srn036}$ semble beaucoup moins essentiel chez les souches appartenant au ST23, nous avons étudié le rôle de CovR dans ces souches également. Dans un premier temps, les expériences d'interférence, réalisées avec l'ensemble des mutants du système CovRS, ont montré que la capacité d'interférence du système CRISPR1-Cas des souches de ST17 était meilleure en l'absence de CovR. En effet, le système est apparu capable de se défendre contre un spacer très anciennement acquis, ce qui n'était pas le cas avec la

souche sauvage. Dans un second temps, nous avons testé l'influence du système CovRS sur la capacité d'intégration de nouveaux spacers. La souche délétée de *covR* est la seule à être apparue capable d'acquérir de nouveaux spacers en réponse au « challenge » plasmidique, alors qu'aucune acquisition n'a pu être détectée chez le mutant de « surexpression » de CovR (BM110 CovS^{T282A}) et le mutant de substitution de CovR (BM110 CovR^{D53A}). CovR semble donc pouvoir réguler l'expression du système CRISPR1-Cas chez les souches de ST17, par le biais de sa fixation à P_{srn036}.

Nous avons confirmé cette hypothèse à l'aide d'expériences de fusion transcriptionnelle. Lorsque le plasmide contenant P_{srn036} en amont de *lacZ* a été électroporé chez BM110Δ*covR*, l'activité de P_{srn036} était 4 fois plus importante que chez la souche sauvage. De même, *cas9* est apparu 50 fois plus transcrit chez la souche délétée de *covR* par rapport à la souche sauvage. Ainsi, CovR réprime la transcription des gènes *cas* à partir de P_{srn036}. Alors que les résultats obtenus avec le mutant BM110 CovS^{T282A} sont concordants avec ce schéma de régulation, étonnamment, le mutant de délétion et le mutant de substitution de CovR ne semblent pas produire les mêmes effets sur P_{srn036}. La forme non phosphorylée semble suffisante pour réprimer P_{srn036}. La phosphorylation favorise la dimérisation de CovR et sa fixation à l'ADN, mais la forme non phosphorylée est tout à fait capable de se fixer à l'ADN. Ceci a déjà été observé auparavant (Jiang *et al*., 2008) et semble dépendre de la combinaison de plusieurs facteurs comme l'affinité de CovR pour ses sites de liaison, les mécanismes de liaison à l'ADN ou l'architecture du promoteur.

Par ailleurs, aucune différence de fonctionnement du système CRISPR1-Cas n'a été observée entre les mutants du système CovRS et la souche sauvage NEM316 (ST23), que ce soit dans la capacité d'interférence du système ou dans sa capacité d'acquisition de nouveaux spacers. Ces résultats sont tout à fait concordants avec l'absence de rôle de P_{srn036} chez les souches de cette lignée phylogénétique. Ainsi la régulation de la transcription du système CRISPR1-Cas de *S. agalactiae* par CovRS, le régulateur majeur de la virulence, semble spécifique aux souches appartenant au ST17, lui-même qualifié d'hypervirulent au sein de l'espèce. Cette spécificité de régulation, dépendante des lignées a d'ailleurs déjà été montrée (Jiang *et al*., 2008). De plus, deux études récentes ont montré que Cas9 elle-même pouvait intervenir dans la régulation de plusieurs facteurs impliqués dans l'adhérence de la bactérie aux épithéliums, le franchissement des barrières physiologiques et plus généralement dans la virulence de *S. agalactiae*, dont une étude réalisée justement chez des souches appartenant au ST17 (Ma *et al*., 2018 ; Spencer *et al*., 2019). Par conséquent, CovRS, en régulant le niveau de production de Cas9 au sein de la cellule, pourrait participer à l'hypervirulence des souches de *S. agalactiae* de ST17.

D'autre part, nous avons également émis l'hypothèse que la régulation de la transcription de *cas9* par CovR chez les souches appartenant au ST17, puisse permettre à la cellule de prévenir une certaine forme d'auto-immunité. En effet, un nouveau mécanisme de régulation de l'expression des gènes *cas* a récemment été mis en évidence chez *S. pyogenes* (Workman *et al*., 2021). Celui-ci fait intervenir une forme longue du tracrRNA, *tracr-L*, qui est capable de jouer le rôle d'ARN guide et dont la cible est le promoteur des gènes *cas*. En effet, *tracr-L* est capable de se fixer à P_{cas}, entraînant ainsi une diminution de la transcription des gènes *cas,* par Cas9 elle-même et prévenant l'auto-immunité due à une trop grande quantité de protéines Cas dans la cellule. Alors que ce mécanisme régulateur pourrait exister chez les souches appartenant au ST23, chez les souches de ST17, la délétion des 103 pb en amont de l'opéron *cas* supprime cette possibilité d'auto-régulation. On peut alors imaginer que CovR puisse jouer un rôle similaire à celui de *tracr-L* sur la transcription des gènes *cas*. Cette hypothèse sera testée par des expériences de compétition en milieu liquide entre la souche sauvage BM110 et les différents mutants du système CRISPR1-Cas et du système CovRS.

En dehors de toute stimulation, le système CRISPR1-Cas de *S. agalactiae* apparaît faiblement transcrit et en particulier dans la lignée ST17. Malgré cela, nous avons montré qu'il restait parfaitement fonctionnel. Or, nous venons de mettre en évidence un mécanisme pouvant réguler l'expression du système CRISPR1-Cas de *S. agalactiae*. De plus, l'importance de la régulation du système, en fonction des conditions dans lesquelles se trouvent les bactéries, afin de maintenir leur intégrité ou leur permettre d'infecter un hôte de manière plus efficace, a déjà été montrée (Louwen *et al.*, 2014). Nous avons ainsi sélectionné divers stimuli pour étudier leur impact sur l'expression du système, chez trois souches représentatives de différentes lignées phylogénétiques, NEM316 (ST23), A909 (ST7) et BM110 (ST17). Nous avons choisi d'étudier la réponse du gène *cas9*, acteur majeur du système de type II-A de *S. agalactiae*. En premier lieu, nous avons comparé la transcription du système en pH neutre et en pH acide, ce dernier mimant le pH vaginal, lieu de colonisation par *S. agalactiae*. De plus, il a déjà été montré que l'activation de CovR pouvait être influencée par des variations physiologiques, dont un pH acide (Park, Jiang, et Wessels, 2012). Ainsi, il était intéressant de connaître l'impact de cet environnement particulier sur les souches de ST17, comparativement aux autres STs. Ces expériences ne nous ont pas permis d'observer de différence significative dans la transcription du système en pH acide par rapport au pH neutre, quelle que soit la souche. Le pH ne semble donc pas influencer sur la transcription des gènes *cas*, que ce soit par le biais de CovR ou par un autre mécanisme de régulation. Toutefois, ces expériences mériteraient d'être réitérées après un temps d'application du stress acide plus long, de l'ordre de 30 minutes, en se basant sur les expériences réalisées par Park *et al*. en 2012.

Ensuite, nous nous sommes intéressés à la transcription du gène *cas9* en réponse à la présence d'éléments génétiques mobiles étrangers. Pour cela nous avons utilisé le plasmide pG1^{ts} vide, jouant le rôle d'un EGM jamais rencontré par la bactérie et le plasmide pG1^{ts}-protospacer, portant un protospacer spécifique du locus CRISPR1 de la souche étudiée et mimant donc un EGM déjà rencontré. Pour la condition « absence de plasmide », les bactéries ont tout de même été rendues compétentes et un choc électrique leur a été appliqué, afin de s'affranchir d'un éventuel impact de ces actions. La présence de ces EGM, qu'ils soient connus ou inconnus, n'a eu aucun impact sur la transcription de *cas9* par rapport à la condition de référence « choc électrique » et cela pour les trois souches testées. Nous avons alors souhaité connaître l'impact de ce choc électrique et du traitement glycine/saccharose par rapport à la croissance de la bactérie en bouillon TH, sans aucun stimulus. Les résultats obtenus ont montré un impact important de l'application de ce choc sur la transcription de *cas9*. La perméabilisation transitoire des membranes bactériennes par l'application du choc s'apparente à un stress important pour la bactérie. De même le traitement chimique par glycine/saccharose, visant à fragiliser les membranes bactériennes, impacte de façon plus modérée la transcription de *cas9*. Ainsi, la présence d'un EGM, par rapport à l'application du choc électrique seul ou du traitement chimique, n'engendre pas d'augmentation supplémentaire de la transcription du système *in vitro*. Cependant, nous ne pouvons pas conclure de manière formelle que cette rencontre bactérie/EGM *in vivo* ne stimule pas le système puisque le choc électrique et le traitement chimique peuvent potentiellement masquer cette stimulation. Nous pourrions également nous interroger sur les effets de la glycine seule ou du saccharose seul sur la transcription de *cas9*, et sur l'implication du système CovRS dans ces voies de stimulation, des études ayant déjà montré une modulation de la réponse du système CovRS en présence d'importantes concentrations de glucose (Palo *et al*., 2013).

Enfin, nous avons étudié la réponse du système CRISPR1-Cas en présence d'antibiotiques. En se basant sur les résultats précédemment décrits et obtenus, nous avons choisi de tester des antibiotiques dont le mode d'action cible la synthèse de la paroi bactérienne. Nous avons ainsi utilisé la pénicilline G et la vancomycine, à doses subinhibitrices, afin de créer un stress pour la bactérie mais sans être bactéricide. Ces conditions ne nous ont pas permis de mettre en évidence de différence dans la transcription de *cas9* en présence des antibiotiques, quelle que soit la souche étudiée. En revanche, nous avons systématiquement remarqué que *cas9* était significativement moins transcrit chez la souche de ST17 (BM110), par rapport à la souche de ST23 (NEM316), dans l'ensemble des conditions testées, ce qui est en accord avec les résultats de la PCR quantitative effectuée en dehors de toute stimulation. Nous pourrions également étudier la transcription des gènes *cas* suite à l'application de plusieurs stimuli simultanément. Récemment, il a été montré que les antibiotiques bactériostatiques favorisaient l'immunité antivirale de bactéries infectées par un phage. En effet, les antibiotiques

bactériostatiques, en ralentissant la croissance bactérienne, ralentissent également la multiplication du phage, ceci laissant plus d'opportunités à la bactérie d'acquérir de nouveaux spacers, dirigés contre le phage (Dimitriu *et al*., 2021).

Dans une dernière partie, nous avons souhaité nous intéresser à l'étude de la traduction de la protéine Cas9, la régulation de l'expression du système CRISPR1-Cas pouvant être post-transcriptionnelle. Ainsi, il semble important d'essayer de quantifier la traduction de la protéine pour voir si des différences peuvent être observées, qu'elles confirment les résultats obtenus pour la transcription ou fassent apparaître de nouvelles différences. Pour cela nous avons ajouté un marquage 3Xflag à la protéine Cas9 chez les trois souches de référence de *S. agalactiae*. Le Western Blot réalisé par la suite nous a permis de valider la spécificité de notre marquage. Néanmoins, il ne nous a pas permis d'observer de différence significative d'expression de Cas9 en fonction des conditions testées, ni des souches étudiées. De plus, le marquage 3Xflag semble être à l'origine d'une perte de fonctionnalité du système CRISPR1-Cas, bien que la fonctionnalité du système ne soit pas complètement indispensable pour notre objectif d'étudier spécifiquement l'expression de la protéine. Quoi qu'il en soit, nous pourrions envisager la possibilité de refaire nos constructions en introduisant le marquage 3Xflag à l'extrémité C-terminale. Dans ce cas-là, il faudra veiller à ajouter un nouveau RBS dans notre construction, l'insertion du flag engendrant l'éloignement du RBS présent en fin de *cas9* de l'ATG de *cas1*. Nous pourrions envisager également l'utilisation d'un linker protéique, permettant d'éloigner le 3Xflag de la protéine et d'éviter qu'il gène ou modifie le repliement de la protéine (Sabourin *et al*., 2007). Enfin, étant donné la conservation des protéines Cas9 dans les systèmes de type II-A et la relative proximité entre les espèces *S. pyogenes* et *S. agalactiae*, l'utilisation des anticorps commerciaux monoclonaux anti-SpCas9 devra être envisagée, afin d'étudier l'expression de la protéine Cas9 chez *S. agalactiae,* après l'application des divers stimuli.

Nous nous interrogeons également sur la coexistence possible de plusieurs systèmes de défense anti-EGMs dans les souches de *S. agalactiae* et sur leur possibilité d'agir en coopération. En effet, comme mentionné précédemment, les bactéries ont développé des mécanismes de défense variés, majoritairement dirigés contre les phages, mais pouvant agir également contre d'autres acides nucléiques étrangers. Une étude a déjà montré que les systèmes de Restriction-Modification et les systèmes CRISPR-Cas étaient capables d'agir conjointement et ainsi, d'augmenter la résistance globale de la bactérie à l'EGM ciblé par les deux systèmes (Dupuis *et al*., 2013). Or, nous avons rencontré à plusieurs reprises, lors de différentes expériences de clonage, des difficultés de transformation des souches appartenant au ST17 avec différents plasmides. De plus, lors des expériences d'interférence,

nous avons constaté que les efficacités de transformation chez BM110 (ST17) avec le plasmide vide étaient systématiquement inférieures à celles observées avec la souche NEM316 (ST23). Ainsi, la souche de ST17 apparaît naturellement moins permissive et l'absence de transformant en culture après électroporation du protospacer chez la souche BM110, pourrait également être due à l'action cumulée de plusieurs mécanismes de défense. Cette absence de permissivité, observée à plusieurs reprises chez les souches de ST17, incite à chercher l'existence d'autres systèmes anti-plasmide, ou plus généralement anti-EGM, dans cette espèce.

Conclusion

Chez *S. agalactiae*, pathogène majeur en médecine humaine et en particulier en période néonatale, deux systèmes CRISPR-Cas ont été décrits. Du fait de son polymorphisme et de sa présence chez la quasi-totalité des souches, un seul semble dynamique. Il s'agit du système CRISPR1-Cas, qui est un système de type II-A, type fréquemment retrouvé chez les bactéries commensales et pathogènes humaines. Des travaux antérieurs ont permis d'observer des différences dans la composition de ce système en fonction du groupe phylogénétique auquel les souches appartiennent. Ainsi, les souches appartenant au ST17, qui constituent un clone hypervirulent au sein de l'espèce *S. agalactiae*, sont caractérisées par un nombre significativement plus faible de spacers au sein de leur locus CRISPR1. Cette moindre diversité pourrait être due à un système CRISPR-Cas moins fonctionnel, ce qui pourrait également constituer un avantage pour cette lignée, la rendant plus permissive à l'acquisition d'EGM, souvent porteurs de facteurs de virulence.

Nous avons utilisé différentes approches méthodologiques complémentaires, afin de comparer la fonctionnalité du système CRISPR1-Cas des souches de ST17 par rapport à celle de souches d'autres lignées phylogénétiques. Dans un premier temps, l'alignement de la région promotrice de l'opéron *cas* a mis en évidence une délétion de 103 pb chez les souches de ST17, entraînant la suppression du promoteur de cet opéron, confirmée par les expériences de fusion transcriptionnelle qui ont suivi. Par contre, l'analyse par RT-PCR quantitative a permis de révéler la transcription de cet opéron *cas*, même si l'expression de *cas9* était à la limite de la détection, dans les conditions testées. Alors que ces résultats confortaient l'hypothèse d'un système CRISPR1-Cas moins actif chez les souches de ST17, les tests de fonctionnalité réalisés par la suite, ont montré que le système a conservé sa capacité de défense vis-à-vis d'un spacer déjà intégré dans le locus CRISPR1 et qu'il est également capable d'acquérir de nouveaux spacers au sein de ce locus.

Ces résultats, qui suggéraient fortement l'intervention d'autres éléments pour le fonctionnement du système, nous ont conduits, en tout premier lieu, à chercher l'existence d'un promoteur alternatif pour la transcription des gènes *cas*. Nous avons identifié P_{srn036}, qui pourrait être renommé $P_{2\text{-}cas}$, situé juste en amont de l'opéron *cas* et conservé dans l'ensemble des lignées phylogénétiques. Les tests réalisés par la suite, visant à caractériser l'implication de P_{srn036} dans la fonctionnalité du système, ont montré des résultats divergents en fonction des lignées phylogénétiques. Ainsi, chez les souches appartenant au ST23, P_{cas} apparaît indispensable au fonctionnement du système et le rôle de P_{srn036} ne semble que très accessoire voire inexistant. A l'inverse, P_{srn036} est essentiel au fonctionnement du système CRISPR1-Cas des souches de ST17, puisque

P_{cas} est absent. Nous avons donc démontré l'existence de spécificités de fonctionnement du système CRISPR1-Cas de *S. agalactiae*, en fonction des lignées phylogénétiques.

Dans un second temps, nous nous sommes intéressés à la régulation de ces systèmes, afin de mieux comprendre leur fonctionnement et apporter éventuellement des éléments de réponse sur la virulence particulière des souches de ST17. Pour cela, deux approches complémentaires ont été étudiées qui sont d'une part, la transcription du gène *cas9* et d'autre part, la traduction de la protéine Cas9, la régulation pouvant s'effectuer à ces deux niveaux. Parmi les divers stimuli testés, l'application d'un certain stress sur la paroi bactérienne, pouvant porter atteinte à son intégrité, a montré une stimulation du système. Cela est en accord avec le principe même du système, dont l'objectif est de permettre à la bactérie de se défendre vis-à-vis des différentes attaques extérieures auxquelles elle peut être soumise. Or, la première étape nécessaire pour un EGM pour entrer dans la bactérie est de se fixer puis de traverser la paroi bactérienne. Nous n'avons pas identifié de condition dans laquelle le système CRISPR1-Cas des souches de ST17 serait surexprimé par rapport aux souches des autres lignées. Au contraire, celui-ci était systématiquement sous-exprimé, dans l'ensemble des conditions testées. Le système des souches de ST17 apparaît donc fonctionnel, mais globalement transcrit à un niveau plus faible, ce qui pourrait avoir un impact sur la virulence observée de ces souches et leur capacité d'adaptation.

Par ailleurs, nous avons montré que le régulateur majeur de virulence CovR était capable de réguler la transcription des gènes *cas* à partir de P_{srn036} et par conséquent, l'expression du système CRISPR1-Cas. Une fois encore, nous avons montré que ce mécanisme de régulation était observé uniquement chez les souches appartenant à la lignée ST17, renforçant leur spécificité au sein de l'espèce *S. agalactiae*. Le CC17, composé d'isolats beaucoup plus homogènes entre eux en comparaison aux autres CCs, constitue un groupe relativement stable au sein de l'espèce *S. agalactiae*. Cette relative stabilité pourrait être due, en partie, à la limitation du transfert horizontal de gènes, à laquelle la perte de P_{cas} pourrait également participer. Ces spécificités de fonctionnement et de régulation des systèmes CRISPR1-Cas vont au-delà de leur rôle de mécanisme de défense anti-phages, semblant les impliquer plus généralement dans l'évolution de l'espèce.

Il semble évident que tous les éléments intervenant dans le fonctionnement et la régulation des systèmes CRISPR-Cas n'ont pas encore été élucidés, comme en témoignent les découvertes perpétuelles à leur propos. Ainsi, l'implication de CovR dans la régulation de l'expression du système CRISPR1-Cas apparaît comme une avancée majeure chez *S. agalactiae*. Enfin, nous nous interrogeons sur le rôle exact du petit ARN Srn036 dans la fonctionnalité du système CRISPR1-Cas. En effet, sa position, en antisens du tracrRNA, n'apparaît pas anodine, d'autant que le tracrRNA est un élément

essentiel au fonctionnement des systèmes CRISPR-Cas de type II-A (Deltcheva *et al*., 2011 ; Heler *et al.*, 2015 ; Jinek *et al.*, 2012). Ceci pourrait suggérer que Srn036 agisse comme un petit ARN régulateur, interagissant avec le tracrRNA et ses fonctions biologiques (Rosinski-Chupin *et al*., 2015), ou qu'il ait un rôle apparenté à celui du scaRNA de *F. novicida*, qui intervient dans la régulation de la virulence de la bactérie (Ratner *et al.*, 2019 ; Sampson *et al*., 2013). Ceci constituerait alors la découverte d'un nouveau mécanisme de régulation du système CRISPR1-Cas de *S. agalactiae* et potentiellement d'autres systèmes de type II-A.

Références bibliographiques

Al-Attar S., Westra E. R., Van der Oost J., Brouns S. J. J. « Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes ». *Biol. Chem.* avril 2011. Vol. 392, n°4, p. 277-289.

Armistead B., Oler E., Adams Waldorf K., Rajagopal L. « The double life of group B *Streptococcus*: Asymptomatic colonizer and potent pathogen ». *J. Mol. Biol.* 26 juillet 2019. Vol. 431, n°16, p. 2914-2931.

Baker C. J., Barrett F. F., Gordon R. C., Yow M. D. « Suppurative meningitis due to streptococci of lancefield group B: A study of 33 infants ». *J. Pediatr.* 1 avril 1973. Vol. 82, n°4, p. 724-729.

Ballard M. S., Schønheyder H. C., Knudsen J. D., Lyytikäinen O., Dryden M., Kennedy K. J., Valiquette L., Pinholt M., Jacobsson G., Laupland K. B., Collaborative On behalf of the I. B. S. « The changing epidemiology of group B *Streptococcus* bloodstream infection: a multi-national population-based assessment ». *Infect. Dis.* 3 mai 2016. Vol. 48, n°5, p. 386-391.

Barcaite E., Bartusevicius A., Tameliene R., Kliucinskas M., Maleckiene L., Nadisauskiene R. « Prevalence of maternal group B streptococcal colonisation in European countries ». *Acta Obstet. Gynecol. Scand.* 2008. Vol. 87, n°3, p. 260-271.

Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D. A., Horvath P. « CRISPR provides acquired resistance against viruses in prokaryotes ». *Science*. 23 mars 2007. Vol. 315, n°5819, p. 1709-1712.

Barrangou R., Horvath P. « A decade of discovery: CRISPR functions and applications ». *Nat. Microbiol.* 5 juin 2017. Vol. 2, p. 17092.

Barrangou R., Marraffini L. A. « CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity ». *Mol. Cell*. 24 avril 2014. Vol. 54, n°2, p. 234-244.

Barton L. L., Feigin R. D., Lins R. « Group B beta hemolytic streptococcal meningitis in infants ». *J. Pediatr.* 1 avril 1973. Vol. 82, n°4, p. 719-723.

Beauruelle C., Pastuszka A., Horvath P., Perrotin F., Mereghetti L., Lanotte P. « CRISPR: A useful genetic feature to follow vaginal carriage of group B *Streptococcus* ». *Front. Microbiol.* 11 octobre 2017. Vol. 8, p. 1981.

Beauruelle C., Pastuszka A., Mereghetti L., Lanotte P. « Group B *Streptococcus* vaginal carriage in pregnant women as deciphered by Clustered Regularly Interspaced Short Palindromic Repeat analysis ». *J. Clin. Microbiol.* 1 juin 2018. Vol. 56, n°6, p. 1-9.

Beauruelle C., Treluyer L., Pastuszka A., Cochard T., Lier C., Mereghetti L., Glaser P., Poyart C., Lanotte P. « CRISPR typing increases the discriminatory power of *Streptococcus agalactiae* typing methods ». *Front. Microbiol.* 2021. Vol. 12, p. 675597.

Beres S. B., Musser J. M. « Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome ». *PloS One*. 29 août 2007. Vol. 2, n°8, p. e800.

Bikard D., Barrangou R. « Using CRISPR-Cas systems as antimicrobials ». *Curr. Opin. Microbiol.* juin 2017. Vol. 37, p. 155-160.

Bikard D., Euler C. W., Jiang W., Nussenzweig P. M., Goldberg G. W., Duportet X., Fischetti V. A., Marraffini L. A. « Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials ». *Nat. Biotechnol.* novembre 2014. Vol. 32, n°11, p. 1146-1150.

Bikard D., Marraffini L. A. « Control of gene expression by CRISPR-Cas systems ». *F1000Prime Rep*. 1 novembre 2013. Vol. 5, n°47.

Bingen E., Denamur E., Lambert-Zechovsky N., Aujard Y., Brahimi N., Geslin P., Elion J. « Analysis of DNA restriction fragment length polymorphism extends the evidence for breast milk transmission in *Streptococcus agalactiae* late-onset neonatal infection ». *J. Infect. Dis.* mars 1992. Vol. 165, n°3, p. 569-573.

Biswas I., Gruss A., Ehrlich S. D., Maguin E. « High-efficiency gene inactivation and replacement system for Gram-positive bacteria. » *J. Bacteriol.* 1 juin 1993. Vol. 175, n°11, p. 3628-3635.

Bolotin A., Quinquis B., Sorokin A., Ehrlich S. D. « Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin ». *Microbiology (Reading).* août 2005. Vol. 151, n°Pt 8, p. 2551-2561.

Boyer K. M., Gotoff S. P. « Strategies for chemoprophylaxis of GBS early-onset infections ». *Antibiot. Chemother.* 1985. Vol. 35, p. 267-280.

Brochet M., Couvé E., Zouine M., Vallaeys T., Rusniok C., Lamy M.-C., Buchrieser C., Trieu-Cuot P., Kunst F., Poyart C., Glaser P. « Genomic diversity and evolution within the species *Streptococcus agalactiae* ». *Microbes Infect.* 1 avril 2006. Vol. 8, n°5, p. 1227-1243.

Broeders M., Herrero-Hernandez P., Ernst M. P. T., Ploeg A. T. Van der, Pijnappel W. W. M. P. « Sharpening the molecular scissors: Advances in gene-editing technology ». *iScience*. 24 janvier 2020. Vol. 23, n°1.

Chylinski K., Le Rhun A., Charpentier E. « The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems ». *RNA Biol.* mai 2013. Vol. 10, n°5, p. 726-737.

Chylinski K., Makarova K. S., Charpentier E., Koonin E. V. « Classification and evolution of type II CRISPR-Cas systems ». *Nucleic Acids Res.* juin 2014. Vol. 42, n°10, p. 6091-6105.

Da Cunha V., Davies M. R., Douarre P.-E., Rosinski-Chupin I., Margarit I., Spinali S., Perkins T., Lechat P., Dmytruk N., Sauvage E., Ma L., Romi B., Tichit M., Lopez-Sanchez M.-J., Descorps-Declere S., Souche E., Buchrieser C., Trieu-Cuot P., Moszer I., Clermont D., Maione D., Bouchier C., McMillan D. J., Parkhill J., Telford J. L., Dougan G., Walker M. J., DEVANI Consortium, Holden M. T. G., Poyart C., Glaser P. « *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline ». *Nat. Commun.* 4 août 2014. Vol. 5, p. 4544.

Dang T. N. D., Zhang L., Zöllner S., Srinivasan U., Abbas K., Marrs C. F., Foxman B. « Uropathogenic *Escherichia coli* are less likely than paired fecal *E. coli* to have CRISPR loci ». *Infect. Genet. Evol.* octobre 2013. Vol. 19, p. 212-218.

Deltcheva E., Chylinski K., Sharma C. M., Gonzales K., Chao Y., Pirzada Z. A., Eckert M. R., Vogel J., Charpentier E. « CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III ». *Nature*. 31 mars 2011. Vol. 471, n°7340, p. 602-607.

Deveau H., Barrangou R., Garneau J. E., Labonté J., Fremaux C., Boyaval P., Romero D. A., Horvath P., Moineau S. « Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus* ». *J. Bacteriol.* février 2008. Vol. 190, n°4, p. 1390-1400.

Dimitriu T., Kurilovich E., Lapinska U., Severinov K., Pagliara S., Szczelkun M. D., Westra E. R. « Bacteriostatic antibiotics promote the evolution of CRISPR-Cas immunity ». *bioRxiv*. 4 mai 2021. p. 2021.05.04.442568.

Doran K. S., Liu G. Y., Nizet V. « Group B streptococcal β-hemolysin/cytolysin activates neutrophil signaling pathways in brain endothelium and contributes to development of meningitis ». *J. Clin. Invest.* 1 septembre 2003. Vol. 112, n°5, p. 736-744.

Doran K. S., Nizet V. « Molecular pathogenesis of neonatal group B streptococcal infection: no longer in its infancy ». *Mol. Microbiol.* octobre 2004. Vol. 54, n°1, p. 23-31.

Doudna J. A., Charpentier E. « The new frontier of genome engineering with CRISPR-Cas9 ». *Science*. 28 novembre 2014. Vol. 346, n°6213.

Dower W. J., Miller J. F., Ragsdale C. W. « High efficiency transformation of *E.coli* by high voltage electroporation ». *Nucleic Acids Res.* 11 juillet 1988. Vol. 16, n°13, p. 6127-6145.

Dupuis M.-È., Villion M., Magadán A. H., Moineau S. « CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance ». *Nat. Commun.* 2 juillet 2013. Vol. 4, p. 2087.

Edmond K. M., Kortsalioudaki C., Scott S., Schrag S. J., Zaidi A. K., Cousens S., Heath P. T. « Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis ». *The Lancet*. 11 février 2012. Vol. 379, n°9815, p. 547-556.

Eickhoff T. C., Klein J. O., Daly A. K., Ingall D., Finland M. « Neonatal sepsis and other infections due to group B beta-hemolytic streptococci ». *N. Engl. J. Med.* 10 décembre 1964. Vol. 271, p. 1221-1228.

Fage C., Lemire N., Moineau S. « Delivery of CRISPR-Cas systems using phage-based vectors ». *Curr. Opin. Biotechnol.* avril 2021. Vol. 68, p. 174-180.

Feil E. J., Li B. C., Aanensen D. M., Hanage W. P., Spratt B. G. « eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data ». *J. Bacteriol.* mars 2004. Vol. 186, n°5, p. 1518-1530.

Filleron A., Lombard F., Jacquot A., Jumas-Bilak E., Rodière M., Cambonie G., Marchandin H. « Group B streptococci in milk and late neonatal infections: an analysis of cases in the literature ». *Arch. Dis. Child. - Fetal Neonatal Ed.* 1 janvier 2014. Vol. 99, n°1, p. F41-F47.

Firon A., Tazi A., Cunha V. D., Brinster S., Sauvage E., Dramsi S., Golenbock D. T., Glaser P., Poyart C., Trieu-Cuot P. « The Abi-domain protein Abx1 interacts with the CovS histidine kinase to control virulence gene expression in group B *Streptococcus* ». *PLoS Pathog.* 21 février 2013. Vol. 9, n°2, p. e1003179.

Florindo C., Ferreira R., Borges V., Spellerberg B., Gomes J. P., Borrego M. J. « Selection of reference genes for real-time expression studies in *Streptococcus agalactiae* ». *J. Microbiol. Methods*. septembre 2012. Vol. 90, n°3, p. 220-227.

Franciosi R. A., Knostman J. D., Zimmerman R. A. « Group B streptococcal neonatal and infant infections ». *J. Pediatr.* 1 avril 1973. Vol. 82, n°4, p. 707-718.

Garneau J. E., Dupuis M.-È., Villion M., Romero D. A., Barrangou R., Boyaval P., Fremaux C., Horvath P., Magadán A. H., Moineau S. « The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA ». *Nature*. 4 novembre 2010. Vol. 468, n°7320, p. 67-71.

Gasiunas G., Barrangou R., Horvath P., Siksnys V. « Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria ». *Proc. Natl. Acad. Sci. U. S. A.* 25 septembre 2012. Vol. 109, n°39, p. E2579-E2586.

Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L., Msadek T., Zouine M., Couvé E., Lalioui L., Poyart C., Trieu-Cuot P., Kunst F. « Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease ». *Mol. Microbiol.* septembre 2002. Vol. 45, n°6, p. 1499-1513.

Green M. R., Sambrook J. « Isolation of high-molecular-weight DNA using organic solvents ». *Cold Spring Harb. Protoc.* 4 janvier 2017. Vol. 2017, n°4, p. pdb.prot093450.

Grissa I., Vergnaud G., Pourcel C. « The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats ». *BMC Bioinformatics*. 23 mai 2007. Vol. 8, p. 172.

Gunderson F. F., Cianciotto N. P. « The CRISPR-associated gene *cas2* of *Legionella pneumophila* is required for intracellular infection of amoebae ». *mBio*. 12 mars 2013. Vol. 4, n°2, p. e00074-00013.

Gusa A. A., Gao J., Stringer V., Churchward G., Scott J. R. « Phosphorylation of the group A streptococcal CovR response regulator causes dimerization and promoter-specific recruitment by RNA polymerase ». *J. Bacteriol.* 1 juillet 2006. Vol. 188, n°13, p. 4620-4626.

Hale C., Kleppe K., Terns R. M., Terns M. P. « Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus* ». *RNA*. 12 janvier 2008. Vol. 14, n°12, p. 2572-2579.

Hall J., Adams N. H., Bartlett L., Seale A. C., Lamagni T., Bianchi-Jassir F., Lawn J. E., Baker C. J., Cutland C., Heath P. T., Ip M., Le Doare K., Madhi S. A., Rubens C. E., Saha S. K., Schrag S., Sobanjo-ter Meulen A., Vekemans J., Gravett M. G. « Maternal disease with group B *Streptococcus* and serotype distribution worldwide: Systematic review and meta-analyses ». *Clin. Infect. Dis.* 6 novembre 2017. Vol. 65, n°suppl_2, p. S112-S124.

Hansen S. M., Uldbjerg N., Kilian M., Sørensen U. B. S. « Dynamics of *Streptococcus agalactiae* colonization in women during and after pregnancy and in their infants ». *J. Clin. Microbiol.* janvier 2004. Vol. 42, n°1, p. 83-89.

Hardie J. M., Whiley R. A. « The genus *Streptococcus* ». In : Wood BJB, Holzapfel WH (éd.). *Genera Lact. Acid Bact.* Boston, MA : Springer US, 1995. p. 55-124.

Hatoum-Aslan A., Marraffini L. A. « Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens ». *Curr. Opin. Microbiol.* février 2014. Vol. 17, p. 82-90.

He F., Vestergaard G., Peng W., She Q., Peng X. « CRISPR-Cas type I-A Cascade complex couples viral infection surveillance to host transcriptional regulation in the dependence of Csa3b ». *Nucleic Acids Res.* 28 février 2017. Vol. 45, n°4, p. 1902-1913.

Heler R., Samai P., Modell J. W., Weiner C., Goldberg G. W., Bikard D., Marraffini L. A. « Cas9 specifies functional viral targets during CRISPR-Cas adaptation ». *Nature*. 12 mars 2015. Vol. 519, n°7542, p. 199-202.

Heussler G. E., O'Toole G. A. « Friendly fire: Biological functions and consequences of chromosomal targeting by CRISPR-Cas systems ». *J. Bacteriol.* 15 mai 2016. Vol. 198, n°10, p. 1481-1486.

Hoch J. A. « Two-component and phosphorelay signal transduction ». *Curr. Opin. Microbiol.* avril 2000. Vol. 3, n°2, p. 165-170.

Horvath P., Barrangou R. « CRISPR/Cas, the immune system of bacteria and archaea ». *Science*. 8 janvier 2010. Vol. 327, n°5962, p. 167-170.

Horvath P., Barrangou R. « RNA-guided genome editing à la carte ». *Cell Res.* juin 2013. Vol. 23, n°6, p. 733-734.

Horvath P., Romero D. A., Coûté-Monvoisin A.-C., Richards M., Deveau H., Moineau S., Boyaval P., Fremaux C., Barrangou R. « Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus* ». *J. Bacteriol.* février 2008. Vol. 190, n°4, p. 1401-1412.

Høyland-Kroghsbo N. M., Muñoz K. A., Bassler B. L. « Temperature, by controlling growth rate, regulates CRISPR-Cas activity in *Pseudomonas aeruginosa* ». *mBio*. 21 décembre 2018. Vol. 9, n°6.

Høyland-Kroghsbo N. M., Paczkowski J., Mukherjee S., Broniewski J., Westra E., Bondy-Denomy J., Bassler B. L. « Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system ». *Proc. Natl. Acad. Sci.* 3 janvier 2017. Vol. 114, n°1, p. 131-135.

Hynes A. P., Labrie S. J., Moineau S. « Programming native CRISPR arrays for the generation of targeted immunity ». *mBio*. 03 2016. Vol. 7, n°3, p. e00202-16.

Hynes A. P., Lemay M.-L., Trudel L., Deveau H., Frenette M., Tremblay D. M., Moineau S. « Detecting natural adaptation of the *Streptococcus thermophilus* CRISPR-Cas systems in research and classroom settings ». *Nat. Protoc.* 2017. Vol. 12, n°3, p. 547-565.

Imperi M., Pataracchia M., Alfarone G., Baldassarri L., Orefici G., Creti R. « A multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of *Streptococcus agalactiae* ». *J. Microbiol. Methods*. février 2010. Vol. 80, n°2, p. 212-214.

Ishino Y., Shinagawa H., Makino K., Amemura M., Nakata A. « Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product ». *J. Bacteriol.* décembre 1987. Vol. 169, n°12, p. 5429-5433.

Jansen R., Embden J. D. A. Van, Gaastra W., Schouls L. M. « Identification of genes that are associated with DNA repeats in prokaryotes ». *Mol. Microbiol.* mars 2002. Vol. 43, n°6, p. 1565-1575.

Jauréguy F., Carton M., Teboul J., Butel M.-J., Panel P., Ghnassia J.-C., Doucet-Populaire F. « Risk factors and screening strategy for group B streptococcal colonization in pregnant women: results of a prospective study ». *J. Gynecol. Obstet. Biol. Reprod. (Paris)*. avril 2003. Vol. 32, n°2, p. 132-138.

Jiang S.-M., Ishmael N., Dunning Hotopp J., Puliti M., Tissi L., Kumar N., Cieslewicz M. J., Tettelin H., Wessels M. R. « Variation in the group B *Streptococcus* CsrRS regulon and effects on pathogenicity ». *J. Bacteriol.* mars 2008. Vol. 190, n°6, p. 1956-1965.

Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. « A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity ». *Science*. 17 août 2012. Vol. 337, n°6096, p. 816-821.

Jones N., Bohnsack J. F., Takahashi S., Oliver K. A., Chan M.-S., Kunst F., Glaser P., Rusniok C., Crook D. W. M., Harding R. M., Bisharat N., Spratt B. G. « Multilocus sequence typing system for group B *Streptococcus* ». *J. Clin. Microbiol.* juin 2003. Vol. 41, n°6, p. 2530-2536.

Joubrel C., Tazi A., Six A., Dmytruk N., Touak G., Bidet P., Raymond J., Trieu Cuot P., Fouet A., Kernéis S., Poyart C. « Group B *Streptococcus* neonatal invasive infections, France 2007-2012 ». *Clin. Microbiol. Infect.* octobre 2015. Vol. 21, n°10, p. 910-916.

Katz V., Bowes W. A. « Perinatal group B streptococcal infections across intact amniotic membranes ». *J. Reprod. Med.* mai 1988. Vol. 33, n°5, p. 445-449.

Krohn M. A., Hillier S. L., Baker C. J. « Maternal peripartum complications associated with vaginal group B streptococci colonization ». *J. Infect. Dis.* juin 1999. Vol. 179, n°6, p. 1410-1415.

Kunin V., Sorek R., Hugenholtz P. « Evolutionary conservation of sequence and secondary structures in CRISPR repeats ». *Genome Biol.* 18 avril 2007. Vol. 8, n°4, p. R61.

Labrie S. J., Samson J. E., Moineau S. « Bacteriophage resistance mechanisms ». *Nat. Rev. Microbiol.* mai 2010. Vol. 8, n°5, p. 317-327.

Lamy M.-C., Zouine M., Fert J., Vergassola M., Couve E., Pellegrini E., Glaser P., Kunst F., Msadek T., Trieu-Cuot P., Poyart C. « CovS/CovR of group B *Streptococcus*: a two-component global regulatory system involved in virulence ». *Mol. Microbiol.* 1 décembre 2004. Vol. 54, n°5, p. 1250-1268.

Lancefield R. C. « A serological differentiation of human and other groups of hemolytic streptococci ». *J. Exp. Med.* 1 avril 1933. Vol. 57, n°4, p. 571-595.

Lancefield R. C., Freimer E. H. « Type-specific polysaccharide antigens of group B streptococci ». *J. Hyg. (Lond.)* juin 1966. Vol. 64, n°2, p. 191-203.

Le Doare K., Heath P. T. « An overview of global GBS epidemiology ». *Vaccine*. 28 août 2013. Vol. 31 Suppl 4, p. D7-12.

Le Gallou B. *Colonisation et infection materno-foetales à Streptococcus agalactiae : étude épidémiologique au CHRU de Tours de 2000 à 2018*. Université de Tours, 2020.

Libster R., Edwards K. M., Levent F., Edwards M. S., Rench M. A., Castagnini L. A., Cooper T., Sparks R. C., Baker C. J., Shah P. E. « Long-term outcomes of group B streptococcal meningitis ». *Pediatrics*. juillet 2012. Vol. 130, n°1, p. e8-15.

Lier C., Baticle E., Horvath P., Haguenoer E., Valentin A.-S., Glaser P., Mereghetti L., Lanotte P. « Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages ». *Front. Genet.* 2015. Vol. 6, p. 214.

Lin T.-L., Pan Y.-J., Hsieh P.-F., Hsu C.-R., Wu M.-C., Wang J.-T. « Imipenem represses CRISPR-Cas interference of DNA acquisition through H-NS stimulation in *Klebsiella pneumoniae* ». *Sci. Rep.* 17 août 2016. Vol. 6, p. 31644.

Lin W.-J., Walthers D., Connelly J. E., Burnside K., Jewell K. A., Kenney L. J., Rajagopal L. « Threonine phosphorylation prevents promoter DNA binding of the group B *Streptococcus* response regulator CovR ». *Mol. Microbiol.* mars 2009. Vol. 71, n°6, p. 1477-1495.

Lindenstrauss A. G., Pavlovic M., Bringmann A., Behr J., Ehrmann M. A., Vogel R. F. « Comparison of genotypic and phenotypic cluster analyses of virulence determinants and possible role of CRISPR elements towards their incidence in *Enterococcus faecalis* and *Enterococcus faecium* ». *Syst. Appl. Microbiol.* décembre 2011. Vol. 34, n°8, p. 553-560.

Lopez-Sanchez M.-J., Sauvage E., Da Cunha V., Clermont D., Ratsima Hariniaina E., Gonzalez-Zorn B., Poyart C., Rosinski-Chupin I., Glaser P. « The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome ». *Mol. Microbiol.* septembre 2012. Vol. 85, n°6, p. 1057-1071.

Louwen R., Horst-Kreft D., De Boer A. G., Van der Graaf L., De Knegt G., Hamersma M., Heikema A. P., Timms A. R., Jacobs B. C., Wagenaar J. A., Endtz H. P., Van der Oost J., Wells J. M., Nieuwenhuis E. E. S., Van Vliet A. H. M., Willemsen P. T. J., Van Baarlen P., Van Belkum A. « A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré syndrome ». *Eur. J. Clin. Microbiol. Infect. Dis.* février 2013. Vol. 32, n°2, p. 207-226.

Louwen R., Staals R. H. J., Endtz H. P., Van Baarlen P., Van der Oost J. « The role of CRISPR-Cas systems in virulence of pathogenic bacteria ». *Microbiol. Mol. Biol. Rev.* mars 2014. Vol. 78, n°1, p. 74-88.

Ma K., Cao Q., Luo S., Wang Z., Liu G., Lu C., Liu Y. « *cas9* enhances bacterial virulence by repressing the *regR* transcriptional regulator in *Streptococcus agalactiae* ». *Infect. Immun.* 1 mars 2018. Vol. 86, n°3, p. e00552-17.

Mackow N. A., Shen J., Adnan M., Khan A. S., Fries B. C., Diago-Navarro E. « CRISPR-Cas influences the acquisition of antibiotic resistance in *Klebsiella pneumoniae* ». *PLoS One*. 2019. Vol. 14, n°11, p. e0225131.

Madrid L., Seale A. C., Kohli-Lynch M., Edmond K. M., Lawn J. E., Heath P. T., Madhi S. A., Baker C. J., Bartlett L., Cutland C., Gravett M. G., Ip M., Le Doare K., Rubens C. E., Saha S. K., Sobanjo-ter Meulen A., Vekemans J., Schrag S., Group For the I. G. D. I., Agarwal R., Da Silva A. R. A., Bassat Q., Berkley J. A., Dangor Z., Dhaded S., Giannoni E., Hammoud M., Kobayahsi M., O'Sullivan C., Sakata H., Sridhar S., Sigaúque B., Tyrrell G., Paul V. « Infant group B streptococcal disease incidence and serotypes worldwide: Systematic review and meta-analyses ». *Clin. Infect. Dis.* 6 novembre 2017. Vol. 65, n°suppl_2, p. S160-S172.

Maisey H. C., Doran K. S., Nizet V. « Recent advances in understanding the molecular basis of group B *Streptococcus* virulence ». *Expert Rev. Mol. Med.* 22 septembre 2008. Vol. 10, p. e27.

Makarova K. S., Grishin N. V., Shabalina S. A., Wolf Y. I., Koonin E. V. « 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 ». *Biol. Direct*. 16 mars 2006. Vol. 1, p. 7.

Makarova K. S., Haft D. H., Barrangou R., Brouns S. J. J., Charpentier E., Horvath P., Moineau S., Mojica F. J. M., Wolf Y. I., Yakunin A. F., Van der Oost J., Koonin E. V. « Evolution and classification of the CRISPR-Cas systems ». *Nat. Rev. Microbiol.* 2011. Vol. 9, n°6, p. 467-477.

Makarova K. S., Wolf Y. I., Iranzo J., Shmakov S. A., Alkhnbashi O. S., Brouns S. J. J., Charpentier E., Cheng D., Haft D. H., Horvath P., Moineau S., Mojica F. J. M., Scott D., Shah S. A., Siksnys V., Terns M. P., Venclovas Č., White M. F., Yakunin A. F., Yan W., Zhang F., Garrett R. A., Backofen R., Van der Oost J., Barrangou R., Koonin E. V. « Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants ». *Nat. Rev. Microbiol.* février 2020. Vol. 18, n°2, p. 67-83.

Mali P., Esvelt K. M., Church G. M. « Cas9 as a versatile tool for engineering biology ». *Nat. Methods*. octobre 2013. Vol. 10, n°10, p. 957-963.

Mali P., Yang L., Esvelt K. M., Aach J., Guell M., DiCarlo J. E., Norville J. E., Church G. M. « RNA-guided human genome engineering via Cas9 ». *Science*. 15 février 2013. Vol. 339, n°6121, p. 823-826.

Manning S. D., Springman A. C., Lehotzky E., Lewis M. A., Whittam T. S., Davies H. D. « Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada ». *J. Clin. Microbiol.* avril 2009. Vol. 47, n°4, p. 1143-1148.

Mazzuoli M. V. *The CovR signaling pathway in hypervirulent Group B Streptococcus*. Université de Paris, 2020.

McCracken G. H. « Group B streptococci: The new challenge in neonatal infections ». *J. Pediatr.* 1 avril 1973. Vol. 82, n°4, p. 703-706.

McKenzie R. E., Almendros C., Vink J. N. A., Brouns S. J. J. « Using CAPTURE to detect spacer acquisition in native CRISPR arrays ». *Nat. Protoc.* 1 mars 2019. Vol. 14, n°3, p. 976-990.

Medina-Aparicio L., Rebollar-Flores J. E., Gallego-Hernández A. L., Vázquez A., Olvera L., Gutiérrez-Ríos R. M., Calva E., Hernández-Lucas I. « The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in *Salmonella enterica* serovar Typhi ». *J. Bacteriol.* mai 2011. Vol. 193, n°10, p. 2396-2407.

Van der Mee-Marquet N., Fourny L., Arnault L., Domelier A.-S., Salloum M., Lartigue M.-F., Quentin R. « Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites ». *J. Clin. Microbiol.* septembre 2008. Vol. 46, n°9, p. 2906-2911.

Mian G. F., Godoy D. T., Leal C. a. G., Yuhara T. Y., Costa G. M., Figueiredo H. C. P. « Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia ». *Vet. Microbiol.* 14 avril 2009. Vol. 136, n°1-2, p. 180-183.

Miller J. H. *Experiments in molecular genetics*. [s.l.] : Cold Spring Harbor Laboratory, 1972. 494 p.ISBN : 978-0-87969-106-6.

Mojica F. J., Díez-Villaseñor C., Soria E., Juez G. « Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria ». *Mol. Microbiol.* avril 2000. Vol. 36, n°1, p. 244-246.

Mojica F. J. M., Díez-Villaseñor C., García-Martínez J., Almendros C. « Short motif sequences determine the targets of the prokaryotic CRISPR defence system ». *Microbiology*. 1 mars 2009. Vol. 155, n°3, p. 733-740.

Mojica F. J. M., Díez-Villaseñor C., García-Martínez J., Soria E. « Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements ». *J. Mol. Evol.* 1 février 2005. Vol. 60, n°2, p. 174-182.

Morinis J., Shah J., Murthy P., Fulford M. « Horizontal transmission of group B *Streptococcus* in a neonatal intensive care unit ». *Paediatr. Child Health*. juin 2011. Vol. 16, n°6, p. e48-50.

Noble K., Lu J., Guevara M. A., Doster R. S., Chambers S. A., Rogers L. M., Moore R. E., Spicer S. K., Eastman A. J., Francis J. D., Manning S. D., Rajagopal L., Aronoff D. M., Townsend S. D., Gaddy J. A. « Group B *Streptococcus cpsE* is required for serotype V capsule production and aids in biofilm formation and ascending infection of the reproductive tract during pregnancy ». *ACS Infect. Dis.* 2 juin 2021.

Nocard E., Mollereau R. « Sur une mammite contagieuse des vaches laitières ». *Ann. Inst. Pasteur*. 1887. Vol. 1, p. 109-126.

Nozawa T., Furukawa N., Aikawa C., Watanabe T., Haobam B., Kurokawa K., Maruyama F., Nakagawa I. « CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes* ». *PLoS One*. 6 mai 2011. Vol. 6, n°5, p. e19543.

Van der Oost J., Westra E. R., Jackson R. N., Wiedenheft B. « Unravelling the structural and mechanistic basis of CRISPR–Cas systems ». *Nat. Rev. Microbiol.* juillet 2014. Vol. 12, n°7, p. 479-492.

Palmer K. L., Gilmore M. S. « Multidrug-resistant enterococci lack CRISPR-cas ». *mBio*. 12 octobre 2010. Vol. 1, n°4.

Palo B. D., Rippa V., Santi I., Brettoni C., Muzzi A., Metruccio M. M. E., Grifantini R., Telford J. L., Paccani S. R., Soriani M. « Adaptive response of group B *Streptococcus* to high glucose conditions: New insights on the CovRS regulation network ». *PLoS One*. 9 avril 2013. Vol. 8, n°4, p. e61294.

Park S. E., Jiang S., Wessels M. R. « CsrRS and environmental pH regulate group B *Streptococcus* adherence to human epithelial cells and extracellular matrix ». *Infect. Immun.* novembre 2012. Vol. 80, n°11, p. 3975-3984.

Patron K., Gilot P., Rong V., Hiron A., Mereghetti L., Camiade E. « Inductors and regulatory properties of the genomic island-associated *fru2* metabolic operon of *Streptococcus agalactiae* ». *Mol. Microbiol.* 2017. Vol. 103, n°4, p. 678-697.

Patterson A. G., Chang J. T., Taylor C., Fineran P. C. « Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference ». *Nucleic Acids Res.* 13 juillet 2015. Vol. 43, n°12, p. 6038-6048.

Patterson A. G., Jackson S. A., Taylor C., Evans G. B., Salmond G. P. C., Przybilski R., Staals R. H. J., Fineran P. C. « Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems ». *Mol. Cell*. 15 décembre 2016. Vol. 64, n°6, p. 1102-1108.

Perez-Rodriguez R., Haitjema C., Huang Q., Nam K. H., Bernardis S., Ke A., DeLisa M. P. « Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli* ». *Mol. Microbiol.* 2011. Vol. 79, n°3, p. 584-599.

Phares C. R., Lynfield R., Farley M. M., Mohle-Boetani J., Harrison L. H., Petit S., Craig A. S., Schaffner W., Zansky S. M., Gershman K., Stefonek K. R., Albanese B. A., Zell E. R., Schuchat A., Schrag S. J., Active Bacterial Core surveillance/Emerging Infections Program Network. « Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005 ». *JAMA*. 7 mai 2008. Vol. 299, n°17, p. 2056-2065.

Plainvert C., Hays C., Touak G., Joubrel-Guyot C., Dmytruk N., Frigo A., Poyart C., Tazi A. « Multidrug-resistant hypervirulent group B *Streptococcus* in neonatal invasive infections, France, 2007-2019 ». *Emerg. Infect. Dis.* novembre 2020. Vol. 26, n°11, p. 2721-2724.

Pougach K., Semenova E., Bogdanova E., Datsenko K. A., Djordjevic M., Wanner B. L., Severinov K. « Transcription, processing and function of CRISPR cassettes in *Escherichia coli* ». *Mol. Microbiol.* septembre 2010. Vol. 77, n°6, p. 1367-1379.

Pourcel C., Salvignol G., Vergnaud G. « CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies ». *Microbiology*. 1 mars 2005. Vol. 151, n°3, p. 653-663.

Poyart C., Réglier-Poupet H., Tazi A., Billoët A., Dmytruk N., Bidet P., Bingen E., Raymond J., Trieu-Cuot P. « Invasive group B streptococcal infections in infants, France ». *Emerg. Infect. Dis.* octobre 2008. Vol. 14, n°10, p. 1647-1649.

Poyart C., Tazi A., Réglier-Poupet H., Billoët A., Tavares N., Raymond J., Trieu-Cuot P. « Multiplex PCR assay for rapid and accurate capsular typing of group B streptococci ». *J. Clin. Microbiol.* juin 2007. Vol. 45, n°6, p. 1985-1988.

Poyart C., Trieu-Cuot P. « A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to beta-galactosidase in Gram-positive bacteria ». *FEMS Microbiol. Lett.* 15 novembre 1997. Vol. 156, n°2, p. 193-198.

Pul U., Wurm R., Arslan Z., Geissen R., Hofmann N., Wagner R. « Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS ». *Mol. Microbiol.* mars 2010. Vol. 75, n°6, p. 1495-1512.

Pursey E., Dimitriu T., Paganelli F. L., Westra E. R., Houte S. Van. « CRISPR-Cas is associated with fewer antibiotic resistance genes in bacterial pathogens ». *bioRxiv*. 12 avril 2021. p. 2021.04.12.439454.

Ratner H. K., Escalera-Maurer A., Le Rhun A., Jaggavarapu S., Wozniak J. E., Crispell E. K., Charpentier E., Weiss D. S. « Catalytically active Cas9 mediates transcriptional interference to facilitate bacterial virulence ». *Mol. Cell*. 8 août 2019. Vol. 75, n°3, p. 498-510.e5.

Ratner H. K., Sampson T. R., Weiss D. S. « I can see CRISPR now, even when phage are gone: a view on alternative CRISPR-Cas functions from the prokaryotic envelope ». *Curr. Opin. Infect. Dis.* juin 2015. Vol. 28, n°3, p. 267-274.

Ricci M. L., Manganelli R., Berneri C., Orefici G., Pozzi G. « Electrotransformation of *Streptococcus agalactiae* with plasmid DNA ». *FEMS Microbiol. Lett.* 1 juin 1994. Vol. 119, n°1-2, p. 47-52.

Richter C., Chang J. T., Fineran P. C. « Function and regulation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) / CRISPR associated (Cas) systems ». *Viruses*. octobre 2012. Vol. 4, n°10, p. 2291-2311.

Rosinski-Chupin I., Sauvage E., Sismeiro O., Villain A., Da Cunha V., Caliot M.-E., Dillies M.-A., Trieu-Cuot P., Bouloc P., Lartigue M.-F., Glaser P. « Single nucleotide resolution RNA-seq uncovers new regulatory mechanisms in the opportunistic pathogen *Streptococcus agalactiae* ». *BMC Genomics*. 30 mai 2015. Vol. 16, p. 419.

Russell N. J., Seale A. C., O'Driscoll M., O'Sullivan C., Bianchi-Jassir F., Gonzalez-Guarin J., Lawn J. E., Baker C. J., Bartlett L., Cutland C., Gravett M. G., Heath P. T., Le Doare K., Madhi S. A., Rubens C. E., Schrag S., Sobanjo-ter Meulen A., Vekemans J., Saha S. K., Ip M., Group For the G. M. C. I., Asturias E., Gaind R., Kumar P., Anthony B., Madrid L., Bassat Q., Zhu C., Luo M., Nagarjuna D., Majumder S. « Maternal colonization with group B *Streptococcus* and serotype distribution worldwide: Systematic review and meta-analyses ». *Clin. Infect. Dis.* 6 novembre 2017. Vol. 65, n°suppl_2, p. S100-S111.

Sabourin M., Tuzon C. T., Fisher T. S., Zakian V. A. « A flexible protein linker improves the function of epitope-tagged proteins in *Saccharomyces cerevisiae* ». *Yeast Chichester Engl.* janvier 2007. Vol. 24, n°1, p. 39-45.

Safadi R. A., Mereghetti L., Salloum M., Lartigue M.-F., Virlogeux-Payant I., Quentin R., Rosenau A. « Two-component system RgfA/C activates the *fbsB* gene encoding major fibrinogen-binding protein in highly virulent CC17 clone group B *Streptococcus* ». *PLoS One*. 4 février 2011. Vol. 6, n°2, p. e14658.

Sampson T. R., Saroj S. D., Llewellyn A. C., Tzeng Y.-L., Weiss D. S. « A CRISPR/Cas system mediates bacterial innate immune evasion and virulence ». *Nature*. 9 mai 2013. Vol. 497, n°7448, p. 254-257.

Sampson T. R., Weiss D. S. « CRISPR-Cas systems: new players in gene regulation and bacterial physiology ». *Front. Cell. Infect. Microbiol.* 4 avril 2014. Vol. 4, p. 37.

Seale A. C., Bianchi-Jassir F., Russell N. J., Kohli-Lynch M., Tann C. J., Hall J., Madrid L., Blencowe H., Cousens S., Baker C. J., Bartlett L., Cutland C., Gravett M. G., Heath P. T., Ip M., Le Doare K., Madhi S. A., Rubens C. E., Saha S. K., Schrag S. J., Sobanjo-ter Meulen A., Vekemans J., Lawn J. E. « Estimates of the burden of group B streptococcal disease worldwide for pregnant women, stillbirths, and children ». *Clin. Infect. Dis.* 6 novembre 2017. Vol. 65, n°suppl_2, p. S200-S219.

Shariat N., Dudley E. G. « CRISPRs: molecular signatures used for pathogen subtyping ». *Appl. Environ. Microbiol.* janvier 2014. Vol. 80, n°2, p. 430-439.

Six A., Bellais S., Bouaboud A., Fouet A., Gabriel C., Tazi A., Dramsi S., Trieu-Cuot P., Poyart C. « Srr2, a multifaceted adhesin expressed by ST-17 hypervirulent Group B *Streptococcus* involved in binding to both fibrinogen and plasminogen ». *Mol. Microbiol.* septembre 2015. Vol. 97, n°6, p. 1209-1222.

Six A., Joubrel C., Tazi A., Poyart C. « Maternal and perinatal infections to *Streptococcus agalactiae* ». *Presse Med*. juin 2014. Vol. 43, n°6 Pt 1, p. 706-714.

Skoff T. H., Farley M. M., Petit S., Craig A. S., Schaffner W., Gershman K., Harrison L. H., Lynfield R., Mohle-Boetani J., Zansky S., Albanese B. A., Stefonek K., Zell E. R., Jackson D., Thompson T., Schrag S. J. « Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990-2007 ». *Clin. Infect. Dis.* 1 juillet 2009. Vol. 49, n°1, p. 85-92.

Solovyev V., Salamov A. « Automatic annotation of microbial genomes and metagenomic sequences ». In : *Metagenomics Its Appl. Agric. Biomed. Environ. Stud.* New-York : Nova Science Publishers, 2011. p. 61-78.ISBN : 978-1-61668-682-6.

Song J. Y., Lim J. H., Lim S., Yong Z., Seo H. S. « Progress toward a group B streptococcal vaccine ». *Hum. Vaccines Immunother.* 2018. Vol. 14, n°11, p. 2669-2681.

Spencer B. L., Deng L., Patras K. A., Burcham Z. M., Sanches G. F., Nagao P. E., Doran K. S. « Cas9 contributes to group B streptococcal colonization and disease ». *Front. Microbiol.* 2019. Vol. 10, p. 1930.

Stern A., Keren L., Wurtzel O., Amitai G., Sorek R. « Self-targeting by CRISPR: gene regulation or autoimmunity? ». *Trends Genet.* 1 août 2010. Vol. 26, n°8, p. 335-340.

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. « MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0 ». *Mol. Biol. Evol.* 1 décembre 2013. Vol. 30, n°12, p. 2725-2729.

Tazi A., Disson O., Bellais S., Bouaboud A., Dmytruk N., Dramsi S., Mistou M.-Y., Khun H., Mechler C., Tardieux I., Trieu-Cuot P., Lecuit M., Poyart C. « The surface protein HvgA mediates group B *Streptococcus* hypervirulence and meningeal tropism in neonates ». *J. Exp. Med.* 25 octobre 2010. Vol. 207, n°11, p. 2313-2322.

Tazi A., Plainvert C., Anselem O., Ballon M., Marcou V., Seco A., El Alaoui F., Joubrel C., El Helali N., Falloukh E., Frigo A., Raymond J., Trieu-Cuot P., Branger C., Le Monnier A., Azria E., Ancel P.-Y., Jarreau P. H., Mandelbrot L., Goffinet F., Poyart C. « Risk factors for infant colonization by hypervirulent CC17 group B *Streptococcus*: Toward the understanding of late-onset disease ». *Clin. Infect. Dis.* 30 octobre 2019. Vol. 69, n°10, p. 1740-1748.

Tettelin H., Masignani V., Cieslewicz M. J., Donati C., Medini D., Ward N. L., Angiuoli S. V., Crabtree J., Jones A. L., Durkin A. S., Deboy R. T., Davidsen T. M., Mora M., Scarselli M., Margarit y Ros I., Peterson

J. D., Hauser C. R., Sundaram J. P., Nelson W. C., Madupu R., Brinkac L. M., Dodson R. J., Rosovitz M. J., Sullivan S. A., Daugherty S. C., Haft D. H., Selengut J., Gwinn M. L., Zhou L., Zafar N., Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K. J. B., Smith S., Utterback T. R., White O., Rubens C. E., Grandi G., Madoff L. C., Kasper D. L., Telford J. L., Wessels M. R., Rappuoli R., Fraser C. M. « Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome" ». *Proc. Natl. Acad. Sci. U. S. A.* 27 septembre 2005. Vol. 102, n°39, p. 13950-13955.

Tettelin H., Masignani V., Cieslewicz M. J., Eisen J. A., Peterson S., Wessels M. R., Paulsen I. T., Nelson K. E., Margarit I., Read T. D., Madoff L. C., Wolf A. M., Beanan M. J., Brinkac L. M., Daugherty S. C., DeBoy R. T., Durkin A. S., Kolonay J. F., Madupu R., Lewis M. R., Radune D., Fedorova N. B., Scanlan D., Khouri H., Mulligan S., Carty H. A., Cline R. T., Van Aken S. E., Gill J., Scarselli M., Mora M., Iacobini E. T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D., Rinaudo D., Rappuoli R., Telford J. L., Kasper D. L., Grandi G., Fraser C. M. « Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae* ». *Proc. Natl. Acad. Sci. U. S. A.* 17 septembre 2002. Vol. 99, n°19, p. 12391-12396.

Toro M., Cao G., Ju W., Allard M., Barrangou R., Zhao S., Brown E., Meng J. « Association of clustered regularly interspaced short palindromic repeat (CRISPR) elements with specific serotypes and virulence potential of shiga toxin-producing *Escherichia coli* ». *Appl. Environ. Microbiol.* février 2014. Vol. 80, n°4, p. 1411-1420.

Vale P. F., Lafforgue G., Gatchitch F., Gardan R., Moineau S., Gandon S. « Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus* ». *Proc. Biol. Sci.* 7 août 2015. Vol. 282, n°1812, p. 20151270.

Verani J. R., McGee L., Schrag S. J., Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention (CDC). « Prevention of perinatal group B streptococcal disease--revised guidelines from CDC, 2010 ». *MMWR Recomm. Rep.* 19 novembre 2010. Vol. 59, n°RR-10, p. 1-36.

Vuillemin X., Hays C., Plainvert C., Dmytruk N., Louis M., Touak G., Saint-Pierre B., Adoux L., Letourneur F., Frigo A., Poyart C., Tazi A. « Invasive group B *Streptococcus* infections in non-pregnant adults: a retrospective study, France, 2007-2019 ». *Clin. Microbiol. Infect.* janvier 2021. Vol. 27, n°1, p. 129.e1-129.e4.

Wei Y., Chesne M. T., Terns R. M., Terns M. P. « Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus* ». *Nucleic Acids Res.* 18 février 2015. Vol. 43, n°3, p. 1749-1758.

Wei Y., Terns R. M., Terns M. P. « Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation ». *Genes Dev.* 15 février 2015. Vol. 29, n°4, p. 356-361.

Westra E. R., Buckling A., Fineran P. C. « CRISPR–Cas systems: beyond adaptive immunity ». *Nat. Rev. Microbiol.* mai 2014. Vol. 12, n°5, p. 317-326.

Westra E. R., Pul U., Heidrich N., Jore M. M., Lundgren M., Stratmann T., Wurm R., Raine A., Mescher M., Van Heereveld L., Mastop M., Wagner E. G. H., Schnetz K., Van Der Oost J., Wagner R., Brouns S. J. J. « H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO ». *Mol. Microbiol.* septembre 2010. Vol. 77, n°6, p. 1380-1393.

Workman R. E., Pammi T., Nguyen B. T. K., Graeff L. W., Smith E., Sebald S. M., Stoltzfus M. J., Euler C. W., Modell J. W. « A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression ». *Cell*. 4 février 2021. Vol. 184, n°3, p. 675-688.e19.

Yosef I., Goren M. G., Qimron U. « Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli* ». *Nucleic Acids Res.* juillet 2012. Vol. 40, n°12, p. 5569-5576.

Young J. C., Dill B. D., Pan C., Hettich R. L., Banfield J. F., Shah M., Fremaux C., Horvath P., Barrangou R., Verberkmoes N. C. « Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus* ». *PLoS One*. 2012. Vol. 7, n°5, p. e38077.

Zheng Z., Zhang Y., Liu Z., Dong Z., Xie C., Bravo A., Soberón M., Mahillon J., Sun M., Peng D. « The CRISPR-Cas systems were selectively inactivated during evolution of *Bacillus cereus* group for adaptation to diverse environments ». *ISME J.* juin 2020. Vol. 14, n°6, p. 1479-1493.

Valorisation de la recherche

- Publications scientifiques

Beauruelle C, Treluyer L, **Pastuszka A**, Cochard T, Lier C, Mereghetti L, Glaser P, Poyart C, Lanotte P. « CRISPR typing increases discriminatory power of *Streptococcus agalactiae* typing methods ». *Front. Microbiol.* **2021** 19;12:675597

Pastuszka A, Beauruelle C, Camiade E, Rousseau G, Moineau S, Mereghetti L, Horvath P, Lanotte P. « Functional study of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* hypervirulent strains ». CRISPR J. **2021** 4(2):233-242.

Beauruelle C, **Pastuszka A**, Mereghetti L, Lanotte P. **"**Group B *Streptococcus* vaginal carriage in pregnant women as deciphered by Clustered Regularly Interspaced Short Palindromic Repeat analysis". *J Clin Microbiol*. **2018;** 25;56

Beauruelle C, **Pastuszka A**, Horvath P, Perrotin F, Mereghetti L, Lanotte P. "CRISPR: a useful genetic feature to follow vaginal carriage of Group B *Streptococcus*". *Front Microbiol*. **2017** 11;8:1981

- Communications orales

Pastuszka A, 2019. "Typage des souches invasives de *Streptococcus agalactiae* par méthode CRISPR". **39ème Réunion Interdisciplinaire de Chimiothérapie Anti Infectieuse (RICAI), Paris.**

Pastuszka A, 2019. "Functional study of type II-A CRISPR-Cas system of *Streptococcus agalactiae* hypervirulent strains". **Journées d'animation scientifique de la Fédération de Recherche en Infectiologie (FéRI) en Région Centre-Val de Loire, Tours**

- Communications affichées

Lemaire C, Lanotte P, **Pastuszka A**, 2021. "Regulation of *Streptococcus agalactiae* type II-A CRISPR-Cas system". **CRISPR 2021, Paris**

Pastuszka A, Taybi N, Beauruelle C, Lartigue M-F, Mereghetti L, Lanotte P, **2019**. "Typage des souches invasives de *Streptococcus agalactiae* par méthode CRISPR". **Journée recherche Tours Angers, Tours**

Pastuszka A, Camiade E, Rousseau G, Moineau S, Mereghetti L, Lanotte P, **2019**. "Functional study of type II-A CRISPR-Cas system of *Streptococcus agalactiae* hypervirulent strains". **CRISPR 2019, Québec, QC, Canada**

Adeline PASTUSZKA

Le système CRISPR-Cas de type II-A des souches hypervirulentes de *Streptococcus agalactiae* : rôle et spécificités

Résumé

Streptococcus agalactiae, pathogène humain majeur, possède un système CRISPR-Cas de type II-A. Au sein de l'espèce, les souches appartenant au clone hypervirulent ST17, possèdent significativement moins de spacers dans leur locus CRISPR que les souches des autres lignées, ce qui pourrait être la conséquence d'un système CRISPR-Cas moins fonctionnel. Nous avons mis en évidence une délétion dans la région promotrice de l'opéron *cas* chez les souches de ST17 et nous avons évalué son impact sur la transcription des gènes *cas* et plus généralement, sur la fonctionnalité du système. Nous avons ainsi démontré que les capacités d'interférence et d'acquisition de nouveaux spacers des souches de ST17 étaient fonctionnelles *in vitro*, malgré l'absence du promoteur *cas* habituel. Nous avons attribué la fonctionnalité du système CRISPR-Cas des souches ST17 à la présence d'un autre promoteur, P_{srn036}, situé en amont de l'opéron *cas*. De plus, nous avons montré que ce promoteur était régulé par CovR, le régulateur majeur de virulence chez *S. agalactiae*. Nous avons donc démontré l'existence de spécificités dans la fonctionnalité et la régulation du système CRISPR-Cas de *S. agalactiae*, en fonction des lignées phylogénétiques, qui pourraient participer à l'adaptation et à l'évolution de l'espèce.

Mots clés : *Streptococcus agalactiae*, CRISPR-Cas, ST17, hypervirulence, fonctionnalité, régulation

Abstract

Streptococcus agalactiae, a major human pathogen, encodes a type II-A CRISPR-Cas system. Interestingly, strains belonging to the hypervirulent ST17 clone contain significantly fewer spacers in their CRISPR locus than other lineages, which could be the result of a less functional CRISPR-Cas system. Here, we revealed one large deletion in the ST17 *cas* promoter region and we evaluated its impact on the transcription of *cas* genes as well as the functionalities of the CRISPR-Cas system. We demonstrated that Cas9 interference is functional and that the CRISPR-Cas system of ST17 strains can still acquire new spacers *in vitro*, despite the absence of the regular *cas* promoter. We attributed the functionality of the ST17 CRISPR-Cas system to another promoter, P_{srn036}, located upstream *cas* operon. In addition, we demonstrated that this promoter is regulated by CovR, the major regulator of *S. agalactiae* virulence. Thus, we have demonstrated specificities in the *S. agalactiae* CRISPR-Cas system functionality and regulation, depending on phylogenetic lineages, which might participate in species adaptation and evolution.

Keywords : *Streptococcus agalactiae*, CRISPR-Cas, ST17, hypervirulence, functionality, regulation