UNIVERSITÉ DE TOURS

ÉCOLE DOCTORALE SSBCV

UMR 1282 – Infectiologie et Santé Publique

Equipe 5 : Bactéries et Risque Materno-Foetal

THÈSE présentée par :

Clémence BEAURUELLE

soutenue le : **14 juin 2019**

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

Discipline/ Spécialité : Microbiologie / Bactériologie

Locus CRISPR de *Streptococcus agalactiae* : marqueur génétique de la phylogénie de l'espèce et de l'évolution récente des isolats

THÈSE dirigée par :

Pr Lanotte Philippe	PU – PH, Université François Rabelais de Tours

RAPPORTEURS :

Pr Marchandin Hélène	PU – PH, Université de Montpellier
Pr Le Hello Simon	PU – PH, Université de Caen, Normandie

JURY :

Pr Poyart Claire	PU – PH, Université Sorbonne Paris Descartes (Présidente du Jury)
Dr Hervé Virginie	MCF, Université François Rabelais de Tours
Dr Horvath Philippe	Senior Scientist, DuPont Nutrition and Health, Dangé-Saint-Romain
Pr Lanotte Philippe	PU – PH, Université François Rabelais de Tours
Pr Le Hello Simon	PU – PH, Université de Caen, Normandie
Pr Marchandin Hélène	PU – PH, Université de Montpellier

« En essayant continuellement on finit par réussir. Donc : plus ça rate, plus on a de chance que ça marche ». Les Shadoks (Azay-le-Rideau, 2017)

Remerciements

Madame le Professeur Hélène Marchandin,

Vous avez accepté de rapporter et de juger mon travail. Soyez assurée de ma gratitude et de mon profond respect.

Monsieur le Professeur Simon Le Hello,

Merci de me faire l'honneur de juger ma thèse. Soyez assuré de mon profond respect.

Madame le Professeur Claire Poyart,

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

Madame le Docteur Virginie Hervé,

Un grand merci d'avoir accepté de juger mon travail. Soyez assuré de mon respect et de ma reconnaissance.

Monsieur le Docteur Philippe Horvath,

Merci d'avoir accepté de juger mon travail qui constitue pour moi un grand honneur. Merci pour vos précieux conseils. Soyez assuré de mon profond respect.

Monsieur le Professeur Philippe Lanotte,

Merci d'avoir accepté de m'encadrer pour cette thèse. Merci pour ce sujet passionnant. Merci pour votre aide, vos conseils et votre gentillesse. Merci de m'avoir fait confiance. Soyez assuré de ma reconnaissance, de mon estime et de mon profond respect.

Merci à toute l'équipe « Bactéries et risque materno-fœtal » et à l'UMR 1282 – Infectiologie et Santé Publique pour la bonne humeur constante et la richesse des échanges scientifiques.

Merci à toute l'équipe du laboratoire de Bactériologie-Virologie de Bretonneau du CHRU de Tours pour leur soutien technique et amical.

- Monsieur le Professeur Laurent Mereghetti, merci pour votre accueil au sein de vos équipes. Merci de m'avoir permis et aidé à réaliser ce projet. Merci pour vos conseils et votre gentillesse.
- Monsieur le Professeur Goudeau. Merci pour votre soutien et vos conseils.
- Adeline Pastuszka, merci pour ces quelques années passées ensemble et nos échanges sur CRISPR-Cas bien sûr, mais aussi sur bien d'autres choses. Merci pour ta bonne humeur constante, merci pour ton soutien.
- A Clément Lier et Elodie Baticle d'avoir mis à disposition les résultats de leurs travaux.
- A Catherine Gaudy-Graffin et Gaelle Baty, merci pour vos conseils et votre soutien amical.
- Merci aux techniciennes, tout particulièrement à Catherine pour sa gentillesse et son aide (quotidienne ?) en biologie moléculaire. A Elodie, Amélie, Nathalie, Aurore pour les bons moments au laboratoire et en dehors.
- Merci aux internes du service pour les bons moments partagés.

Merci à l'équipe « infections bactériennes et immunité des ruminants », notamment à Thierry Cochard pour son aide et son expertise dans l'analyse des données ; à Franck Biet et Maxime Branger pour le séquençage des génomes complets et leur expertise pour l'interprétation des données.

Merci à toute l'équipe du département de Bactériologie-Virologie et Parasitologie-Mycologie du CHRU de Brest. Merci pour votre soutien constant.

- Monsieur le Professeur Christopher Payan, merci pour votre accueil au sein de votre laboratoire.
- Geneviève Héry-Arnaud, un énome merci pour ton soutien, tant pour ce travail que pour tout le reste. Merci pour l'énergie et la bonne humeur que tu insuffles à tes équipes. C'est un vrai bonheur de travailler avec toi.

- Didier Tandé, un grand merci pour tout ce que tu m'as transmis, tes connaissances et ta passion pour la Bactériologie. Merci pour ton amitié.
- Hervé le Bars, merci pour ton amitié et pour nos échanges, aussi bien au labo qu'à l'internat.
- Léa Pilorgé, merci pour ton amitié et ton soutien.
- Adissa, Sophie, Elodie, Dorothée, Gilles Nevez et Valérie merci pour votre gentillesse et votre soutien.
- Aux techniciennes, secrétaires et internes du service pour ces années de collaboration fructueuses et conviviales.

Merci aux membres de l'équipe INSERM UMR1078, et plus particulièrement à l'axe microbiota ; tout particulièrement à Stéphanie, Rozenn, Sophie, Charles-Antoine et Rosine, merci pour votre soutien, votre bonne humeur et nos échanges amicaux et scientifiques.

A tous les miens pour votre soutien si précieux.

- A mes amis de Brest, de Rouen, de Tours. Pour tous les bons moments passés ensemble, pour ceux à venir.
- A mes parents, merci pour votre soutien, votre amour et votre présence
- A ma belle famille, merci pour votre gentillesse et vos encouragements
- A Pauline, merci pour ton soutien et ton aide. Merci pour tes conseils précieux de maman que je découvre de jour en jour.
- A ma famille, mes tantes, oncles, cousins et cousines. Que de bons moments passés ensemble ! J'espère tant d'autres.

A Timothée, merci pour ton immense soutien là encore. Merci pour tout ce que tu me donnes. Merci pour ce merveilleux cadeau.

A Gabrielle, mon trésor.

Résumé

Nous avons étudié l'intérêt du locus CRISPR1 (associé au système CRISPR-Cas [clustered regularly interspaced short palindromic repeats – CRISPR associated] de type II-A) comme marqueur épidémiologique pour le typage et l'analyse phylogénétique des souches de *Streptococcus agalactiae*, le streptocoque du groupe B (SGB). Par ce travail de thèse, nous avons pu mettre en évidence i) l'activité du système CRISPR-Cas *in vivo* ainsi que sa faible vitesse d'évolution ; ii) une conservation des marqueurs ancestraux, permettant d'obtenir des informations sur les lignées phylogénétiques (congruence entre le typage CRISPR et le MLST[muli locus sequence typing]) ; iii) un polymorphisme du locus CRISPR1 (notamment des spacers d'acquisition récente), offrant une méthode de typage très discriminante (séparation des isolats au sein d'un même séquence type défini par MLST). L'analyse de ces spacers nous donne également des informations sur l'évolution récente des isolats, notamment de leurs contacts avec les éléments génétiques mobiles. Nous avons montré l'intérêt de cet outil pour le suivi de souches de portage ou l'étude d'une population. Ainsi, à l'issue de ce travail de thèse nous proposons le typage CRISPR comme méthode de référence pour le typage et l'analyse phylogénétique des souches de SGB.

Mots clés : *Streptococcus agalactiae,* Streptocoque du groupe B, CRISPR, Typage

Résumé en anglais

We studied the relevance of the CRISPR1 array (associated to a type II-A CRISPR-Cas system [clustered regularly interspaced short palindromic repeats – CRISPR associated]) as an epidemiological marker for genotyping and phylogenetic analyses of *Streptococcus agalactiae* (or Group B *Streptococcus* (GBS)) isolates. We demonstrate that i) spacer acquisition events occurred *in vivo* which strongly suggest that the CRISPR1-Cas system is functionally active for adaptation; ii) ancestral markers (terminal direct repeat and ancestral spacers) are highly conserved and reflect the phylogenetic structure of the GBS population (in congruence with MLST [Multi Locus Sequence Typing]); iii) CRISPR1 arrays presented a high degree of polymorphism (especially for leader end spacers) offering a highly discriminatory typing method (allowing to separate isolates within a same sequence type defined by MLST). Leader end analysis also provides specific evidence on isolats recent evolution, especially encounters with mobile genetic elements. The CRISPR1 array appears as a useful genetic feature to follow vaginal carriage of GBS in women and to evaluate the diversity of GBS vaginal carriage population. On the basis of these data, we propose that this method could become a reference method for GBS phylogenetic typing.

Keywords: *Streptococcus agalactiae*, Group B Streptococcus, CRISPR, Typing

Table des matières

Remerciements 5
Résumé 9
Résumé en anglais 10
Table des matières 11
Liste des tableaux 13
Liste des figures 14
Liste des abréviations 16
Revue de la littérature 17
I. *S. agalactiae* : Epidémiologie 19
A. Histoire naturelle 19
B. Pouvoir pathogène et physiopathologie 20
1. Un commensal chez l'homme 20
2. Un pathogène majeur chez le nouveau-né 21
3. Un pathogène reconnu chez la femme enceinte 26
4. Un pathogène émergent chez l'adulte (hors grossesse) 27
C. Epidémiologie de *S. agalactiae* 30
1. Outils actuels pour l'étude épidémiologique des souches de *S. agalactiae* 31
2. En fonction de l'origine géographique 31
3. En fonction de l'origine anatomique et des manifestations cliniques 34
4. Au cours du temps 38
5. Répartition des complexes clonaux en fonction du sérotype 40
II. Phylogénie de *S. agalactiae :* les techniques de typage 42
A. Sérotypage capsulaire 42
B. L'Etude du polymorphisme allélique des isoenzymes : Multilocus Enzyme Electrophoresis (MLEE) 42
C. Méthodes génotypiques basées sur l'analyse de l'ADN bactérien 44
1. Restriction enzymatique de l'ADN chromosomique 44
2. Séquençage des génomes bactériens 46
D. Méthodes génotypiques basées sur l'amplification de séquences d'ADN 49
1. Amplification aléatoire d'ADN polymorphe : Random Amplified Polymorphic DNA (RAPD) 49
2. Amplification et séquençage – Etude du polymorphisme des loci 49

3. Amplification des éléments génétiques mobiles insérés dans le génome 50
4. Amplification des séquences répétées 56
III. Le système CRISPR-Cas 60
A. Généralités 60
1. Histoire de la découverte 62
2. Composition du système CRISPR-Cas 63
3. Classification 66
4. Mode de fonctionnement 70
5. Application comme outil de typage 74
B. CRISPR-Cas chez *S. agalactiae* 77
1. Organisation du système CRISPR-Cas de type II-A 77
2. Fonctionnalité du locus CRISPR1 80
3. Diversité du locus CRISPR1 81
Objectifs du travail 86
Travail expérimental 92
I. Suivi du portage vaginal de *S. agalactiae* et évolution *in vivo* du locus CRISPR1 94
II. Etude de la diversité génétique des souches de *S. agalactiae* au sein d'une population de portage vaginal. 128
A. Etude de la diversité par CRISPR/RFLP et séquençage du locus CRISPR1 128
B. Comparaison des génomes des clones présents en co-culture. 141
III. L'approche CRISPR pour le typage des souches de *S. agalactiae* : comparaison aux méthodes usuelles et pouvoir discriminant 152
Discussion et perspectives 210
Conclusion 218
Références bibliographiques 222
Valorisation de la recherche 246

Liste des tableaux

Tableau 1. Prévalence de la colonisation maternelle en fonction des régions du monde 20
Tableau 2. Répartition des principaux complexes clonaux en fonction de l'origine géographique et clinique des souches. 33
Tableau 3. Répartition des complexes clonaux dans les infections néonatales en fonction des manifestations cliniques. 38
Tableau 4. Génotypage bactérien basé sur le *locus* CRISPR 77
Tableau 5. Liste et répartition des différents DR au sein du locus CRISPR1 de *S. agalactiae*. 78
Tableau 6. Sélection des enzymes de restriction. 130

Liste des figures

Figure 1. Physiopathologie des infections néonatales à *S. agalactiae.* 23
Figure 2. Nombre estimé de cas d'infections invasives néonatales à *S. agalactiae* en France entre 1996 et 2017 d'après les données de l'INVS (réseau Epibac). 25
Figure 3. Evolution de la répartition des cas d'infections invasives néonatales à SGB. 26
Figure 4. Evolution des cas de bactériémie en fonction du sexe et de l'âge dans une population issue de plusieurs nations 27
Figure 5. Epidémiologie des infections invasives à *S. agalactiae* en France en 2016 en fonction de l'âge (Laboratoire de microbiologie du réseau EPIBAC, InVS 2016) 28
Figure 6. Facteurs de risques des infections invasives à *S. agalactiae* chez l'adulte (hors grossesse). 29
Figure 7. Manifestations cliniques des infections de l'adulte à SGB en France de 2011 à 2016. 29
Figure 8. Répartition des sérotypes en fonction de l'origine anatomique des souches et des manifestations cliniques. 35
Figure 9. Distribution des séquences types et des complexes clonaux de souches invasives isolées chez l'adulte et le nouveau-né. 37
Figure 10. Evolution de la répartition des complexes clonaux au cours du temps dans les infections invasives de l'adulte en Islande. 39
Figure 11. Evolution de la répartition des sérotypes capsulaires au cours du temps dans les infections invasives de l'adulte en France (CNR des Streptocoques, 2016). 40
Figure 12. Structure de la population établie par MLEE. 44
Figure 13. Type RDP et manifestations cliniques d'isolats d'infection néonatale 45
Figure 14. Arbre phylogénétique des génomes *de S. agalactiae* séquencés disponibles sur la base NCBI. 48
Figure 15. Relation entre complexes clonaux et EGM. 53
Figure 16. Organisation de la région intergénique *scpB-lmb* et relation avec l'origine phylogénétique des souches. 53
Figure 17. Prévalence de trois fragments prophagiques en fonction de l'origine clinique des souches. 54
Figure 18. Distribution des souches de SGB en groupes prophagiques..... 55
Figure 19. Position des séquences répétées en tandem dans les génomes des 3 souches de référence entièrement séquencées (2603V/R, A909 et NEM316). 57

Figure 20. Cycle de réplication d'un bactériophage. 61
Figure 21. Organisation générale du système CRISPR-Cas. 64
Figure 22. Classification fonctionnelle des protéines Cas et des systèmes CRISPR-Cas associés. 66
Figure 23. Représentation de l'architecture de l'opéron *cas* permettant le sous-typage des systèmes CRISPR-Cas. 69
Figure 24. Mécanisme d'action du système CRISPR-Cas. 70
Figure 25. Mécanisme d'action des étapes d'expression et d'interférence du système CRISPR-Cas. 72
Figure 26. Modèle d'organisation du système CRISPR-Cas. 78
Figure 27. Correspondance des spacers avec les bases de données. 79
Figure 28. Efficacité de transformation des plasmides en fonction de la nature des protospacers et les PAM. 81
Figure 29. Relation entre le locus CRISPR et le MLST chez *S. agalactiae*. 83
Figure 30. Nombre de spacers par locus CRISPR1 en fonction de l'origine phylogénétique des souches chez *S. agalactiae* 84
Figure 31. Culure d'un prélèvement vaginal de dépistage de SGB sur gélose Granada®..... 129
Figure 32. Différence de pigmentation observée sur gélose Granada® entre les deux types de colonies de la patiente W10. 142

Liste des abréviations

ADN	Acide désoxyribonucléique
ANAES	Agence Nationale d'Accréditation et d'Evaluation en Santé
ARN	Acide ribonucléique
ARNr	ARN ribosomaux
Cas	CRISPR associated
Cascade	CRISPR associated complex for antiviral defense
CC	Complexe clonal
CDC	Center for disease control and prevention
CNGOF	Collège national des gynécologues obstétriciens
CRISPR	Clustered regularly interspaced short palindromic repeat
crRNA	CRISPR RNAs
crRNP	Complexe crRNA-effecteur
dNTP	2′-desoxynucleoside 5′-triphosphate
DR	Direct repeat
DRT	DR terminal
EGM	Eléments génétiques mobiles
ET	électrophorétype
ICE	éléments intégratifs conjugatifs
IMF	Infection materno-foetale
InVS	Institut de veille sanitaire
IS	Insertion sequence
MLEE	Multilocus enzyme electrophoresis
MLST	Multi locus sequence typing
MLVA	Multiple locus variable-number of tandem repeat analysis
OMS	Organisation mondiale de la santé
PAM	Protospacer adjacent motif
pb	paire de bases
PCR	polymerase chain reaction
PFGE	Electrophorèse en champ pulsé : Pulsed Field Gel Electrophoresis
RAPD	Amplification aléatoire d'ADN polymorphe : Random Amplified Polymorphic DNA
RDP	Restriction digest pattern ou Profil de restriction sur gel
rep-PCR	Repetitive element based PCR
RFLP	Polymorphisme de longueur de fragments de restriction
SGB	Streptocoque du groupe B
SiRNA	small interfering RNA
SRSR	Short regularly spaced repeats
spoligotyping	Spacer-oligonucleotide typing
ST	Sequence type
tracrRNA	trans-activating CRISPR RNA
VNTR	Variable number of tandem repeat

Revue de la littérature

I. *S. agalactiae* : Epidémiologie

A. Histoire naturelle

Streptococcus agalactiae est une bactérie appartenant à la famille des *Streptococcaceae* et au genre *Streptococcus*. Le genre *Streptococcus* fut attribué en 1874 par Billroth qui décrit un microorganisme sous forme de coques disposées en chaînette dans des lésions d'érysipèle. La première classification des streptocoques a été établie en 1933 par Rebecca Lancefield en se basant sur les propriétés antigéniques d'un composé pariétal, le polyoside C, et identifie *S. agalactiae* comme un streptocoque de groupe B (SGB) (Lancefield 1934).

L'espèce *S. agalactiae* a été initialement décrite en médecine vétérinaire. A la fin du XIXe siècle, Nocard et Mollereau identifient cette bactérie comme responsable de la « mammite contagieuse des vaches laitières » responsable d'une altération de la lactation, ce qui lui valut le nom de « *agalactiae* » ou « absence de lait ».

En 1938, Fry décrit les premiers cas d'infection à *S. agalactiae* chez l'homme chez trois patientes décédées d'une fièvre puerpérale (Fry 1938). Durant les trois décennies suivantes, des cas sporadiques ont été reportés mais cette bactérie restera méconnue en médecine humaine jusque dans les années 1960, période à laquelle Théodore Eickhoff constate un nombre particulièrement important d'infections à SGB et place cette bactérie comme la principale cause de sepsis néonatal au Boston City Hospital (Eickhoff et *al.* 1964). Deux ans plus tard, les premiers cas de méningites néonatales à SGB sont décrits en Europe (Kexel et Schönbohm 1965). *S. agalactiae* devient alors la première cause d'infection néonatale bactérienne dans les pays développés. L'incidence des infections néonatales s'est par la suite stabilisée alors qu'elle est toujours en augmentation chez l'adulte. Parallèlement à la diminution des cas de mastites bovines, les cas d'infections humaines sont apparus (Keefe 1997) ; l'utilisation massive de tétracyclines étant incriminée dans la sélection des souches responsables d'infection humaine (Da Cunha et *al.* 2014).

S. agalactiae a depuis été isolé chez d'autres animaux mais reste sporadique, excepté chez le poisson où *S. agalactiae* est une cause importante de morbi-mortalité et une source de zoonose pour l'homme (Leal et *al.* 2019; Chau et *al.* 2017, G. Liu et *al.* 2016).

B. Pouvoir pathogène et physiopathologie

Chez l'homme, *S. agalactiae* a un double comportement. C'est à la fois un commensal et un redoutable pathogène, particulièrement pour le nouveau-né. C'est en effet la première cause d'infection materno-fœtale dans les pays développés depuis les années 1970 et un pathogène émergent chez l'adulte depuis les années 1990.

1. Un commensal chez l'homme

Chez l'Homme, *S. agalactiae* à un statut commensal où il est retrouvé principalement dans le tractus digestif mais également en portage vaginal (Bliss et *al.* 2002). Le tractus gastro-intestinal est le réservoir naturel de SGB. Il est également à l'origine de la colonisation vaginale (Dillon et *al.* 1982).

a) Prévalence du portage

La prévalence moyenne du portage recto-vaginal dans le monde est estimée à 17,9% et varie selon les études, le portage étant plus fréquent en Afrique, en Amérique puis en Europe (Tableau 1) (Kwatra et *al.* 2016). Aux Etats-Unis, le taux de portage chez la femme enceinte est estimé entre 10 et 30% (Regan et *al.* 1991; Schuchat 1995; Yancey et *al.* 1994). En Europe, la prévalence chez la femme enceinte est comprise entre 6,5 et 36 % selon les études et semble plus importante dans les pays de l'Est et du Nord (Barcaite et *al.* 2008). En ce qui concerne l'Europe de l'ouest, elle est comprise entre 11 et 21%. Pour la France, plus particulièrement, le portage chez la femme enceinte se situe entre 10% et 15% (ANAES 2001).

Tableau 1. Prévalence de la colonisation maternelle en fonction des régions du monde (Kwatra et *al.*, 2016).

	Number of studies	Number of women	Median number of women per study (IQR)	Number of women with GBS	Estimated mean prevalence (95% CI)	I^{2*}	p value
Africa	4	2735	599 (415–867)	619	22·4 (18·1–26·7)	82·7%	0·005
Americas	29	23 163	316 (222–962)	4360	19·7 (16·7–22·7)	98·4%	<0·0001
Eastern Mediterranean	7	2729	300 (248–318)	443	16·7 (11·7–21·7)	91·9%	<0·0001
Europe	23	31 642	500 (275–942)	6113	19·0 (16·1–22·0)	97·2%	<0·0001
Southeast Asia	7	3749	406 (310–727)	389	11·1 (6·8–15·3)	95·1%	<0·0001
Western Pacific	8	9773	1049 (404–1700)	1176	13·3 (7·8–18·8)	98·7%	<0·0001
Total	78	73791	391 (251–1011)	13100	17·9 (16·2–19·7)	98·1%	<0·0001

GBS=group B streptococcus. *Within-region test for heterogeneity.

Table 2: Meta-analysis of studies reporting prevalence of maternal colonisation by region

Ces variations de prévalence s'expliquent en partie par une différence dans les modalités de prélèvement, de transport et de technique de mise en évidence de la bactérie. Les prévalences sont plus fortes quand le prélèvement est à la fois rectal et vaginal, quand un milieu de transport est utilisé, quand l'acheminement du prélèvement est rapide et quand un milieu d'enrichissement est utilisé (Badri et *al.* 1977; Håkansson et *al.* 2008; Jauréguy et *al.* 2003; Schrag et *al.* 2002). Cependant, ces facteurs n'expliquent pas toutes les différences de prévalence ; ces différences étant également observées dans des études internationales multicentriques utilisant les mêmes techniques (Whitney et *al.* 2004). D'autre part, en plus des variations géographiques, des différences ethniques ont été observées au sein d'un même pays (Valkenburg-van den Berg et *al.* 2006; Regan et *al.* 1991). Ainsi la prévalence est plus élevée chez les populations d'origine africaine par rapport aux populations européennes asiatiques.

b) Evolution du portage

La colonisation vaginale n'est pas constante chez la femme. Elle peut être transitoire, intermittente ou persistante (Anthony et *al.*1978; Verani et *al.* 2010; Lewin & Amstey 1981; Dillon et *al.* 1982; Hansen et *al.* 2004). Ainsi, le statut d'une femme enceinte vis-à-vis du SGB peut changer au cours de la grossesse (Verani et *al.* 2010). Seules 30 à 40% des femmes colonisées en début de grossesse le sont encore à l'accouchement (Dillon et *al.* 1982). Bien que le portage vaginal pour une femme donnée ne soit pas constant, une même souche est généralement retrouvée dans des prélèvements vaginaux ou recto-vaginaux itératifs (Dillon et *al.* 1982; Foxman 2006; Hansen et *al.* 2004; Manning et *al.* 2008).

2. Un pathogène majeur chez le nouveau-né

a) Mode de transmission et physiopathologie

Les infections précoces

Les infections précoces surviennent pendant les 7 premiers jours de vie et le plus souvent dans les 48 premières heures. Elles résultent d'une transmission verticale de la bactérie de la mère colonisée à l'enfant, soit par contamination du liquide amniotique à partir des souches vaginales, soit le plus souvent, par inhalation de sécrétions vaginales contenant du SGB au cours de la délivrance (Figure 1) (Six et *al.* 2014; Verani et *al.* 2010). Ainsi, la transmission en *per-partum* est évaluée à 45% dans le cas des accouchements par voie basse

contre 25% dans le cas des accouchements par césarienne en urgence (Håkansson et *al.* 2008; Schrag et *al.* 2002). D'autre part, la transmission materno-foetale est favorisée notamment par les facteurs de risques habituels d'infections materno-foetales (rupture prématurée des membranes, ouverture du col utérin) et par la densité de la colonisation maternelle. Parmi les nouveau-nés colonisés, 1 à 5% développent une infection (Schuchat et *al.* 1994; Siegel et *al.* 1980). L'infection précoce se manifeste généralement par une détresse respiratoire révélatrice d'une pneumopathie pouvant se compliquer d'un sepsis (73% des cas) ou plus rarement d'une méningite (27% des cas) (Baker 1978; Poyart et *al.* 2008).

Les infections tardives

Les infections tardives surviennent entre le 7e et le 89e jour de vie, généralement dans un délai de 4 à 8 semaines, sous forme de bactériémies fréquemment associées à une méningite (53,3% des cas) (Figure 1) (Poyart et *al.* 2008; Six et *al.* 2014). La physiopathologie de cette infection est encore incertaine. La contamination du nouveau-né se ferait par translocation digestive ou par transmission horizontale par manuportage ou par l'allaitement (A Tazi et al. 2011a; Bingen et al. 1992).

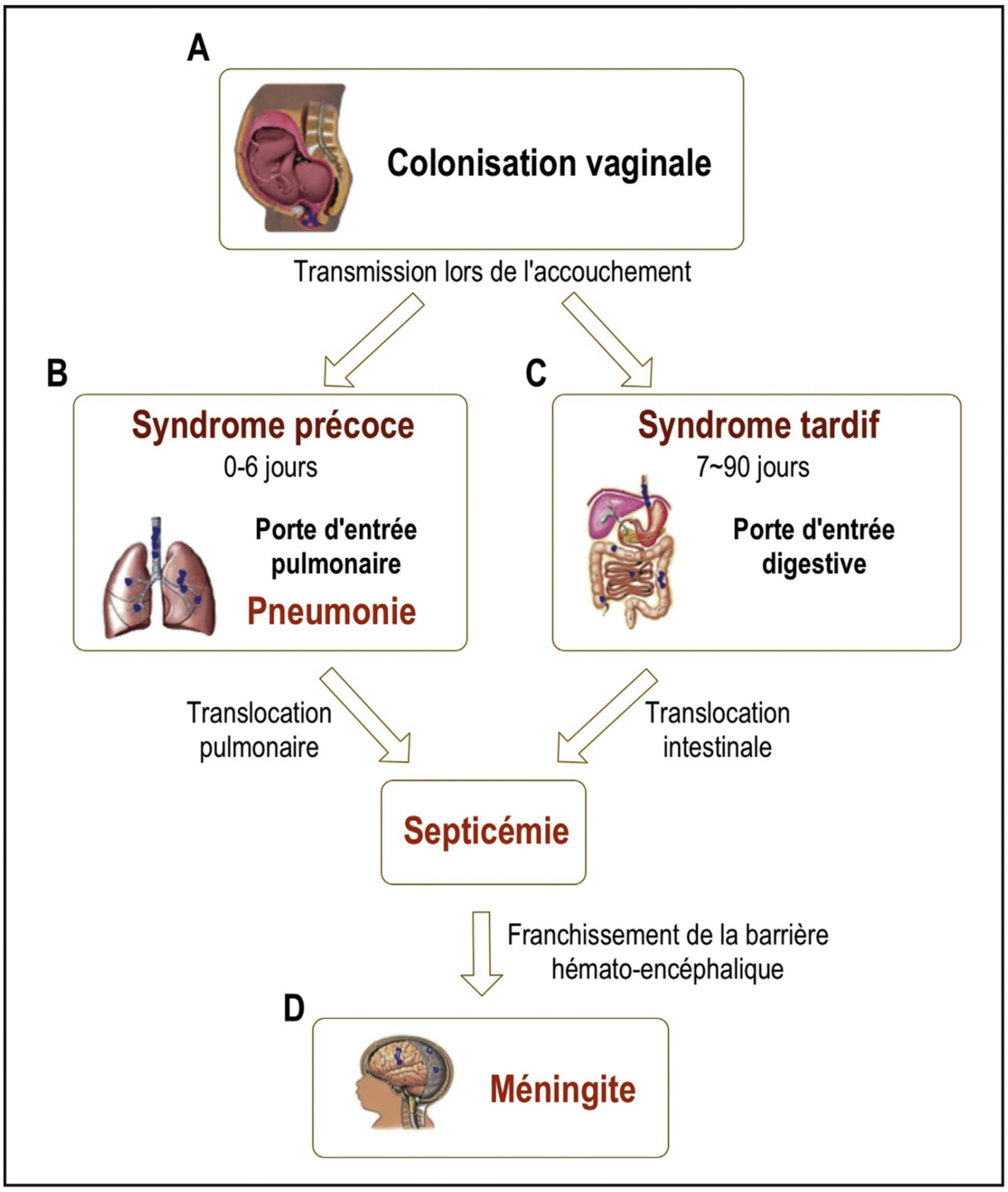


Figure 1. Physiopathologie des infections néonatales à *S. agalactiae* (Six et *al.* 2014).

b) Stratégie de prévention et efficacité

Du fait de la prévalence du portage de SGB et des conséquences de l'infection néonatale, certains pays ont mis en place une antibioprophylaxie *per-partum* dès les années 1980. Cette stratégie est actuellement recommandée en France. L'objectif de cette antibioprophylaxie est de diminuer l'inoculum bactérien présent en portage vaginal chez la femme au moment de

l'accouchement, afin de diminuer le risque de transmission mère – enfant. Cette prophylaxie a démontré une réduction de plus de ¾ du risque infectieux néonatal (Boyer & Gotoff 1986).
Sur la base de ces données, de nombreux pays ont émis des recommandations nationales sur la prévention du risque infectieux bactérien néonatal précoce. Les premières recommandations ont été émises en 1996 aux Etats-Unis par le CDC (Center for Disease Control and prevention). En France, le Collège National des Gynécologues Obstétriciens (CNGOF) recommande un dépistage systématique dès 1997 (CNGOF 1997). En 2001 l'ANAES (Agence Nationale d'Accréditation et d'Evaluation en Santé) s'appuie sur des critères de l'OMS pour proposer un dépistage de masse, compromis entre fiabilité, coût, sensibilité et spécificité (ANAES 2001). L'ANAES recommande un dépistage systématique du portage de *S. agalactiae* associé à une antibioprophylaxie *per-partum* chez les porteuses de cette bactérie. Ce dépistage est recommandé en fin de grossesse (entre 34 et 38 semaines d'aménorrhée) par prélèvement vaginal et par culture. L'association d'un prélèvement ano-rectal au prélèvement vaginal, ainsi que les méthodes d'enrichissement ne sont pas recommandées en France, du fait d'une absence significative d'amélioration dans la prévention des infections materno-foetales et du surcoût engendré. Depuis 2017, les nouvelles recommandations émises pour la prise en charge des nouveau-nés à risque d'infection bactérienne néonatale précoce autorise la recherche de colonisation maternelle à SGB par PCR rapide en *per-partum* (Société Française de Néonatologie et Société Française de Pédiatrie 2017). L'ANAES recommande également la recherche des bactéries à haut risque infectieux, dont le SGB, pendant la grossesse en cas de menace d'accouchement prématuré ou de rupture prématurée des membranes. Ces recommandations visent à diminuer le risque de transmission verticale et ainsi à diminuer les infections materno-fœtales précoces mais cela n'a pas d'effet sur les infections tardives (Li et *al.* 2017; Edmond et *al.* 2012; Poyart et *al.* 2008; Moore et *al.* 2003). On a pu noter en France, depuis l'application de ces recommandations, une efficacité de cette prévention avec une diminution de 39% dans les infections de l'enfant de moins de 1 an et de 70% des infections materno-fœtales précoces (Figure 2).

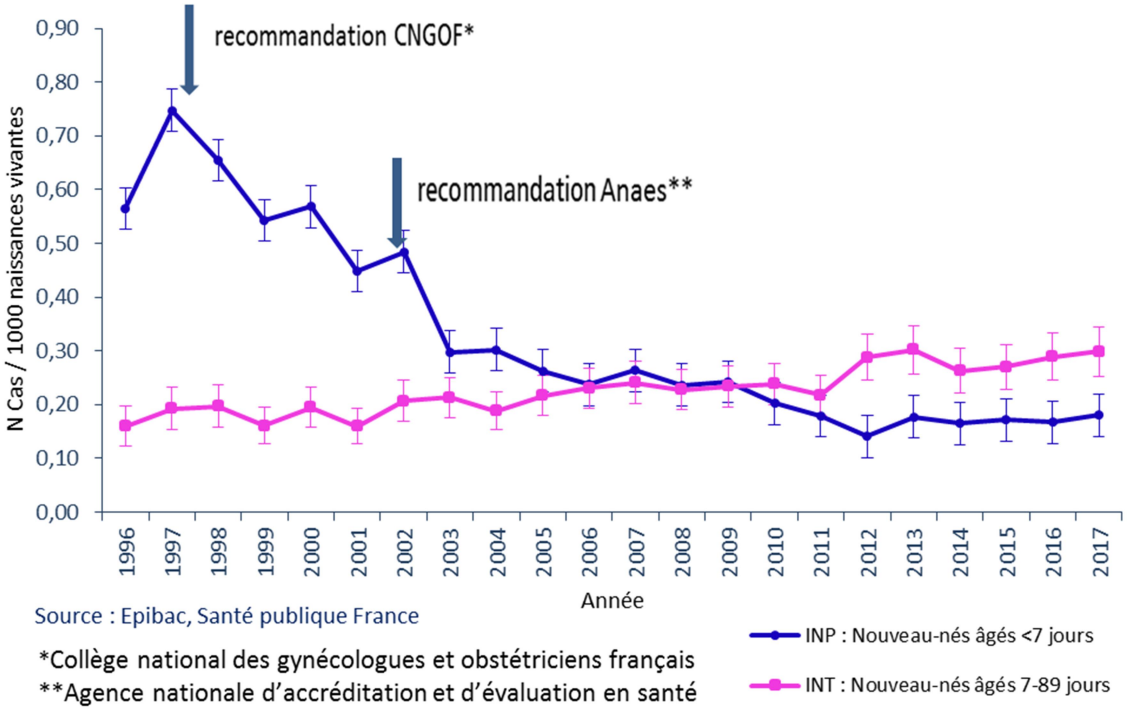


Figure 2. Nombre estimé de cas d'infections invasives néonatales à *S. agalactiae* en France entre 1996 et 2017 d'après les données de l'INVS (réseau Epibac).
INP (infections néonatales précoces), INT (infections néonatales tardives). Les recommandations du CNGOF en 1997 et de l'ANAES en 2001, portant sur le dépistage et l'antibioprophylaxie per-partum, ont permis une diminution de l'incidence des infections néonatales précoces (laboratoire de microbiologie du réseau EPIBAC, INVS 2016).

L'incidence globale des infections néonatales est aujourd'hui stabilisée chez les nouveau-nés avec une diminution des infections précoces et une augmentation des infections tardives (Bekker et *al.* 2014; CNR des Streptocoques 2016). Il est à noter que la diminution de la fréquence des infections précoces observée en France jusqu'en 2014 connait depuis un infléchissement avec une re-augmentation du nombre de cas sans explication évidente (CNR des Streptocoques 2016) (Figure 3).

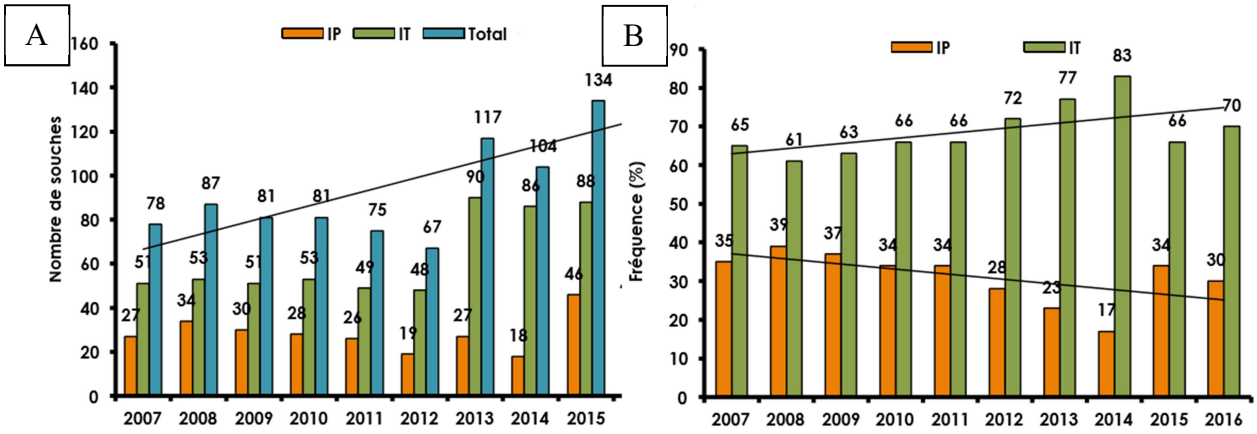


Figure 3. Evolution de la répartition des cas d'infections invasives néonatales à SGB.
D'après le rapport 2016 du CNR des Streptocoques. IP : infection précoce ; IT : infection tardive. ***A.*** *Répartition en nombre de cas des infections invasives néonatales à SGB en fonction du type d'infection.* ***B.*** *Evolution de la répartition en pourcentage des infections invasives néonatales à SGB en fonction du type d'infection.*

3. Un pathogène reconnu chez la femme enceinte

S. agalactiae est responsable d'infection chez la femme enceinte ou en *post-partum*. Il serait en cause dans 15% des infections de la femme enceinte en *post-partum* ou en *post-abortum* (Baker 1997). *S. agalactiae* serait impliqué dans 1 à 5% des infections du tractus urinaire, 20% des endométrites et 30% des chorioamniotites (Krohn et *al.* 1999; Schwartz et *al.* 1991).

4. Un pathogène émergent chez l'adulte (hors grossesse)

Depuis quelques années, un changement radical dans l'épidémiologie des infections invasives à *S. agalactiae* a été observé. 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 augmente, due au doublement de l'incidence de ces infections chez l'adulte de plus de 60 ans (Figure 4) (Schuchat 1998; Trivalle et *al.* 1998; Farley 2001; Blancas et *al.* 2004; Phares et *al.* 2008; Lamagni et *al.* 2013; Skoff et *al.* 2009; Ballard et *al.* 2016).

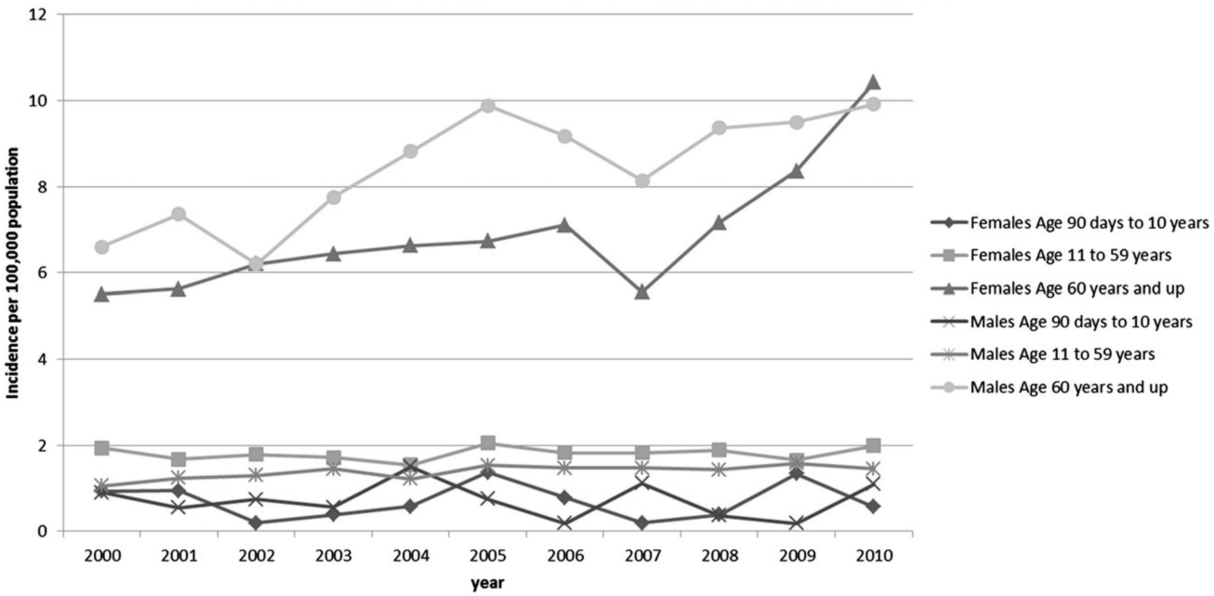


Figure 4. Evolution des cas de bactériémie en fonction du sexe et de l'âge dans une population issue de plusieurs nations (Australie, Canada, Danemark, Suède, Finlande et Royaume-Uni) (Ballard et *al.*, 2016).

En France, en 2016, les infections invasives à *S. agalactiae* concernaient 3019 infections (2865 bactériémies et 154 méningites) dont 2541 chez l'adulte de plus de 20 ans et 1744 chez l'adulte de plus de 65 ans (chiffres corrigés pour la couverture du réseau ; données Epibac, INVS)(laboratoire de microbiologie du réseau EPIBAC, INVS 2016) (Figure 5).

	Streptococcus agalactiae
<1 an	466
1-4 ans	7
5-9 ans	0
10-14 ans	0
15-19 ans	5
20-24 ans	25
25-29 ans	47
30-34 ans	54
35-39 ans	73
40-44 ans	49
45-49 ans	95
50-54 ans	114
55-59 ans	141
60-64 ans	198
65-69 ans	310
70-74 ans	278
75-79 ans	261
80-84 ans	349
85-89 ans	305
>89 ans	241
Total	3019

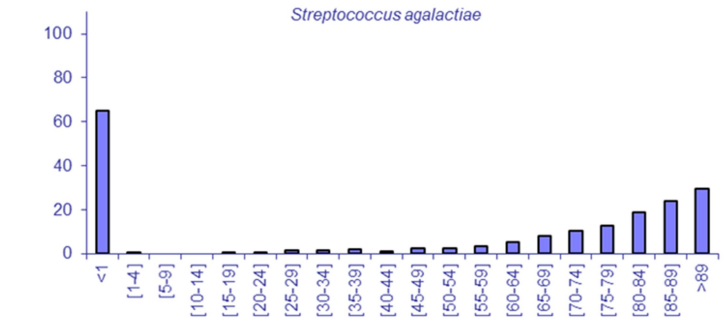


Figure 5. Epidémiologie des infections invasives à *S. agalactiae* en France en 2016 en fonction de l'âge (Laboratoire de microbiologie du réseau EPIBAC, InVS 2016)

a) Facteurs de risques

Les infections de l'adulte touchent préférentiellement le sujet âgé ou atteint de maladie chronique. Le diabète de type 2 est la maladie sous-jacente la plus fréquemment retrouvée, suivie des maladies néoplasiques, l'alcoolisme chronique et la cirrhose, ou encore les maladies cardio-vasculaires (Ballard et *al.* 2016; Skoff et *al.* 2009; Phares et *al.* 2008; Lambertsen et *al.* 2010; Farley 2001; Jackson et *al.* 1995) (Figure 6).

Variable	Cases, by age group				
	All (n = 1546)	18–39 years (n = 157)	40–64 years (n = 699)	≥65 years (n = 690)	P[a]
Underlying condition[l]					
≥1 Condition	1360 (88.0)	120 (76.4)	626 (89.6)	614 (89.0)	<.001
Diabetes mellitus	687 (44.4)	42 (26.8)	343 (49.1)	302 (43.8)	<.001
ASCVD	325 (21.0)	4 (2.5)	96 (13.7)	225 (32.6)	<.001
Obesity	259 (16.8)	22 (14.0)	161 (23.0)	76 (11.0)	<.001
Cancer	229 (14.8)	8 (5.1)	87 (12.4)	134 (19.4)	<.001
Liver disease[m]	108 (7.0)	6 (3.8)	71 (10.2)	31 (4.5)	<.001
Alcoholism	102 (6.6)	6 (3.8)	72 (10.3)	24 (3.5)	<.001
Asthma	68 (4.4)	4 (2.5)	38 (5.4)	26 (3.8)	.156
Substance abuse	31 (2.0)	11 (7.0)	19 (2.7)	1 (0.1)	<.001
Heart failure and/or congestive heart failure	220 (14.2)	8 (5.1)	57 (8.2)	155 (22.5)	<.001
Renal disease[j]	204 (13.2)	15 (9.6)	73 (10.4)	116 (16.8)	.001
Current smoker	194 (12.5)	30 (19.1)	124 (17.7)	40 (5.8)	<.001
COPD	151 (9.8)	1 (0.6)	57 (8.2)	93 (13.5)	<.001
Neurologic disease[k]	149 (9.6)	13 (8.3)	54 (7.7)	82 (11.9)	.026
Immunosuppression[l]	142 (9.2)	17 (10.8)	79 (11.3)	46 (6.7)	.009

ASCVD, atherosclerotic cardiovascular disease; COPD, Chronic obstructive pulmonary disease (includes emphysema and chronic bronchitis)

Figure 6. Facteurs de risques des infections invasives à *S. agalactiae* chez l'adulte (hors grossesse). D'après Skoff et *al.* 2009.

b) Manifestations cliniques

Les infections de l'adulte à *S. agalactiae* sont très variées. Les bactériémies primaires étant la cause la plus fréquente puis les infections de la peau et des tissus mous et les infections ostéo-articulaires, les pneumonies, les infections urinaires, et plus rarement les méningites et les endocardites (Figure 7) (Phares et *al.* 2008; Lambertsen et *al.* 2010; Skoff et *al.* 2009; Farley 2001).

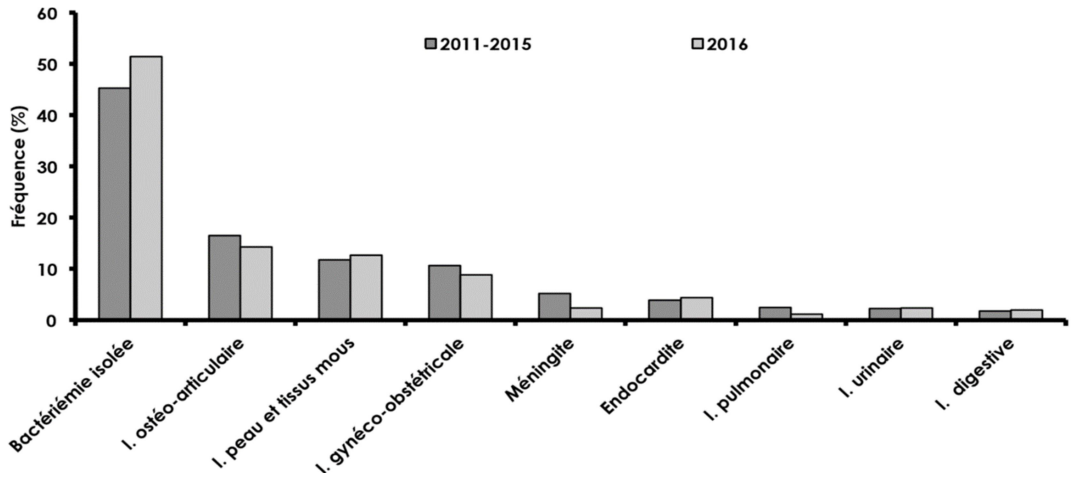


Figure 7. Manifestations cliniques des infections de l'adulte à SGB en France de 2011 à 2016. (CNR des Streptocoques 2016).

C. Epidémiologie de *S. agalactiae*

Deux techniques de typage sont aujourd'hui largement utilisées pour étudier l'épidémiologie de *S. agalactiae* : le sérotypage et le MLST (voir

Phylogénie *de S. agalactiae* : les techniques de typage).

1. Outils actuels pour l'étude épidémiologique des souches de *S. agalactiae*

a) Détermination du sérotype capsulaire (sérotypage)

La capsule présente un rôle majeur dans la virulence par son pouvoir anti-phagocytaire (Avci & Kasper 2010). Elle est composée d'unités polysaccharidiques répétées, codées par l'opéron *cps*. Une variabilité intra-espèce de sa composition a été mise en évidence dès les années 1930, permettant de définir les différents sérotypes (Lancefield 1934).
Actuellement, dix sérotypes capsulaires ont été définis (Ia, Ib, II, III, IV, V, VI, VII, VIII, IX), parmi lesquels des pouvoirs pathogènes différents semblent exister.

b) Le MLST : Multi *Locus* Sequence Typing

Dans cette stratégie de typage, les souches sont caractérisées par l'indexation de variations basées sur la diversité génétique de gènes de ménage codant pour des enzymes métaboliques. Pour chacun de ces gènes, tous les allèles sont identifiés arbitrairement par un numéro unique. La combinaison des numéros d'allèle pour chaque locus donne un profil allélique appelé Sequence Type (ST). Les souches étudiées au sein d'une population microbienne peuvent être regroupées selon leur profil allélique en complexes clonaux (CC) (Feil et *al.* 2004). Un complexe clonal regroupe l'ensemble des souches ayant en commun 6 allèles sur 7 avec au moins une autre souche appartenant à ce complexe clonal. Le génotype « fondateur » d'un complexe clonal est défini comme le ST qui détient en commun 6 allèles sur 7 avec le plus grand nombre de ST affiliés au complexe clonal. Ainsi, le génotype « fondateur » peut changer au cours du temps en fonction de l'enrichissement de la base de données.

2. En fonction de l'origine géographique

a) Répartition des sérotypes

Parmi les dix sérotypes décrits, cinq (Ia, Ib, II, III and V) représentent la grande majorité des isolats (> 90%) rencontrés chez l'homme, aussi bien en portage que dans les infections (Phares et *al.* 2008; Kwatra et *al.* 2016). Les études épidémiologiques réalisées dans différents pays montrent qu'il existe peu de différences de répartition des sérotypes à travers le monde (Phares et *al.* 2008; Le Doare & Heath 2013; Kwatra et *al.* 2016). Quelques spécificités géographiques sont néanmoins à noter. En Amérique et en Europe, les sérotypes Ia, Ib et III

représentent une plus faible proportion des isolats de portage que dans le reste du monde, probablement dû à une plus grande proportion de sérotype II (Kwatra et *al.* 2016). Aux Emirats Arabes Unis, le sérotype IV est le sérotype le plus fréquent dans les isolats de portage (Amin et *al.* 2002). Au Japon, ce sont les sérotypes VI et VIII, alors que le sérotype VIII, comme le VII, sont habituellement rares et ne sont décrits que dans quelques études (Lee et *al.* 2010; Tsolia et *al.* 2003; Lachenauer et *al.* 1999; Matsubara et *al.* 2013; Ueno et *al.* 2012). Dans certains pays africains, le sérotype V est le sérotype le plus fréquent (Suara 1994; Moyo et *al.* 2002).

b) Répartition des complexes clonaux

Chez l'Homme, cinq complexes clonaux prédominent à travers le monde : les CC-1 (ou CC-2), CC-10 (ou CC-12, CC-9, CC-8), CC-17, CC-19 et CC-23 (Jones et *al.* 2003; Bohnsack et *al.* 2008; Manning et *al.* 2009; Luan et *al.* 2005; Huber et *al.* 2011). Les séquences types ST-1, ST-17, ST-19 et ST-23 représentent les 2/3 des 152 isolats sélectionnés dans le monde pour la mise au point du MLST chez *S. agalactiae* (Jones et *al.* 2003).

En ce qui concerne les complexes clonaux, la structure de la population de *S. agalactiae* est similaire dans les différentes zones géographiques étudiées (Jones et *al.* 2003; Bisharat 2005; Morozumi et *al.* 2016; Huber et *al.* 2011; Björnsdóttir et *al.* 2016). Peu de différences de répartition des souches à travers le monde semblent exister, y compris dans les pays à faible prévalence d'infection néonatale (Bohnsack et *al.* 2008). La diffusion des cinq principaux complexes clonaux est mondiale, y compris pour le clone hyper-invasif ST-17 (Tableau 2).

Tableau 2. Répartition des principaux complexes clonaux en fonction de l'origine géographique et clinique des souches.

	CC1 (%)	CC10 (%)	CC17 (%)	CC19 (%)	CC23 (%)	pays	n=	Références
Adulte, Souches invasives	22	41	9	9	14	Monde	22	(Jones et *al.* 2003)
	11	16	19	14	29	Kenya	74	(Huber et *al.* 2011)
	25	18	12	12	28	Irlande	60	(Meehan et *al.* 2014)
	35	40	2	10	9	Japon	157	(Morozumi et *al.* 2016)
	23	23	7	30	11	Suède	53	(Luan et *al.* 2005)
Adulte, Souches de colonisation	25	16	13	25	11	Monde	67	(Jones et *al.* 2003)
	14	16	23	14	25	Kenya	95	(Huber et *al.* 2011)
	20	13	9	20	29	Irlande	45	(Meehan et *al.* 2014)
	25	16	8	19	11	Israël	104	(Bohnsack et *al.* 2008)
	28	18	6	17	25	Canada	464	(Manning et *al.* 2009)
	29	21	11	9	21	France	66	(Van der Mee-Marquet et *al.* 2008)
Nouveau né Souches invasives	3	12	59	12	10	Monde	59	(Jones et *al.* 2003)
	15	8	46	7	12	Irlande	72	(Meehan et *al.* 2014)
	28	6	22	14	12	Israël	50	(Bohnsack et *al.* 2008)
	16	9	28	24	17	Canada	424	(Manning et *al.* 2009)
	12	8	33	28	15	Suède	102	(Luan et *al.* 2005)

3. En fonction de l'origine anatomique et des manifestations cliniques

c) Répartition des sérotypes

Une répartition préférentielle des sérotypes est observée en fonction de l'origine anatomique des souches (Figure 8). Les souches de portage génital appartiennent préférentiellement aux sérotypes Ia, Ib, II, III, and V pour plus de 90% des isolats à travers le monde. Parmi ces sérotypes, les sérotypes III et Ia sont les plus représentés (Barcaite et *al.* 2008; Kwatra et *al.* 2016; Phares et *al.* 2008; Van der Mee-Marquet et *al.* 2008). Cette même distribution (sérotypes Ia et III puis Ib, II and V) est observée dans les infections néonatales précoces avec une prédominance des souches de sérotype III dans les méningites (Phares et *al.* 2008; Edmond et *al.* 2012). Les infections néonatales tardives sont causées principalement par des souches de sérotype III (Figure 8A et B) (Phares et *al.* 2008). Chez l'adulte (à l'exclusion des femmes enceintes), trois sérotypes représentent la grande majorité des isolats : les sérotypes V, III et Ia (Figure 8C) (Bergseng et *al.* 2008; Lambertsen et *al.* 2010; Phares et *al.* 2008; Skoff et *al.* 2009; Tazi et *al.* 2011b).

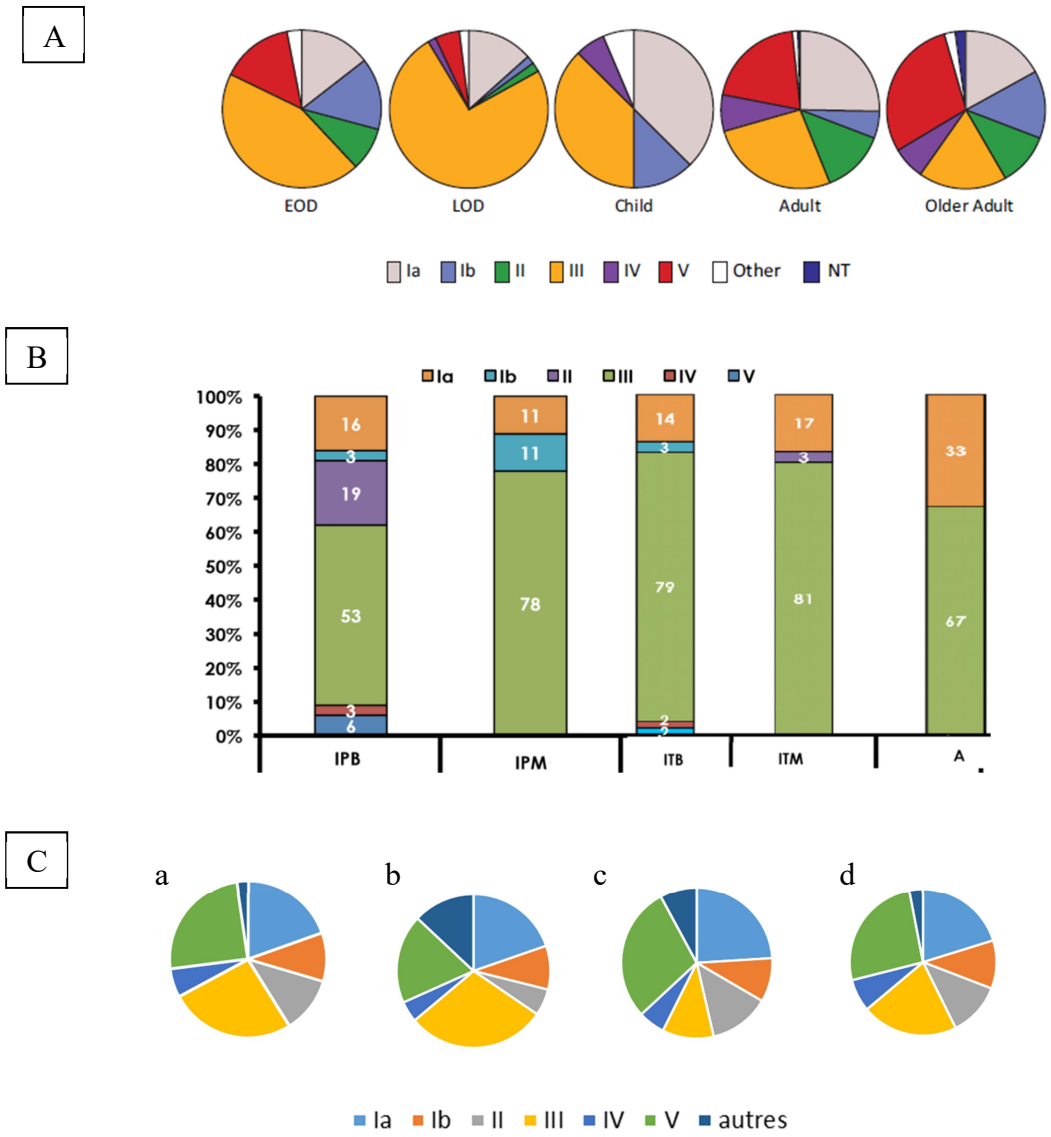


Figure 8. Répartition des sérotypes en fonction de l'origine anatomique des souches et des manifestations cliniques.
A. *Répartition des souches invasives en fonction de l'âge (caractère précoce ou tardif pour les infections néonatales) au Canada* (S. Teatero et *al.* 2014). *EOD (Early Onset Disease, infection néonatale précoce) ; LOD (Late Onset Disease, infection néonatale tardive).* ***B.*** *Répartition des sérotypes des souches invasives chez l'enfant en France en 2016 (IP : infection précoce, IT : infection tardive, M : méningites, B : bactériémies, A : Autres) (CNR des Streptocoques 2016).* ***C.****: Répartition des sérotypes des souches invasives de l'adulte.* ***a.*** *En France entre 2008 et 2014 (d'après le rapport d'activité (CNR des Streptocoques 2014)).* ***b.*** *Au Danemark entre 1999 et 2004 (Lambertsen et al. 2010)* ***c.*** *Aux Etats Unis entre 1990 et 2007 (d'après Skoff et al. 2009).* ***d.*** *Au Canada de 2009 à 2012 (d'après Teatero et al. 2014).*

D'autre part, des pouvoirs pathogènes différents semblent exister entre les sérotypes. Le sérotype Ia est associé à un taux de mortalité plus élevé, le sérotype III est particulièrement impliqué dans les méningites néonatales, alors que le sérotype V est particulièrement retrouvé dans les infections de l'adulte (Phares et *al.* 2008).

d) Répartition des lignées phylogénétiques

La répartition des complexes clonaux diffère également en fonction de l'origine anatomique des souches et des manifestations cliniques observées (Figure 9, Tableau 2 et Tableau 3). Les souches de portage génital appartiennent préférentiellement aux cinq complexes clonaux prédominants à travers le monde : les CC-1, CC-10, CC-17, CC-19 et CC-23 (Tableau 2) (Bohnsack et *al.* 2008; Huber et *al.* 2011; Jones et *al.* 2003; Luan et *al.* 2005; Manning et *al.* 2009; Meehan et *al.* 2014; Van der Mee-Marquet et *al.* 2008).

Quelle que soit l'incidence des infections néonatales à *S. agalactiae,* toutes les études identifient le clone ST-17 de serotype III comme le clone hyper-invasif associé aux souches néonatales (Bisharat 2005; Bohnsack et *al.* 2008; Jones et *al.* 2006; Lin et *al.* 2006; Jones et *al.* 2003; Luan et *al.* 2005). Ce clone est responsable de la majorité des infections néonatales tardives et des méningites néonatales, alors qu'il est sous représenté parmi les isolats cliniques de l'adulte (Tableau 2). En France, sur la période 2007-2015, les souches de CC-17 de sérotype III étaient responsables de la majorité des infections néonatales, de 79 % des infections néonatales tardives et de 70% des méningites néonatales (précoces et tardives) (CNR des Streptocoques 2016). Ainsi, bien que les infections néonatales tardives soient plus particulièrement dues à des souches de ST-17, les infections néonatales précoces sont dues à des souches appartenant à des lignées plus variées correspondant au panel de souches rencontrées en portage génital (Tableau 3) (Bohnsack et *al.* 2008; Bisharat 2005; Jones et *al.* 2003; Luan et *al.* 2005). Contrairement au clone ST-17,le potentiel infectieux de ces différentes lignées varie en fonction des études, certaines les reliant essentiellement au portage asymptomatique (Jones et *al.* 2003; Lin et *al.* 2006), alors que dans d'autres études les souches de ST-19 et de ST-23 semblent particulièrement représentées dans les infections précoces, indépendamment de la fréquence du portage (Manning et *al.* 2009; Martins et *al.* 2007; Bohnsack et *al.* 2008). Dans les infections de l'adulte, le clone hyper-virulent de ST-17 est particulièrement peu représenté au profit notamment, depuis les années 1990, des souches de sérotype V appartenant au CC-1 (Tableau 2). (Skoff et *al.* 2009; Teatero et *al.* 2014; Björnsdóttir et *al.* 2016).

Concernant la région Centre, une enquête d'incidence menée entre 2007 et 2015 (3 mois/an) met en évidence une forte représentation des souches de CC-17 dans les bactériémies néonatales tardives, de CC-23 dans les bactériémies néonatales précoces et de CC-1 dans les bactériémies à *S. agalactiae* de l'adulte (Barbera et *al.* 2016).

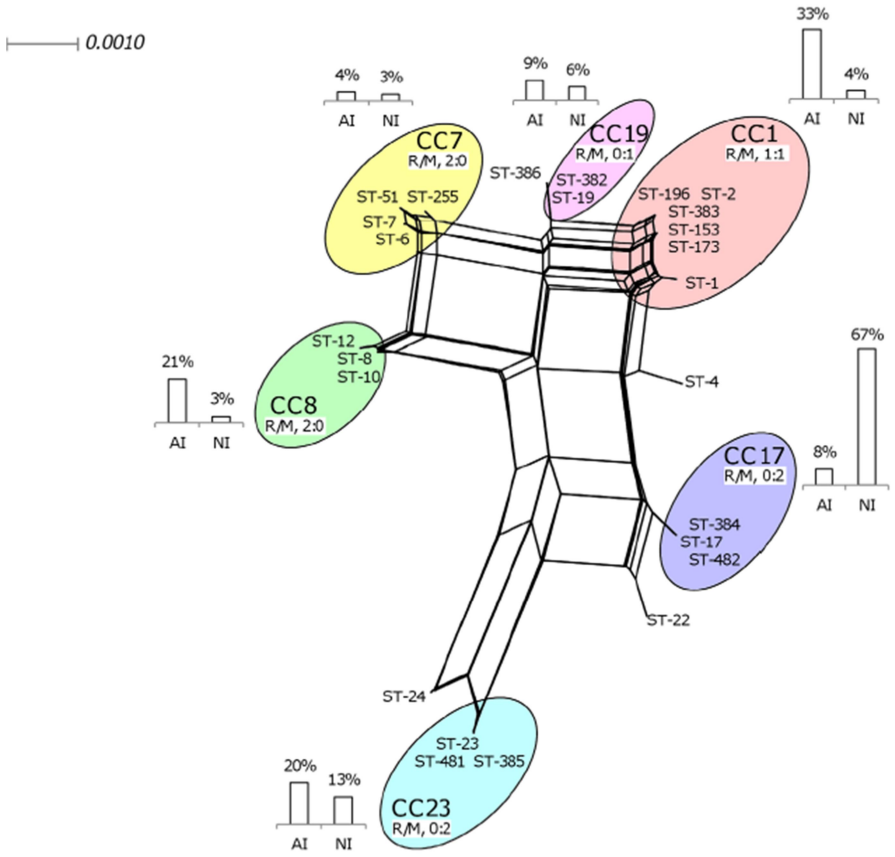


Figure 9. Distribution des séquences types et des complexes clonaux de souches invasives isolées chez l'adulte (AI) et le nouveau-né (NI) (Salloum et *al.* 2011).

Tableau 3. Répartition des complexes clonaux dans les infections néonatales en fonction des manifestations cliniques.

	CC1 (%)	CC10 (%)	CC17 (%)	CC19 (%)	CC23 (%)	Références
Infections	19	8	17	27	24	(Manning et *al.* 2009)
néonatales	21	10	37	8	13	(Meehan et *al.* 2014)
précoces	-	-	56	-	-	(Poyart et *al.* 2008)
Infections	8	11	57	20	4	(Manning et *al.* 2009)
néonatales	9	6	56	6	12	(Meehan et *al.* 2014)
tardives	-	-	76	-	-	(Poyart et *al.* 2008)

4. Au cours du temps

Peu d'évolution des sérotypes sont à noter ces 30 dernières années (Le Doare & Heath 2013). Cependant, l'émergence récente (années 1990) de deux sérotypes (IV et V) est à souligner. Bien que ces deux sérotypes étaient considérés comme non typables jusqu'en 1985 (date d'attribution de ces sérotypes), un faible pourcentage de souches non typables étaient décrites jusqu'à cette date (Jelínková & Motlová 1985; Baker & Barrett 1974). Depuis les années 1990, le serotype V a été associé à une augmentation marquée des cas d'infections invasives de l'enfant et de l'adulte (Farley 2001; Harrison et *al*. 1995; Tyrrell et *al.* 2000; Lamagni et *al.* 2013). Le typage de ces souches isolées dans différents pays, avant et après l'augmentation du nombre de cas, met en évidence une clonalité entre les différents isolats ; l'émergence de ces souches semblant être due à un clone particulier préexistant (Elliott et *al.* 1998; Le Thomas-Bories et *al.* 2001). Sa prévalence dans les infections invasives de l'adulte varie actuellement de 23% en France et en Angleterre (Tazi et *al.* 2011b; Lamagni et *al.* 2013) à plus de 30% aux USA (Skoff et *al.* 2009) et 42% en Suède (Persson et *al.* 2004). Ce sérotype est actuellement le sérotype prédominant dans les infections de l'adulte en Amérique du Nord, Taiwan ou encore en Suède (Persson et *al.* 2004; Harrison et *al.* 1995; Ko et *al.* 2001), alors que le sérotype III reste prédominant en France (Tazi et *al.* 2011b). D'autre part, il semble associé à une plus faible sensibilité aux antibiotiques de la famille des macrolides et lincosamides (Domelier et *al.* 2008; Bergseng et *al.* 2009; Tazi et *al.* 2011b).

Depuis quelques années, on note également l'émergence de souches de sérotype IV, un sérotype précédemment peu fréquent, devenant aujourd'hui une cause importante d'infection de l'adulte et du nouveau-né à travers le monde (Teatero et *al.* 2015; Florindo et *al.* 2014;

Ferrieri et *al.* 2013; Kiely et *al.* 2011; Diedrick et *al.* 2010; Amin et *al.* 2002; Skoff et *al.* 2009). Le typage de ces souches chez l'adulte relie la majorité des souches entre elles, mettant en évidence un clone responsable de l'augmentation des cas (Teatero et *al.* 2015; Florindo et *al.* 2014). On note également une augmentation de la proportion des souches de sérotype IV en portage ces dernières années (1% (2/148) en 2006 à 20% (19/97) en 2012) (Florindo et *al.* 2014).

Parallèlement, on observe une augmentation des souches de CC-1 (notamment des deux clones ST-1 de serotype V et ST-196 de serotype IV), et de CC23 (notamment des clones ST-23 et ST-24 de serotype Ia) qui s'accompagne d'une diminution relative du clone CC-17 (Figure 10) (Björnsdóttir et *al.* 2016; Martins et *al.* 2012). Ce constat est observé également en France avec une diminution relative des CC-17 et des sérotypes III depuis 5 ans (données CNR) (Figure 11). Enfin, il est a noté depuis 2015 des cas sporadiques dus à des souches de ST-283 et de sérotype III en tant qu'agent de zoonose via la consommation de poisson cru (Chau et *al.* 2017; Leal et *al.* 2019).

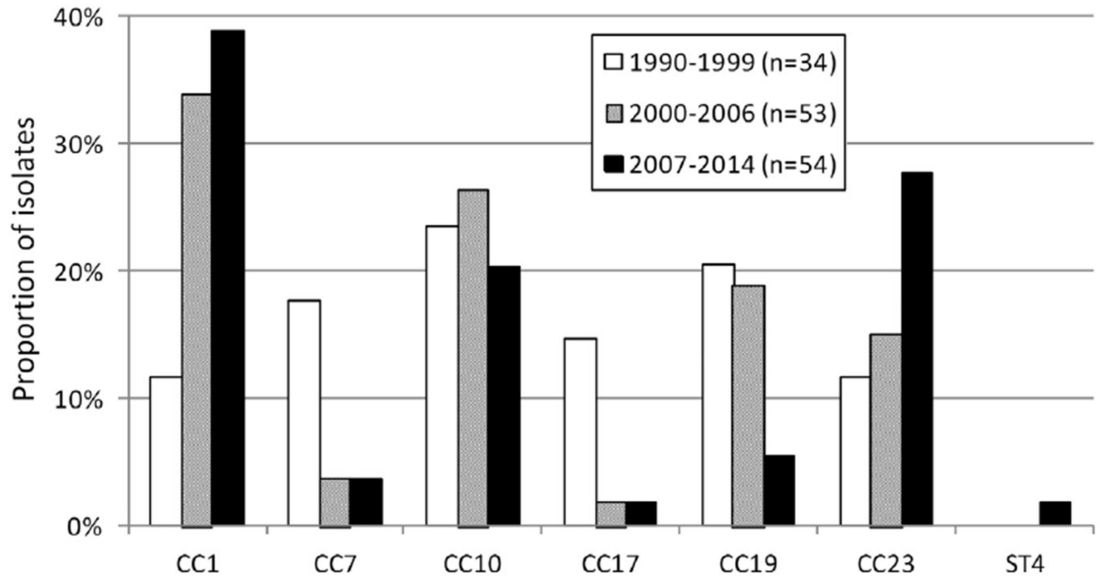


Figure 10. Evolution de la répartition des complexes clonaux au cours du temps dans les infections invasives de l'adulte en Islande (Björnsdóttir et *al.*, 2016).

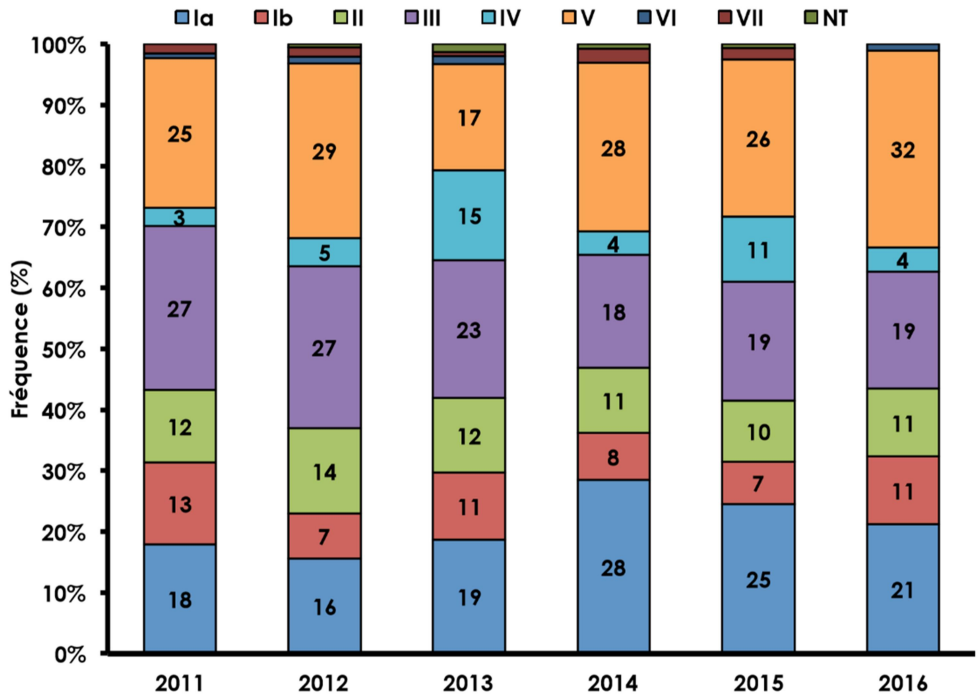


Figure 11. Evolution de la répartition des sérotypes capsulaires au cours du temps dans les infections invasives de l'adulte en France (CNR des Streptocoques, 2016).

5. Répartition des complexes clonaux en fonction du sérotype

La majorité des isolats de sérotypes Ia, Ib et V sont contenus dans un seul complexe clonal, les CC-23, CC-12 (ou CC-10) et CC-1 respectivement (Bohnsack et *al.* 2008; Morozumi et *al.* 2016; Meehan et *al.* 2014). Les sérotypes II et III sont répartis différemment en fonction des populations étudiées ; ils sont répartis environ à part égale entre les CC-17 et CC-19 chez le nouveau-né aux Etats-Unis (Bohnsack et *al.* 2008), ou majoritairement dans le CC-1 (sérotype II) ou le CC-19 (sérotype III) chez l'adulte au Japon (Morozumi et *al.* 2016). Les souches de ST-17 ne comprennent que des souches de sérotype III. En revanche, les ST-19 et ST-1 sont plus diversifiés et sont composés de différents sérotypes capsulaires (Bohnsack et *al.* 2008; Jones et *al.* 2003; Huber et *al.* 2011).

II. Phylogénie de *S. agalactiae* : les techniques de typage

Les premières techniques utilisées pour différencier les souches de *S. agalactiae* ont reposé sur des techniques phénotypiques de sérotypie, biotypie, antibio-résistance ou encore de lysotypie (Christensen & Christensen 1978; Stringer & Maxted 1979). Cependant, ces méthodes sont souvent peu discriminantes pour comparer les isolats et tracer la diffusion des clones (Denning et *al.* 1989; Noya et *al.* 1987). Par la suite, l'apparition des techniques moléculaires à largement contribué à la caractérisation des souches et de nombreuses méthodes génotypiques ont été développées (Tompkins 1992).

A. Sérotypage capsulaire

La détermination du sérotype peut être réalisée à l'aide de méthodes conventionnelles basées sur les antisérums ou de méthodes moléculaires basées sur l'étude de l'opéron *cps* (Imperi et *al.* 2010; Poyart et *al.* 2007). Les méthodes moléculaires tendent aujourd'hui à remplacer les méthodes classiques et ont permis de mettre en évidence de nouveaux sérotypes (Slotved et *al.* 2007). Actuellement, dix sérotypes capsulaires ont été définis (Ia, Ib, II, III, IV, V, VI, VII, VIII, IX), parmi lesquels cinq (Ia, Ib, II, III and V) représentent la grande majorité des isolats rencontrés chez l'homme (Phares et *al.* 2008; Kwatra et *al.* 2016) (voir partie Epidémiologie de *S. agalactiae*). Le typage capsulaire est aujourd'hui largement utilisé dans l'épidémiologie descriptive de SGB (Phares et *al.* 2008; Lamagni et *al.* 2013; Alhhazmi et *al.* 2016; Poyart et *al.* 2008).

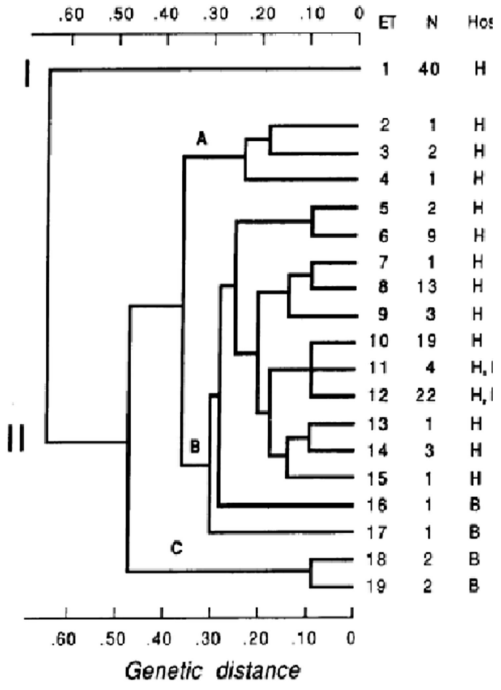
B. L'Etude du polymorphisme allélique des isoenzymes : Multilocus Enzyme Electrophoresis (MLEE)

Cette technique repose sur les variations de mobilité électrophorétique d'un grand nombre d'enzymes. La distance de migration de la protéine dépend de sa séquence en acide aminés, ce qui permet de différencier les allèles correspondants au *locus* étudié en fonction des variants de mobilité (Selander et *al.* 1986). Chaque isolat est caractérisé par la combinaison des allèles

correspondant aux différentes enzymes. La combinaison des allèles définit un génotype multilocus désigné « électrophorétype (ET) ». Cette méthode fournit des informations sur la distance génétique entre les isolats et permet une classification phylogénétique des souches. Chez *S. agalactiae*, cette méthode de typage a mis en évidence une structure clonale de la population. La population des souches de *S. agalactiae* isolées aux Etats-Unis et appartenant à différents sérotypes a pu être séparée en deux divisions phylogénétiques (Figure 12A) (Musser et *al.* 1989). La première division, très homogène, est composée d'un seul ET (ET1) regroupant uniquement des souches de sérotype III. La deuxième division, plus hétérogène, est composée de trois lignées majeures regroupant 18 ET et des souches de différents sérotypes, notamment de sérotypes III. Ils ont ainsi pu mettre en évidence que les souches de sérotype III appartenaient à deux lignées évolutives distinctes qui différaient dans leur capacité à envahir le LCR et qu'un seul clone particulièrement virulent était responsable, en majeur partie, de la forte morbi-mortalité des souches de sérotype III.

Les mêmes résultats ont été obtenus en Europe, confirmant l'existence de ces deux divisions (Figure 12B) (Quentin et *al.* 1995; Helmig et *al.* 1993). Des différences semblent cependant exister entre les deux continents : i) les souches appartenant à la première division sont moins homogènes en France qu'aux Etats unis ; ii) les souches invasives françaises sont reparties à parts égales entre les deux divisions, tandis que 81% des souches invasives américaines appartiennent à la première division, laissant penser que plus d'événements génétiques ont eu lieu dans la population européenne (Quentin et *al.* 1995).

A



B

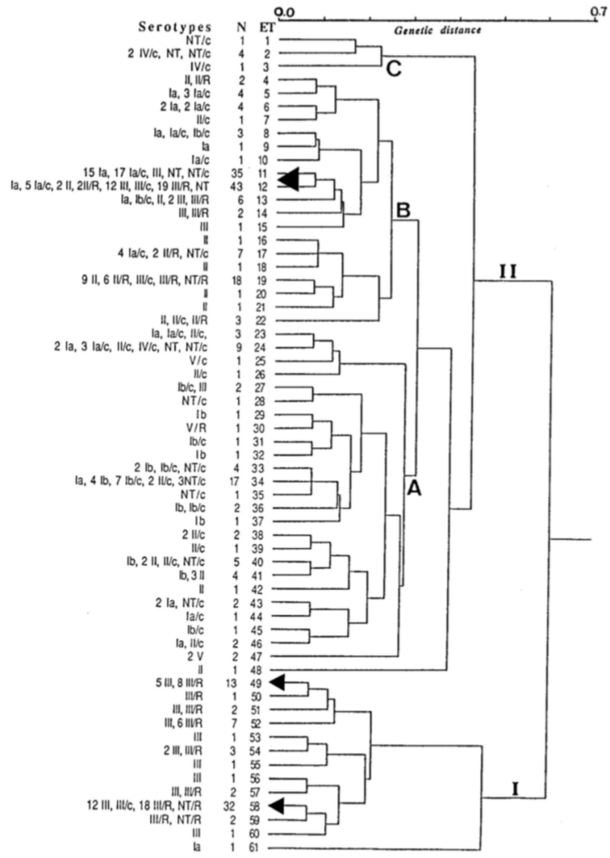


Figure 12. Structure de la population établie par MLEE.
A. *Aux Etats-Unis* (Musser et *al.* 1989). *Deux division primaires (I et II) sont observées et trois lignées majoritaires (A, B et C) au sein de la division II. La première division ne contient qu'un seul ET (ET1) et regroupe la majorité des souches invasives. H : souche d'origine humaine, B : souche d'origine bovine ; N : nombre de souches.* ***B.*** *En France* (Quentin et *al.* 1995). *Deux division primaires (I et II) sont observées et quatre lignées majoritaires (A, B, C et ET48) au sein de la division II. Les isolats de LCR sont reliés à trois familles clonales désignées par des flèches (ET11, ET49 et ET58). N : nombre de souches.*

C. Méthodes génotypiques basées sur l'analyse de l'ADN bactérien

1. Restriction enzymatique de l'ADN chromosomique

Les premières approches moléculaires ont reposé sur l'utilisation d'enzymes de restriction directement sur l'ADN génomique purifié. Cependant, ces techniques sont parfois d'interprétation difficile et les différences séparant les isolats parfois subtiles.

a) Profil de restriction sur gel : Restriction digest pattern (RDP)

Cette technique, basée sur l'utilisation d'enzyme ayant une forte fréquence de sites de coupure, a permis de séparer les isolats au sein d'un même sérotype (Nagano et *al.* 1991;

Denning et *al.* 1989). Deux lignées à risque pour les bactériémies néonatales précoces ont pu être identifiées par cette technique en utilisant l'enzyme HindIII : RDP III-3 et RDP Ia-3 (Figure 13) (Y. Nagano et *al.* 1991). La grande majorité (91%) des isolats responsables d'infection néonatale précoce au sein du sérotype III possède un même profil de restriction (en utilisant deux enzymes [HindIII et Sse83871]) : RDP III-3. Au contraire, ce profil n'étant retrouvé que dans 1/3 des souches de colonisation, suggérant que les souches du groupe RDP III-3 possèdent des facteurs de virulence particuliers (Takahashi et *al.* 1998; Nagano et *al.* 1991). Par la suite, des différences génomiques spécifiques au niveau de gènes de virulence des souches du groupe RDP III-3 ont été mises en évidence (Bohnsack 2002; Takahashi et *al.* 2002). D'autre part, certains profils sont retrouvés dans des contextes épidémiologiques et des zones géographiques distinctes, suggérant une dissémination d'un nombre restreint de clones (Denning et *al.* 1989).

Strain no.	RDP type	Age of* patient (days)	Disease pattern	
			Septicaemia	meningitis
11	Ia–3	0	+	–
12	Ia–3	0	+	–
14	Ia–3	2	+	–
13	Ia–3	3	+	–
15	Ia–3	3	+	–
5	Ia–2	5	–	+
6	Ia–2	8	–	+
19	Ia–4	23	–	+
20	Ia–4	48	–	+
42	III–3	0	+	–
46	III–3	0	+	–
32	III–2	1	+	+
41	III–3	3	+	–
47	III–3	3	+	–
22	III–1	3	–	+
44	III–3	7	+	+
29	III–2	9	+	+
43	III–3	16	–	+
28	III–2	17	–	+
45	III–3	19	–	+
34	III–2	22	–	+
35	III–2	24	–	+
36	III–2	38	–	+
30	III–2	37	–	+
33	III–2	71	–	+
31	III–2	92	–	+

* Age at onset of infection.

Figure 13. Type RDP et manifestations cliniques d'isolats d'infection néonatale
D'après Nagano et al. 1991. Deux lignées à risque pour les bactériémies néonatales précoces ont pu être identifiées (I-a et III-3), parmi lesquelles le RDP III-3 regroupe 91% des isolats responsables d'infection néonatale précoce (encadré en rouge) au sein du sérotype III.

b) Typage par transfert et hybridation (Southern blot).

Afin de simplifier la lecture, des techniques d'hybridation sélective avec des sondes spécifiques ont été mises au point. L'ADN génomique est digéré par des enzymes de restriction puis les fragments d'ADN sont séparés sur gel d'agarose et transférés sur

membrane. L'hybridation est ensuite réalisée avec des sondes ciblant des facteurs de virulence, des EGMs ou encore les opérons codant les ARNr (Ribotypage) (Hauge et *al.* 1996; Takahashi et *al.* 2002; Bohnsack 2002; Huet et *al.* 1993; Chatellier et *al.* 1996). La ribotypie a confirmé les groupes définis précédemment par MLEE et l'homogénéité de la lignée hautement invasive (Chatellier et *al.* 1996).

c) Electrophorèse en champ pulsé: Pulsed Field Gel Electrophoresis (PFGE)

Le PFGE utilise des enzymes à faible fréquence de coupure et permet ainsi de réduire le nombre de fragments à analyser. La séparation de ces fragments de grande taille est assurée par l'utilisation d'un courant dont la polarité est modifiée à intervalles réguliers et alternant une migration selon trois axes différents permettant une « reptation » des fragments d'ADN de grande taille (Schwartz & Cantor 1984). Les bandes à analyser sont moins nombreuses et mieux séparées que dans le RDP et les profils de restrictions plus faciles à interpréter. Dans de nombreuses publications, l'enzyme utilisée pour *S. agalactiae* est SmaI. Une grande diversité génomique a ainsi pu être observée, notamment au sein d'un même sérotype (Gordillo et *al.* 1993; Fasola et *al.* 1993). Cependant, parmi les souches de sérotype III, les souches responsables de méningite ou les souches appartenant à la lignée hyper-invasive RDPIII-3 ou ET1 en MLEE, présentent une forte homologie de profil en PFGE, ce qui confirme la grande homogénéité de la lignée néonatale hyper invasive (Rolland et *al.* 1999; Bidet et *al.* 2003). L'étude des populations par PFGE a permis de mettre en évidence la diversité génétique des souches circulantes dans un même pays ou provenant de sites anatomiques distincts (Von Both et *al.* 2008; Moyo et *al.* 2002; Bidet et *al.* 2003; Van der Mee-Marquet et *al.* 2009). Elle a permis également d'expliquer des phénomènes localisés comme l'émergence du clone invasif de sérotype V aux Etats-Unis ou d'un clone résistant aux β-lactamines au Japon (Elliott et *al.* 1998; Nagano et *al.* 2012; Al-Maani et *al.* 2014). Cette technique qui possède un fort pouvoir discriminant est particulièrement adaptée aux études épidémiques locales. Elle reste toujours utilisée aujourd'hui, notamment pour mettre en évidence des liens de clonalité en cas de suspicion de cas groupés (CNR des Streptocoques n.d.). Elle est cependant peu exportable entre laboratoires et peu reproductible.

2. Séquençage des génomes bactériens

Depuis la publication du premier génome bactérien, celui d'*Haemophilus influenzae*, en 1995 (Fleischmann et *al.* 1995), des progrès spectaculaires ont été réalisés au niveau des techniques

de séquençage et d'analyse bio-informatique. Aujourd'hui, près de 150 000 génomes bactériens ont été séquencés entièrement (*http://www.genomesonline.org/*). Ce séquençage de génomes complets, ou WGS (Whole Genome Sequencing) est de plus en plus abordable, de plus en plus rapide et de plus en plus utilisé.

Chez *S. agalactiae,* 1124 génomes sont aujourd'hui assemblés et annotés (https://www.ncbi.nlm.nih.gov/genome/?term=Streptococcus+agalactiae) dont 108 complets (Figure 14). La disponibilité de ces séquences a permis de mettre en évidence la taille moyenne des génomes de *S. agalactiae* (2,08 Mb) et leur composition. Le génome de SGB comprend un « core genome », représentant 80% du génome total, et un « dispensable genome » (accessoire) composé de gènes plus ou moins spécifiques de certaines souches (Tettelin et *al.* 2002; Brochet et *al.* 2006).

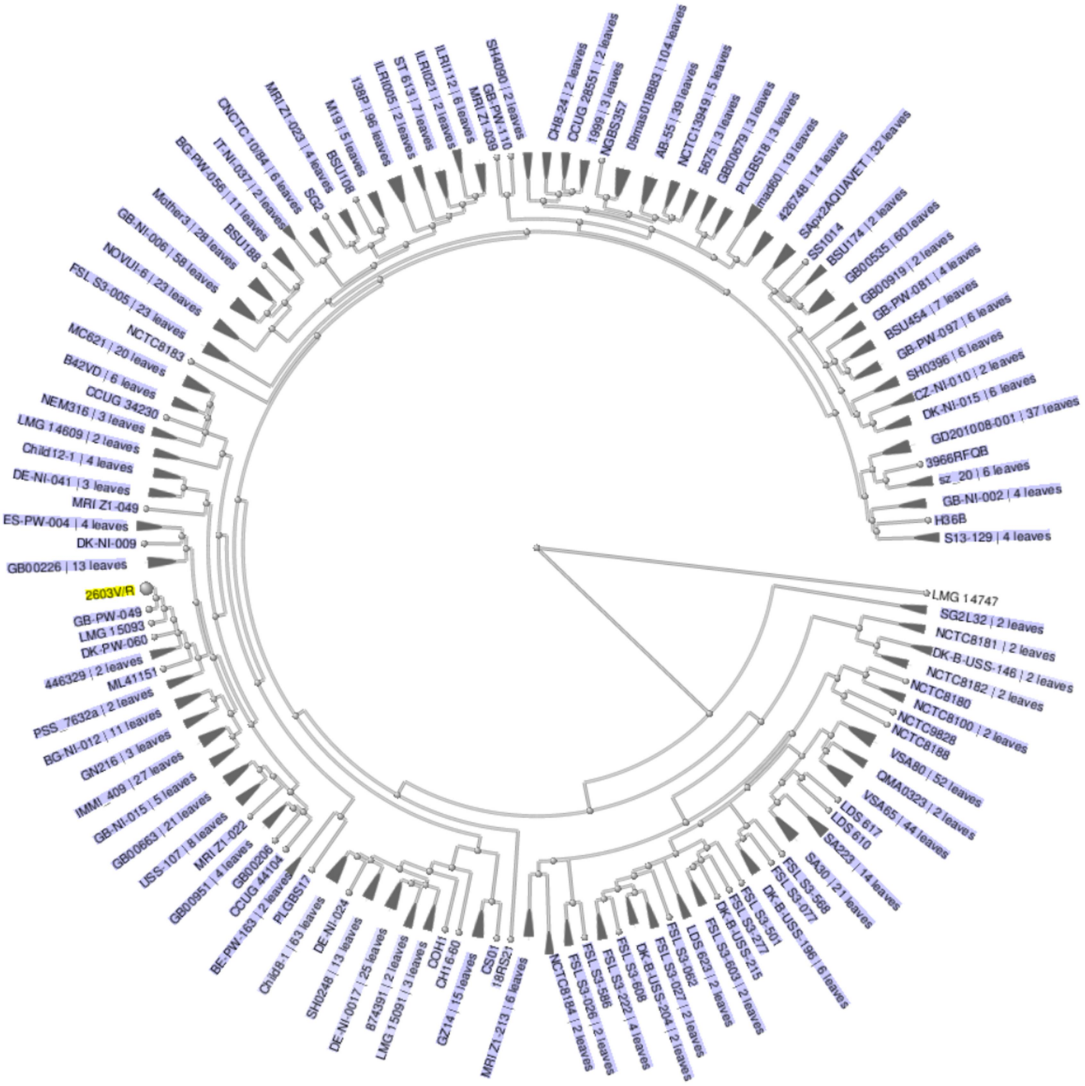


Figure 14. Arbre phylogénétique des génomes *de S. agalactiae* séquencés disponibles sur la base NCBI. *(https://www.ncbi.nlm.nih.gov/genome/tree/186).*

Cette technologie est puissante pour comparer les isolats. Son intérêt a été démontré pour la surveillance des souches responsables d'épidémie afin de la détecter et de comprendre son origine. De nombreuses bactéries ont ainsi été séquencées en ce but comme *Vibrio cholerae* (Chin et *al.* 2011), *Mycobacterium tuberculosis* (Gardy et *al.* 2011; Schurch et *al.* 2010), *Salmonella* (Den Bakker et *al.* 2014; Zhang et *al.* 2015), *Escherichia coli* producteur de verotoxines (Dallman et *al.* 2015; Joensen et *al.* 2014), *Staphylococcus aureus, Clostridioides (Clostridium). difficile* (Eyre et *al.* 2012; Köser et *al.* 2012). Bien que ces techniques de séquençage haut débit soient de plus en plus abordables, elles restent aujourd'hui techniquement exigeantes, aussi bien en matériel qu'en connaissances bio-informatiques indispensables pour extraire et interpréter le grand nombre de données fournies.

D. Méthodes génotypiques basées sur l'amplification de séquences d'ADN

1. Amplification aléatoire d'ADN polymorphe : Random Amplified Polymorphic DNA (RAPD)

Le principe repose sur l'utilisation d'amorces aléatoires courtes (environ 10 pb) pour amplifier par PCR différentes portion du génome sans connaitre sa séquence (Welsh & McClelland 1990; Williams et *al.* 1990). Le nombre et la localisation des sites d'hybridation varient en fonction des souches au sein d'une espèce donnée, permettant ainsi de différencier les isolats. La RAPD a permis de mettre en évidence les mêmes groupes de souches définies par MLEE et ainsi de distinguer les souches responsable de méningites (Chatellier et *al.* 1997; Quentin et *al.* 1995). Cette technique apparait plus rapide, plus accessible et plus discriminante que le MLEE puisqu'elle permet de distinguer certaines souches au sein de chaque cluster. D'autre part, les clusters définis par RAPD sont corrélés aux sérotypes et permettent également de différencier des souches au sein des différents sérotypes (Zhanget *al.* 2002).

2. Amplification et séquençage – Etude du polymorphisme des loci

a) Multi Locus Sequence Typing (MLST)

Avec l'avènement de l'automatisation des méthodes de séquençage, une nouvelle approche multi locus s'est développée. Cette technique très discriminante offre une très bonne reproductibilité et la possibilité de consulter librement des banques de données internationales. Ceci permet une comparaison des séquences dans le monde entier et constitue donc une ressource importante pour l'épidémiologie mondiale (Maiden et *al.* 1998). Le MLST a été utilisé pour la première fois chez *Neisseria meningitidis* et est devenu une technique moléculaire clé dans l'épidémiologie descriptive des bactéries pathogènes (Maiden et *al.* 1998). Conceptuellement, le MLST est identique au MLEE : les souches sont caractérisées par l'indexation de variations basées sur la diversité génétique de gènes « neutres » ou gènes « de ménage » codant pour des enzymes métaboliques. L'identification des allèles se fait par séquençage nucléotidique d'un fragment d'ADN (d'environ 500 pb) et détecte ainsi tous les polymorphismes nucléotidiques (pas uniquement ceux altérant la composition en acides aminés comme en MLEE). A chaque séquence unique est affecté un

numéro d'allèle et la combinaison des numéros d'allèle de chaque locus (en général sept loci) donne un profil allélique appelé sequence type (ST). La recherche des numéros d'allèle et de ST est réalisée via une base de données internationale qui regroupe tous les allèles et profils alléliques décrits (http://pubmlst.org/sagalactiae/). Cette banque regroupe aujourd'hui 1349 ST. Les souches étudiées au sein d'une population microbienne peuvent être regroupées en complexes clonaux (CC) grâce à l'algorithme eBURST (http://eburst.mlst.net/) (Feil et *al.* 2004). Cette méthode a permis i) d'établir des relations phylogénétiques entre les différentes souches ; ii) de mettre en évidence qu'une seule lignée homogène de sérotype III, la lignée ST-17, était significativement associée aux cas d'infections néonatales (Jones et *al.* 2003) (voir partie Epidémiologie de *S. agalactiae*).

b) Etude du polymorphisme unilocus, diversité allélique, profil allélique des gènes de virulence

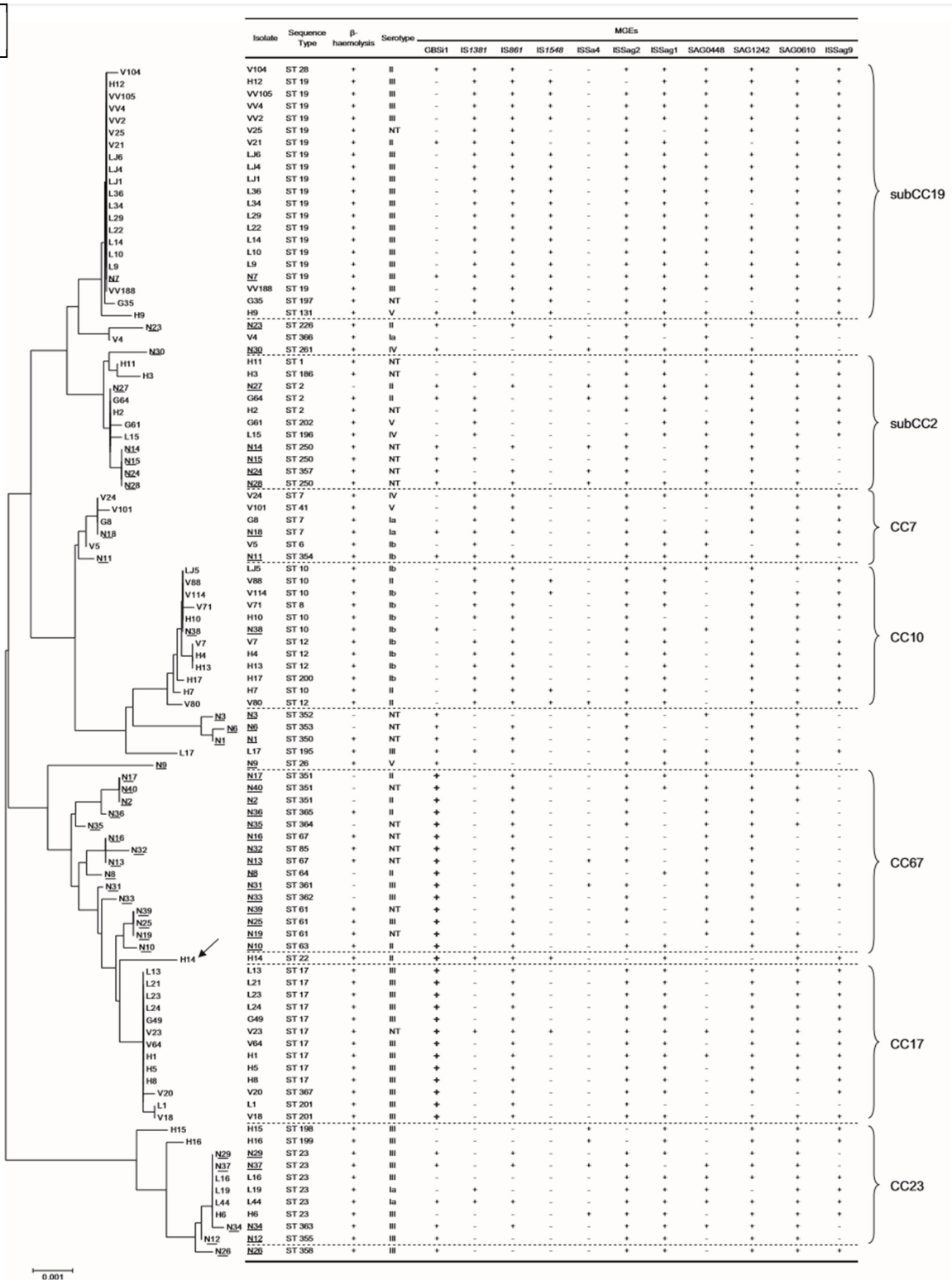
La comparaison des génomes bactériens, notamment l'analyse du polymorphisme de putatifs gènes de virulence, a permis de mettre en évidence des allèles spécifiques de lignées, notamment pour les souches de ST-17 (Springman et *al.* 2009; Brochet et *al.* 2006). Le ST-17 représente une lignée très homogène, possédant un répertoire de protéines de surface spécifiques et conservées. Ainsi, les gènes *gbs2018, sip, infB, fbsB* ou encore *fbsA* présentent des allèles spécifiques des souches de ST-17 (Brochet et *al.* 2006; Hedegaard et *al.* 2000; Rosenau et *al.* 2007; Springman et *al.* 2009; Takahashi et *al.* 2002). Le clone hyper-virulent ST-17 peut ainsi être recherché par détection par PCR d'allèle spécifique de ces gènes comme l'allèle *gbs2018C* du gène *gbs2018,* codant pour la protéine de paroi HvgA (Hypervirulent GBS adhesin) (CNR des Streptocoques n.d.; Lamy et *al.* 2006; Tazi et *al.* 2010). D'autre part, l'utilisation de puces à ADN ciblant des gènes de virulence ou de résistance a montré une bonne concordance avec le MLST (Nitschke et *al.* 2014). Une forte corrélation est également observée entre les ST et les gènes codant pour des protéines de surface (α-like proteins) ou pour des pili (Björnsdóttir et *al.* 2016).

3. Amplification des éléments génétiques mobiles insérés dans le génome

L'analyse comparative des génomes bactériens, par des techniques de restriction enzymatique (RDP, PFGE ou RAPD), d'analyse phénotypique (MLEE) ou plus récemment par séquençage total ont permis d'identifier la présence de fragments d'ADN ou d'ilots génomiques spécifiques de lignée (Chatellier et *al.* 1996; Rolland et *al.* 2002; Van der Mee-

Marquet et *al.* 2006; Herbert et *al.* 2005; Brochet et *al.* 2006). Un lien a pu être établi entre les lignées phylogénétiques et les éléments insérés dans le génome (Figure 15) (Takahashi et *al.* 2002; Bohnsack 2002; Granlund et *al.* 1998; Rolland et *al.* 1999). Ainsi, la séquence d'insertion *IS1548* dans le gène *hylB* est un marqueur associé à un groupe homogène défini en RDP (RDP III-2) ou MLEE (ET12) ou au CC-19 en MLST (Granlund et *al.* 1998; Rolland et *al.* 1999; Bohnsack 2002; Hery-Arnaud et *al.* 2005). L'intégration d'EGM au sein du fragment intergénique *scpB-lmb* est à l'origine de trois structures différentes, en lien avec la lignée phylogénétique de *S. agalactiae* (Al Safadi et *al.* 2010; Bohnsack 2002; Hery-Arnaud et *al.* 2005, 2007; Takahashi et *al.* 2002). Le complexe clonal CC-17 est caractérisé par la présence de GBSi1. Les souches de CC-19 sont divisées en deux groupes en fonction du sérotype : les souches de CC-19 de sérotype II sont caractérisées par la présence de GBSi1, tandis que les souches de CC-19 et de serotype III sont caractérisées par la présence de *IS1548*. Les souches de CC-7, CC-23 et CC-10 sont dépourvues de ces deux EGMs (Figure 16).

A



Isolate	Sequence Type	β-haemolysis	Serotype	MGEs										
				GBSi1	IS*1381*	IS*861*	IS*1548*	ISSa4	ISSag2	ISSag1	SAG0448	SAG1242	SAG0610	ISSag9	
V104	ST 28	+	II	+	+	+	-	-	+	+	+	+	+	+	subCC19
H12	ST 19	+	III	-	+	+	+	-	-	+	+	+	+	+	
VV105	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
VV4	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
VV2	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
V25	ST 19	+	NT	-	+	+	-	-	+	-	+	+	+	+	
V21	ST 19	+	II	+	+	+	-	-	+	+	+	-	+	+	
LJ6	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
LJ4	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
LJ1	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
L36	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
L34	ST 19	+	III	-	+	+	+	-	+	+	+	-	+	+	
L29	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
L22	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
L14	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
L10	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
L9	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
N7	ST 19	+	III	+	+	+	+	-	+	+	+	+	+	-	
VV188	ST 19	+	III	-	+	+	+	-	+	+	+	+	+	+	
G35	ST 197	+	NT	-	+	+	+	-	+	+	+	-	+	+	
H9	ST 131	+	V	+	+	+	+	-	+	+	+	+	+	+	
N23	ST 226	+	II	+	-	+	-	-	+	+	+	+	+	+	subCC2
V4	ST 366	+	Ia	-	-	-	+	-	+	-	+	-	+	-	
N30	ST 261	+	IV	+	-	-	-	+	+	+	+	+	+	-	
H11	ST 1	+	NT	-	-	-	-	-	+	+	+	+	+	+	
H3	ST 186	+	NT	-	+	-	-	-	+	+	-	+	+	+	
N27	ST 2	-	II	+	-	+	-	+	+	+	+	+	+	+	
G64	ST 2	+	II	+	+	-	-	+	+	+	+	+	+	+	
H2	ST 2	+	NT	-	+	-	-	-	+	+	-	+	+	+	
G61	ST 202	+	V	-	+	-	-	-	-	+	+	+	+	+	
L15	ST 196	+	IV	-	+	-	-	-	+	+	+	+	+	+	
N14	ST 250	+	NT	+	-	+	-	+	+	-	+	+	+	-	
N15	ST 250	+	NT	+	+	-	-	-	+	-	+	+	+	-	
N24	ST 357	+	NT	+	-	+	-	+	+	-	+	+	+	-	
N28	ST 250	+	NT	+	+	+	-	+	+	+	+	+	+	-	
V24	ST 7	+	IV	-	+	+	-	-	+	+	+	+	+	+	CC7
V101	ST 41	+	V	-	+	+	-	-	+	-	-	+	+	+	
G8	ST 7	+	Ia	-	+	+	-	-	+	-	-	+	+	+	
N18	ST 7	+	Ia	+	+	+	-	-	+	+	+	+	+	+	
V5	ST 6	+	Ib	-	+	-	-	-	+	+	+	-	+	+	
N11	ST 354	+	Ib	+	+	+	-	-	+	+	+	+	+	-	
LJ5	ST 10	+	Ib	-	+	+	-	-	+	+	+	+	+	+	CC10
V88	ST 10	+	II	-	+	+	+	-	+	+	-	+	-	+	
V114	ST 10	+	Ib	-	+	+	+	-	+	+	-	+	+	+	
V71	ST 8	+	Ib	-	+	+	-	-	+	+	-	+	+	+	
H10	ST 10	+	Ib	-	+	+	-	-	+	-	-	+	+	+	
N38	ST 10	+	Ib	+	-	+	-	-	+	+	+	+	+	-	
V7	ST 12	+	Ib	-	+	+	-	-	+	+	-	+	+	+	
H4	ST 12	+	Ib	-	+	+	-	-	+	+	-	+	+	+	
H13	ST 12	+	Ib	-	+	+	-	-	-	+	-	+	+	+	
H17	ST 200	+	Ib	-	+	+	-	-	+	+	-	+	+	+	
H7	ST 10	+	II	-	+	+	+	-	+	+	-	+	+	+	
V80	ST 12	+	II	-	+	+	+	+	+	+	-	+	+	+	
N3	ST 352	-	NT	+	-	-	-	-	+	-	+	+	+	-	CC67
N6	ST 353	-	NT	+	-	+	-	-	+	-	-	+	+	-	
N1	ST 350	+	NT	+	-	+	-	-	+	-	-	+	+	-	
L17	ST 195	+	III	+	+	+	-	-	+	+	+	+	+	+	
N9	ST 26	+	V	+	-	-	-	-	+	+	+	+	+	-	
N17	ST 351	-	II	+	-	+	-	-	+	+	+	+	+	-	
N40	ST 351	-	NT	+	-	+	-	-	+	+	+	+	+	-	
N2	ST 351	-	II	+	-	+	-	-	+	-	+	+	+	-	
N36	ST 365	+	II	+	-	+	-	-	+	-	+	+	-	-	
N35	ST 364	-	NT	+	-	+	-	-	+	-	+	+	+	-	
N16	ST 67	+	NT	+	-	+	-	-	-	-	+	+	-	-	
N32	ST 85	+	NT	+	-	+	-	-	+	-	+	+	-	-	
N13	ST 67	+	NT	+	-	+	-	+	+	-	+	+	-	-	
N8	ST 64	-	II	+	-	+	-	-	-	+	+	+	-	-	
N31	ST 361	-	III	+	-	+	-	+	+	-	+	+	+	+	
N33	ST 362	-	III	+	-	+	-	-	+	-	+	+	-	-	
N39	ST 61	+	NT	+	-	+	-	-	+	-	+	+	+	-	
N25	ST 61	+	III	+	-	+	-	-	+	-	+	+	-	-	
N19	ST 61	+	NT	+	-	+	-	-	-	-	+	+	+	-	
N10	ST 63	+	II	+	-	+	-	-	+	+	-	+	+	-	
H14	ST 22	+	II	+	+	+	+	-	-	+	-	-	+	+	
L13	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	CC17
L21	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	
L23	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	
L24	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	
G49	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	
V23	ST 17	+	NT	+	+	+	+	-	+	+	+	+	+	+	
V64	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	
H1	ST 17	+	III	+	-	+	-	-	+	+	+	+	+	+	
H5	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	
H8	ST 17	+	III	+	-	+	-	-	+	+	-	+	+	+	
V20	ST 367	+	III	+	-	+	-	-	+	+	-	+	-	+	
L1	ST 201	+	III	+	-	+	-	-	+	-	-	-	-	-	
V18	ST 201	+	III	+	-	+	-	-	+	+	-	+	+	+	
H15	ST 198	+	III	-	-	-	-	+	-	+	-	+	+	+	CC23
H16	ST 199	+	III	-	-	-	-	+	-	+	-	+	+	+	
N29	ST 23	+	III	+	-	+	-	-	+	+	-	+	+	-	
N37	ST 23	+	III	+	-	+	-	+	+	-	+	+	+	-	
L16	ST 23	+	III	-	-	-	-	-	+	+	+	+	+	+	
L19	ST 23	+	Ia	-	+	-	-	-	+	+	+	-	+	+	
L44	ST 23	+	Ia	+	+	+	-	-	+	+	+	+	+	+	
H6	ST 23	+	III	-	-	-	-	+	+	+	-	+	+	+	
N34	ST 363	+	III	+	-	+	-	-	+	+	+	+	+	-	
N12	ST 355	+	III	+	-	-	-	-	+	+	-	+	+	-	
N26	ST 358	+	III	+	-	-	-	-	+	+	-	+	+	+	

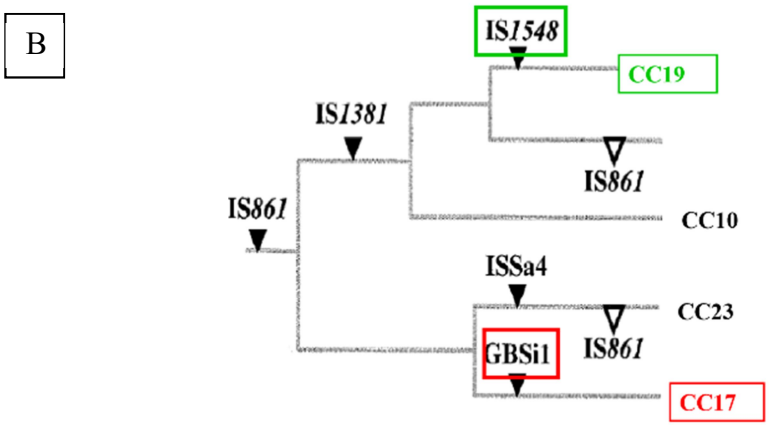


Figure 15. Relation entre complexes clonaux et EGM. D'après Hery-Arnaud et al. 2007.
A. *Pour chaque isolat est précisé : le ST, la présence d'une beta-hémolyse, le sérotype, la présence (+) ou l'absence (-) de 11 EGMs. Les isolats proviennent de prélèvements humains (Liquide cerebrospinal [L ou LJ], hémoculture de l'adulte [H], portage vaginal [V ou VV], liquide gastrique [G] et animal (N).* ***B****. Schéma évolutif montrant la chronologie d'acquisition des EGMs au cours de l'évolution de SGB, déduite de la prévalence de chaque EGM au sein des complexes clonaux. Le complexe clonal CC-17 (en rouge) est caractérisé par la présence de* GBSi1. *Les souches de CC-19 de serotype III (en vert) sont caractérisées par la présence de IS*1548.

	N° de souches	CC17	CC19	CC23	CC10	CC7	CC1	Autres CC
		n (%)						
	n = 109	n = 38	n = 21	n = 12	n = 15	n = 6	n = 7	n = 10
scpB3382 → scpB lmb ← rlmb51	40	2 (5.3)	0 (0.0)	11 (91.7)	11 (73.3)	6 (100)	5 (71.4)	5 (50.0)
scpB3382 → scpB GBSi1 lmb ← rGBSi143 ← rlmb51	51	36 (94.7)	8 (38.1)	1 (8.3)	2 (13.3)	0 (0.0)	2 (28.6)	2 (20.0)
scpB3382 → scpB IS1548 lmb ← rIS1548 ← rlmb51	18	0 (0.0)	13 (61.9)	0 (0.0)	2 (13.3)	0 (0.0)	0 (0.0)	3 (30.0)

Figure 16. Organisation de la région intergénique *scpB-lmb* et relation avec l'origine phylogénétique des souches. *D'après Al Safadi et al. 2010.*

De même, des différences de contenu prophagique sont retrouvées entre les différents sérotypes et ST ou encore en fonction du pouvoir infectieux des souches. Dès les années 1980, Stringer & Maxted ont mis au point un système de typage basé sur les bacteriophages (lysotypie) (Stringer & Maxted 1979). Par la suite, l'analyse des génomes entièrement

séquencés a mis en évidence que les gènes associés à des phages représentaient 10% des gènes séquencés (Tettelin et *al.* 2005). Les comparaisons des analyses RAPD ont révélé que la présence de trois fragments d'ADN prophagique était significativement associée aux souches à haut risque pour le nouveau-né (Figure 17) (Van der Mee-Marquet et *al.* 2006). D'autre part, l'analyse de la diversité du contenu prophagique basée sur la recherche de 10 fragments prophagiques dans le génome bactérien a permis de définir différents groupes prophagiques au sein de l'espèce (Salloum et *al.* 2011). Un lien a été mis en évidence entre l'origine d'une souche et son appartenance à un groupe prophagique ; les prophages F5, F7 et F10 étant associés aux souches à haut risque infectieux (Figure 17). Parmi les 5 groupes majeurs d'ADN prophagique décrits, deux (A et B) représentent 84% des souches de ST-17 et sérotype III (Figure 188) (Salloum et *al.* 2011).

Strain group[a]	No. of strains	No. of strains (%) with positive detection of[b]:					
		F5	F7	F10	At least:		
					1 fragment	2 fragments	3 fragments
All strains							
HR	63	17 (27)	19 (30)	38 (60)	42 (67)	28 (44)	3 (5)
LR	46	4 (9)	4 (9)	12 (26)	17 (37)	3 (6)	
Total	109	21 (19)	23 (21)	50 (48)	59 (54)	31 (28)	3 (3)

Figure 17. Prévalence de trois fragments prophagiques en fonction de l'origine clinique des souches.
D'après Van der Mee-Marquet et al. 2006. HR : Higher-risk group, LR : Lower-risk group. Les trois fragments prophagiques F5, F7 et F10 sont significativement associés aux souches à haut risque infectieux.

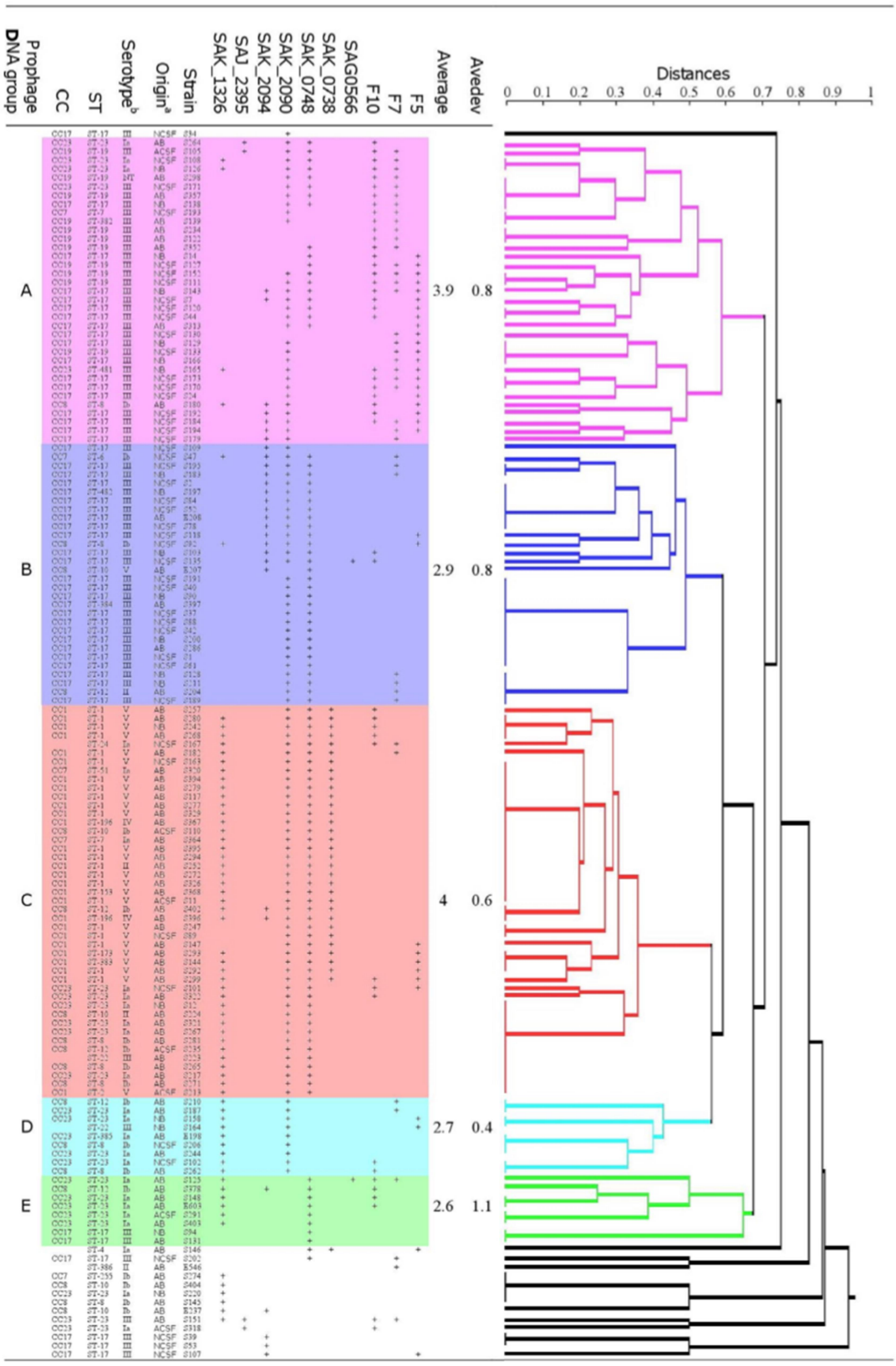


Figure 18. Distribution des souches de SGB en groupes prophagiques.
D'après Salloum et *al.* 2011. *Parmi les 5 groupes majeurs d'ADN prophagique décrits, deux (A et B) représentent 84% des souches de ST-17 et sérotype III.*

4. Amplification des séquences répétées

L'analyse des génomes bactériens entièrement séquencés a permis de mettre en évidence l'existence de séquences répétées en tandem qui varient en terme de taille, de localisation ou de type de répétition (Lindstedt 2005). Ces structures sont constituées d'une série de séquences répétées constituées de 1 à plus de 200 nucléotides, d'identité parfaite ou non, de localisation intra ou inter-génique. Ces séquences sont retrouvées à la fois chez les eucaryotes et les procaryotes (Nakamura et *al.* 1987; Lunel et *al.* 1998; Bulle et *al.* 2002). En fonction de la taille de la séquence répétée, on distingue les microsatellites (< 10 pb) des minisatellites (≥ 10 pb) (Lindstedt 2005). Ces séquences répétées ont tendance à varier en nombre de répétitions d'une souche à l'autre et sont ainsi qualifiées de VNTR (« variable number of tandem repeats ») (Nakamura et *al.* 1987). Ces variations du nombre d'unités répétées peuvent être utilisées par les bactéries comme un moyen rapide d'adaptation génomique et phénotypique à leur environnement (van Belkum 1999). Par exemple, chez *S. agalactiae,* le nombre de répétitions dans le gène *fbsA* est lié à la capacité d'adhésion des souches (Rosenau et *al.* 2007).

a) Repetitive element based PCR (rep-PCR)

Cette méthode utilise des amorces ciblant de multiples séquences répétées non-codantes dispersées sur tout le génome bactérien. Les amplicons générés produisent un profil de bandes spécifiques. Les techniques non automatisées ont été utilisées avec succès pour le typage de différentes bactéries mais restent très sensibles aux conditions expérimentales et donc peu reproductibles (Deplano et *al.* 2000; Tyler et *al.* 1997). Une technique semi automatisée, visant à standardiser la méthode est aujourd'hui commercialisée (Diversilab®, Biomérieux) mais semble moins appropriée que le MLST ou le PFGE pour comparer la diversité génétique des populations de *S. agalactiae* (Al Nakib et *al.* 2011).

b) Multiple Locus Variable-Number Tandem Repeat Analysis (MLVA)

Une méthode de typage moléculaire basée sur la variabilité des VNTR a été développée et appliquée au typage de nombreuses espèces bactériennes (Lindstedt 2005). La combinaison de plusieurs allèles retrouvés au niveau de plusieurs *loci* de VNTR permet l'étude des populations bactériennes selon la technique de MLVA (Multi *Locus* VNTR Analysis). Les premières illustrations de cette méthode de typage chez les bactéries ont été réalisées chez *Haemophilus influenzae* (Van Belkum et *al.* 1997)*, Bacillus anthracis* (Jackson et *al.* 1997) et

Mycobacterium tuberculosis (Frothingham & Meeker-O'Connell 1998; Supply et *al.* 2001). Cette méthode a depuis été utilisée pour le génotypage de nombreux pathogènes de l'homme comme *Streptococcus pneumoniae* (Koeck et *al.* 2005), *Legionella pneumophila* (Pourcel et *al.* 2007), *Pseudomonas aeruginosa* (Vu-Thien et *al.* 2007) ou *Staphylococcus aureus* (Pourcel et *al.* 2009). Cette technique présente plusieurs avantages : i) elle repose sur une différence de nombre de répétitions mesurée par la taille d'un fragment d'ADN, opération très simple à réaliser par simple migration sur gel d'agarose ; ii) elle ne nécessite pas de séquençage, d'où un coût, un temps de réalisation et un plateau technique modeste ; iii) sa fiabilité, sa reproductibilité et l'analyse des résultats en profils alléliques permet la constitution de bases de données échangeables. Pour les espèces bactériennes avec un haut niveau de diversité génétique, l'étude de six à huit marqueurs est suffisante pour obtenir une bonne discrimination entre les souches (Pourcel et *al.* 2007). Contrairement au MLST, le MLVA cible différents types de marqueurs. Chez SGB, les gènes cibles sont des gènes associés au métabolisme, à la virulence et à un ilot génomique.

Chez *S. agalactiae,* la recherche de VNTR au sein des génomes de référence entièrement séquencés (2603V/R, A909 et NEM316) à l'aide du programme Tandem Repeat Finder® a permis de mettre en évidence de nombreuses séquences répétées dispersées sur tout le génome bactérien (Figure 19).

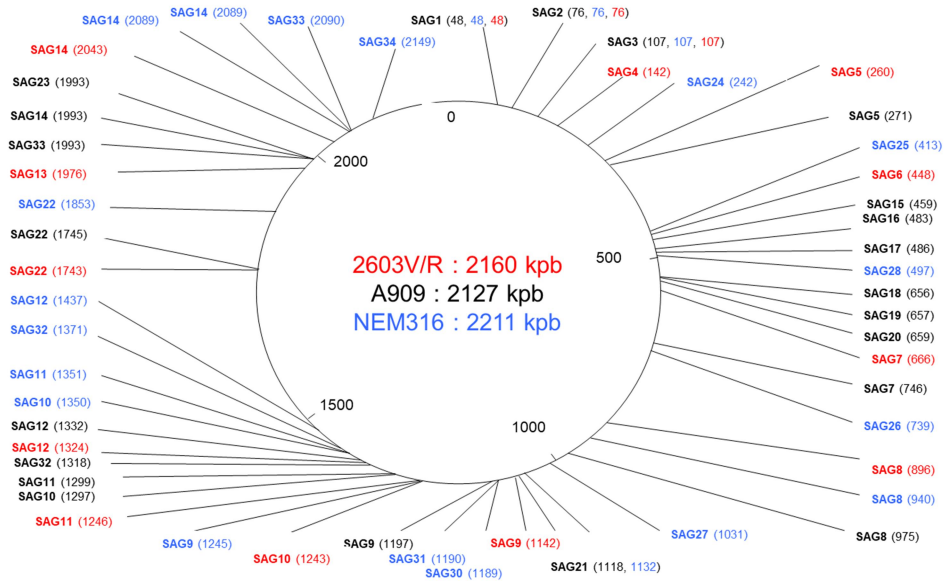


Figure 19. Position des séquences répétées en tandem dans les génomes des 3 souches de référence entièrement séquencées (2603V/R, A909 et NEM316).

Les VNTR sont ensuite sélectionnés en fonction de leur polymorphisme, leur taille et leur présence dans tous les génomes testés. Deux schémas de MLVA, reposant sur cinq (Radtke et *al.* 2010) ou six (Haguenoer et *al.* 2011) VNTR ont ainsi été décrits. Ces deux schémas ont en commun trois VNTR (SATR1 ou SAG4, SATR2 ou SAG7, SATR5 ou SAG21). Le typage MLVA défini par ces deux études a montré que le pouvoir de discrimination du MLVA était supérieur à celui du MLST. Les résultats obtenus avec le schéma à six loci permet de distinguer les souches au sein d'un même complexe clonal. Les souches de CC-23 sont divisées en deux groupes correspondant aux deux sérotypes distincts (un groupe comprenant les souches de sérotype III, un groupe comprenant les souches de sérotype Ia). Les souches de CC-17 sont également séparées en deux groupes, un groupe correspondant aux souches humaines, l'autre groupe aux souches animales. Les souches humaines appartenant au CC-17 restent cependant difficile à discriminer.

III. Le système CRISPR-Cas

A. Généralités

La diversité des souches de *S. agalactiae* décrite précédemment peut s'expliquer, en partie, par les réarrangements génomiques nécessaires pour l'évolution et l'adaptation de la bactérie. L'analyse des génomes de *S. agalactiae* entièrement séquencés montre une organisation mosaïque. En effet, le core génome est entrecoupé par de multiples ilots génomiques retrouvés uniquement dans une partie des génomes séquencés (Lopez-Sanchez et *al.* 2012; Tettelin et *al.* 2005). Ces ilots renferment fréquemment des séquences issues d'EGM comme des séquences prophagiques acquises par lysogénie.

Les EGMs, notamment les bactériophages (ou phages), sont fréquemment vecteurs de facteurs de virulence et sont des acteurs majeurs dans l'adaptation des pathogènes à leurs hôtes (Brussow et *al.* 2004). Toutefois, les phages, organismes prépondérants sur la planète, sont également une menace constante pour la survie des bactéries. Ainsi, afin de maintenir un équilibre, les bactéries ont développé divers mécanismes de défense contre les bactériophages, pouvant concerner différentes étapes du cycle d'infection phagique (Figure 20) (Labrie et *al.* 2010).

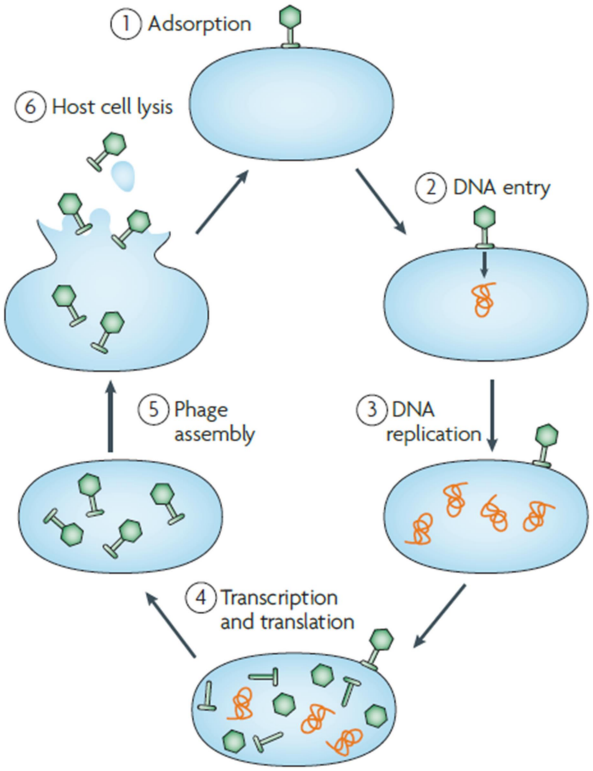


Figure 20. Cycle de réplication d'un bactériophage.
Chaque étape peut être la cible d'un mécanisme anti-phage bactérien (Labrie et *al.* 2010).

Les bactéries peuvent bloquer la première étape d'infection phagique qui consiste en l'adsorption du phage à la surface bactérienne (Labrie et *al.* 2010). Elles peuvent ainsi modifier le récepteur spécifique du bactériophage, synthétiser une matrice extracellulaire afin de créer une barrière physique entre le phage et son récepteur, ou encore produire des inhibiteurs compétitifs qui se fixent au récepteur du phage. La seconde étape du cycle de réplication des phages peut aussi être inhibée. L'entrée de l'ADN phagique dans la cellule bactérienne peut être inhibée par les systèmes d'exclusion de la superinfection (Sie). La multiplication phagique peut ensuite être inhibée par les systèmes Abi (Abortive infection). D'autres mécanismes permettent la dégradation des acides nucléiques étrangers, comme les systèmes de Restriction-Modification, présents chez la quasi-totalité des bactéries (Labrie et *al.* 2010). Ces systèmes associent une enzyme de restriction, qui reconnaît et clive l'ADN étranger, et une méthylase qui protège l'ADN bactérien.

Récemment, un rôle similaire a été attribué aux systèmes CRISPR-Cas (Barrangou et *al.* 2007). L'acronyme CRISPR, pour Clustered Regularly Interspaced Short Palindromic Repeats, est dû à la composition du *locus* constitué de séquences répétées, de taille et de séquences très conservées pour un génome donné, séparées par des séquences variables mais

de longueur relativement conservée. Ces séquences sont systématiquement associées à un opéron *cas* (CRISPR associated genes) pour former le système CRISPR-Cas. Ces systèmes sont présents dans de nombreux génomes procaryotes (90% des Archaea et 47% des bactéries) et codent pour un véritable système immunitaire adaptatif pour se défendre contre les EGMs, notamment les bactériophages (Barrangou et *al.* 2007; Horvath & Barrangou 2010; Van Der Oost et *al.* 2009).

1. Histoire de la découverte

La découverte de ces *loci* particuliers a été faite il y a 30 ans par Ishino *et al.* qui ont mis en évidence des séquences répétées, palindromiques, séparées par des séquences uniques au sein du génome d'*E. coli* (Ishino et *al.* 1987). Une dizaine d'années plus tard, une analyse *in silico* menée par Mojica *et al.* a mis en évidence la présence de ces fragments d'ADN répétés dans de nombreux génomes de groupes phylogénétiques éloignés et une large distribution chez les procaryotes (Mojica et *al.* 2000). Ces séquences répétées en tandem sont alors qualifiées de Short Regularly Spaced Repeats (SRSRs). Ces SRSRs sont généralement organisés en cluster, sont de nature palindromique, sont très homogènes pour un génome donné (la majorité d'entre eux étant identiques) et sont séparés de manière régulière. Ces observations suggèrent une fonction commune à ces loci chez les procaryotes. Deux ans plus tard, ces structures ont été rebaptisées CRISPR pour *Clustered Regularly Interspaced Short Palindromic Repeats* par Jansen *et al.* (Jansen et *al.* 2002a). Dans ce même travail, les auteurs décrivent, associée à ces structures, la présence d'une séquence *leader* de 300 à 500 pb, conservée pour une espèce donnée et la présence de gènes associés à proximité du locus CRISPR : les gènes *cas* (pour CRISPR associated genes). Aidé par la disponibilité de génomes entièrement séquencés, ils identifient ces séquences dans la majorité des génomes des Archaea et dans la moitié des génomes bactériens. Ces structures sont exclusivement retrouvées chez les procaryotes. Une des particularités de ces séquences répétées est l'existence de spacers de taille similaire, non répétés séparant les séquences répétées. La signification biologique de ces séquences est alors inconnue.

Le rôle biologique de ce locus a été élucidé au milieu des années 2000. En 2005, 3 groupes ont rapporté une homologie entre les séquences uniques (les spacers) séparant les séquences répétées et des EGMs, notamment des phages et des plasmides (Bolotin et *al.* 2005; Pourcel et *al.* 2005; Mojica et *al.* 2005). Ces données ont conduit à l'hypothèse que CRISPR pouvait fonctionner comme un système immunitaire contre les bactériophages, et plus

généralement, contre l'ADN étranger. D'autre part, les gènes *cas* semblent systématiquement associés au locus CRISPR et sont situés à proximité, suggérant une fonction associée. Makarova et *al.* propose que ces gènes codent pour les protéines de la machinerie cellulaire d'un système de défense siRNA-like (Makarova et *al.* 2006). Ce système fonctionnerait par l'intégration de fragments de gènes étrangers au sein du chromosome du procaryote, conduisant à une immunité héréditaire à cet agent.

En 2007, 20 ans après la découverte du locus CRISPR, la fonction du système CRISPR-Cas comme système immunitaire a été mise en évidence par l'équipe de Philippe Horvath chez *Streptococcus thermophilus* (Barrangou *et al.* 2007). Ces auteurs ont démontré que le système CRISPR-Cas permettait l'acquisition d'une immunité contre les phages. Cette immunité est spécifique d'un phage via l'acquisition d'un spacer de séquence identique à une séquence présente au niveau du génome du phage à l'origine de l'infection. Après un « challenge » avec un bactériophage, la bactérie exposée acquiert un (ou plusieurs) nouveau(x) spacer(s) dérivant du génome phagique. Le phénotype de résistance phagique de la bactérie est modifié par le retrait ou l'addition d'un spacer particulier. D'autre part, la présence des gènes *cas* est indispensable à l'acquisition d'un nouveau spacer. CRISPR, associé aux protéines Cas, confère une résistance contre les bactériophages et cette résistance est spécifique et déterminée par l'homologie de séquence entre le spacer et le bactériophage (la présence d'un polymorphisme nucléotidique entre la séquence du spacer et du bactériophage ne confère pas de résistance). De plus, une bactérie ayant acquis ce spacer et étant devenue résistante à ce phage, peut redevenir sensible lorsque ce spacer est expérimentalement ôté (Barrangou et *al.* 2007). Cette acquisition de nouveau spacer se fait de manière polarisée comme suggéré précédemment (Pourcel et *al.* 2005). Chaque nouvelle acquisition de spacer, en réponse à une attaque de phage, survient au niveau de l'extrémité du *locus* CRISPR adjacente au leader.

2. Composition du système CRISPR-Cas

Les premières analyses comparatives des loci CRISPR ont révélé des différences majeures entre les séquences répétées, les séquences et l'architecture des gènes *cas* (Kunin et *al.* 2007; Jansen et *al.* 2002b; Makarova et *al.* 2006). Cependant, les systèmes CRISPR-Cas présentent généralement la même organisation. Ils sont composés d'un *locus* CRISPR flanqué en amont par une région *leader* et en aval par une région *trailer*, ainsi que des gènes *cas* associés regroupés en opéron (Figure 21).

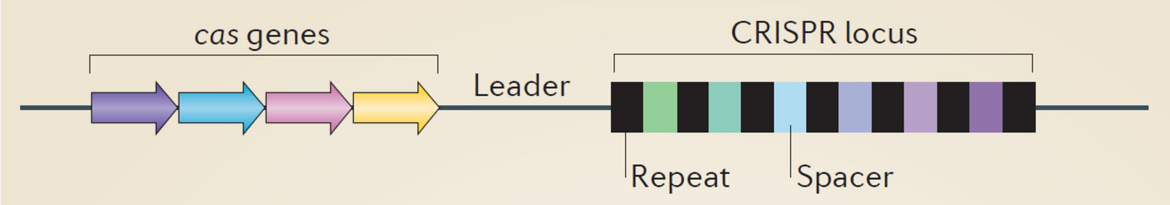


Figure 21. Organisation générale du système CRISPR-Cas.
D'après Van der Oost et al. 2014. Le système CRISPR-Cas est composé de deux à trois régions. Le locus CRISPR constitué d'une alternance de repeats (rectangle noirs) et de spacers (rectangles colorés) et l'opéron cas *situé en amont, constitué de différents gènes. Ces deux régions peuvent être séparées par le leader.*

Plusieurs systèmes CRISPR-Cas peuvent être présents dans un même génome. Cependant, la majorité des organismes ne possèdent que un ou deux loci CRISPR (Grissa et *al.* 2007). La localisation des systèmes CRISPR-Cas est classiquement chromosomique mais peut également être plasmidique (Mojica et *al.* 2005). Les systèmes CRISPR-Cas sont très variables probablement du fait d'une évolution rapide du système immunitaire en réponse à la dynamique de pression de sélection imposée par les EGMs.

a) Le locus CRISPR

Les loci CRISPR sont constitués d'une alternance de courts « tronçons » d'ADN répétés non contigus, appelés repeats ou Direct Repeats (DR), séparés par des spacers de séquences variées.

Les repeats ou DR

Une large diversité de repeats est observée entre les espèces microbiennes. Ces séquences de 20 à 47 pb selon les espèces peuvent être répétées jusqu'à environ 600 fois dans un seul génome (*Haliangium ochraceum* DSM 14365). Cependant, la plupart des *loci* contiennent moins de 50 unités (Barrangou & Marraffini 2014). Bien que ces séquences soient très variables selon les espèces, elles sont généralement très conservées pour un locus CRISPR donné (Jansen et *al.* 2002a; Grissa et *al.* 2007).

Comme indiqué dans l'acronyme CRISPR, les DR sont généralement partiellement palindromiques et adaptés pour former des structures secondaires stables et très conservées (Kunin et *al.* 2007). Chaque locus CRISPR commence et se termine par un DR. Le DR situé à la fin du locus, appelé DR terminal (DRT), est le plus souvent tronqué ou dégénéré par rapport au DR interne (DR consensus) (Barrangou et *al.* 2007).

Les spacers

Les spacers sont généralement uniformes en taille (classiquement 21 à 72 pb selon les espèces) mais possèdent des séquences très variables (Horvath et *al.* 2008). Ces séquences sont généralement uniques dans le locus, bien que certains spacers puissent être dupliqués. Les spacers déterminent la spécificité du mécanisme de défense et possèdent pour la plupart une origine extra-chromosomique, comme les phages ou les plasmides. Toutefois certains spacers sont d'origine chromosomique et proviennent du chromosome hébergeant le locus CRISPR. Le polymorphisme d'un système CRISPR-Cas donné, pour une espèce bactérienne donnée, est essentiellement dû à la nature et au nombre de spacers.

b) Régions flanquantes du locus CRISPR

La plupart des loci CRISPR sont flanqués d'un côté par la séquence *leader*, située directement en amont du premier repeat du locus CRISPR, et de l'autre côté par la séquence *trailer*, située après le DRT. Le *leader* est une séquence riche en A/T, non codante, de taille variable (entre 100 et 550 pb) (Jansen et *al.* 2002b; Lillestøl et *al.* 2006). On observe des séquences *leader* très différentes parmi les espèces microbiennes, mais cette séquence est le plus souvent hautement conservée pour un système CRISPR-Cas donné. La région *leader* contient le promoteur nécessaire à la transcription du locus CRISPR (Hale et *al.* 2008; Lillestøl et *al.* 2006).

c) Les gènes *cas*

Les gènes *cas* sont généralement situés à proximité du locus CRISPR et en amont du *leader* (Barrangou 2013). Ces gènes n'ont été décrits que sur des génomes porteurs d'un locus CRISPR, d'où leur nom pour « CRISPR associated ». Les gènes *cas* sont organisés en opéron dont l'architecture est très variable en fonction des systèmes CRISPR-Cas considérés (Makarova et *al.* 2011, Makarova et *al.* 2015). La nature des gènes *cas* a permis une classification des différents systèmes actuellement connus (Makarova et *al.* 2011, Makarova et *al.* 2015).

Les protéines codées par les gènes *cas* fournissent l'équipement enzymatique nécessaire au fonctionnement du système et peuvent être divisées en quatre modules fonctionnels distincts (Makarova et *al.* 2015). Trois des quatre modules correspondent aux étapes fonctionnelles du système, le 4e module correspond à une fonction auxiliaire de régulation ou d'autres fonctions associées (Makarova et *al.* 2015). Parmi toutes les protéines Cas décrites,

les protéines Cas1 et Cas2 sont présentent dans la quasi-totalité des systèmes et forment le module d'adaptation impliqué dans l'acquisition de nouveaux spacers.

3. Classification

Avec l'augmentation de la quantité de données sur la composition et le fonctionnement des systèmes CRISPR-Cas, la première classification de ces systèmes a été réévaluée (Haft et *al.* 2005, Makarova et *al.* 2011, Makarova et *al.* 2015; Koonin et *al.* 2017). Bien que le complexe d'adaptation montre peu de variations, le complexe d'interférence est plus variable et a permis de distinguer les différents systèmes. Chaque séquence de spacer est transcrite et transformée en court ARN appelé CRISPR-RNA (crRNA) qui va former un complexe effecteur avec des protéines Cas qui diffèrent en fonction des systèmes CRISPR-Cas.

Aujourd'hui, deux classes de système CRISPR-Cas ont été définies: les systèmes de classe 1 pour lesquels le complexe crRNA-effecteur est composé de plusieurs sous-unités (CRISPR-associated complex for antiviral defense [Cascade]) et les systèmes de classe 2 pour lesquels toutes les fonctions du complexe effecteur sont portées par une seule protéine (Cas9 par exemple pour les systèmes de type II) (Figure 22).

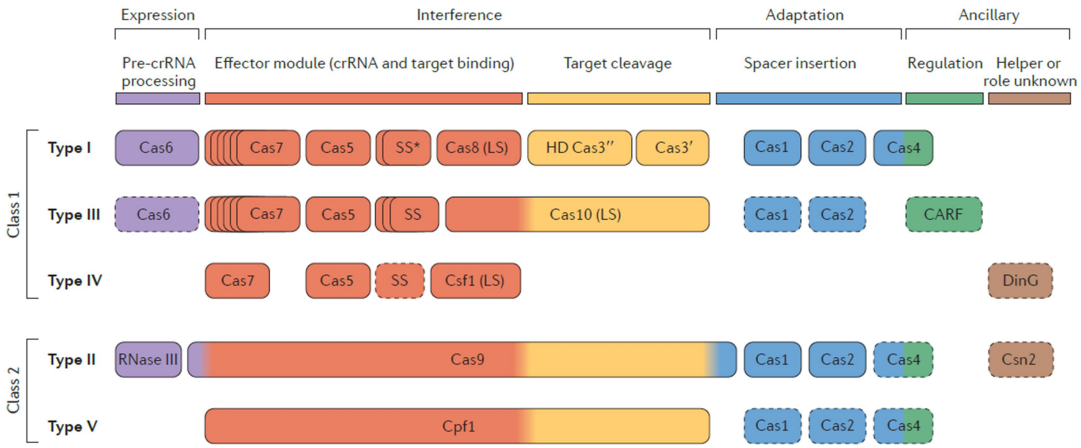


Figure 22. Classification fonctionnelle des protéines Cas et des systèmes CRISPR-Cas associés (Makarova et *al.* 2015).

Chaque classe est divisée en types, caractérisés par un gène *cas* signature, et en sous-types définis par les autres gènes *cas* et par leur organisation. Aujourd'hui, six types et de nombreux sous-types ont été définis (Figure 23) (Makarova et *al.* 2015; Shmakov et *al.* 2017; Koonin et *al.* 2017).

Les systèmes de classe 1 sont composés des systèmes de type I et III et le putatif nouveau type IV.

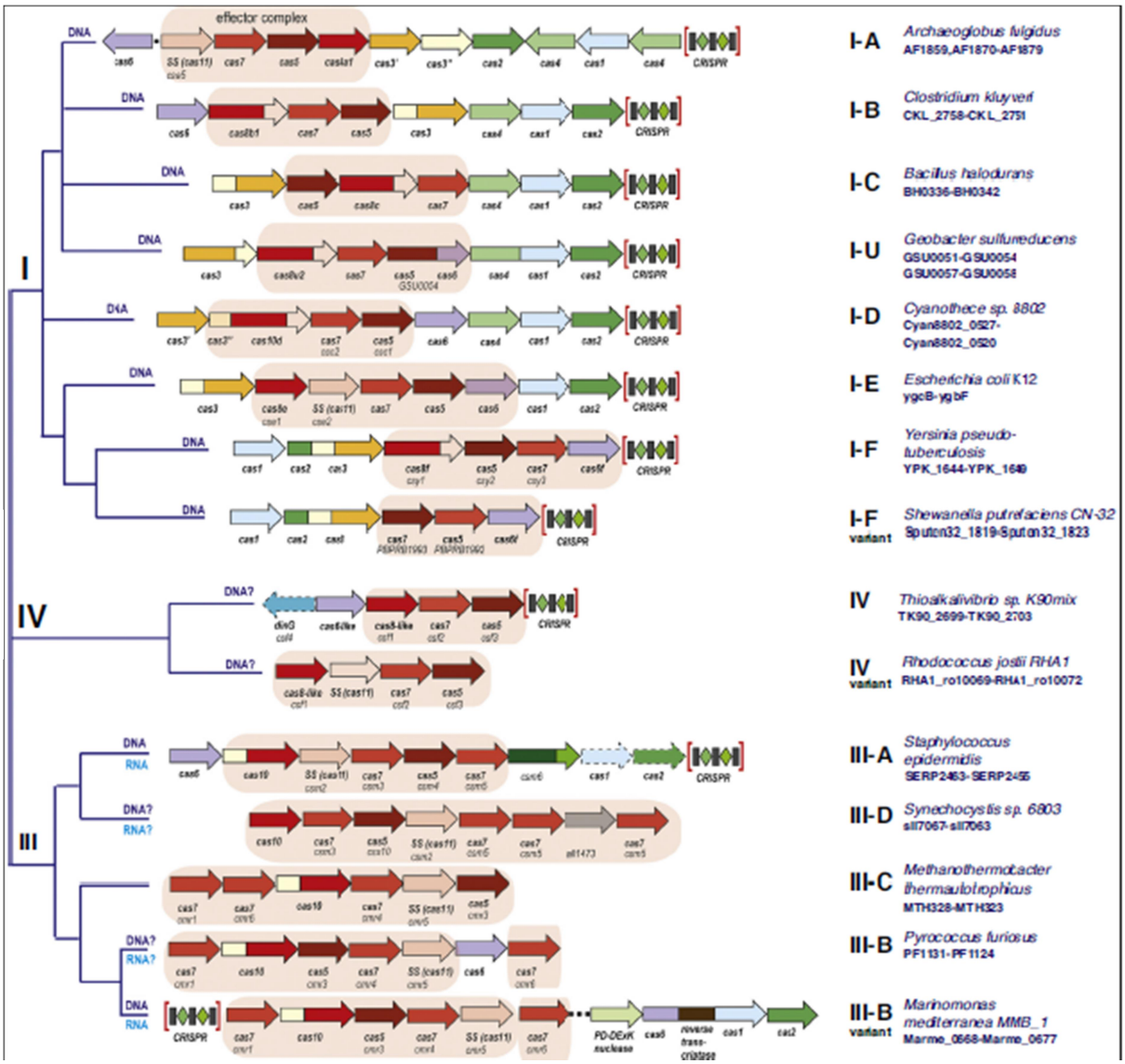
- Les systèmes de type I sont caractérisés par la présence du gène signature *cas3* et sont divisés en sept sous-types (I-A à I-F et I-U) ayant des gènes *cas* et des organisations en opéron différentes.
- Les systèmes de type III sont caractérisés par la présence du gène signature *cas10* et contiennent deux sous-types (III-A et III-B).
- Les systèmes putatifs de type IV, sont caractérisés par la présence du gène signature *csf1* et sont composés de deux sous-types. Dans ce système, les gènes *cas* sont souvent situés à distance du locus CRISPR ou dans des génomes sans locus CRISPR. La fonction de ce système Cas (plus que CRISPR-Cas) est aujourd'hui non caractérisée.

Les systèmes de classe 2 comportent le système de type II et les putatifs nouveaux types V et VI.

- Les systèmes de type II sont caractérisés par la présence du gène signature *cas9.* Ils diffèrent largement des types I et III, notamment en termes de nombre de gènes. Il s'agit des systèmes les plus simples puisqu'ils ne sont composés que de 3 ou 4 gènes *cas*. La protéine Cas9 forme un complexe avec le crRNA pour former le complexe crRNA-effecteur qui clive l'ADN cible et contribue également à l'adaptation. Trois sous-types composent les systèmes de type II (II-A à II-C) en fonction de la composition en gènes *cas.* Par exemple, le type II-A se compose de l'enchainement *cas9, cas1, cas2, csn2*, le type II-B contient *cas9, cas1, cas2* et *cas4*, et le type II-C ne contient que les trois premiers gènes. Un autre gène signature est présent dans les systèmes de type II : le *trans*-activating CRISPR RNA (tracrRNA). Ce gène est impliqué dans la maturation du precrRNA en crRNA avec l'aide d'une RNAse III (gène de ménage bactérien) et forme un composé essentiel au complexe de clivage Cas9 (Deltcheva et *al.* 2011; Jinek et *al.* 2012).

- Les systèmes de type V et VI sont de description récente et sont caractérisés respectivement par les gènes signatures *cas12* et *cas13* (Shmakov et *al.* 2017; Koonin et *al.* 2017).

Classe 1


I
IV
III
DNA
DNA?
RNA
RNA?
effector complex
CRISPR
I-A Archaeoglobus fulgidus AF1859,AF1870-AF1879
I-B Clostridium kluyveri CKL_2758-CKL_2751
I-C Bacillus halodurans BH0336-BH0342
I-U Geobacter sulfureducens GSU0051-GSU0054 GSU0057-GSU0058
I-D Cyanothece sp. 8802 Cyan8802_0527-Cyan8802_0520
I-E Escherichia coli K12 ygcB-ygbF
I-F Yersinia pseudo-tuberculosis YPK_1644-YPK_1649
I-F variant Shewanella putrefaciens CN-32 Sputcn32_1819-Sputcn32_1823
IV Thiosulfativibrio sp. K90mix TK90_2699-TK90_2703
IV variant Rhodococcus jostii RHA1 RHA1_ro10069-RHA1_ro10072
III-A Staphylococcus epidermidis SERP2453-SERP2455
III-D Synechocystis sp. 6803 sll7067-sll7063
III-C Methanothermobacter thermautotrophicus MTH328-MTH323
III-B Pyrococcus furiosus PF1131-PF1124
III-B variant Marinomonas mediterranea MMB_1 Marme_0668-Marme_0677

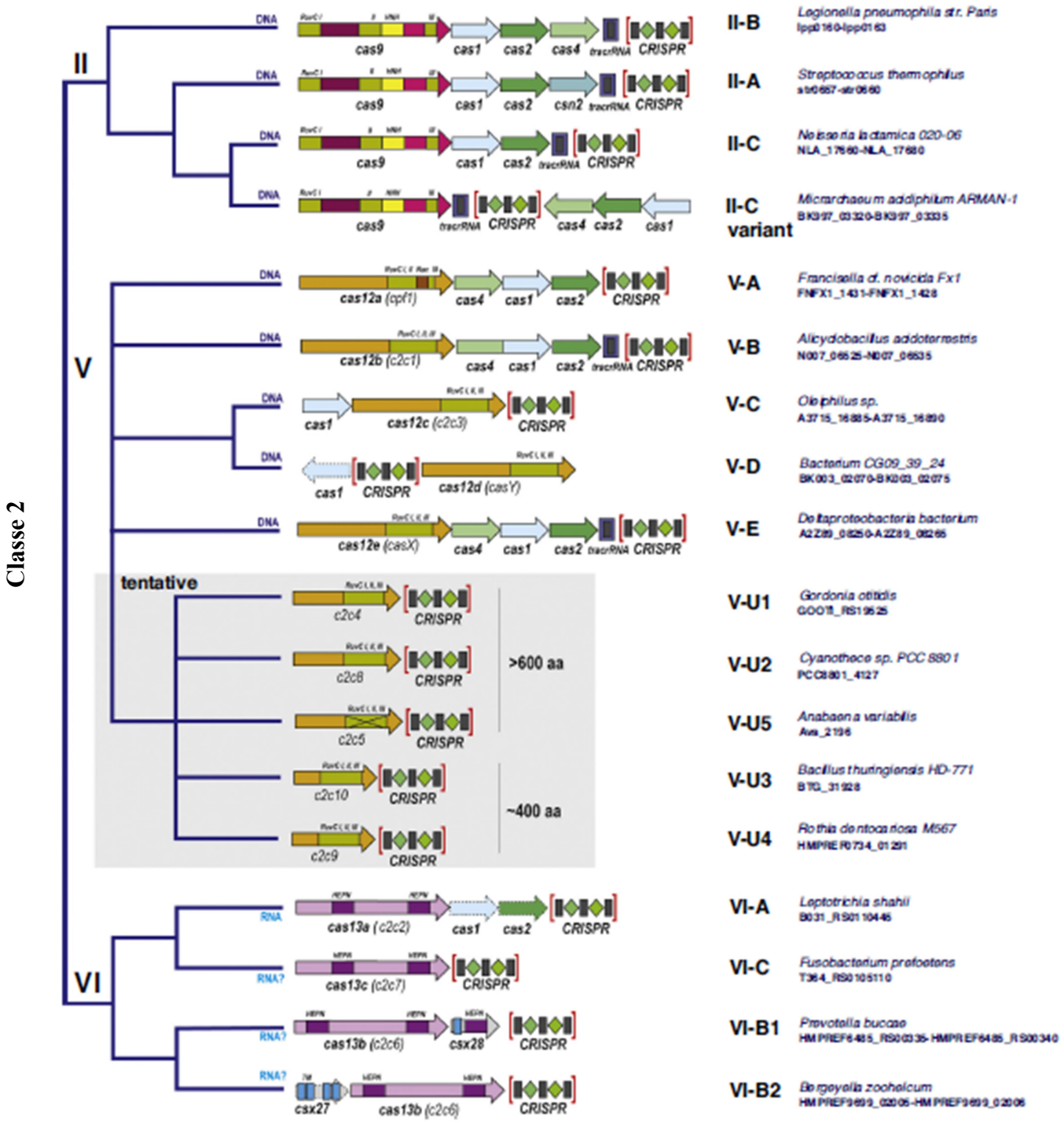


Figure 23. Représentation de l'architecture de l'opéron *cas* permettant le sous-typage des systèmes CRISPR-Cas (Koonin et *al.* 2017).

4. Mode de fonctionnement

L'immunité adaptative via le système CRISPR-Cas se produit en trois étapes : adaptation, expression et interférence (Figure 24) (van der Oost et *al.* 2009; Al-Attar et *al.* 2011; Barrangou & Marraffini 2014; Barrangou & Horvath 2017). Ces différentes étapes reposant sur l'action des protéines Cas, les différences de composition en gène *cas* des systèmes entraînent des différences de mécanisme d'action (Richter et *al.* 2012). Cependant, malgré la grande diversité des gènes *cas*, tous les systèmes présentent un même principe moléculaire : les crRNA contiennent un spacer partiel unique dérivant de « l'envahisseur » qui guide une ou plusieurs protéines Cas vers l'acide nucléique apparenté pour éventuellement le détruire après reconnaissance spécifique de sa séquence (Charpentier et *al.* 2015).

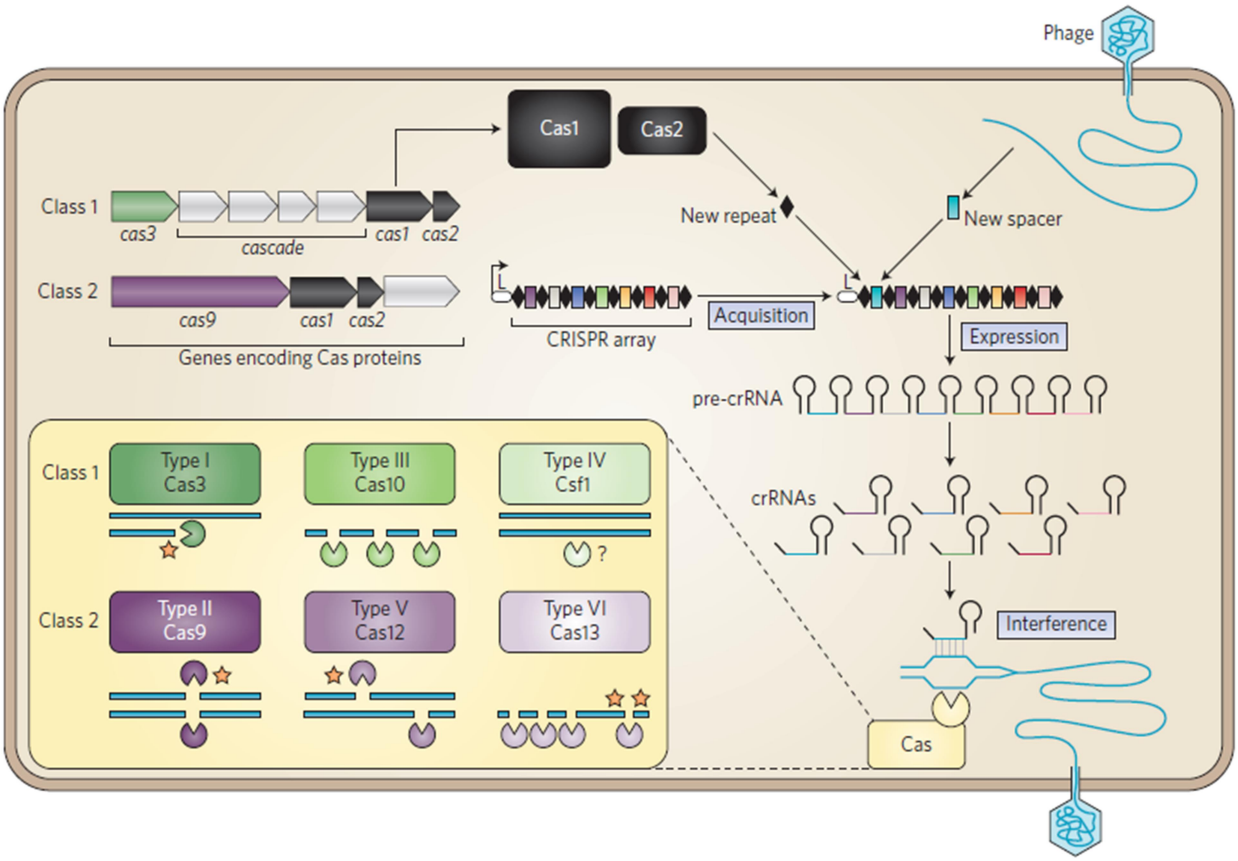


Figure 24. Mécanisme d'action du système CRISPR-Cas (Barrangou & Horvath 2017).

a) Phase d'adaptation

La phase d'adaptation ou « immunisation » correspond à l'incorporation d'un fragment d'ADN étranger provenant généralement d'un EGM (appelé « protospacers ») au sein du locus CRISPR sous forme d'un nouveau spacer. Ces spacers sont à l'origine de la mémoire immunitaire, nécessaire à une défense ciblée contre les invasions ultérieures par l'EGM correspondant. Les protéines Cas1 et Cas2, seules protéines conservées au sein des systèmes

CRISPR identifiés, sont nécessaires et dans certains systèmes suffisants à cette étape d'acquisition (Figure 24) (Amitai & Sorek 2016; Koonin et *al.* 2017; Makarova et *al.* 2015; Yosef et *al.* 2012).

L'acquisition d'un nouveau spacer se fait de manière polarisée à l'extrémité de la séquence CRISPR adjacente au *leader* et s'accompagne de la duplication du DR situé à l'extrémité *leader*. Cette intégration polarisée fournit donc une vision chronologique des immunisations de la bactérie. La présence du *leader*, associée à la présence d'un repeat est indispensable à l'acquisition de nouveaux spacers (Wei et *al.* 2015; Yosef et *al.* 2012). La séquence clé du *leader* est située à proximité immédiate du 1er DR. Les 10 nucléotides correspondant à la jonction « *leader*-repeat » fournissent l'information cruciale à l'acquisition de nouveaux spacers (Van Orden et *al.* 2017; Wang et *al.* 2016; Wei et *al.* 2015; Yosef et *al.* 2012). L'intégration polarisée repose sur la jonction *leader*-repeat dans les systèmes de type I et II avec cependant des différences entre les systèmes. Dans les systèmes de type I, les protéines Cas1 et Cas2 reconnaissent une séquence interne entre le *leader* et le repeat, alors que dans les systèmes de type II, la séquence reconnue est la fin du repeat (Wright & Doudna 2016).

b) Phase d'expression

La phase d'expression correspond à la transcription du locus CRISPR, à partir d'un promoteur situé dans la région leader, en un précurseur pre-crRNA suivi du clivage au niveau des repeats de ce long pre-crRNA en petits ARNs CRISPR matures (crRNAs). Les crRNAs se composent d'une séquence dérivant d'un spacer flanquée de fragments des DR adjacents : un fragment de 8 à 9 nucléotides en 5' et un fragment plus long et plus hétérogène en 3' (van der Oost et *al.* 2009).

Le mécanisme d'action et les enzymes impliquées pour ce clivage varient selon les types de systèmes CRISPR-Cas. Dans les systèmes de type I et de type III, une nucléase Cas6-like est responsable de la transformation du pre-crRNA (Chylinski et *al.* 2014). Dans les systèmes de type II, la transformation du pre-crRNA implique une RNAse III bactérienne et le tracrRNA (Deltcheva et *al.* 2011). Le tracrRNA contient une séquence complémentaire de la région repeat du pre-crRNA. La structure double brin obtenue est ensuite clivée par la RNAse III (Deltcheva et *al.* 2011; van der Oost et *al.* 2014).

Pendant la phase d'expression, les crRNA matures sont associés à des protéines Cas formant le complexe crRNA-effecteur (ou crRNP) qui permet la surveillance immunitaire. Dans les systèmes CRISPR-Cas de type I, le complexe effecteur, appelé « Cascade », est constitué de multiples sous-unités formées par des protéines Cas. Dans les systèmes CRISPR-Cas de type III, le complexe effecteur présente des similarités structurales avec le complexe Cascade et est appelé Csm et Cmr pour les types III-A et les types III-B, respectivement. Pour les systèmes CRISPR-Cas de type II, le complexe crRNA-effecteur est formé uniquement de la protéine Cas9 et des deux ARN (crRNA-tracrRNA) (Figure 25).

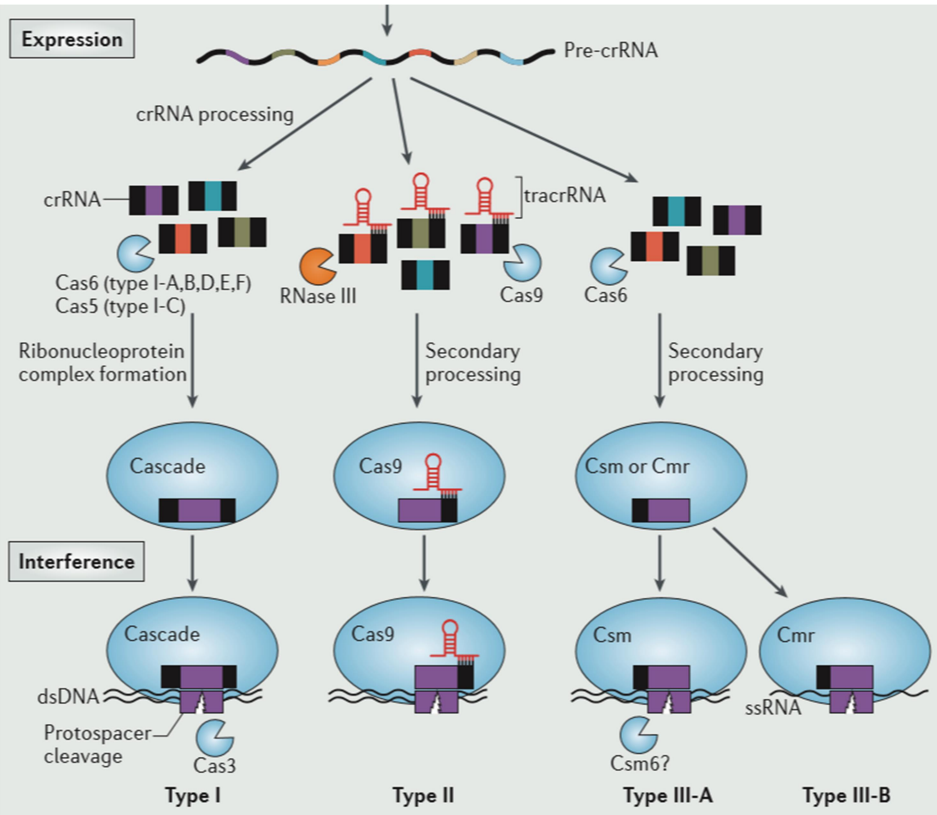


Figure 25. Mécanisme d'action des étapes d'expression et d'interférence du système CRISPR-Cas (Westra et *al.* 2014).

c) Phase d'interférence

La phase d'interférence ou « immunité » correspond à la reconnaissance et au clivage de l'ADN étranger par les nucléases Cas guidées par les crRNA. Pendant cette étape, le crRNA mature reste lié à la protéine Cas9 pour les systèmes de type II, ou au complexe effecteur à plusieurs sous-unités pour les autres systèmes. Les crRNA ciblent spécifiquement une séquence apparentée du génome envahisseur et guident les nucléases vers l'ADN ou l'ARN complémentaire (Deltcheva et *al.* 2011; Gasiunas et *al.* 2012). Différentes nucléases Cas sont

décrites et varient en fonction du type de système CRISPR-Cas. Par exemple, pour les systèmes de type I, l'exonucléase Cas3 coupe et rogne le brin d'ADN cible, tandis que dans les systèmes de type II, l'endonucléase Cas9 génère deux coupures conduisant à une cassure du double brin d'ADN (Figure 24) (Barrangou & Horvath 2017).

d) Soi versus non-soi : rôle du Protospacer Adjacent Motif (PAM)

La discrimination du soi et du non-soi par les systèmes CRISPR-Cas est possible grâce à la présence d'un court motif adjacent au sein de l'ADN cible appelé PAM (Protospacer Adjacent Motif) (Deveau et *al.* 2008; F. J. M. Mojica et *al.* 2009; Shah et *al.* 2013). Cette courte séquence est très conservée et spécifique d'un système donné mais sa séquence et sa localisation varient en fonction des types de systèmes CRISPR-Cas. La présence de la séquence PAM permet la reconnaissance de l'acide nucléique étranger durant la phase d'interférence. Une fois assemblé, le complexe de surveillance crRNA-effecteur « scanne » la cellule à la recherche de séquences PAM. En cas de reconnaissance, lors de la fixation, le complexe évalue l'importance de l'appariement de bases entre la séquence en aval (protospacer) et la portion du crRNA (spacer). Si un nombre suffisamment important d'appariements de bases a lieu, la séquence cible est clivée et dégradée par les protéines Cas. La présence du PAM est indispensable à la reconnaissance de la séquence cible. En cas d'absence de séquence PAM, la protéine effectrice « n'interroge » pas la séquence en aval même en cas de complémentarité parfaite avec le crRNA (Leenay & Beisel 2017).

Outre son rôle dans l'étape d'interférence, le PAM joue également un rôle essentiel dans la sélection du protospacer durant la phase d'acquisition (Leenay & Beisel 2017). Concernant les systèmes de type II de deux espèces de *Streptococcus* (*S. pyogenes* et *S. thermophilus*), cette reconnaissance des séquences PAM se fait via la protéine Cas9 (Heler et *al.* 2015). Cas9 est donc impliquée dans le clivage des acides nucléiques et est également nécessaire pour déterminer la séquence PAM et intégrer les spacers au sein du *locus* CRISPR. Cette propriété semble propre à la protéine Cas9. Dans d'autres systèmes, comme le système de type I d'*E. coli*, les protéines Cas1 et Cas2 sont nécessaires et suffisantes pour sélectionner un protospacer contenant un PAM, cette intégration de nouveau spacer étant améliorée par la présence de la protéine Cas4 (Kieper et *al.* 2018; Yosef et *al.* 2012).

5. Application comme outil de typage

Le système immunitaire adaptatif CRISPR-Cas est à l'origine de la « vaccination » contre les EGMs envahisseurs. Cette immunité est acquise par capture et intégration d'ADN étranger au cours du temps sous forme de nouveau spacer au sein du locus CRISPR. Les événements immunitaires sont enregistrés génétiquement, les uns après les autres, de manière chronologique. Du fait de l'acquisition polarisée de nouveaux spacers à l'extrémité *leader* du locus CRISPR, les spacers situés côté *leader* donnent des informations sur les événements récents, tandis que les spacers situé à l'extrémité opposée (côté *trailer*) témoignent de contacts anciens avec des EGMs et offrent un aperçu de l'origine phylogénétique des souches (spacers ancestraux). L'organisation temporelle et le polymorphisme des spacers des loci CRISPR en font des marqueurs phylogénétiques attractifs pour le génotypage et l'analyse phylogénétique (Shariat & Dudley 2014).

a) Spoligotyping

Les premières méthodes de typage utilisant les *loci* CRISPR ont vu le jour avant même de connaitre la fonctionnalité du système. La première méthode développée utilisant ce locus a été appliquée à *Mycobacterium tuberculosis* (Groenen et *al.* 1993; Kamerbeek et *al.* 1997). Le haut niveau de polymorphisme dû à la variabilité des spacers a permis la mise au point du spoligotyping (*Sp*acer-*oligo*nucleotide typing), permettant de séparer les isolats au sein du complexe *tuberculosis* (Kamerbeek et *al.* 1997). Cette technique repose sur l'amplification du *locus* CRISPR avec des amorces ciblant les DR, suivit de l'hybridation des produits de PCR à une membrane sur laquelle ont été préalablement fixées des sondes portant les séquences des spacers. Un des inconvénients de cette méthode est qu'elle recherche uniquement la présence de spacers de séquences connues. Cette méthode a par la suite été appliquée au typage d'autres bactéries comme *Corynebacterium diphteriae* (Mokrousov et *al.* 2007; Mokrousov et *al.* 2009) ou *Legionella pneumophila* (Ginevra et *al.* 2012).

b) Typage CRISPR par séquençage

Peu de temps après le développement du spoligotyping, Hoe et *al* amplifie une région chromosomique de *Streptococcus pyogenes* de serotype M1 contenant des séquences répétées polymorphes (Hoe et *al.* 1999). L'analyse de la taille de la région chez 14 souches permet de mettre en évidence des différences correspondant aux types RFLP de deux séquences d'insertion (IS*1548* et IS*1562*). Le séquençage de cette région chez une quarantaine de

souches montre des variations entre les souches, à la fois en nombre et en séquence de spacers (Hoe et *al.* 1999). Cette étude est la première à mettre en évidence l'intérêt du séquençage du locus CRISPR pour le typage bactérien.

Suite à la description du rôle du système CRISPR-Cas chez *Streptococcus thermophilus,* les systèmes CRISPR chez cette bactérie ont été particulièrement étudiés. Ces systèmes montrent une forte diversité (principalement des systèmes associés aux loci CRISPR1 et CRISPR3) et une forte occurrence parmi les différentes souches étudiées. La composition en spacers de ces loci donne ainsi des informations sur l'origine de la souche et leur type génétique, laissant suggérer leur intérêt pour le typage et l'analyse des populations (Horvath et *al.* 2008). Par la suite, l'analyse des loci CRISPR de *Yersinia pestis* (3 loci) a permis de séparer des isolats entre eux, y compris des isolats de même profil MLVA (Pourcel et *al.* 2005; Cui et *al.* 2008). Cette analyse permettait également de distinguer les isolats en fonction de l'origine géographique de la souche (Cui et *al.* 2008).

Une des espèces clé dans la mise en évidence de l'intérêt du *locus* CRISPR comme outil de typage est *Salmonella enterica* (Fabre et *al.* 2012; Pettengill et *al.* 2014; Fricke et *al.* 2011; Liu et *al.* 2011; Liu, et *al.* 2011; Timme et *al.* 2013; Wehnes et *al.* 2014; Shariat et *al.* 2013a; Shariat et *al.* 2013b; Shariat et *al.* 2013c; Shariat & Dudley 2014). Chez cette espèce, le typage CRISPR permet :

- de comparer les souches au cours d'une épidémie : l'étude des deux loci CRISPR des *Salmonella* serovar Enteritidis ou Typhimurium isolées au cours de 10 épidémies a prouvé le lien épidémiologique entre les souches (Fabre et *al.* 2012).
- de typer et sous-typer les souches. La composition en spacers des deux systèmes CRISPR de *Salmonella* est fortement corrélée au sérotype et au MLST (Fabre et *al.* 2012; Liu et *al.* 2011a). D'autre part, le haut niveau de polymorphisme des loci CRISPR de souches appartenant à un même serovar permet le sous-typage des souches.

c) Etude des populations

L'étude des loci CRISPR peut également servir à l'analyse des populations présentes au sein d'une niche écologique. L'analyse de ces loci par des études de métagénomique peut donner des informations sur les interactions bactérie-bactériophage et sur leur dynamique *in vivo* (Andersson & Banfield 2008; Heidelberg et *al.* 2009; Held et *al.* 2010; Held & Whitaker 2009; Tyson & Banfield 2007; Hidalgo-Cantabrana et *al.* 2018). L'analyse des loci CRISPR

permet de reconstruire le génome viral et leur cible en fonction de l'homologie avec les spacers acquis (Andersson & Banfield 2008; Sun et *al.* 2016; Tyson & Banfield 2007; Hidalgo-Cantabrana et *al.* 2018). D'autre part, ces analyses permettent de différencier les lignées au sein d'une population et dans différents environnements. Les souches de *Leptospirillum* ont pu être séparées en deux sous-populations par rapport à leurs spacers ancestraux dans une population présente en biofilm (Andersson & Banfield 2008; Tyson & Banfield 2007). Ces analyses ont également mis en évidence une évolution du nombre de spacers au sein des loci CRISPR de streptocoques au cours du temps, *in vivo,* à partir de prélèvements salivaires (Pride et *al.* 2011).

Ces approches sont possibles du fait de la présence fréquente des systèmes CRISPR-Cas dans les populations bactériennes et du polymorphisme des spacers. Cependant, le typage via le locus CRISPR n'est pas possible chez toutes les bactéries. En effet, les *loci* CRISPR ne sont actuellement documentés que pour 47% des bactéries (Grissa et *al.* 2007; Makarova et *al.* 2011, Makarova et *al* 2015). D'autre part, certains systèmes évoluent trop vite en réponse à des EGMs et peuvent varier au cours d'une enquête épidémiologique. Des modifications internes de spacers (acquisitions ou délétions) ou SNP peuvent apparaitre fréquemment et gêner la comparaison entre les souches. D'autres systèmes CRISPR sont au contraire peu adaptatifs et montrent peu de variabilité entre les souches. Les systèmes CRISPR doivent avoir une capacité à acquérir suffisamment de spacers à l'extrémité *leader* pour offrir un polymorphisme suffisant permettant de séparer les isolats entre eux, alors que les spacers ancestraux doivent être conservés afin de mettre en évidence leur appartenance phylogénétique. Aujourd'hui, seulement quelques systèmes CRISPR ont été caractérisés par leur habilité à acquérir un nouveau spacer en réponse à un bactériophage ou un plasmide. Le polymorphisme des spacers (nombre et séquence) est directement corrélé à l'activité du *locus* CRISPR (Barrangou & Dudley 2016). Les bactéries pour lesquelles un typage CRISPR a été appliqué sont représentées dans le Tableau 4 (Barrangou & Dudley 2016).

Tableau 4. Génotypage bactérien basé sur le *locus* CRISPR. D'après Barrangou & Dudley, 2016.

Gram positif	Gram negatif	Autres
Bifidobacterium animalis		*Microcystis aeruginosa*
Clostridioides difficile	*Campylobacter jejuni*	*Mycobacterium tuberculosis*
Corynebacterium diphteriae	*Erwinia amylovora*	
Enterococcus faecalis	*Escherichia coli*	
Enterococcus faecium	*Legionella pneumophila*	
Lactobacillus buchneri	*Salmonella enterica*	
Lactobacillus casei	*Vibrio parahemolyticus*	
Cutibacterium acnes	*Xanthomonas oryzae*	
Streptococcus agalactiae	*Yesrsinia pestis*	
Streptococcus thermophilus		
Staphylococcus aureus		

B. CRISPR-Cas chez *S. agalactiae*

Le génome de *S. agalactiae* comporte deux systèmes CRISPR-Cas (Glaser et *al.* 2002; Tettelin et *al.* 2002, Tettelin et *al.* 2005; Richards et *al.* 2011; Lopez-Sanchez et *al.* 2012) :

- Le système CRISPR-Cas de type II-A, décrit initialement sous le nom de *locus* CRISPR1. Ce système est ubiquitaire, polymorphe et fonctionnel.

- Le système CRISPR-Cas de type I-C, décrit initialement sous le nom de *locus* CRISPR2. Ce système n'est présent que chez 20% des souches, il est moins polymorphe, évoquant une faible ou une absence d'activité.

Compte-tenu de ces observations, les travaux ultérieurs se sont orientés sur le *locus* CRISPR1.

1. Organisation du système CRISPR-Cas de type II-A

a) Le locus CRISPR1

Le locus CRISPR1 est constitué de repeats possédant une séquence très conservée de 36 pb, séparés par des spacers de séquence variable constituée généralement de 30 pb (Figure 26).

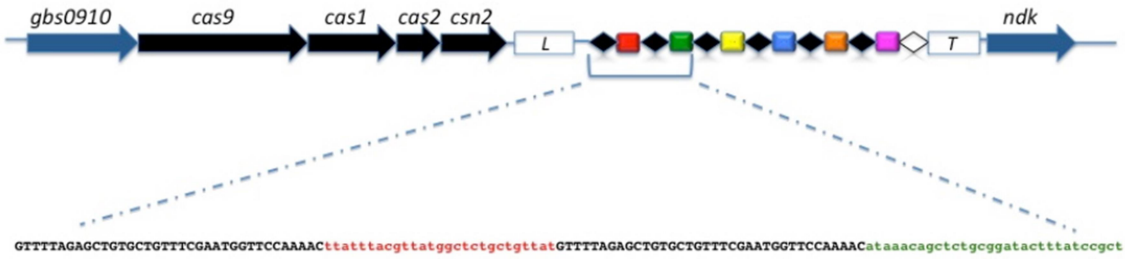


Figure 26. Modèle d'organisation du système CRISPR-Cas chez *S. agalactiae.*
Les séquences répétées (DR) (losanges noirs) sont entrecoupées par des séquences très variables ou « spacers » (rectangles colorés) (Lier et *al.* 2015).

Bien que les DR soient très conservés, le DRT montre une séquence dégénérée par rapport au DR, principalement en 3'. Parmi les 126 souches analysées par Lier et *al.*, quatre DRT différents ont été décrits (Lier et *al.* 2015). Les séquences des différents DR et leur distribution apparaissent dans le
Tableau 5.

Tableau 5. Liste et répartition des différents DR au sein du locus CRISPR1 de *S. agalactiae* (Lier et *al.*, 2015).

Type	Repeat sequence (5'- 3')	Frequency (%)	Clonal complex or ST (No. of strains)
Internal repeat			
Typical repeat	GTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAAC	98.5	CC-1 (26) CC-10 (25) CC-17 (30) CC-19 (14) CC-23 (26) ST-4 (1) ST-388 (2) ST-130 (1) ST-24 (1)
Repeat variants	GTTTTAGTGCTGTGCTGTTTCGAATGGTTCCAAAAC	0.1	CC-17 (2)*
	GTTTTAGAGCTGTGCTATTTCGAATGGTTCCAAAAC	0.1	CC-23 (1)*
	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC	1.2	CC-23 (4)*
	GTTTTAAAGCTGTGCTGTTTCGAATGGTTCCAAAAC	0.1	CC-19 (1)*
Terminal repeat			
Typical terminal repeat	GTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCA	52	CC-1 (26) CC-10 (25) CC-19 (14) ST-4 (1)
Terminal repeat variants	GTTTTAGAGCTGTGCGGTTATTATGCTAGGGCACCG	25	CC-17 (30) ST-130 (1)
	GTTTTAAAGCTGTGCTGTTATTATGCTAGGGCACCA	21.5	CC-23 (26) ST-24 (1)
	GTTTTAGAGCTGTGCTGTTATTATGCTAGGGCACCA	1.5	ST-388 (2)

Nucleotide polymorphisms as compared to the typical CRISPR repeat sequence are indicated in red.
**All CRISPR1 arrays with at least a repeat variant contain also the typical repeat.*

Le nombre de spacers est très variable selon les isolats. Il est en moyenne de 11,5 selon Lopez-Sanchez et *al* et de 13,6 selon Lier et *al* avec un minimum d'un seul spacer et un maximum de 33 spacers (Lier et *al.* 2015; Lopez-Sanchez et *al.* 2012). Les spacers présents dans le *locus* CRISPR1 ciblent généralement des EGMs et des phages tempérés appartenant au genre *Streptococcus.* La grande majorité des spacers pour lesquels une homologie a été retrouvée présentaient une homologie avec des séquences présentes dans les génomes de SGB, principalement des élements integratifs conjugatifs (ICE) et des prophages, plus rarement avec le core génome. La plupart des spacers qui ne présentaient pas d'homologie avec un génome de SGB présentent une homologie avec des EGM insérés dans des génomes de *Streptococcus* (Figure 27).

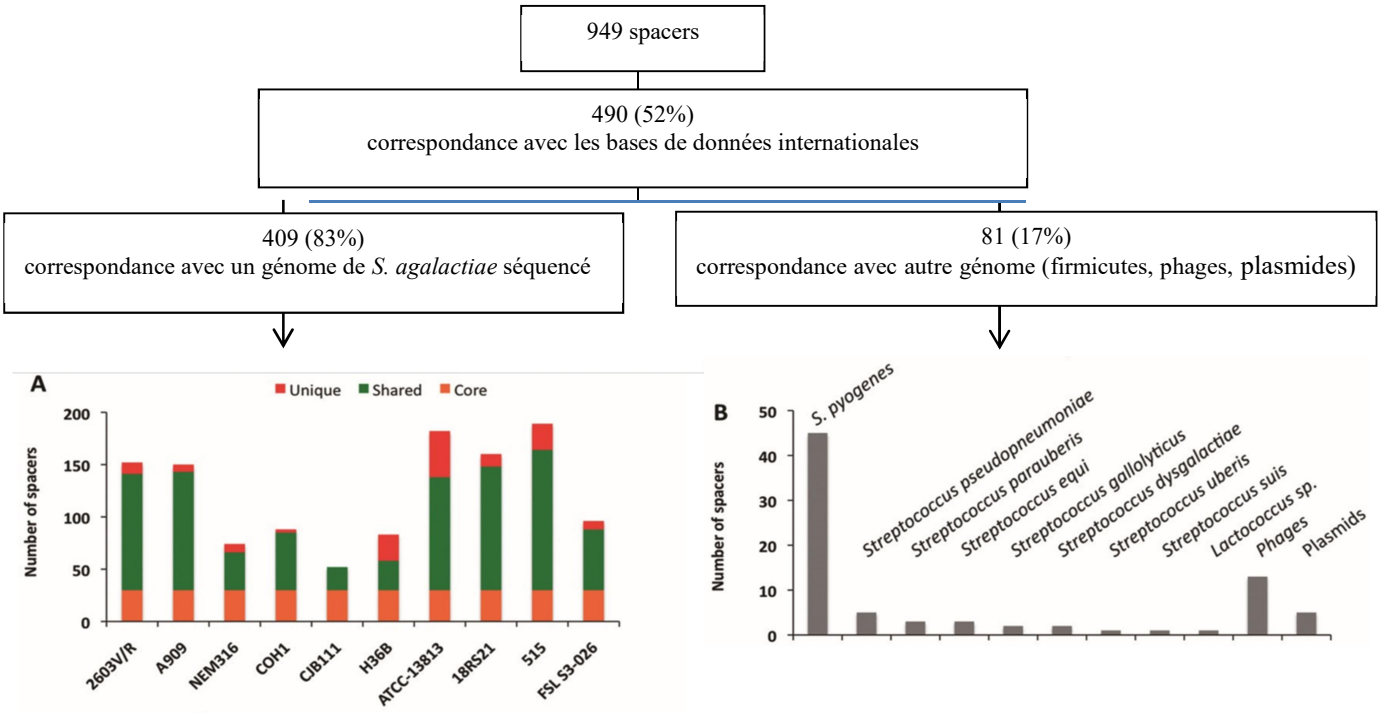


Figure 27. Correspondance des spacers avec les bases de données. D'après Lopez-Sanchez et *al.* 2012.
Pour 83% des spacers présentant une correspondance avec les bases de données, l'homologie était retrouvé avec des séquences présentes dans les génomes de SGB. Les spacers pour lesquels une correspondance était observée avec le core génome sont représentés en orange, une correspondance avec plusieurs génomes sont représentés en vert, une correspondance avec un seul génome sont représentés en rouge. Pour 17%, l'homologie était retrouvée avec d'autres génomes et concernait principalement des EGM insérés dans des génomes de Streptococcus.

b) Régions flanquantes et gènes *cas*

Le *leader*, très conservé, est une séquence riche en A/T (63%) comme décrit pour d'autres organismes, et est composé de 103 pb (Lopez-Sanchez et *al.* 2012; Lier et *al.* 2015). Des variations du *leader* ont cependant été décrites, principalement pour les souches appartenant au CC-10 dont la moitié (13/24) présentait une base manquante à l'extrémité adjacente au 1er DR (Lier et *al.* 2015). Le *trailer* est une séquence non codante, très conservée composée de 540 pb (Lier et *al.* 2015).

En amont de la séquence leader se trouve l'opéron *cas,* constitué des gènes *cas9, cas1, cas2* et *csn2,* typique des systèmes de type II-A (Makarova et *al.* 2015) (Figure 26). En amont de l'opéron *cas* et en aval du gène *gbs0910* se trouve le tracrRNA (Lopez-Sanchez et *al.* 2012)

2. Fonctionnalité du locus CRISPR1

Le caractère ubiquitaire au sein de l'espèce, le faible niveau de polymorphisme des repeats, l'absence de séquences d'insertion, la conservation des régions flanquantes, l'homogénéité de la taille des spacers et leur grande diversité sont autant d'arguments en faveur de l'activité du *locus* CRISPR1 (Horvath et *al.* 2008; Lopez-Sanchez et *al.* 2012; Lier et *al.* 2015). La fonctionnalité du système a été démontrée par Lopez-Sanchez et *al.*, qui ont mis en évidence une forte réduction du transfert par conjugaison du transposon *TnGBS2* en présence de spacers le ciblant (Lopez-Sanchez et *al.* 2012). En revanche, aucune acquisition de spacer en réponse à un EGM n'a été mise en évidence au sein du locus CRISPR1 chez *S. agalactiae.*

a) Séquence PAM

Le système CRISPR-Cas de type II-A de *S. agalactiae* est capable de distinguer le soi du non-soi. En effet, un plasmide contenant un protospacer dérivant d'un EGM (*TnGBS2, gbs1146*) conduit à une immunité, contrairement à un plasmide contenant un protospacer dérivant du « core genome » (*polA*) ; la transformation n'étant possible que pour le plasmide contenant la région génomique (*polA*) et le vecteur vide (Figure 28) (Lopez-Sanchez et *al.* 2012). Le protospacer *polA* est reconnu comme du soi, aussi bien dans sa localisation chromosomique qu'après introduction par transformation et n'est pas ciblé par la machinerie CRISPR-Cas.

L'étude du protospacer et des bases avoisinantes a permis de mettre en évidence la présence du dinucléotide GG dans 90% des séquences PAM des protospacers des EGM (position +2 et +3) et l'absence de ce dinucléotide dans les séquences PAM des protospacers du core genome. Ce dinucléotide est indispensable pour l'activité du système. En effet, la modification de ces bases permet de restaurer ou de supprimer l'activité immunitaire du système (Figure 28). Le dinucléotide GA permet également d'obtenir une efficacité.

	Recipient strain	
Plasmid	A909	NEM316
pTCVerm	5.1×10^3	3×10^3
pTCVerm polA [b]CA	3.2×10^3	3.4×10^3
pTCVerm polA GG	0	4.4×10^3
pTCVerm gbs1146 [b]GG	0	6.6×10^3
pTCVerm gbs1146 AC	8.3×10^3	7.1×10^3
pTCVerm gbs1146 CA	3×10^3	1.8×10^3
pTCVerm gbs1146 AG	7.5×10^3	6.6×10^3
pTCVerm gbs1146 GA	0	8.5×10^3
pTCVerm gbs1146 GT	1.9×10^3	6.5×10^3

a. Number of transformants per microgram of DNA; the average value of three transformations is given.
b. Corresponds to the chromosomal sequence.

Figure 28. Efficacité de transformation des plasmides en fonction de la nature des protospacers et les PAM.
Efficacité de transformation des plasmides portant les protospacers polA *(core génome) et gbs1146 (EGM) avec modification du PAM* (Lopez-Sanchez et *al.* 2012)*. La présence du protospacer GG dérivant d'un EGM empêche la transformation, contrairement à la présence de CA dérivant du core genome. D'après Lopez-Sanchez et al. 2012. Le dinucléotide GG est suffisant pour la discrimination du soi et du non-soi durant la phase d'interférence et peut également être remplacé par le dinucléotide GA.*

3. Diversité du locus CRISPR1

L'analyse comparative des séquences des spacers montre une très forte diversité due à l'acquisition, la délétion et la duplication de nouveaux spacers (Lopez-Sanchez et *al.* 2012; Lier et *al.* 2015). Cette variabilité n'est pas homogène et un gradient de conservation est observé au sein du *locus* CRISPR1. Les spacers à proximité du *trailer* sont très stables et correspondent aux spacers ancestraux. Les spacers situés à proximité du *leader* sont généralement uniques et peu représentés, ce qui est cohérent avec une acquisition polarisée et une forte instabilité des spacers nouvellement acquis (Lopez-Sanchez et *al.* 2012; Lier et *al.* 2015).

Certaines souches conservent les spacers récents mais présentent des différences parmi les spacers plus anciens. Ces modifications peuvent concerner une seule unité spacer-repeat ou de larges portions et correspondent à des délétions des additions ou des duplications de spacer. Le spacer ancestral en contact avec le DRT n'est jamais concerné. *In vitro*, ces modifications sont observées en l'absence de pression de sélection. Compte-tenu de la diversité des réarrangements et la faible taille des séquences répétées, elles sont probablement dues à des recombinaisons homologues et/ou à des glissements de la polymérase pendant la réplication (Lopez-Sanchez et *al.* 2012).

Les premières études portant sur les systèmes CRISPR-Cas de *S. agalactiae* ont montré un lien entre la composition du *locus* CRISPR1, notamment les séquences des spacers ancestraux, et la structure de la population définie par MLST (Lopez-Sanchez et *al.* 2012). Outre les spacers ancestraux, une congruence a également été observée entre le DRT et le typage MLST (Lier et *al.* 2015). Les loci CRISPR1 des souches appartenant aux CC-17, CC-23 et ST-388 possèdent un DRT caractéristique et ceux des souches appartenant aux CC-1, CC-19 et CC-10 présentent un DRT commun (Lier et *al.* 2015). Les spacers ancestraux, ainsi que le DRT, apparaissent très conservés au sein d'un complexe clonal et pourraient prédire l'appartenance des souches à un complexe clonal donné (Figure 29). Les souches appartenant aux CC-17, CC-23 et CC-10 peuvent ainsi être séparées sur la base de leurs spacers ancestraux et de leur DRT, contrairement aux souches appartenant aux CC-1 et CC-19 qui présentent le même DRT et les même spacers ancestraux (Lier et *al.* 2015; Lopez-Sanchez et *al.* 2012).

A

Clonal complex	Sequence type		CRISPR1 trailer end spacers
CC17		Cluster 1	102 49 53 54 55 56
		Cluster 2	49 50 51 52 53 54 55 56
		Cluster 3	293 53 54 55 56
CC1/19	ST1-196		45 46 8
	ST19-328-182		89 46 8
	ST28		373 8
CC6/7	ST6		35 36 37
	ST7-10-12		180 36
CC23	ST23	Cluster 1	192 30 31
		Cluster 2	249 30 31
		Cluster 3	684 31
CC26			19 20
	ST61		560 147
			174 147
CC67	ST67-63		7 147
CC103	ST103		211 212
CC340	ST340		1 32
CC260			98

B

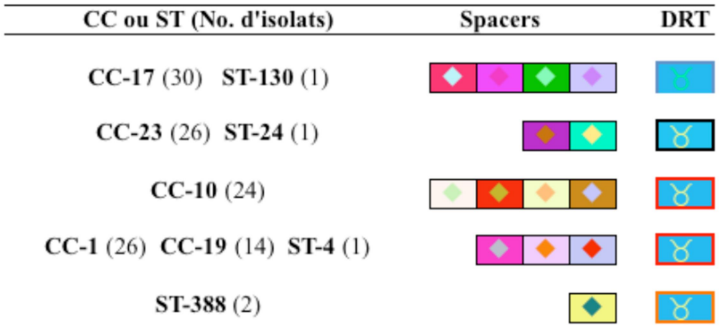


Figure 29. Relation entre le locus CRISPR et le MLST chez *S. agalactiae*.
Les spacers ancestraux sont spécifiques des différents complexes clonaux ou de complexes clonaux proches. ***A.*** *Selon Lopez-Sanchez et al.* 2012. ***B.*** *Selon Lier et al. 2015.*

Contrairement aux spacers ancestraux, les spacers situés à l'extrémité *leader* sont très variables et permettent de séparer les souches appartenant à un même ST. Ainsi, le locus CRISPR1 semble présenter un intérêt pour l'étude de la population de *S. agalactiae.* Le typage par le locus CRISPR1 apparait plus discriminant que le MLST, technique actuelle de référence (Lier et *al.* 2015).

Bien que le polymorphisme des *loci* CRISPR1 semble être un atout pour le typage des souches, cette diversité semble varier en fonction de l'appartenance phylogénétique des souches. Les souches de CC-17 et, dans une moindre mesure, de CC-23 montrent une plus faible variabilité des loci CRISPR1. Les loci CRISPR1 des souches de CC-17 sont moins polymorphes avec un moins grand nombre de spacers uniques, d'arrangements de spacers uniques et de nombre de spacer par locus (Figure 30) (Lier et *al.* 2015).

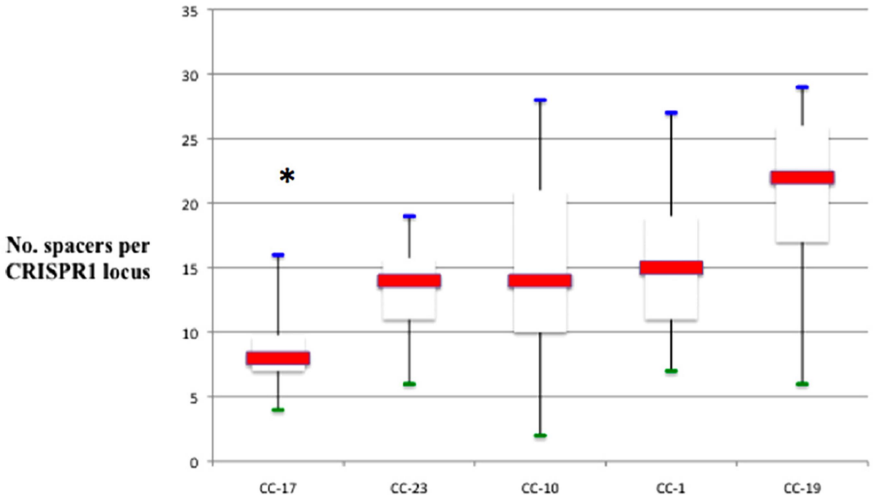


Figure 30. Nombre de spacers par locus CRISPR1 en fonction de l'origine phylogénétique des souches chez *S. agalactiae* (Lier et *al.* 2015).

Ces deux travaux portant sur un total de 474 souches (351 souches (Lopez-Sanchez et *al.* 2012) et 123 souches (Lier et *al.* 2015)), suggèrent l'intérêt majeur du *locus* CRISPR1 pour le typage des souches de *S. agalactiae*. Le typage CRISPR montre une étroite corrélation avec le MLST, une meilleure discrimination et apparaît très attractif pour les études phylogénétiques.

Objectifs du travail

Streptococcus agalactiae est une bactérie au pouvoir pathogène complexe : initialement décrite comme agent de mammite des bovidés, commensale des tractus génito-urinaires et gastro-intestinaux de l'adulte sain, pathogène majeur en gynécologie-obstétrique, et pathogène émergent chez les personnes âgées. L'impact récent de cette bactérie en santé humaine (années 1960) en fait un pathogène émergent qui a su évoluer dans sa capacité à infecter l'homme et a su s'adapter à différents environnements.

A ce jour, la phylogénie de l'espèce n'est encore que partiellement comprise. Toutes les souches ne possèdent pas la même capacité invasive et certains sérotypes ou génotypes ont pu être identifiés comme plus souvent associés à certaines pathologies, en particulier dans les infections néonatales (ST-17). Un changement du spectre de cette bactérie semble avoir lieu depuis quelques années avec l'augmentation des cas d'infections invasives chez l'adulte et l'augmentation des cas d'infection néonatales tardives. La caractérisation des souches de *S. agalactiae* reste donc primordiale aujourd'hui. Différentes approches de typage de *S. agalactiae* ont été développées, comme évoqué précédemment, avec pour but de surveiller les infections et de comprendre la dynamique des populations bactériennes. Néanmoins, ces techniques sont lourdes et peu exportables entre laboratoires (PFGE, MLEE…), longue et couteuse (MLST) ou peu discriminante (sérotypage). Le séquençage de génome total apparaît comme le *gold standard* de demain mais cette technique, bien que de plus en abordable, reste encore coûteuse et nécessite une expertise, notamment bio-informatique, et un équipement approprié. Durant cette phase de transition, une technique hautement discriminante et de haute résolution semble nécessaire. Celle-ci doit utiliser un marqueur épidémiologique suffisamment discriminant pour permettre la différenciation des isolats au sein de l'espèce, sans l'être trop pour ne pas compromettre le lien épidémiologique. D'autre part, la technique utilisée doit être exportable entre différents laboratoires afin d'identifier et de comparer avec le plus de précision possible les souches isolées à l'échelle locale ou internationale.

L'organisation temporelle et le polymorphisme des spacers des loci CRISPR en font des marqueurs phylogénétiques séduisants pour le génotypage et l'analyse phylogénétique (Shariat & Dudley 2014). Cependant, le typage via le locus CRISPR n'est pas possible chez toutes les bactéries. En effet, ces loci ne sont pas présents chez toutes bactéries et un équilibre doit être trouvé quant à leur vitesse d'évolution (Grissa et *al.* 2007; Makarova et *al.* 2011, Makarova et *al.* 2015). Ils doivent évoluer suffisamment rapidement pour offrir un polymorphisme permettant de séparer les isolats entre eux, mais pas trop rapidement pour conserver des marqueurs communs entre les isolats. Ainsi, l'utilisation d'un locus CRISPR

comme marqueur épidémiologique impose l'évaluation au préalable de sa prévalence au sein de l'espèce concernée, de son polymorphisme et de sa vitesse d'évolution. Concernant *S. agalactiae*, les études précédentes ont permis de mettre en évidence i) l'ubiquité du locus CRISPR1 ; ii) sa grande variabilité ; iii) un lien entre sa composition et l'origine phylogénétique des souches ; trois caractéristiques en faveur d'un intérêt du locus CRISPR1 comme marqueur épidémiologique chez *S. agalactiae.*

L'objectif principal de ce travail était d'évaluer l'intérêt du système CRISPR-Cas de type II-A et plus particulièrement du locus CRISPR1, comme marqueur épidémiologique chez *S. agalactiae*. Pour cela, notre travail a consisté à :

i) évaluer l'activité du système et estimer sa vitesse d'évolution *in vivo.*

ii) évaluer le polymorphisme du locus CRISPR1 chez un nombre relativement important de souches de différentes origines anatomiques et géographiques isolées chez l'homme (souches de portage et invasives, isolées chez le nouveau-né ou l'adulte dans différents centres hospitaliers).

iii) confirmer le lien entre composition du locus CRISPR1 et origine phylogénétique des souches.

iv) comparer le typage CRISPR1 aux méthodes de références actuelles pour le typage de SGB (MLST et sérotypage).

v) étudier des différences du système (composition et/ou fonctionnalité) en fonction de l'appartenance phylogénétique des souches.

vi) développer des approches de typage basées sur le locus CRISPR1 : amplification et séquençage, amplification et restriction enzymatique du locus CRISPR1 (CRISPR-RFLP).

En utilisant cet outil, l'objectif secondaire était d'obtenir un aperçu des souches présentes en portage vaginal dans différents contextes :

i) au sein de différents microbiotes vaginaux (patientes différentes) : étude de la diversité des clones de SGB chez différentes patientes.

ii) au sein d'un même microbiote vaginal (même patiente à un temps donné) : étude d'une population de SGB ; étude du portage à un temps donné et de la cohabitation des différentes souches de SGB, notamment de différentes lignées clonales.

iii) au sein d'un microbiote vaginal « en évolution » (même patiente au cours du temps) : étude de l'évolution des différentes lignées clonales et des loci CRISPR1 pour une souche donnée au cours du temps. Cette analyse offrant un aperçu des différents contacts bactéries – EGMs (notamment bacteriophages) *in vivo* et ainsi de la coévolution bactéries-bactériophages *in vivo* dans leur environnement naturel.

- La première partie de ce travail a consisté à évaluer l'évolution du locus CRISPR1 au cours du temps, dans son environnement naturel, afin :
  - de confirmer l'activité du système CRISPR-Cas *in vivo.*
  - d'estimer sa vitesse d'évolution *in vivo.*
  - de mettre en évidence les éventuels contacts entre les souches de *S. agalactiae* et les EGMs dans leur environnement naturel et ainsi d'investiguer les interactions bactéries - EGMs, notamment bactériophages.

Il s'agissait également d'évaluer l'analyse du locus CRISPR1 comme outil épidémiologique pour le typage des souches de *S. agalactiae* à différents temps de portage sur des prélèvements itératifs espacés dans le temps. Dans cette première partie, les loci CRISPR1 de souches isolées à différent temps de portage chez une même femme sur une cohorte de 100 patientes ont été séquencés. L'analyse de la composition en spacer des loci CRISPR1, de ces variations et de ces acquisitions, notamment polarisées, nous permettaient de répondre à ces objectifs.

- La deuxième partie de notre travail a consisté à évaluer l'analyse du locus CRISPR1 comme outil de caractérisation des souches de *S. agalactiae* présentes au sein d'un même prélèvement vaginal. L'étude de la diversité clonale au sein d'un microbiote vaginal donné était réalisée par une première approche de « screening » portant sur le locus CRISPR1. L'approche utilisée était une technique de restriction enzymatique du locus CRISPR (CRISPR-RFLP). Les loci CRISPR1 présentant des profils de restrictions différents étaient séquencés. Cette étude nous permettait d'avoir un aperçu de la diversité des souches de *S. agalactiae* au sein d'une population de portage dans un même contexte clinique (femme enceinte entre 34 et 38 semaines d'aménorrhées).

- La troisième partie de ce travail avait pour objectif de confirmer l'intérêt du locus CRISPR1 pour le typage des souches de *S. agalactiae* en le comparant aux autres méthodes : le typage capsulaire et le MLST. Des souches de différentes origines étaient caractérisées par ces trois techniques (séquençage du locus CRISPR1, typage capsulaire et MLST). Le but de cette étude était d'évaluer la capacité de l'analyse du locus CRISPR1 à i) séparer des isolats au sein de l'espèce et évaluer son caractère discriminant par rapport aux autres méthodes de typage ; ii) mettre en évidence des liens phylogénétiques entre les différents isolats. Dans cet objectif, les *loci* CRISPR1 étaient séquencées puis regroupées en clusters en fonction de leur composition. Deux sélections de marqueurs (DRT et spacers) étaient ensuite effectuées afin de mettre en évidence des marqueurs spécifiques de lignées. Le pouvoir discriminant de ces deux sélections était ensuite évalué et comparé à celui des autres méthodes de typage.

Travail expérimental

I. Suivi du portage vaginal de *S. agalactiae* et évolution *in vivo* du locus CRISPR1

Publication n°1 : Clémence Beauruelle, Adeline Pastuszka, Philippe Horvath, Franck Perrotin, Laurent Mereghetti, Philippe Lanotte. (2017) **CRISPR: a useful genetic feature to follow vaginal carriage of Group B Streptococcus.** *Frontiers in Microbiology* (IF 4,3) doi: 10.3389/fmicb.2017.01981.

Cette étude a été réalisée avec l'aide de Philippe Horvath, Christophe Fremaux et Sabine Van Dillen de DuPont Nutrition and Health de Dangé-Saint-Romain.

Ce travail a consisté à évaluer l'évolution du locus CRISPR1 au cours du temps *in vivo* et d'étudier cette méthode génotypique pour le suivi des souches de portage génital de *S. agalactiae.* Cette étude était basée en particulier sur l'analyse de la composition en *spacers* du locus CRISPR1. Les différentes modifications de composition en *spacers* ont été analysées et permettaient de répondre aux différents objectifs :

i) confirmer l'activité du système chez *S. agalactiae.*
ii) estimer sa vitesse d'évolution.
iii) évaluer l'environnement phagique de SGB *in vivo* en fonction de la nature des *spacers* acquis et obtenir un aperçu de la cohabitation de ces deux entités.
iv) étudier des différences de composition et de fonctionnalité du système en fonction de l'appartenance phylogénétique des souches.
v) évaluer l'intérêt du locus CRISPR1 comme outil de typage pour le suivi des souches de portage.

Les souches sélectionnées correspondaient à des souches de portage génital isolées à différents temps de portage (minimum de 3 mois, maximum de 6,9 ans entre les prélèvements) chez 100 femmes, soit un total de 205 souches (Annexe A-1). Les loci CRISPR1 de chacune de ces souches étaient analysés et comparés entre eux, particulièrement

la composition en spacers des loci CRISPR1 des souches isolées d'une même patiente. Pour chaque spacer acquis de manière polarisée, une homologie était recherchée entre la séquence de ce spacer et les bases de données internationales.
Parallèlement au séquençage des loci CRISPR1 et à l'étude de leur évolution, chacune des souches était caractérisée par deux techniques de typage discriminantes : le MLST et le MLVA.

Cette étude nous a permis :

i) ***d'améliorer notre compréhension de l'évolution* in vivo *du système CRISPR-Cas de* S. agalactiae*, et plus généralement de l'évolution d'un système CRISPR-Cas* in vivo.**
ii) ***de confirmer l'activité du système et d'évaluer sa vitesse d'évolution.***
iii) ***d'obtenir un aperçu de la dynamique de l'interaction bactérie – EGM* in vivo.**

En effet, nous avons pu mettre en évidence :

- une évolution *in vivo* des loci CRISPR1, notamment des acquisitions polarisées de spacers en réponse à des EGMs (plasmides ou bacteriophages), permettant de conclure que le système était actif *in vivo* et que les souches de *S. agalactiae* étaient en contact avec des EGMs dans leur environnement naturel et «s'immunisaient» contre ceux-ci. Cette étude est, à notre connaissance, la première à s'intéresser au suivi et à l'évolution d'un locus CRISPR *in vivo* chez l'homme.
- des modifications des loci CRISPR1 au cours du temps pour 18% des souches (moyenne de 2,3 ans entre les prélèvements), dont 4% correspondaient à des acquisitions polarisées (moyenne de 3,1 ans entre les prélèvements), soit une vitesse d'évolution du locus CRISPR1 relativement faible.
- un lien entre l'origine phylogénétique des souches et leur capacité à acquérir de nouveaux spacers, suggérant un rapport entre leur passé génétique et leur capacité d'adaptation.

iv) ***de confirmer l'intérêt du locus CRISPR1 comme outil de typage pour le suivi des souches de portage.***

En effet,

- Une forte congruence entre les différentes techniques de typage était observée. Une concordance entre le typage CRISPR et les autres techniques de typages (MLST et MLVA) était observée dans 94% des cas.

- Cette méthode nous a permis de décrire le portage de SGB au cours du temps. Ce portage était relativement homogène pour une femme donnée (même CRISPR type pour 83% des femmes, même clone pour 78% des femmes toutes techniques de typage confondues). En revanche, une grande variabilité des souches était retrouvée parmi les 100 femmes de l'étude (CRISPR type « personnel » dans 96% des cas), suggérant un fort pouvoir discriminant de cette technique de typage.

v) ***de mettre en évidence des différences dans la capacité d'évolution du locus CRISPR1 entre les lignées phylogénétiques.***

En effet,

- Les souches présentant des acquisitions polarisées de spacers appartenaient aux souches de ST-1, ST-19 et ST-6.
- Aucune modification du locus CRISPR1 n'était observée pour les souches de CC-17 qui, de plus, présentent une plus faible variabilité de spacers, suggérant la possibilité d'une plus faible activité du système CRISPR-Cas chez les souches de CC-17.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 October 2017
doi: 10.3389/fmicb.2017.01981

CRISPR: A Useful Genetic Feature to Follow Vaginal Carriage of Group B *Streptococcus*

Clémence Beauruelle[1,2,3], Adeline Pastuszka[1,2,3], Philippe Horvath[4], Franck Perrotin[5,6], Laurent Mereghetti[1,2,3] and Philippe Lanotte[1,2,3]*

[1] Université François Rabelais de Tours, UMR1282 Infectiologie et Santé Publique, Tours, France, [2] INRA, UMR1282 Infectiologie et Santé Publique, Nouzilly, France, [3] Service de Bactériologie-Virologie, Centre Hospitalier Universitaire de Tours, Tours, France, [4] DuPont Nutrition and Health, Dangé-Saint-Romain, France, [5] Inserm U930, Université François Rabelais de Tours, Tours, France, [6] Département d'Obstétrique de Gynécologie et de Médecine Fœtale, Centre Hospitalier Universitaire de Tours, Tours, France

Clustered regularly interspaced short palindromic repeats (CRISPR) and Cas (CRISPR-associated proteins) play a critical role in adaptive immunity against mobile genetic elements, especially phages, through their ability to acquire novel spacer sequences. Polarized spacer acquisition results in spacer polymorphism and temporal organization of CRISPR loci, making them attractive epidemiological markers. Group B *Streptococcus* (GBS), a genital commensal for 10 to 30% of healthy women and a major neonatal pathogen, possesses a ubiquitous and functional CRISPR1 locus. Our aim was to assess the CRISPR1 locus as an epidemiological marker to follow vaginal carriage of GBS in women. This study also allowed us to observe the evolution of the CRISPR1 locus in response to probable phage infection occurring *in vivo*. We followed carriage of GBS among 100 women over an 11-year period, with a median duration of approximately 2 years. The CRISPR1 locus was highly conserved over time. The isolates that show the same CRISPR1 genotype were collected from 83% of women. There was an agreement between CRISPR genotyping and other typing methods [MLVA (multilocus variable number of tandem repeat Analysis) and MLST (multilocus sequence typing)] for 94% of the cases. The CRISPR1 locus of the isolates from 18 women showed modifications, four of which acquired polarized spacer, highlighting the *in vivo* functionality of the system. The novel spacer of one isolate had sequence similarity with phage, suggesting that phage infection occurred during carriage. These findings improve our understanding of CRISPR-Cas evolution in GBS and provide a glimpse of host-phage dynamics *in vivo*.

Keywords: CRISPR, *Streptococcus agalactiae*, carriage, genotyping, diversity

OPEN ACCESS

Edited by:
Fumito Maruyama, Kyoto University, Japan

Reviewed by:
Ryohei Nomoto, Kobe Institute of Health, Japan
Masatoshi Okura, National Institute of Animal Health (NARO), Japan

***Correspondence:**
Philippe Lanotte
philippe.lanotte@univ-tours.fr

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *28 June 2017*
Accepted: *25 September 2017*
Published: *11 October 2017*

Citation:
Beauruelle C, Pastuszka A, Horvath P, Perrotin F, Mereghetti L and Lanotte P (2017) CRISPR: A Useful Genetic Feature to Follow Vaginal Carriage of Group B Streptococcus. Front. Microbiol. 8:1981. doi: 10.3389/fmicb.2017.01981

INTRODUCTION

Streptococcus agalactiae, or Group B *Streptococcus* (GBS), is a major pathogen in humans and is the leading cause of neonatal infections in industrialized countries (Stoll et al., 2011; Edmond et al., 2012). Early-onset disease (EOD) (days 0–6) is the result of vertical transmission from a colonized mother during, or just before, delivery (Colbourn and Gilbert, 2007), whereas the pathogenesis of

late-onset disease (LOD) (days 7–89) is less well understood (Berardi et al., 2013). GBS belongs to the commensal microbiota that colonize the gastrointestinal and genital tracts of 10–30% of healthy humans and can cause serious infections in neonates (Regan et al., 1991; Yancey et al., 1994). It is now universal practice to screen pregnant women for vaginal and/or rectal colonization with GBS to prevent early-onset infection (ANAES, 2001; Verani et al., 2010; Di Renzo et al., 2015). The aim of this screening strategy is to limit bacterial transmission and prevent EOD by the administration of intrapartum antibiotic prophylaxis. The GBS colonization status can change over the course of pregnancy. Thus, colonization status late in the third trimester has been used as a proxy for intrapartum colonization (at 35–37 weeks of gestation) (Anthony et al., 1978; Lewin and Amstey, 1981; Dillon et al., 1982; Hansen et al., 2004; Verani et al., 2010). In France, GBS screening is performed by swabbing the lower vagina followed by cultivation. This screening strategy must be repeated for each pregnancy, with the exception of a few special cases (women with GBS isolated from the urine during the current pregnancy or who had a previous infant with invasive GBS disease) (ANAES, 2001; Verani et al., 2010; Di Renzo et al., 2015). Moreover, screening for pathogenic bacteria, including GBS, must be performed in cases of imminent delivery risk (suspicion of chorioamnionitis, prolonged membrane rupture, or preterm labor) (ANAES, 2001).

Several methods have been developed to improve the diagnostic and prognostic classification of GBS isolates, such as serotyping, based on capsule polysaccharides, or, more recently, multilocus sequence typing (MLST), based on the sequencing of seven genes, which is expensive and time consuming (Jones et al., 2003; Lindahl et al., 2005; Slotved et al., 2007). These methods have highlighted the involvement of serotype III and sequence-type ST17 in causing more invasive neonate disease (Manning et al., 2009). Other molecular typing methods have been recently developed, such as multilocus variable number of tandem repeat (VNTR) analysis (MLVA), a molecular typing method based on VNTR variability (Haguenoer et al., 2011). This typing method targets six dispersed chromosomal loci. For GBS, it was shown to be more discriminant than MLST, especially for ST17 isolates (Haguenoer et al., 2011). CRISPR (clustered regularly interspaced short palindromic repeats) loci have also been described in GBS and used to genotypically characterize GBS isolates (Lopez-Sanchez et al., 2012; Lier et al., 2015).

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins form the CRISPR-Cas system, which protects bacteria against bacteriophages (phages), and more generally, against exogenic mobile genetic elements (MGEs) (Barrangou et al., 2007). This system provides adaptive immunity through specific immunization based on nucleic acid sequences. CRISPR arrays are composed of short (25–40 bp) direct repeats (DR) interspaced by non-repetitive similar-sized sequences called spacers. DRs are highly conserved within a given CRISPR array and are often partially palindromic, having the potential to form hairpin structures (Kunin et al., 2007). Most CRISPR arrays are flanked on one side by an AT-rich sequence, called the leader, and on the other side by a trailer-end sequence, which is located downstream of the terminal direct repeat (TDR), a DR which is often degenerate and/or truncated. The CRISPR-Cas system acts in three stages (Barrangou and Marraffini, 2014). In the adaptation stage, a new spacer, usually derived from a foreign nucleic acid, is incorporated at the leader end of the CRISPR locus, concomitantly to the duplication of the leader-end DR. The expression stage consists on CRISPR RNA (crRNA) biogenesis, starting with transcription of the CRISPR array from a promoter located within the leader, followed by cleavage within the DRs of this long pre-crRNA molecule into short, mature crRNAs. Finally, during the interference stage, proteins target complementary foreign nucleic acids for cleavage within the sequence corresponding to the spacers. Thus, the CRISPR-Cas system provides sequence-specific immunity, whereby specificity is dictated by the sequence of the spacers that have been integrated within the CRISPR array. CRISPR loci thus show a dynamic, rapidly evolving, and polymorphic composition due to the ability of the system to capture novel spacers derived mostly from MGEs, reflecting past encounters with MGEs present in the environment. Spacer acquisition is polarized, as novel spacers are inserted at the leader end of the CRISPR locus. Consequently, sequences present at the trailer end can be considered to be ancestral, providing evidence of ancient MGE invasion events (Barrangou et al., 2007). Overall, spacer polymorphism and polarized acquisition of novel spacers result in the temporal organization of CRISPR loci, making them attractive epidemiological markers for genotyping and phylogenetic analyses (Shariat and Dudley, 2014).

Two CRISPR-Cas systems have been characterized in a collection of GBS strains (Lopez-Sanchez et al., 2012). A type II-A system, associated with the CRISPR1 locus, is ubiquitous and functional, whereas a type I-C system, associated with the CRISPR2 locus, is rare and most often incomplete, suggesting little or no activity (Lopez-Sanchez et al., 2012; Makarova et al., 2015). The CRISPR1 locus contains highly conserved 36-bp DRs separated by spacers of approximately 30 bp. Similarities between CRISPR1 spacer sequences and MGEs have been previously reported. Comparative sequence analysis of CRISPR1 loci across numerous strains revealed extensive diversity, due to the acquisition of new spacers, spacer duplications, and spacer deletions, illustrating the dynamics of the system (Lopez-Sanchez et al., 2012; Lier et al., 2015).

Previous studies reported the use of CRISPR1 spacer content and TDRs for the genotyping of GBS strains, showing a strong correlation between MLST and CRISPR-based genotyping (Lopez-Sanchez et al., 2012; Lier et al., 2015). Here, we used this genotyping method to follow vaginal carriage of GBS from 100 women with a minimum 3-month time interval between samples, and show that CRISPR1 polymorphism-based genotyping represents a rapid and discriminating tool for epidemiological studies. Furthermore, as spacer acquisition events occurred *in vivo* in multiple and independent cases during carriage, notably in response to a probable phage infection, our results strongly suggest that the CRISPR1-Cas system is functionally active for adaptation.

MATERIALS AND METHODS

Bacterial Isolates

Streptococcus agalactiae isolates were collected between January 2004 and December 2014 in the Department of Obstetrics, Gynecology and Fetal Medicine of the CHRU of Tours, France, where approximately 4,000 deliveries are occurring each year. Isolates originated from gynecological (vaginal and endocervix) swabs, and for some patients, the gastric fluid of the neonate was kept as well. Isolates from gastric fluid aspirate samples originated from amniotic fluid and vaginal secretions ingested by the newborn during labor (Lanotte and Seme, 2012). The protocol was reviewed and approved by the Ethics Committee in Human Research of the University Hospital Center of Tours (approval number 2017-060) (Tours, France). Samples were considered in this study only if GBS isolates were obtained from the same women with a time interval of at least 3 months between two samplings. Among these women, 100 were randomly selected using the RAND function in Excel (Microsoft Office version 1997–2003). Isolates were stored at −80°C and grown on blood agar plates (TSH, BioMerieux®).

DNA Extraction

Genomic DNA was extracted following enzymatic lysis with mutanolysin (Sigma). A bacterial suspension of 1.5 McFarland was prepared in 500 μL water containing 50 U of mutanolysin. The suspension was incubated for 1 h at 56°C, followed by 10 min at 100°C, leading to cell lysis. Lysates were centrifuged for 3 min at 1500 × *g* and the supernatants containing DNA collected.

CRISPR1 Locus Sequencing

Clustered regularly interspaced short palindromic repeats1 locus amplification was performed in a T3000 thermocycler (Biometra) using CRISPR1 PCR-F and CRISPR1 PCR-R primers which target the CRISPR1-flanking regions, as previously described (Lopez-Sanchez et al., 2012). PCR amplification was performed in a total volume of 25 μL, containing 0.5 μM of each primer, 0.2 μM deoxynucleoside triphosphate (dNTP), 2 mM $MgCl_2$, 0.02 U/μL Q5 high-fidelity DNA polymerase (New England Biolabs), 1× PCR buffer, and 5 μL extracted DNA. The PCR mixtures were heated at 98°C for 5 min, followed by 40 cycles of a denaturation step at 98°C for 30 s, an annealing step at 56°C for 30 s, an elongation step at 72°C for 120 s, and ending with a final extension step at 72°C for 10 min. PCR amplification was verified by electrophoretic migration in 1% agarose gel. PCR products were purified using centrifugal filter units (Millipore) as recommended by the manufacturer. The purified products were sequenced with BigDye Terminator Mix v3.1 on a Hitachi 3130xl Genetic Analyzer (Applied Biosystems), using the internal sequencing primers CRISPR1SEQ-F and CRISPR1SEQ-R, as previously described (Lopez-Sanchez et al., 2012). For CRISPR1 regions exceeding 1.3 kb, the use of primers targeting internal spacers was necessary to complete the sequencing of PCR products (Supplementary Table S1).

CRISPR1 Sequence Analysis

DNA sequences were analyzed and assembled using the BioEdit sequence alignment editor version 7.2.5[1]. Spacers, repeats, and flanking regions for each sequence were identified using a macro-enabled Excel tool (P. Horvath, DuPont), which allows the identification and extraction of CRISPR features from nucleotide sequences, and the graphic representation of spacers as colored cells in Excel spreadsheets. Spacer sequences were compared to the dictionary of spacers established earlier (Lopez-Sanchez et al., 2012). New spacers identified in this study expanded the dictionary and were numbered incrementally. The spacer dictionary is available on the website[2]. Loci were labeled "identical" if all spacers were identical (same spacers in the same order) between two CRISPR1 loci. Two CRISPR1 loci were considered "close" if at least 70% of spacers were identical (same spacers in the same order), while they were considered "distinct" if at least 50% of spacers were different.

We performed a similarity search for each novel CRISPR1 spacer found in our isolates using BLAST[3] with the Nucleotide collection (nr/nt) database and the default parameters of a prior analysis. All matches with a bit score above 40.0 and a query cover above 80% (corresponding to 100% identity) were retained. We performed a second search for spacer sequences without significant similarity using the whole genome shotgun contig (wgs) database limited to "*Streptococcaceae*" or "Firmicutes."

Multilocus Sequence Typing (MLST)

Multilocus sequence typing was carried out as previously described (Jones et al., 2003). Allelic profiles and sequence type (ST) were assigned using the international MLST database[4]. Clonal complexes (CC) were defined using the stringent group definition (6/7 shared alleles) and eBurst analysis[5].

Multilocus VNTR (Variable Number of Tandem Repeats) Analysis (MLVA)

In parallel, MLVA typing, targeting six loci, was carried out as previously described (Haguenoer et al., 2011). The number of repeats for each VNTR was deduced from amplicon size by electrophoretic migration in an agarose gel.

RESULTS

Samples and Isolates Studied

Among the 8,190 isolates of *S. agalactiae* collected between January 2004 and December 2014 at the CHRU of Tours, France, 205 isolates from 100 women were selected for the study (**Figure 1**). The average age of the cohort was 29.8 years at the time of sampling (median: 30 years, minimum: 19 years, maximum: 44 years). We analyzed two isolates for 95 women, and three for the remaining five. The isolates originated from

[1]http://www.mbio.ncsu.edu/bioedit/bioedit.html

[2]http://crispr.i2bc.paris-saclay.fr/CRISPRcompar/Dict/Dict.php

[3]https://blast.ncbi.nlm.nih.gov/

[4]http://pubmlst.org/sagalactiae/

[5]http://eburst.mlst.net/v3/mlst_datasets/

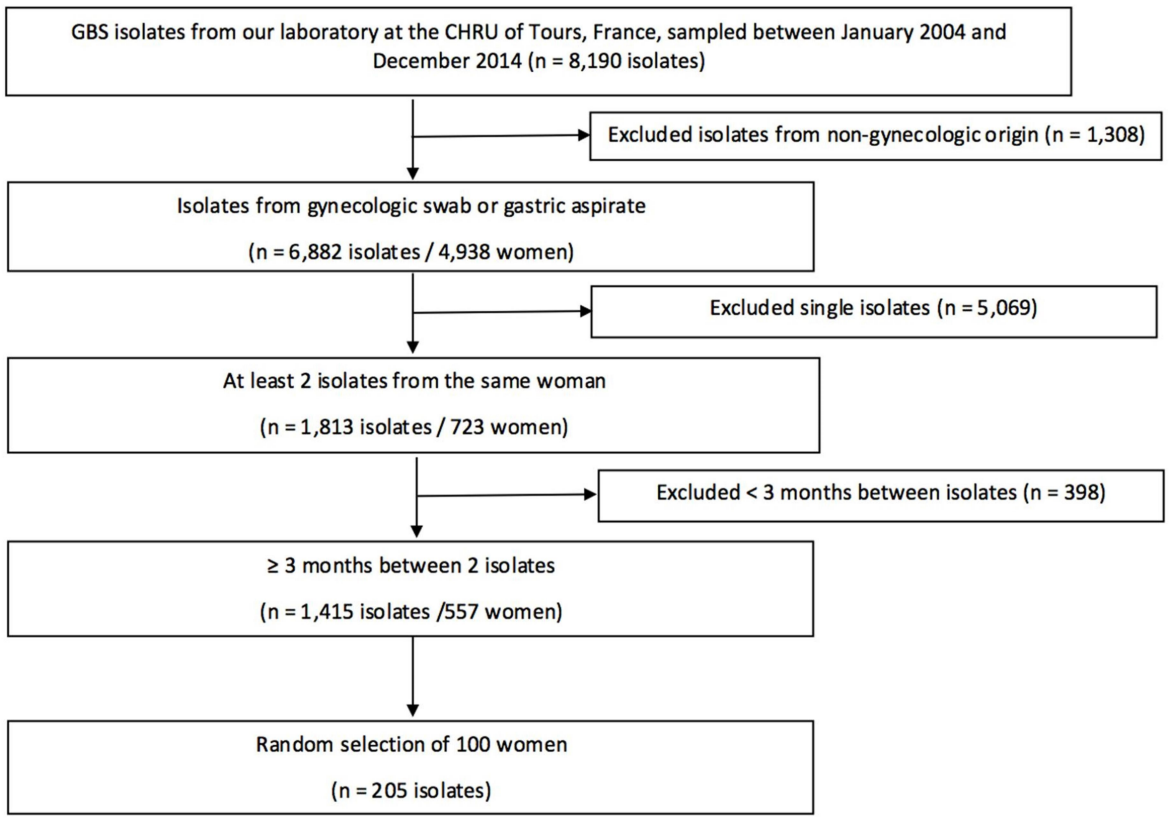


FIGURE 1 | Study flow chart. Among the 8,190 isolates of our laboratory, multiple isolates originating from the same women were selected, and those from 100 women were further selected randomly to follow carriage evolution.

vaginal swabs in 84% of cases (173/205), endocervical swabs in 11% (23/205), and gastric fluid samples from the neonate in 5% (9/205). The average time period between two samples was 2.4 years (median: 2.1 years, minimum: 3 months, maximum: 6.9 years). MLST analysis indicated that 51 isolates belong to CC1 (including 33 ST1), 46 to CC19 (including 24 ST19), 44 to CC23 (including 32 ST23), 26 to CC17 (including 24 ST17), and 21 to CC10 (including 5 ST10), whereas the remaining 17 isolates were ungrouped.

CRISPR1 Locus Analysis

We obtained a complete CRISPR1 sequence for all isolates. The external primers (CRISPR1SEQ-F and CRISPR1SEQ-R) were sufficient for sequencing the entire locus for 123 isolates. Primers targeting internal spacer sequences were designed for the remaining 82. A total of 2,727 spacers were sequenced, including 240 not previously described in *S. agalactiae*. The average number of spacers per CRISPR1 locus was 13.8, with a minimal of three (isolates 43 and 44), and a maximal of 29 (isolates 24, 25, and 39). Detailed results obtained for each isolate are presented in Supplementary Figure S1.

We compared CRISPR1 loci of isolates originating from the same women and observed "distinct" CRISPR1 loci between isolates for 18% (18/100) (**Figure 2** and Supplementary Figure S1A). GBS were isolated at three different dates for one of these women (number 98), and the last carried a different CRISPR1 locus than the first two. This woman was thus classified into both groups, "identical" and "distinct" CRISPR1 (**Figure 2**).

The isolates contained a "close" or "identical" CRISPR1 locus for 83% of the women (83/100) (Supplementary Figure S1B); the isolates of 65 women presented an identical CRISPR1 locus, whereas those for the other 18 contained "close" loci. We observed only slight variations in spacer composition, including additional spacers at the leader end and deletion or acquisition of internal repeat-spacer unit(s) (**Figure 3**). Spacer sequences are provided in Supplementary Figure S2.

The final isolate from four of these 18 women had one (women 11, 51, and 63) or three (woman 75) additional repeat-spacer units at the leader end of CRISPR1 compared to the older isolate (**Figure 3A**). There is no sequence similarity between these six novel spacers and those already described in GBS. In contrast, they share similarities with plasmid or phage sequences. The novel spacer of strain 30 (woman 11) is similar to an *Enterococcus faecalis* pLG2 plasmid sequence (identities: 30 nucleotides /30), whereas that of strain 286 (woman 51) is similar to a *S. pseudopneumoniae* pDRPIS7493 plasmid sequence (identities: 30 nucleotides /30). The spacer acquired by strain 357 (woman 63) is similar to a *S. agalactiae* phage tail gene (identities: 30 nucleotides /30), whereas the three novel spacers of strain 428 (woman 75) do not show any similarity with known sequences.

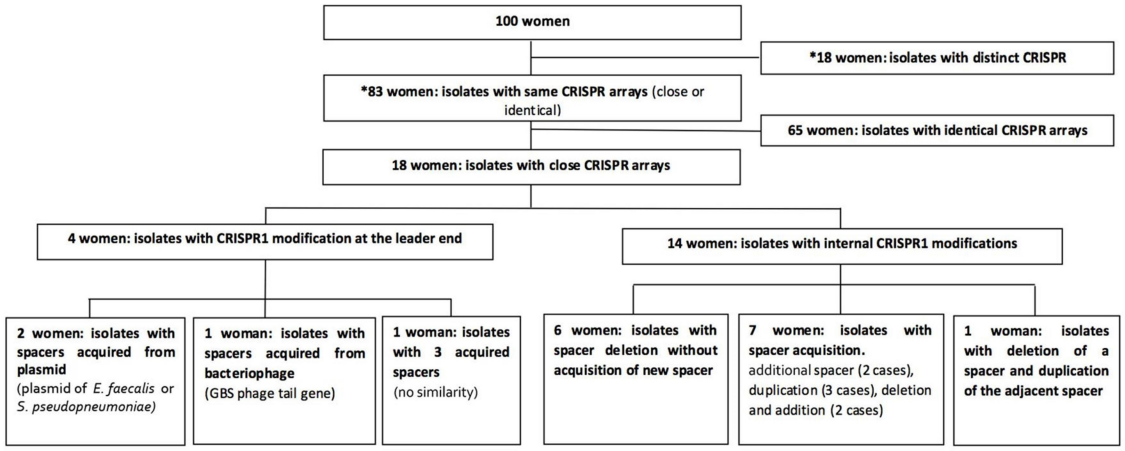


FIGURE 2 | Diagram of case flow up to clustered regularly interspaced short palindromic repeats1 (CRISPR1) analysis. We compared CRISPR1 arrays of GBS isolated from the same women, and examined other regions of the genome by MLST and MLVA analysis to confirm the genetic relationships between isolates. *Three isolates from one women were collected, two with the same CRISPR1 locus, and one with a different one. The corresponding woman was thus classified into two groups (same CRISPR, and different CRISPR).

The average time for the acquisition of these novel spacers was 3.1 years (median: 2.5 years, minimum: 1.9 years, maximum: 5.5 years) (**Table 1**).

We observed internal CRISPR1 locus modifications for isolates originating from 14 women. The final isolate for six of these women has deletions of repeat-spacer units: one-unit deletion in four isolates (women 21, 46, 88, and 97), and two-unit deletion in the two others (women 7 and 30) (**Figure 3B**). The final isolates originating from the seven other women present additional repeat-spacer units in internal regions of the CRISPR1 array: one unit in four cases (women 2, 76, 77, and 79), and two, four, or five units in the other cases (women 42, 14, and 20, respectively) (**Figure 3C**). In three cases, the additional spacers correspond to a single spacer duplication (women 2, 76, and 77). The duplicated spacer is the adjacent one in two cases (women 2 and 76), but a distant spacer in another case (woman 77). In two other cases, the additional spacers were not already present in the locus (women 20 and 79). In two cases (woman 14 and 42), internal spacer acquisition was seemingly combined with the deletion of an adjacent spacer. Spacer acquisition corresponded to the duplication of two spacers (woman 42) or the acquisition of four new spacers (woman 14). Interestingly, a spacer switch occurred in one case (woman 67) (**Figure 3D**). The intermediary (out of three) isolate originating from this woman shows a spacer replacement likely resulting from a duplication of the downstream spacer. As the initial and final isolates contain an identical CRISPR1 locus, it is likely that this strain was present, but not isolated, during the second sampling. The average time for internal modification of the CRISPR1 locus was 2.3 years (median: 1.9 years, minimum: 6 months, maximum: 5.5 years) (**Table 1**). Among the five isolates with spacer duplications (4, 232, 432, 434, and 381), four belong to CC23. Overall, we observed spacer duplication in the CRISPR1 locus for 73 isolates (35.6%). In 24 isolates, the duplication involved spacer 247 in our dictionary, which is CC23-specific and shares sequence similarity with *Streptococcus* phage phi3396. This spacer is present in 82% of CC23 isolates [36/44, including two isolates (293 and 295) where it was truncated], and duplicated in 54% of them (24/44) (Supplementary Figure S1).

We compared modifications of CRISPR1 loci in the isolates with their phylogenetic affiliations as established by MLST (**Table 2**). The polarized acquisition of novel spacer(s) at the leader end involved one isolate belonging to CC1 (ST1), one belonging to CC19 (ST19), and two belonging to CC10 (ST6). We observed no CRISPR1 locus modification for isolates belonging to CC17.

MLST and MLVA Analysis

To confirm the genetic relatedness between isolates, phylogenetic markers dispersed throughout the GBS chromosome were analyzed using MLST and MLVA (**Table 3**). There was concordance between MLST analysis and CRISPR1 typing for 99% of cases (isolates from 99/100 women), whereas 94% of cases showed concordance between CRISPR1 typing and MLVA. We observed a differing CRISPR genotype for one woman (woman 70, isolates 396 and 397), but these isolates actually belong to the same ST (ST19) and have the same MLVA profile (Supplementary Figure S1A). The CRISPR1 locus from these isolates shows similarities at the level of ancestral spacers (trailer end) and TDR, but high variability across leader-adjacent spacers. With the exception of this case, we systematically observed different MLVA profiles (94.5%; 17/18 women) among women for whom isolates possess different CRISPR1 arrays. We observed the same MLVA profile for 94% of women (78/83) among those from whom isolates possess identical or "close" CRISPR1 arrays. In isolates of 6% of women

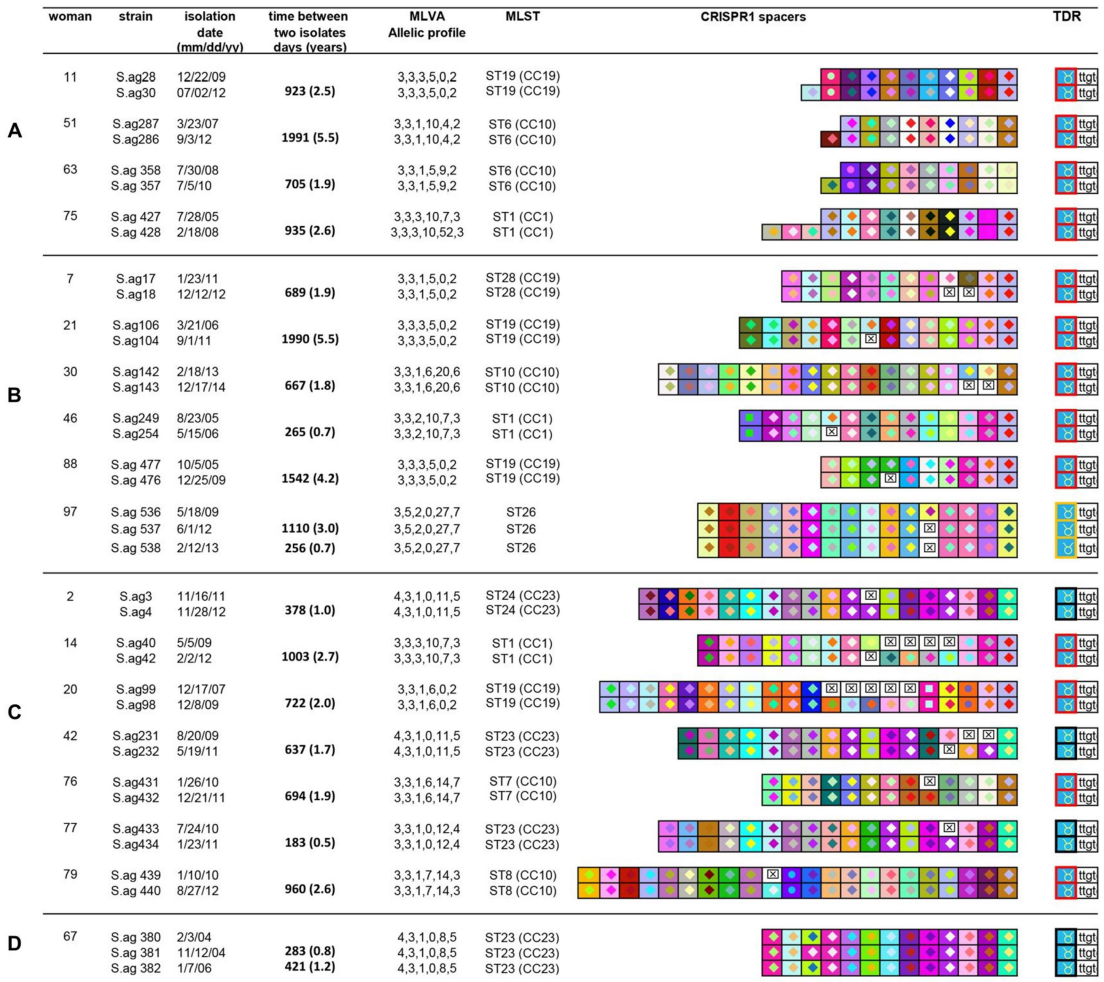


FIGURE 3 | Comparison of the CRISPR1 arrays of isolates from 18 women where CRISPR1 modifications were observed. The CRISPR1 arrays are represented using a macro-enabled Excel tool, whereby spacers are converted into two-color symbols based on spacer sequence. Gaps (=missing spacers) are shown with a boxed cross symbol (⊠) after alignment of identical spacers between strains of the same group. The MLVA allelic profile corresponds to the number of repeats of each VNTR (SAG2, SAG3, SAG4, SAG7, SAG21, and SAG22) (Haguenoer et al., 2011). Terminal Direct Repeats (TDRs) are represented by different colored borders according to their sequence. **(A)** CRISPR1 arrays with leader-end acquisitions; **(B)** CRISPR1 arrays with internal deletions; **(C)** CRISPR1 arrays with internal additions; **(D)** CRISPR1 array with a spacer switch (women 67). The second of three isolates shows a spacer replacement, likely resulting from a duplication of the downstream spacer.

TABLE 1 | Period between sampling of isolates deriving from the same ancestor, among the 83 women with close or identical clustered regularly interspaced short palindromic repeats1 (CRISPR1) arrays.

	No CRISPR1 locus modification (65 women, 133 isolates)	Leader-end modification (4 women, 8 isolates)	Internal locus modification (14 women, 29 isolates)
Average (years)	2.0	3.1	2.3
Median (years)	1.8	2.5	1.9
Minimum (years)	0.3	1.9	0.5
Maximum (years)	5.6	5.5	5.5

(5/83), the MLVA analysis showed a different profile, due exclusively to a modification in the SAG21 VNTR (women 33, 34, 35, and 98 with isolates having identical CRISPR1 arrays, and woman 75 with isolates having "close" CRISPR1 arrays).

Analysis of CRISPR1-Flanking Regions and Internal DRs

Finally, we analyzed the flanking regions of CRISPR1 (leader and trailer ends), as well as internal DRs, as these sequences sometimes show variations between groups of strains. The

TABLE 2 | Phylogenetic affiliation and CRISPR1 array modifications among isolates from 83 women with close or identical CRISPR1 arrays.

	CC1 ($n = 19$)	CC19 ($n = 19$)	CC23 ($n = 17$)	CC17 ($n = 12$)	CC10 ($n = 9$)	Others* ($n = 7$)	Total ($n = 83$)
No modification	16	14	13	12	4	6	65
Middle locus modification	2	4	4	–	3	1 (ST26)	14
Leader-end spacer acquisition	1 (ST1)	1 (ST19)	–	–	2 (ST6)	–	4

**ST22 (isolates from two women), ST130 (isolates from two women), ST26 (isolates from two women), ST243 (isolates from one woman).*

TABLE 3 | Agreement between CRISPR genotyping and other typing methods (MLST and MLVA).

		CRISPR array	
		Different	**Close or identical**
MLVA	Different	17 (94.5%)	5 (6%)
	Identical	1 (5.5%)	78 (94%)
MLST	Different	17 (94.5%)	–
	Identical	1 (5.5%)	83 (100%)
Total		18	83

There was an agreement between the three typing methods for 94% of women. For the others women, one presented isolates with a different CRISPR genotype but belonging to the same ST and have a same MLVA profile (woman 70). The other five women (33, 34, 35, 75, and 98) presented a same CRISPR genotype, a same ST but a different MLVA profile always due to SAG21 VNTR.

CRISPR1 trailer sequence was conserved for all isolates except for ST22 isolates which shared five additional nucleotide following DRT (Supplementary Figure S1B). The CRISPR1 leader sequence was identical for 180 isolates (87.8%) over the 130 bp located immediately upstream of the first DR. Among the 27 isolates showing differences, the last nucleotide was deleted in 15 instances, or duplicated in one case (Supplementary Table S3). This last nucleotide modification concerned CC10 isolates in 87.5% of cases (14/16). For one isolate, 26 additional nucleotides were found at the 3′ end of the leader, corresponding to a partial DR sequence, likely originating from an incorrect tandem duplication of the first repeat. A single nucleotide difference was present in the leader sequence of the 10 remaining isolates. In contrast, DR sequence analysis across all CRISPR1 loci showed a high conservation of the internal DRs, with only six isolates showing SNPs among the 2,727 DRs (isolates 17, 43, 44, 231, 420, and 421). The TDR was represented by five different sequences, including four described previously (Supplementary Table S2) (Lier et al., 2015). TDR sequences were specific for each CC, except for CC1 and CC19 that share the same TDR sequence (Supplementary Figure S1).

DISCUSSION

In this work, we explored the *in vivo* dynamics of human colonization by GBS over several years by comparing the CRISPR1 loci of 205 isolates collected from 100 women over an 11-year period. The aims of this study were to evaluate CRISPR1 locus analysis as an epidemiological marker, to describe the evolution of CRISPR1 arrays in GBS over time during human carriage, and to assess the likelihood of phage infection *in vivo,* highlighted by alterations of the spacer content, notably additions at the leader end of CRISPR1.

Eighteen women (18/100) provided GBS isolates with a "close" CRISPR1 locus over the sampling time. MLST typing confirmed the genetic relationship between these isolates, as well as MLVA analysis, except for one case which differed only at the SAG21 VNTR locus (woman 75), as discussed below. These isolate pairs arguably share common ancestry, and the observed CRISPR locus modifications are the result of *in vivo* evolution within patients. Modifications at the leader end (four women) correspond to spacer acquisition which likely occurred in response to MGE invasion events. Two isolates show similarities between novel spacers and plasmids from related species (*E. faecalis* and *S. pseudopneumoniae*), consistent with previous studies showing similarity between most GBS spacers and streptococcal genome sequences (Lopez-Sanchez et al., 2012). One isolate acquired a novel spacer showing strong similarity with a gene encoding a phage tail protein, probably originating from a Caudovirales/*Siphoviridae* phage, as are most *Streptococcaceae* phages (Domelier et al., 2009). For one woman, the two isolates differed from each other by the polar acquisition of three spacers that show no similarity to any known sequence. These spacers could have been acquired successively, or all at the same time, during the 2.5 years separating the two samplings. Overall, polarized spacer acquisition at the leader end was relatively rare in our population. Indeed, this event occurred in only 4% of the colonized women of the study. The mean time between sampling of GBS isolates showing polarized acquisition was 3.1 years (compared to 2.3 years for all isolates). Metagenomic analyses of the population dynamics of MGE has shown a rapid ability of CRISPR-Cas systems to acquire spacers in both environmental and human niches (Fineran and Charpentier, 2012). In the *S. thermophilus* type II-A CRISPR system, just one *in vitro* challenge with a phage was sufficient to result in the acquisition of new spacers (Barrangou et al., 2013). Similarly, *in vitro* biofilm analysis showed a high rate of evolution for *Leptospirillum* type II CRISPR arrays, whereby the number of spacers increased from 37 to 197 in 5 months (Tyson and Banfield, 2007; Andersson and Banfield, 2008). Pride et al. (2011) analyzed *Streptococcus* spp. (*S. thermophilus, S. mutans, S. pyogenes*, and *S. agalactiae*) CRISPR loci directly from the saliva of four patients over 18 months by evaluating spacer number by PCR amplification with primers directed at common DRs.

These authors showed for each patient a high rate of alteration of spacer number at one CRISPR locus, over time, demonstrating the dynamics of the CRISPR1 system in human. In our study, the rare cases of spacer acquisition might be explained by a low occurrence of MGEs in this environment, or by a low activity of the CRISPR1-Cas system in GBS. Furthermore, as no lytic phage has been isolated yet from GBS, the acquisition of new spacers showing similarity to phage sequences may correspond to adaptation from excised prophages.

In our study, we observed two types of modifications of the internal CRISPR1 array: spacer acquisitions, and deletions. Deletions always occurred in an internal region of the locus or close to the last spacer, as previously described (Deveau et al., 2008; Horvath et al., 2008; Lopez-Sanchez et al., 2012). These deletions never concerned spacers in contact with the leader or TDR. Deletions and spacer duplications may result from slipped misalignment during DNA replication (Lopez-Sanchez et al., 2012). For two isolates (women 20 and 79), the acquired spacers were not already present in the locus, and acquisition could be due to homologous recombination, which has been described in GBS (Bisharat et al., 2004; Brochet et al., 2008; Lopez-Sanchez et al., 2012). In two cases (women 14 and 42), internal spacer acquisition -including a spacer duplication case- was seemingly combined with the deletion of an adjacent spacer. A similar phenomenon probably also occurred for the intermediary isolate of woman 67, from whom three isolates were studied; the first and third isolates were identical, while the intermediate isolate differed by a duplication and deletion of one spacer.

We observed a correlation between GBS phylogenetic lineages and the ability to integrate new spacers in CRISPR1, suggesting an impact of the genetic background on adaptation efficiency. Isolates with polarized acquisition belong only to ST1, ST19, or ST6, whereas internal locus modifications were observed for isolates belonging to CC1, CC19, CC23, CC10, and ST26 (**Table 2**). We observed no CRISPR1 modification for CC17 isolates, neither in at the leader end nor in the middle of the locus, despite their relatively large representation (26 isolates from 18 women). Moreover, CC17 isolates showed less spacer variation at CRISPR1 locus, suggesting lower activity of CC17 CRISPR system. This result is consistent with previous observations which also highlight a significantly lower number of spacers in CC17 isolates in comparison with isolates belonging to other lineages (Lier et al., 2015). Nevertheless, Domelier et al. (2009) demonstrated the unique character of CC17 phages, in addition to a considerable prophage diversity in ST17 isolates. This is consistent with previous studies which compared inter-species CRISPR loci, reporting a strong inverse correlation between the number of spacers and prophages within a given genome (Nozawa et al., 2011). Similarly, CC23 isolates displayed less spacer variation, as previously described, and shared a relatively high number of common spacers (Lier et al., 2015). We found that internal spacer duplication was more frequent among CC23 isolates. Notably, spacer 247 was duplicated in 54% of CC23 isolates and was highly represented (82% of CC23 isolates). This spacer shares similarity with *S. dysgalactiae* phage *phi* 3396, which likely originated from a *S. pyogenes* phage (Davies et al., 2007).

Furthermore, we noticed an association between phylogenetic affiliation and leader sequence. The leader sequence is highly conserved in GBS, as well as in other bacteria (Yosef et al., 2012; Lier et al., 2015; Wei et al., 2015). Notably, the sequence at the leader-repeat junction is critical for adaptation by CRISPR-Cas systems, as shown in *Escherichia coli* type I-E (Yosef et al., 2012) and *S. thermophilus* type II-A (Wei et al., 2015). Among our isolates, the leader-repeat junction was modified in 8% of them (17/205), and isolates with leader-repeat junction modifications mostly belonged to CC10 (82%, 14/17). This modification was frequent and occurred in 67% of the CC10 isolates (14/21), in agreement with prior observations (Lier et al., 2015).

Our study showed a strong congruence between the various typing methods used, namely CRISPR typing, MLST, and MLVA. We explored other regions of the GBS genome by MLVA and MLST (exploring six and seven loci distributed throughout the chromosome, respectively) to confirm the relationship between the isolates. For one woman (woman 70), the two isolates carried different CRISPR1 loci but displayed the same MLST and MLVA profiles, suggesting a genetic link. Their CRISPR1 arrays contained five identical ancestral spacers and the same TDR, confirming common ancestry. However, spacers located near the leader showed high variability, likely due to an independent recent evolution. For these isolates, whole genomic analysis could be applied to evaluate their relationship.

For five patients, isolate pairs having identical or close CRISPR1 loci had the same ST but different MLVA profiles. Among them, the second isolate of one woman (woman 75) acquired three spacers at the leader end of CRISPR1. The only difference between MLVA profiles were found within the SAG21 VNTR and may be due to repeat sequence acquisition or deletion over time. Of note, SAG21 is the most variable VNTR among those used for MLVA (Haguenoer et al., 2011) and is perhaps too variable for MLVA typing for middle- to long-term carriage of GBS, and typing over a long period. The rate of evolution of this locus has not been explored yet.

These results need to be interpreted with caution. Although spacer content of the CRISPR1 locus is generally specific for isolates for a given woman, some isolates from different women (isolates from women 5 and 6, and those from women 68 and 18) and belonging to CC23 had identical CRISPR1 loci, as observed previously (Lier et al., 2015). Similarly, some isolates belonging to ST17 had close CRISPR1 loci, with a change in just the most recent spacer (isolates from women 99, 26, and 87). Identical CRISPR1 loci among strains of different origin and belonging to ST17 were observed previously by Lopez-Sanchez et al. (2012). This may be due to the circulation of highly prevalent clones of CC23 or CC17, even if there is no evidence of an epidemiological link between the different isolates, except pairwise.

Another aim of this study was to evaluate the evolution of GBS during carriage using these three typing methods. Among the 100 women studied, 78% presented carriage of GBS strains with the same profile, irrespective of the molecular method used (CRISPR1 typing, MLVA, or MLST), and 83% presented carriage of the same GBS strain with the same or close CRISPR1 array. Similar GBS carriage has been previously described (Hansen et al., 2004; Manning et al., 2008). Hansen et al. (2004) followed

carriage for 21 months (during pregnancy and up to 1 year after delivery) by pulsed-field gel electrophoresis (PFGE). In their study of 35 carriers, 86% of women were colonized by just one clone, and the others by two clones simultaneously, and no clonal shift was observed (Hansen et al., 2004). Manning et al. (2008) explored perinatal GBS colonization (during pregnancy, and 6 weeks after delivery) by serotyping, GBS capsular gene cluster genotyping, and MLST. Comparison of the three typing methods revealed differences, with clonal replacement in 18% of women, and a change of ST for 10% (Manning et al., 2008). However, these longitudinal studies were performed during a short period (maximum 21 months between two samples). Our study, based on a longer period, extends previous findings on the predominance of one clone during carriage, using highly discriminant typing methods. One limitation of our study is that the stored isolates may represent the predominant strain in cases where two or more coexisted in the sample. This approach does not provide a complete overview of bacterial diversity over time and we cannot exclude the possibility of multiple populations, particularly if a strain was predominant. Indeed, carriage of a double population has been shown previously (Hansen et al., 2004; Perez-Ruiz et al., 2004; Brzychczy-Włoch et al., 2014). This double population may explain our observations for woman 67, where the second isolate carried a different CRISPR1 locus than that of the first and third isolates. The discriminatory power provided by CRISPR1 locus analysis could make this a simple and relatively rapid typing method to explore GBS vaginal carriage.

Clustered regularly interspaced short palindromic repeats-based typing provides a glimpse of GBS vaginal carriage at the molecular level. This method has the advantage of being linked to the genetic lineage of the isolate through analysis of the TDR and ancestral spacers. Furthermore, it allows exploration of the recent evolution of the isolate, especially encounters with MGEs, through analysis of the leader end. The CRISPR-based typing method appears to be a useful tool to compare GBS isolates, thanks to its high discriminatory power and ease of use. However, further studies are required before CRISPR can be used more generally as a typing method for GBS. In addition, it is now evident that the GBS CRISPR1 locus evolves *in vivo*. We observed polarized acquisitions in response to MGEs (plasmids or phages), highlighting the dynamics of the system. The ability of the CRISPR1 array to evolve may be linked to phylogenetic affiliation, exemplified by the apparently lower ability of the hypervirulent strain ST17 to acquire novel spacers, compared to the higher ability of ST6 isolates. This potential link is yet to be explored.

AUTHOR CONTRIBUTIONS

Conceived and designed the work: CB and PL. Performed manipulations: CB and AP. Drafted the paper: CB, PH, and PL. Critically revised the manuscript: PH, FP, and LM. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We are most grateful to Philippe Glaser from providing the initial spacer dictionary. We thank the Institut de Biologie Intégrative de la Cellule (I2BC) for maintaining the CRISPRs Web server (http://crispr.i2bc.paris-saclay.fr). We also thank Coralie Lemaire for excellent technical assistance in MLST analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01981/full#supplementary-material

REFERENCES

ANAES (2001). Antenatal prevention of early neonatal bacterial infection (September 2001). *J. Gynécol. Obstét. Biol. Reprod.* 32, 68–74.

Andersson, A. F., and Banfield, J. F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320, 1047–1050. doi: 10.1126/science.1157358

Anthony, B. F., Okada, D. M., and Hobel, C. J. (1978). Epidemiology of group B *Streptococcus*: longitudinal observations during pregnancy. *J. Infect. Dis.* 137, 524–530. doi: 10.1093/infdis/137.5.524

Barrangou, R., Coûté-Monvoisin, A.-C., Stahl, B., Chavichvily, I., Damange, F., Romero, D. A., et al. (2013). Genomic impact of CRISPR immunization against bacteriophages. *Biochem. Soc. Trans.* 41, 1383–1391. doi: 10.1042/BST20130160

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140

Barrangou, R., and Marraffini, L. A. (2014). CRISPR-cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* 54, 234–244. doi: 10.1016/j.molcel.2014.03.011

Berardi, A., Rossi, C., Lugli, L., Creti, R., Bacchi Reggiani, M. L., Lanari, M., et al. (2013). Group B *Streptococcus* late-onset disease: 2003-2010. *Pediatrics* 131, e361–e368. doi: 10.1542/peds.2012-1231

Bisharat, N., Crook, D. W., Leigh, J., Harding, R. M., Ward, P. N., Coffey, T. J., et al. (2004). Hyperinvasive neonatal Group B *Streptococcus* has arisen from a bovine ancestor. *J. Clin. Microbiol.* 42, 2161–2167. doi: 10.1128/JCM.42.5.2161-2167.2004

Brochet, M., Rusniok, C., Couvé, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., et al. (2008). Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15961–15966. doi: 10.1073/pnas.0803654105

Brzychczy-Włoch, M., Pabian, W., Majewska, E., Zuk, M. G., Kielbik, J., Gosiewski, T., et al. (2014). Dynamics of colonization with group B *Streptococci* in relation to normal flora in women during subsequent trimesters of pregnancy. *New Microbiol.* 37, 307–319.

Colbourn, T., and Gilbert, R. (2007). An overview of the natural history of early onset group B streptococcal disease in the UK. *Early Hum. Dev.* 83, 149–156. doi: 10.1016/j.earlhumdev.2007.01.004

Davies, M. R., McMillan, D. J., Van Domselaar, G. H., Jones, M. K., and Sriprakash, K. S. (2007). Phage 3396 from a *Streptococcus dysgalactiae* subsp. *equisimilis* pathovar may have its origins in *Streptococcus pyogenes*. *J. Bacteriol.* 189, 2646–2652. doi: 10.1128/JB.01590-06

Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., et al. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390–1400. doi: 10.1128/JB.01412-07

Di Renzo, G. C., Melin, P., Berardi, A., Blennow, M., Carbonell-Estrany, X., Donzelli, G. P., et al. (2015). Intrapartum GBS screening and antibiotic prophylaxis: a European consensus conference. *J. Matern. Fetal Neonatal Med.* 28, 766–782. doi: 10.3109/14767058.2014.934804

Dillon, H. C., Gray, E., Pass, M. A., and Gray, B. M. (1982). Anorectal and vaginal carriage of group B *Streptococci* during pregnancy. *J. Infect. Dis.* 145, 794–799. doi: 10.1093/infdis/145.6.794

Domelier, A.-S., van der Mee-Marquet, N., Sizaret, P.-Y., Hery-Arnaud, G., Lartigue, M.-F., Mereghetti, L., et al. (2009). Molecular characterization and lytic activities of *Streptococcus agalactiae* bacteriophages and determination of lysogenic-strain features. *J. Bacteriol.* 191, 4776–4785. doi: 10.1128/JB.00426-09

Edmond, K. M., Kortsalioudaki, C., Scott, S., Schrag, S. J., Zaidi, A. K., Cousens, S., et al. (2012). Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. *Lancet* 379, 547–556. doi: 10.1016/S0140-6736(11)61651-6

Fineran, P. C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 434, 202–209. doi: 10.1016/j.virol.2012.10.003

Haguenoer, E., Baty, G., Pourcel, C., Lartigue, M.-F., Domelier, A.-S., Rosenau, A., et al. (2011). A multi locus variable number of tandem repeat analysis (MLVA) scheme for *Streptococcus agalactiae* genotyping. *BMC Microbiol.* 11:171. doi: 10.1186/1471-2180-11-171

Hansen, S. M., Uldbjerg, N., Kilian, M., and Sorensen, U. B. S. (2004). Dynamics of *Streptococcus agalactiae* colonization in women during and after pregnancy and in their Infants. *J. Clin. Microbiol.* 42, 83–89. doi: 10.1128/JCM.42.1.83-89.2004

Horvath, P., Romero, D. A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., et al. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1401–1412. doi: 10.1128/JB.01415-07

Jones, N., Bohnsack, J. F., Takahashi, S., Oliver, K. A., Chan, M.-S., Kunst, F., et al. (2003). Multilocus sequence typing system for group B *Streptococcus*. *J. Clin. Microbiol.* 41, 2530–2536. doi: 10.1128/JCM.41.6.2530-2536.2003

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8:R61. doi: 10.1186/gb-2007-8-4-r61

Lanotte, P., and Seme, K. (2012). *"Perinatal Infections," in European Manual of Clinical Microbiology (Société Française de Microbiologie), 241–250*. Available at: https://www.escmid.org/escmid_publications/manual_of_microbiology/ [accessed March 28, 2017].

Lewin, E. B., and Amstey, M. S. (1981). Natural history of group B *Streptococcus* colonization and its therapy during pregnancy. *Am. J. Obstet. Gynecol.* 139, 512–515. doi: 10.1016/0002-9378(81)90509-3

Lier, C., Baticle, E., Horvath, P., Haguenoer, E., Valentin, A.-S., Glaser, P., et al. (2015). Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. *Front. Genet.* 6:214. doi: 10.3389/fgene.2015.00214

Lindahl, G., Stalhammar-Carlemalm, M., and Areschoug, T. (2005). Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. *Clin. Microbiol. Rev.* 18, 102–127. doi: 10.1128/CMR.18.1.102-127.2005

Lopez-Sanchez, M.-J., Sauvage, E., Da Cunha, V., Clermont, D., Ratsima Hariniaina, E., Gonzalez-Zorn, B., et al. (2012). The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol. Microbiol.* 85, 1057–1071. doi: 10.1111/j.1365-2958.2012.08172.x

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Manning, S. D., Lewis, M. A., Springman, A. C., Lehotzky, E., Whittam, T. S., and Davies, H. D. (2008). Genotypic diversity and serotype distribution of group B *Streptococcus* isolated from women before and after delivery. *Clin. Infect. Dis.* 46, 1829–1837. doi: 10.1086/588296

Manning, S. D., Springman, A. C., Lehotzky, E., Lewis, M. A., Whittam, T. S., and Davies, H. D. (2009). Multilocus sequence types associated with neonatal group B Streptococcal sepsis and meningitis in Canada. *J. Clin. Microbiol.* 47, 1143–1148. doi: 10.1128/JCM.01424-08

Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., et al. (2011). CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. *PLOS ONE* 6:e19543. doi: 10.1371/journal.pone.0019543

Perez-Ruiz, M., Rodríguez-Granger, J. M., Bautista-Marín, M. F., Romero-Noguera, J., and Rosa-Fraile, M. (2004). Genetic diversity of *Streptococcus agalactiae* strains colonizing the same pregnant woman. *Epidemiol. Infect.* 132, 375–378. doi: 10.1017/S0950268803001845

Pride, D. T., Sun, C. L., Salzman, J., Rao, N., Loomer, P., Armitage, G. C., et al. (2011). Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Res.* 21, 126–136. doi: 10.1101/gr.111732.110

Regan, J. A., Klebanoff, M. A., and Nugent, R. P. (1991). The epidemiology of group B streptococcal colonization in pregnancy. Vaginal infections and prematurity study group. *Obstet. Gynecol.* 77, 604–610.

Shariat, N., and Dudley, E. G. (2014). CRISPRs: molecular signatures used for pathogen subtyping. *Appl. Environ. Microbiol.* 80, 430–439. doi: 10.1128/AEM.02790-13

Slotved, H.-C., Kong, F., Lambertsen, L., Sauer, S., and Gilbert, G. L. (2007). Serotype IX, a proposed new *Streptococcus agalactiae* serotype. *J. Clin. Microbiol.* 45, 2929–2936. doi: 10.1128/JCM.00117-07

Stoll, B. J., Hansen, N. I., Sánchez, P. J., Faix, R. G., Poindexter, B. B., Van Meurs, K. P., et al. (2011). Early onset neonatal sepsis: the burden of group B Streptococcal and *E. coli* disease continues. *Pediatrics* 127, 817–826. doi: 10.1542/peds.2010-2217

Tyson, G. W., and Banfield, J. F. (2007). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* 10, 200–207. doi: 10.1111/j.1462-2920.2007.01444.x

Verani, J. R., McGee, L., Schrag, S. J., and Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, and Centers for Disease Control and Prevention [CDC] (2010). Prevention of perinatal group B streptococcal disease-revised guidelines from CDC, 2010. *MMWR Recomm. Rep. Morb. Mortal. Wkly. Rep. Recomm. Rep. Cent. Dis. Control* 59, 1–36.

Wei, Y., Chesne, M. T., Terns, R. M., and Terns, M. P. (2015). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res.* 43, 1749–1758. doi: 10.1093/nar/gku1407

Yancey, M. K., Duff, P., Clark, P., Kurtzer, T., Frentzen, B. H., and Kubilis, P. (1994). Peripartum infection associated with vaginal group B streptococcal colonization. *Obstet. Gynecol.* 84, 816–819.

Yosef, I., Goren, M. G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 40, 5569–5576. doi: 10.1093/nar/gks216

Conflict of Interest Statement: PH is an employee of DuPont, and co-inventor on several patent applications related to CRISPR-Cas systems.

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

Copyright © 2017 Beauruelle, Pastuszka, Horvath, Perrotin, Mereghetti and Lanotte. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Supplementary Material

CRISPR: a useful genetic feature to follow vaginal carriage of Group B *Streptococcus*.

Clémence Beauruelle, Adeline Pastuszka, Philippe Horvath, Franck Perrotin, Laurent Mereghetti and Philippe Lanotte *

*Corresponding author: philippe.lanotte@univ-tours.fr

Table S1. Internal primers used for CRISPR1 locus sequencing.

Primer name	Sequence (5' → 3')
Sp335F	AAGGGTGTTAGATGATAATACCTTTTTTAA
Sp2166F	TGTTTTTTCTTTCCATATGTGGTGGCTTTC
Sp856F	TATCAAGTCTGCTTTGTCAACAAAGTCAGC
Sp2234F	AGTTTGTTTATGATGAACACCAAAATTTAA
Sp325F	GGATGATTTCGATTATGCGGCGGTGGTTGA
Sp511F	AGTTACTTCTGCTTGGGTTTGATAAGGGTC
Sp40F	TATGTCTTCTAACAGTTGCTTCTTGTGCTT
Sp531F	AGTTTAGAGTTAACGTCTGTATTTTAGATA
Sp109F	ACAAAATCTTTTGTTGCTCCTGGACGTATT
Sp823F	AACACAGCTTCCTCGAAAGGGATATATCTA
Sp2069F	CCGTGCAAGAGTTAAGGAACTGACACAGCA
Sp2201F	GAAAAATTAAATACTTTCCAATAATTATAA
Sp2191F	AACCTAAAAAGATGTATGAGAAGTTGCCTT
Sp650F	ATGCTTGTCAAAGGTAATAAATCTGGAGAT
Sp2228F	ACGTTGAGGGGTTAGTTGCTGCTGGTCTGG
Sp154F	GATATTTGGAAAGATTCTGATTTTGGTAAG
Sp2312F	CTGTTATATCAACAATTACCAAGTCTGCTG
Sp2293F	TTTTCTACTTTATTTCCTACTTCATTTGTT
Sp328F	ATTACTTTCAAGATGTCTATGACTATATGC
Sp107F	ACCATCAAGGCTCTTATCTGCAGATTGTTA
Sp2325F	ATCGGCAATTTCTTGAAGACGTTCAGCGTT

Sp436F	CCATCCAACCATTTGATTTTGTGTAAATAT
Sp2211F	AAAATCTAAAAATACATTTTACCGTTTGAA
Sp2359F	AAGGATTTTACTTCAAAGATTTTTTTGTTG
Sp904bisF	GATATGGTGGACAGGTTCAGCAGGTCATGT
Sp2199F	TCAATATAAATCTGTCCAGTCTTTTTCTCA
Sp351F	ACAAACCTCTAATGGATAATATAGAACAAA
Sp348F	CGTTCTGCTTTAGTCATATGTGTCCTTTCT
Sp75F	AGAGTGTGTCCAAGACCAGAGTTACTGTTT
Sp379F	CATGCCTTTGAAAATAAATTCCGAGCCATT
Sp2132F	ATGGCTAATTATTGCGTGATGTTAGCGGTT
Sp299F	ATGGTCAGAAATCGAAGAAAACGAAGTCGT
Sp305bis	TTGGAAAAAACACGAAAGTGATATTACTTT
Sp263F	AAGTGAAGTTGAATTTTATTTGAGATACTA
Sp295F	TTGGAAAGATGGCAGAGGGGTTAACGCAGC
Sp2259F	TAATTTTGTTATATCTATATTTAAGTCTTT
Sp8R	TATCCATCTCGGTGAGATGAGAATTAGCTT
Sp31R	ATCAGTACCAACAAATGATTTTGTACCATC
Sp52R	TTTATTTTTTCTCAGTTCCTTGATTTTAGA
Sp36R	ACATATCCTTTTGTTAGGTCAAAAGAAGAT
Sp20R	AGATGTCTTGATATCAACTGCTTTAAAAGA
Sp830F	TCAAAAGTTTCCACTAATAGCGTTT
Sp298F	CTTTTAACCGTCTCTCCGCTTC
Sp39F	GCCTTTTCTAACTCTTCAG
Sp246F	GATTAGTTTGCGTACTCGCTC
Sp511F	GCTTGGGTTTGATAAGGG
Sp177R	AGTTACTGTTGAGGGTAGTCC
Sp87R	TATGCAAAGCAGTCACCAGC
Sp42R	TAAGTAACAAGACAGACTTGAA

Table S2. Terminal direct repeat sequences and distribution.

Sequence (5′ → 3′)	Number of isolates n (%)	Clonal complex or ST
GTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCA	122 (59%)	CC1/CC19, CC10, ST103, ST243, ST1001
GTTTTAAAGCTGTGCTGTTATTATGCTAGGGCACCA	43 (21%)	CC23
GTTTTAGAGCTGTGCGGTTATTATGCTAGGGCACCG	30 (15%)	CC17, ST130
GTTTTAGAGCTGCGCGGTTATTATGCTATGCTAGGA	4 (2%)	CC22
GTTTTAGAGCTGTGCTGTTATTATGCTAGGGCACCA	6 (3%)	ST26

Table S3. Altered leader sequences and distribution.

Leader end modification (n=)	CC/ST (n=)	Isolates
Last nucleotide deletion (15)	CC10 (13), CC19 (2)	142, 143, 211, 212, 361, 362, 431, 432, 439, 440, 488, 497, 498, 499, 500
Last nucleotide duplication (1)	CC10 (1)	508
26 additional nucleotides (1)	CC19 (1)	104
Single nucleotide difference (10)	ST130 (4), CC19 (2), CC22 (4)	24, 25, 43, 44, 116, 117, 121, 123, 420, 421

Figure S1. CRISPR1 spacer content of all (205) isolates from all (100) women. The CRISPR1 arrays are represented using a macro-enable Excel tool, whereby spacers are converted into two-color symbols based on spacer sequence. Gaps (= missing spacers) are shown with a boxed cross symbol ☒ after alignment of identical spacers between strains of the same group. MLVA allelic profiles correspond to the number of repeats of each VNTR (SAG2, SAG3, SAG4, SAG7, SAG21, and SAG22). Terminal Direct Repeats (TDRs) are represented using a different color border according to their sequence (table S2). Isolates are grouped by source, and the period between sampling of isolates is indicated. Altered leader sequences are highlighted according to their differences (yellow: last nucleotide deletion, orange: last nucleotide duplication, red: 26 additional nucleotides, green: Single nucleotide difference). The spacer 247 (◆) is CC23 specific, is present in 82% of CC23 isolates and duplicate in 54% of them. **A**, isolates with "distinct" CRISPR1 arrays. **B**, isolates with "close" or "identical" CRISPR1 arrays.

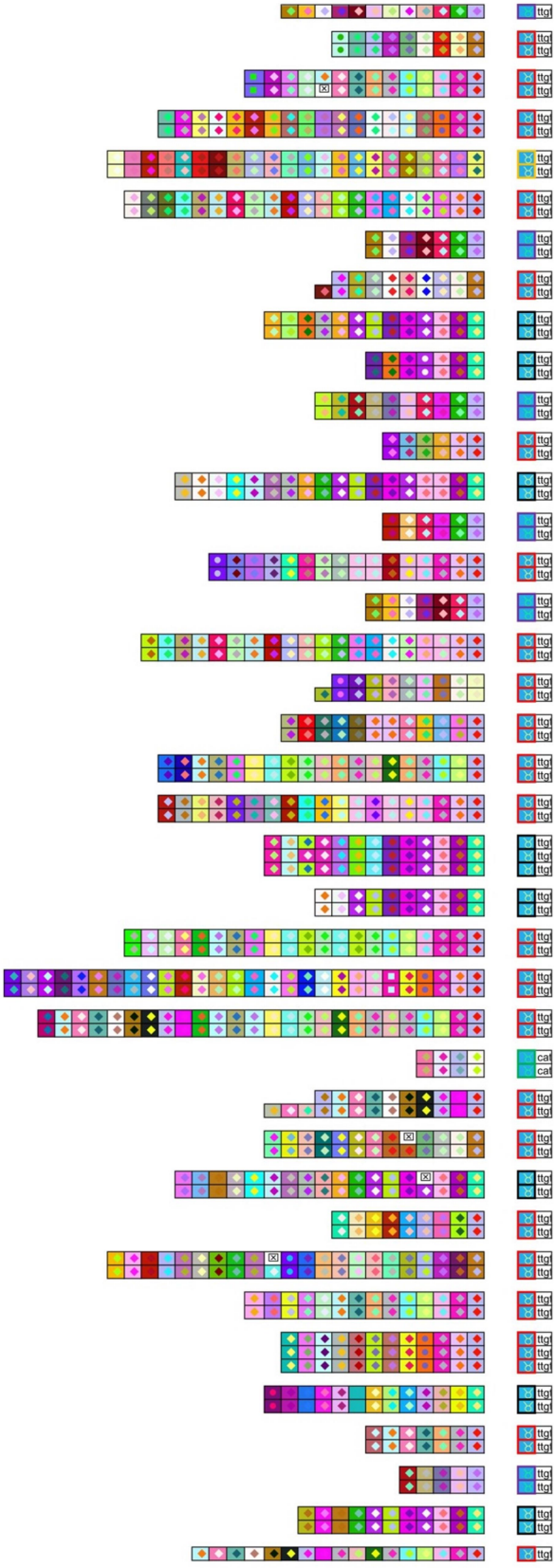
Figure S1A

woman	strain	isolation date (mm/dd/yy)	* time between two isolates days (years)	MLVA Allelic profile	MLST	L	CRISPR1 spacers	TDR
9	S.ag23	05/11/06		3,3,3,5,0,2	ST19 (CC19)	gag		ttg
	S.ag22	06/11/09	**1128 (3.1)**	3,3,1,0,7,4	ST23 (CC23)	gag		ttg
13	S.ag38	08/04/09		2,3,1,0,12,5	ST24 (CC23)	gag		ttg
	S.ag39	04/16/12	**987 (2.7)**	3,3,3,10,48,3	ST1 (CC1)	gag		ttg
16	S.ag45	04/28/11		3,3,3,6,19,3	ST136 (CC2)	gag		ttg
	S.ag46	10/31/13	**918 (2.5)**	3,3,2,0,9,6	ST26	gag		ttg
22	S.ag110	09/06/06		3,3,2,6,11,1	ST1003	gag		ttg
	S.ag109	06/29/10	**1393 (3.8)**	2,3,3,10,6,3	ST17 (CC17)	gag		ttg
27	S.ag136	08/25/06		3,3,1,5,16,3	ST196 (CC2)	gag		ttg
	S.ag135	02/21/13	**2373 (6.5)**	3,3,2,5,0,2	ST28 (CC19)	gag		ttg
29	S.ag141	11/15/07		3,3,3,10,7,3	ST1 (CC1)	gag		ttg
	S.ag140	07/01/13	**2056 (5.6)**	3,3,3,0,18,3	ST2 (CC2)	gag		ttg
32	S.ag153	09/14/07		3,3,1,10,11,3	ST1 (CC1)	gag		ttg
	S.ag152	04/21/10	**951 (2.6)**	4,3,3,0,11,5	ST23 (CC23)	gag		ttg
58	S.ag 329	10/06/05		3,3,3,10,19,3	ST1 (CC1)	gag		ttg
	S.ag 328	01/20/10	**1568 (4.3)**	4,3,1,10,19,5	ST23 (CC23)	gag		ttg
62	S.ag 348	04/29/08		4,3,1,0,11,5	ST23 (CC23)	gag		ttg
	S.ag 346	02/11/11	**1019 (2.8)**	3,3,2,6,0,2	ST327 (CC19)	gag		gtt
70	S.ag 396	04/10/08		3,3,3,5,0,2	ST19 (CC19)	gag		ttg
	S.ag 397	12/08/08	**243 (0.7)**	3,3,3,5,0,2	ST19 (CC19)	gag		ttg
73	S.ag 417	06/11/07		3,3,3,10,7,3	ST1 (CC1)	gag		ttg
	S.ag 416	11/05/10	**1244 (3.4)**	4,3,1,10,11,5	ST23 (CC23)	gag		ttg
89	S.ag 485	07/16/10		3,3,3,5,0,2	ST19 (CC19)	gag		ttg
	S.ag 487	01/06/13	**906 (2.6)**	3,3,3,10,27,3	ST1 (CC1)	gag		ttg
90	S.ag 490	07/05/07		3,3,3,6,2,7	ST27 (CC19)	gag		ttg
	S.ag 488	06/05/09	**702 (1.9)**	3,3,1,6,23,6	ST10 (CC10)	cga		ttg
94	S.ag 506	03/01/06		3,3,1,6,5,7	ST12 (CC10)	gag		ttg
	S.ag 505	02/05/13	**2534 (6.9)**	4,3,1,6,11,5	ST23 (CC23)	gag		ttg
95	S.ag 508	06/06/08		3,3,1,6,0,7	ST12 (CC10)	agg		ttg
	S.ag 507	08/14/12	**1531 (4.3)**	4,3,1,0,11,5	ST24 (CC23)	gag		ttg
96	S.ag 522	12/12/07		3,3,3,10,40,3	ST1 (CC1)	gag		ttg
	S.ag 521	09/09/11	**1368 (3.7)**	2,2,1,10,3,3	ST17 (CC17)	gag		ttg
98	S.ag 541	02/17/10		3,4,3,0,51,3	ST1005 (CC2)	gag		ttg
	S.ag 542	02/22/11	**371 (1.0)**	3,3,1,0,9,4	ST23 (CC23)	gag		ttg
100	S.ag 564	03/30/10		3,4,4,10,18,3	ST297 (CC1)	gag		ttg
	S.ag 565	06/06/14	**1530 (4.2)**	3,4,35,0,2	ST1001	gag		ttg

Figure S1B

woman	strain	isolation date (mm/dd/yy)	* time between two isolates days (years)	MLVA Allelic profile	MLST	L	CRISPR1 spacers	TDR
1	S.ag1	01/21/10		3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
	S.ag2	07/17/12	**908 (2.5)**	3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
2	S.ag3	11/16/11		4,3,1,0,11,5	ST24 (CC23)	gag		ttgt
	S.ag4	11/28/12	**378 (1.0)**	4,3,1,0,11,5	ST24 (CC23)	gag		ttgt
3	S.ag8	02/26/07		3,3,1,0,10,5	ST23 (CC23)	gag		ttgt
	S.ag9	07/11/07	**135 (0.4)**	3,3,1,0,10,5	ST23 (CC23)	gag		ttgt
4	S.ag10	07/02/10		3,3,3,5,0,2	ST182 (CC19)	gag		ttgt
	S.ag12	04/08/14	**1376 (3.8)**	3,3,3,5,0,2	ST182 (CC19)	gag		ttgt
5	S.ag14	01/22/08		5,3,1,0,8,5	ST23 (CC23)	gag		ttgt
	S.ag13	05/31/10	**860 (2.4)**	5,3,1,0,8,5	ST23 (CC23)	gag		ttgt
6	S.ag15	11/16/09		3,3,1,0,10,2	ST23 (CC23)	gag		ttgt
	S.ag16	12/06/11	**750 (2.1)**	3,3,1,0,10,2	ST23 (CC23)	gag		ttgt
7	S.ag17	01/23/11		3,3,1,5,0,2	ST28 (CC19)	gag		ttgt
	S.ag18	12/12/12	**689 (1.9)**	3,3,1,5,0,2	ST28 (CC19)	gag		ttgt
8	S.ag20	09/26/13		3,3,3,5,2,7	ST27 (CC19)	gag		ttgt
	S.ag21	03/17/14	**172 (0.5)**	3,3,3,5,2,7	ST27 (CC19)	gag		ttgt
10	S.ag24	12/06/05		3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
	S.ag25	12/26/06	**385 (1.1)**	3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
11	S.ag28	12/22/09		3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
	S.ag30	07/02/12	**923 (2.5)**	3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
12	S.ag36	03/01/13		3,3,3,10,14,3	ST1 (CC1)	gag		ttgt
	S.ag37	04/03/14	**398 (1.1)**	3,3,3,10,14,3	ST1 (CC1)	gag		ttgt
14	S.ag40	05/05/09		3,3,3,10,7,3	ST1 (CC1)	gag		ttgt
	S.ag42	02/02/12	**1003 (2.7)**	3,3,3,10,7,3	ST1 (CC1)	gag		ttgt
15	S.ag44	04/04/06		2,3,4,0,3,5	ST22 (CC22)	gag		cat
	S.ag43	04/27/09	**1119 (3.1)**	2,3,4,0,3,5	ST22 (CC22)	gag		cat
17	S.ag88	05/23/07		3,3,1,5,0,2	ST28 (CC19)	gag		ttgt
	S.ag90	12/10/07	**201 (0.6)**	3,3,1,5,0,2	ST28 (CC19)	gag		ttgt
18	S.ag94	09/03/07		4,3,2,0,11,5	ST23 (CC23)	gag		ttgt
	S.ag95	12/04/07	**92 (0.3)**	4,3,2,0,11,5	ST23 (CC23)	gag		ttgt
19	S.ag96	11/22/09		3,3,1,5,0,2	ST28 (CC19)	ctct		ttgt
	S.ag97	01/18/12	**787 (2.2)**	3,3,1,5,0,2	ST28 (CC19)	ctct		ttgt
20	S.ag99	12/17/07		3,3,1,6,0,2	ST28 (CC19)	gag		ttgt
	S.ag98	12/08/09	**722 (2.0)**	3,3,1,6,0,2	ST28 (CC19)	gag		ttgt
21	S.ag106	03/21/06		3,3,3,5,0,2	ST28 (CC19)	gag		ttgt
	S.ag104	09/01/11	**1990 (5.5)**	3,3,3,5,0,2	ST28 (CC19)	atg		ttgt
23	S.ag111	09/25/09		3,3,1,0,8,4	ST366 (CC23)	gag		ttgt
	S.ag113	12/30/09	**96 (0.3)**	3,3,1,0,8,4	ST366 (CC23)	gag		ttgt
24	S.ag117	05/23/06		3,3,1,15,9,4	ST130	gag		ttgt
	S.ag116	07/07/11	**1871 (5.1)**	3,3,1,15,9,4	ST130	gag		ttgt
25	S.ag121	11/10/09		3,3,1,25,10,4	ST130	gag		ttgt
	S.ag123	09/23/11	**682 (1.9)**	3,3,1,25,10,4	ST130	gag		ttgt
26	S.ag128	01/05/11		2,2,2,10,6,3	ST17 (CC17)	gag		ttgt
	S.ag132	11/23/11	**322 (0.9)**	2,2,2,10,6,3	ST17 (CC17)	gag		ttgt
28	S.ag138	04/15/05		3,3,3,6,18,3	ST196 (CC2)	gag		ttgt
	S.ag139	02/12/07	**668 (1.8)**	3,3,3,6,18,3	ST196 (CC2)	gag		ttgt
	S.ag137	11/03/09	**995 (2.7)**	3,3,3,6,18,3	ST196 (CC2)	gag		ttgt
30	S.ag142	02/18/13		3,3,1,6,20,6	ST10 (CC10)	cga		ttgt
	S.ag143	12/17/14	**667 (1.8)**	3,3,1,6,20,6	ST10 (CC10)	cga		ttgt
31	S.ag148	06/28/12		3,3,3,10,14,3	ST1 (CC1)	gag		ttgt
	S.ag149	12/21/12	**176 (0.5)**	3,3,3,10,14,3	ST1 (CC1)	gag		ttgt
33	S.ag161	02/15/08		3,3,3,10,28,3	ST1 (CC1)	gag		ttgt
	S.ag159	10/04/13	**2058 (5.6)**	3,3,3,10,25,3	ST1 (CC1)	gag		ttgt
34	S.ag169	01/19/06		3,3,1,0,32,3	ST2 (CC2)	gag		ttgt
	S.ag168	07/15/09	**1273 (3.5)**	3,3,1,0,27,3	ST2 (CC2)	gag		ttgt
35	S.ag171	09/11/06		3,3,1,0,14,5	ST1002 (CC23)	gag		ttgt
	S.ag170	04/08/09	**940 (2.6)**	3,3,1,0,19,5	ST1002 (CC23)	gag		ttgt
36	S.ag183	07/18/13		3,3,1,0,12,4	ST144 (CC23)	gag		ttgt
	S.ag185	11/07/13	**112 (0.3)**	3,3,1,0,12,4	ST144 (CC23)	gag		ttgt
37	S.ag195	08/08/05		3,3,3,6,18,3	ST3 (CC1)	gag		ttgt
	S.ag196	12/08/05	**122 (0.3)**	3,3,3,6,18,3	ST3 (CC1)	gag		ttgt
38	S.ag200	02/07/11		3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
	S.ag202	02/03/14	**1092 (3.0)**	3,3,3,5,0,2	ST19 (CC19)	gag		ttgt
39	S.ag207	03/01/06		2,2,2,10,8,3	ST17 (CC17)	gag		ttgt
	S.ag209	06/29/06	**120 (0.3)**	2,2,2,10,8,3	ST17 (CC17)	gag		ttgt
40	S.ag211	05/10/07		2,2,2,10,8,3	ST12 (CC10)	cga		ttgt
	S.ag212	08/16/07	**98 (0.3)**	3,3,1,6,0,3	ST12 (CC10)	cga		ttgt
41	S.ag228	04/16/09		3,3,3,10,10,3	ST1 (CC1)	gag		ttgt
	S.ag230	10/27/10	**559 (1.5)**	3,3,3,10,10,3	ST1 (CC1)	gag		ttgt
42	S.ag231	08/20/09		4,3,1,0,11,5	ST23 (CC23)	gag		ttgt
	S.ag232	05/19/11	**637 (1.7)**	4,3,1,0,11,5	ST23 (CC23)	gag		ttgt
43	S.ag234	10/01/07		2,2,2,10,8,3	ST17 (CC17)	gag		ttgt
	S.ag233	10/22/10	**1117 (3.1)**	2,2,2,10,8,3	ST17 (CC17)	gag		ttgt
44	S.ag236	03/03/09		2,2,2,12,8,3	ST17 (CC17)	gag		ttgt

	S.ag235	12/03/12	**1371 (3.8)**	2,2,2,12,8,3	ST17 (CC17)	gag
45	S.ag247	06/01/10		3,3,1,6,14,5	ST12 (CC10)	gag
	S.ag248	11/03/10	**155 (0.4)**	3,3,1,6,14,5	ST12 (CC10)	gag
46	S.ag249	08/23/05		3,3,2,10,7,3	ST1 (CC1)	gag
	S.ag254	05/15/06	**265 (0.7)**	3,3,2,10,7,3	ST1 (CC1)	gag
47	S.ag259	03/27/08		3,3,3,5,0,2	ST19 (CC19)	gag
	S.ag258	11/29/11	**1342 (3.7)**	3,3,3,5,0,2	ST19 (CC19)	gag
48	S.ag274	08/20/12		3,3,2,0,9,7	ST26	gag
	S.ag275	12/27/12	**129 (0.4)**	3,3,2,0,9,7	ST26	gag
49	S.ag278	04/29/10		3,3,3,5,0,2	ST19 (CC19)	gag
	S.ag279	06/12/13	**1140 (3.1)**	3,3,3,5,0,2	ST19 (CC19)	gag
50	S.ag280	04/02/05		2,2,2,10,5,3	ST17 (CC17)	gag
	S.ag281	07/25/06	**479 (1.3)**	2,2,2,10,5,3	ST17 (CC17)	gag
51	S.ag287	03/23/07		3,3,1,10,4,2	ST6 (CC10)	gag
	S.ag286	09/03/12	**1991 (5.5)**	3,3,1,10,4,2	ST6 (CC10)	gag
52	S.ag289	02/04/08		4,3,1,0,12,5	ST23 (CC23)	gag
	S.ag288	05/21/12	**1568 (4.3)**	4,3,1,0,12,5	ST23 (CC23)	gag
53	S.ag293	05/03/13		4,3,1,0,11,5	ST23 (CC23)	gag
	S.ag295	08/29/14	**483 (1.3)**	4,3,1,0,11,5	ST23 (CC23)	gag
54	S.ag298	12/02/12		2,2,2,10,6,3	ST17 (CC17)	gag
	S.ag299	09/10/14	**647 (1.8)**	2,2,2,10,6,3	ST17 (CC17)	gag
55	S.ag303	07/17/12		3,2,3,10,3,3,	ST243	gag
	S.ag305	11/26/12	**132 (0.4)**	3,2,3,10,3,3,	ST243	cag
56	S.ag311	07/17/09		3,3,2,0,11,5	ST23 (CC23)	gag
	S.ag313	06/06/12	**1055 (2.9)**	3,3,2,0,11,5	ST23 (CC23)	gag
57	S.ag 325	04/24/05		2,2,2,10,6,2	ST17 (CC17)	gag
	S.ag 324	11/03/09	**1654 (4.5)**	2,2,2,10,6,2	ST17 (CC17)	gag
59	S.ag 330	05/19/11		3,3,3,6,16,10	ST196 (CC2)	gag
	S.ag 331	11/09/12	**540 (1.5)**	3,3,3,6,16,10	ST196 (CC2)	gag
60	S.ag 338	02/16/07		2,2,2,10,5,3	ST17 (CC17)	gag
	S.ag 337	10/04/10	**1326 (3.6)**	2,2,2,10,5,3	ST17 (CC17)	gag
61	S.ag 340	02/15/05		3,3,3,4,0,2	ST19 (CC19)	gag
	S.ag 343	10/19/05	**246 (0.7)**	3,3,3,4,0,2	ST19 (CC19)	gag
63	S.ag 358	07/30/08		3,3,1,5,9,2	ST6 (CC10)	gag
	S.ag 357	07/05/10	**705 (1.9)**	3,3,1,5,9,2	ST6 (CC10)	gag
64	S.ag 362	08/19/08		3,3,1,5,0,2	ST28 (CC19)	cga
	S.ag 361	05/14/09	**268 (0.7)**	3,3,1,5,0,2	ST28 (CC19)	cga
65	S.ag 365	10/12/09		3,3,3,10,23,3	ST1 (CC1)	gag
	S.ag 366	05/04/12	**935 (2.6)**	3,3,3,10,23,3	ST1 (CC1)	gag
66	S.ag 371	01/08/09		3,3,3,6,15,3	ST196 (CC2)	gag
	S.ag 372	05/22/09	**134 (0.4)**	3,3,3,6,15,3	ST196 (CC2)	gag
67	S.ag 380	02/03/04		4,3,1,0,8,5	ST23 (CC23)	gag
	S.ag 381	11/12/04	**283 (0.8)**	4,3,1,0,8,5	ST23 (CC23)	gag
	S.ag 382	01/07/06	**421 (1.2)**	4,3,1,0,8,5	ST23 (CC23)	gag
68	S.ag 387	09/29/11		4,3,2,0,11,5	ST23 (CC23)	gag
	S.ag 388	11/08/12	**406 (1.1)**	4,3,2,0,11,5	ST23 (CC23)	gag
69	S.ag 391	06/29/09		3,3,3,10,22,3	ST1 (CC1)	gag
	S.ag 393	07/08/11	**739 (2.0)**	3,3,3,10,22,3	ST1 (CC1)	gag
71	S.ag 400	12/08/10		3,3,3,6,0,2	ST19 (CC19)	gag
	S.ag 401	02/15/12	**434 (1.2)**	3,3,3,6,0,2	ST19 (CC19)	gag
72	S.ag 409	02/01/08		3,3,3,10,19,3	ST1 (CC1)	gag
	S.ag 408	03/10/10	**768 (2.1)**	3,3,3,10,19,3	ST1 (CC1)	gag
74	S.ag 421	03/20/08		2,3,6,0,4,5	ST22 (CC22)	gag
	S.ag 420	09/15/11	**1274 (3.5)**	2,3,6,0,4,5	ST22 (CC22)	gag
75	S.ag 427	07/28/05		3,3,3,10,7,3	ST1 (CC1)	gag
	S.ag 428	02/18/08	**935 (2.6)**	3,3,3,10,52,3	ST1 (CC1)	gag
76	S.ag431	01/26/10		3,3,1,6,14,7	ST7 (CC10)	cga
	S.ag432	12/21/11	**694 (1.9)**	3,3,1,6,14,7	ST7 (CC10)	cga
77	S.ag433	07/24/10		3,3,1,0,12,4	ST23 (CC23)	gag
	S.ag434	01/23/11	**183 (0.5)**	3,3,1,0,12,4	ST23 (CC23)	gag
78	S.ag 435	07/30/09		3,3,3,6,0,2	ST19 (CC19)	gag
	S.ag 436	11/25/11	**848 (2.3)**	3,3,3,6,0,2	ST19 (CC19)	gag
79	S.ag 439	01/10/10		3,3,1,7,14,3	ST8 (CC10)	cga
	S.ag 440	08/27/12	**960 (2.6)**	3,3,1,7,14,3	ST8 (CC10)	cga
80	S.ag 442	10/02/06		3,3,3,10,7,3	ST1 (CC1)	gag
	S.ag 443	07/02/08	**639 (1.8)**	3,3,3,10,7,3	ST1 (CC1)	gag
81	S.ag 444	11/02/07		3,3,2,6,0,2	ST327 (CC19)	gag
	S.ag 446	09/22/08	**325 (0.9)**	3,3,2,6,0,2	ST327 (CC19)	gag
	S.ag 449	09/11/09	**354 (1.0)**	3,3,2,6,0,2	ST327 (CC19)	gag
82	S.ag 456	12/09/08		2,3,1,0,26,5	ST88 (CC23)	gag
	S.ag 454	07/21/09	**224 (0.6)**	2,3,1,0,26,5	ST88 (CC23)	gag
83	S.ag 459	02/19/08		3,3,3,10,7,3	ST1 (CC1)	gag
	S.ag 457	06/27/11	**1224 (3.4)**	3,3,3,10,7,3	ST1 (CC1)	gag
84	S.ag 463	12/17/08		2,2,2,10,6,3	ST17 (CC17)	gag
	S.ag 462	08/25/10	**616 (1.7)**	2,2,2,10,6,3	ST17 (CC17)	gag
85	S.ag 469	08/30/05		3,3,1,0,12,4	ST23 (CC23)	gag
	S.ag 468	10/05/10	**1862 (5.1)**	3,3,1,0,12,4	ST23 (CC23)	gag
86	S.ag 471	08/12/05		3,3,3,10,8,3	ST1 (CC1)	gag



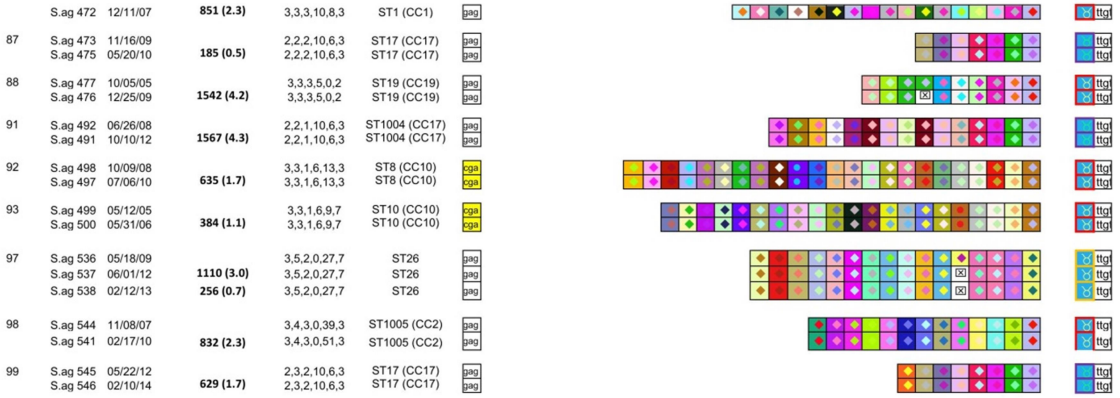

S.ag 472 12/11/07 851 (2.3) 3,3,3,10,8,3 ST1 (CC1) gag
87 S.ag 473 11/16/09 2,2,2,10,6,3 ST17 (CC17) gag
S.ag 475 05/20/10 185 (0.5) 2,2,2,10,6,3 ST17 (CC17) gag
88 S.ag 477 10/05/05 3,3,3,5,0,2 ST19 (CC19) gag
S.ag 476 12/25/09 1542 (4.2) 3,3,3,5,0,2 ST19 (CC19) gag
91 S.ag 492 06/26/08 2,2,1,10,6,3 ST1004 (CC17) gag
S.ag 491 10/10/12 1567 (4.3) 2,2,1,10,6,3 ST1004 (CC17) gag
92 S.ag 498 10/09/08 3,3,1,6,13,3 ST8 (CC10) cga
S.ag 497 07/06/10 635 (1.7) 3,3,1,6,13,3 ST8 (CC10) cga
93 S.ag 499 05/12/05 3,3,1,6,9,7 ST10 (CC10) cga
S.ag 500 05/31/06 384 (1.1) 3,3,1,6,9,7 ST10 (CC10) cga
97 S.ag 536 05/18/09 3,5,2,0,27,7 ST26 gag
S.ag 537 06/01/12 1110 (3.0) 3,5,2,0,27,7 ST26 gag
S.ag 538 02/12/13 256 (0.7) 3,5,2,0,27,7 ST26 gag
98 S.ag 544 11/08/07 3,4,3,0,39,3 ST1005 (CC2) gag
S.ag 541 02/17/10 832 (2.3) 3,4,3,0,51,3 ST1005 (CC2) gag
99 S.ag 545 05/22/12 2,3,2,10,6,3 ST17 (CC17) gag
S.ag 546 02/10/14 629 (1.7) 2,3,2,10,6,3 ST17 (CC17) gag

Figure S2. DNA sequences of CRISPR1 arrays represented in Figure 3. Leader and spacer sequences are represented by lowercase letters, while Direct Repeats (DRs) are represented by uppercase letters.

Woman 2

Isolate 3

AaactgaagtcctgctgagacgaatggcgcgattacgaaagctcaaaagaaaatttctacgagGTTTTAGAGCTGTGCTGTT
TCGAATGGTTCCAAAACaactacagagattgtctttagagccgttaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAAC
gatatggtggacaggttcagcaggtcatgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatcgcaaatgaaacgg
gcgcttttgcactGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaataactgtccgttttgccatttcttgcGTTT
TAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgaagtatactacggttagagattggctcaaGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACgaaacttcgattagtttgcgtactcgctcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACat
tagtcctgttgttatgatggatgattatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaacctctaatggataa
tatagaacaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacacaattaaacataaagaagtgcttatttGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACtcaggagattgtgtgtactcacgaatttttGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaacc
ttgataggcgtactgtatggattccaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctactgtattatctgaaatag
tatcgttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtctcgggaaatgcaactgatgctgcaatcaGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACttaataactgtccgttttgccatttcttgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagaggg
gaaaatatcaatgccgaatgctgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttgg
tactgatGTTTTAAAGCTGTGCTGTTATTATGCTAGGGCACCAttgtggtgttctagtttttgttatactgaaataaatttt
cagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcacccc
aatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaa
attcgaattgtttttcggattcaaagagaatgtgtcgataaaagagatatgaaaggctataatt

Isolate 4

Acaattctgataggcgaaatatggtatactattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctga
gacgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACa
actacagagattgtctttagagccgttaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgatatggtggacaggttc
agcaggtcatgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatcgcaaatgaaacgggcgcttttgcactGTTTT
AGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaataactgtccgttttgccatttcttgcGTTTTAGAGCTGTGCTGTTTCG
AATGGTTCCAAAACgaagtatactacggttagagattggctcaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgaa
acttcgattagtttgcgtactcgctcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattagtcctgttgttatgat
ggatgattatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaacctctaatggataatatagaacaaaGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACacacaattaaacataaagaagtgcttatttGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACtcaggagattgtgtgtactcacgaatttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaatt
gtttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgtaatata
aagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaccttgataggcgtactgtatggattccaGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACCtactgtattatctgaaatagtatcgttttGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACtctcgggaaatgcaactgatgctgcaatcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgt
ttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaataactgtccgttttgccatt
tcttgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagaggggaaaatatcaatgccgaatgctgaGTTTTAGAGCT
GTGCTGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttggtactgatGTTTTAAAGCTGTGCTGTTATTATGCTA
GGGCACCAttgtggtgttctagtttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaat
tcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattgg
cgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgt
cgataaaagagatatgaaaggctataattccaaccttatggttaaagggcaagatgtc

Woman 7

Isolate 17

tttcttgataggcgatatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgaga
cgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtac
gtcattatttgttactaaaggactcatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattggttattttacaaacga
cccggataaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaatccagcttacttttaatcaaaagctGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACaaaacgtcgtaaacgtgtcattgattgtatGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACttttgctaagtctgcaacgcctttaataatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctttg
taaaaatttctttaactaattcatgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtctcgatggaagatttcgaatt
aaaaaccaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtttaacaaaatcataacctaaaggggtattGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACtattatcagatacgctaaaaccagtagttcGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACctacacgcaagattacatggcttctcagttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggttat
acatttactaatccatcagcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgaga
tggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagtttttgttatactgaaataaattttc
agagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcacccca
atttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaa

ttcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggttaaaggg
ctaggtgtctacgc

Isolate 18

tttcttgataggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgaga
cgaatggcgcgattacgaaagctcaaaagaaaatttttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtac
gtcattatttgttactaaaggactcatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattggttattttacaaacga
cccggataaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaatccagcttacttttaatcaaaagctGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACaaaacgtcgtaaacgtgtcattgattgtatGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACttttgctaagtctgcaacgcctttaataatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctttg
taaaaatttctttaactaattcatgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtctcgatggaagatttcgaatt
aaaaaccaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtttaacaaaatcataacctaaaggggtattGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACtggttatacatttactaatccatcagcattGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggt
gttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattca
gaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaaatatg
aacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatat
gaaaggctataattccaaccttatggttaaagggctaggtgtctacgc

Woman 11

Isolate 28

atatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattac
gaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtactggtcttctgaaaaag
ttttataactGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcctccgacttctggcgcagtcattgcaactGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACacatgccaccttactgcatggtatatcgttGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACataatcttctgaaattgttgcttcgtaataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccatc
caaccatttgattttgtgtaaatatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgttgcatgaattgaaagttcaa
tataagcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccgttcaatctgttcttgcttttggtcatcGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACaagggtgttagatgataatacctttttaaGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACacaccgttgcggttgttgtcggtcactcaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaa
ttctcatctcaccgagatggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgtta
tactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaag
ctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtg
tgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattcca
accttatggtaaagggctaggatgtcta

Isolate 30

tatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattacg
aaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaaaatgaagtgcgaaaaaa
actactcgatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtactggtcttctgaaaaagttttataactGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACcctccgacttctggcgcagtcattgcaactGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACacatgccaccttactgcatggtatatcgttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACataatc
ttctgaaattgttgcttcgtaataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccatccaaccatttgattttgtg
taaatatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgttgcatgaattgaaagttcaatataagcaGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACccgttcaatctgttcttgcttttggtcatcGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACaagggtgttagatgataatacctttttaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaccgtt
gcggttgttgtcggtcactcaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagat
ggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttca
gagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaa
tttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaat
tcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggtaaagggct
aggtgtctacg

Woman 14

Isolate 40

tatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattacg
aaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattaagccaaaggcggttaa
aacaggaatgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgataggcttcagttaccaagatagcacgGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACatgttctgagttacctttctcgacacgcggGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACtgtttttattgaaatatccaataggtctgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggaa
ctgaatatcactcgctttttacatcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaatcgtcttcattgttgcgatg
attttgatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaacattagccttttctaactcttcagctgtGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACtatgtcttctaacagttgcttcttgtgcttGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACatatgttccactctatgaatttaggctcatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtttttac
caatgcttccatatcgcttatatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtacttgacgaattgaagatgacgg
aatttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCT
GTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaag

gcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattg
ggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcgg
attcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggt

Isolate 42

gaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgatt
acgaaagctcaaaagaaaatttttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattaagccaaaggcggt
taaaacaggaatgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgataggcttcagttaccaagatagcacgGTTT
TAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatgttctgagttacctttctcgacacgcggGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACtgtttttattgaaatatccaataggtctgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtg
gaactgaatatcactcgctttttacatcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaatcgtcttcattgttgcg
atgattttgatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaacattagccttttctaactcttcagctgtGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACtatgtcttctaacagttgcttcttgtgcttGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACccgtcaaacaagagcgacagcgaaacaagcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACTtaa
cagtttcaagtctgtcttgttacttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACactctaaatgatagttatgag
ttaaatgttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaaattacagtttcgactgattatggaaatGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACtttttaccaatgcttccatatcgcttatatGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACtacttgacgaattgaagatgacggaatttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagcta
attctcatctcaccgagatggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgtt
atactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaa
gctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggt
gtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattcc
aaccttatggtaaagggctaggtgt

Woman 20

Isolate 98

cgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcga
ttacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaaatcttttgtt
gctcctggacgtattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatgtaaggtattgtaaaacttcttcttgcgGT
TTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACactgttcctataattaaaataaaagaggtaGTTTTAGAGCTGTGCTGTT
TCGAATGGTTCCAAAACtgttccagtaaaaagtaattttaaagcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAAC
cgctcgattgatgctatcaactatattgaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttcttcaagagaacttg
tagaacagcttcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaggtacttttagcttgttcttgtggtgttGTTT
TAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacagctactgtAAATTCTGCTTTTACGGTTGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACtagtgcagttgtcaaggagattgtgagcgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtt
taaccttgaaaatgtgaaaggctcgtaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgcgatgatggtaagtcatc
atggacagcgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttttacacacgatgtcagatataatgtcaaGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACagtactgcactaggaattgtagagatcaaaGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACcgtaccatctatcaatttaccgcaagctgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaa
aagatttaaactatcaagcgtcaattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACTtctaaatgctggtgactgct
ttgcataaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACTtgctgctagacccaaacagtttatttttagGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACtccttttttagataataatgtgcgatcacggacGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACttttaccaatgcttccatatcgcttatatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggtta
tacatttactaatccatcagcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgag
atggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaatttt
cagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcacccc
aatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaa
attcgaattgtttttcggattcaaagataatgtgtcaataaaagagatatgaaaggctataattccaaccttatggtaaaggg
ctaggtgt

Isolate 99

cgaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgat
tacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaaatcttttgttg
ctcctggacgtattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatgtaaggtattgtaaaacttcttcttgcgGTT
TTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACactgttcctataattaaaataaaagaggtaGTTTTAGAGCTGTGCTGTTT
CGAATGGTTCCAAAACtgttccagtaaaaagtaattttaaagcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACc
gctcgattgatgctatcaactatattgaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttcttcaagagaacttgt
agaacagcttcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaggtacttttagcttgttcttgtggtgttGTTTT
AGAGCTGTGCTGTTTCGAATGGTTCCAAAACacagctactgtaaattctgcttttacggttGTTTTAGAGCTGTGCTGTTTCG
AATGGTTCCAAAACtagtgcagttgtcaaggagattgtgagcgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttt
aacctttgaaaatgtgaaaggctcgtaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgcgatgatggtaagtcatca
tggacagcgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgctgctagacccaaacagtttatttttagGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACtccttttttagataatgtgcgatcacggacGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACttttaccaatgcttccatatcgcttatatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggtt
atacatttactaatccatcagcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccga
gatggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattt
tcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccc
caatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaa

aattcgaattgtttttcggattcaaagataatgtgtcaataaaagagatatgaaaggctataattccaaccttatggtaaagg
gctaggtgtt

Woman 21

Isolate 104

Tatggtataatattagttaaaagcacagtaataacaaggaatcactcgaaactgaagtcctgctgagacgaatggcgcgatta
cgaaagctcaaaagaaaatttttctacgaggttttagagctgtgctgtttcgaatgGTTTTAGAGCTGTGCTGTTTCGAATGGT
TCCAAAACccgtgcaagagttaaggaactgacacagcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcatactagc
tttcatgactttcgaacctttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaacacagcttcctcgaaagggatata
tctaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaatattgaaatagtttgtgcagatggttttGTTTTAGAGCTGT
GCTGTTTCGAATGGTTCCAAAACtcctcgccgtcgacataaccagccatttcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTC
CAAAACaacgatttgcacaataatccaggacaatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaataggcttag
gcgcatctggttctttactGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACactataagtattttgttttatccctagt
agGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatctacgagacgaaaacaattgccaatgtGTTTTAGAGCTGTGC
TGTTTCGAATGGTTCCAAAACagttatcaatgctatatggtttctcatcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCA
AAACtatttcatcatagaaaatcctgctagtggtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggttatacattt
actaatccatcagcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggata
GTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagtttttgttatactgaaataaattttcagagaa
tgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttat
cagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaa
ttgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggttaaaggggctagg
t

Isolate 106

tgaagtcctgctgagacgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACccgtgcaagagttaaggaactgacacagcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcata
ctagctttcatgactttcgaacctttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaacacagcttcctcgaaaggg
atatatctaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaatattgaaatagtttgtgcagatggttttGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACtcctcgccgtcgacataaccagccatttcaGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACaacgatttgcacaataatccaggacaatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagtgt
gttatcatatcttcctttggttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaataggcttaggcgcatctggtt
ctttactGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACactataagtattttgttttatccctagtagGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACtatCtacgagacgaaaacaattgccaatgtGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACagttatcaatgctatatggtttctcatcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatttcat
catagaaaatcctgctagtggtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggttatacatttactaatccatca
gcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCTG
TGCTGTTATTATGCTAGGACATCAttgtggtgttctagtttttgttatactgaaataaattttcagagaatgtgggggaagg
cggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgg
gaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcgga
ttcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggtaaagggctaggt

Woman 30

Isolate 142

Atatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattac
gaaagctcaaaagaaaattttctacgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaatttccaatacttacgat
atcttcgtcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcatgcctttgaaaataaattccgagccattGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACacttcattgatgttttgttcaaatttctcaGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACcaaattatagctaagtacccaaacagtattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgtt
tagcaacatagtctaatgttgcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaaagatacacctcaaaatgat
acatctgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgttgcttgttcgtattcgtcaggagaagaGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACcaaccctatgtttgataatattttagacgtGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACaagtgaagttgaattttatttgagatactaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaatact
tttacaatatgtgttttcactacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagtgccacagtttgtggctgatt
ggattgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcattcaaggactaccctcaacagtaactctGTTTTAGAGCT
GTGCTGTTTCGAATGGTTCCAAAACataatatattatataataaatatagaaataGTTTTAGAGCTGTGCTGTTTCGAATGGT
TCCAAAACaaataatattgattttcaaaattattttaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttagcagaa
ttaagtatgtattatatgaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctcaaaaccttaattaagatagtagc
attGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcttcttttttaattcttctaacactccatcGTTTTAGAGCTGTG
CTGTTTCGAATGGTTCCAAAACatcttcttttgacctaacaaaaggatatgtGTTTTAGAGCTGTGCTGTTATTATGCTAGGA
CATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattca
agacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcga
ctaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaa
taaaagagatatgaaaggctataattccaaccttatggtaaagggctaggtgtctacgc

Isolate 143

atatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattac
gaaagctcaaaagaaaattttctacgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaatttccaatacttacgat

atcttcgtcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcatgcctttgaaaataaattccgagccattGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACacttcattgatgttttgttcaaatttctcaGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACcaaattatagctaagtacccaaacagtattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgtt
tagcaacatagtctaatgttgcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaaagatacacctcaaaatgat
acatctgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgttgcttgttcgtattcgtcaggagaagaGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACcaaccctatgtttgataatattttagacgtGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACaagtgaagttgaattttatttgagatactaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaatact
tttacaatatgtgttttcactacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagtgccacagtttgtggctgatt
ggattgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcattcaaggactaccctcaacagtaactctGTTTTAGAGCT
GTGCTGTTTCGAATGGTTCCAAAACataatatattatataataaatatagaaataGTTTTAGAGCTGTGCTGTTTCGAATGGT
TCCAAAACaaataatattgattttcaaaattattttaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttagcagaa
ttaagtatgtattatatgaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatcttcttttgacctaacaaaaggata
tgtGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagtttttgttatactgaaataaattttcaga
gaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatt
tatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattc
gaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggtaaagggctag
g

Woman 42

Isolate 231

aaactgaagtcctgctgagacgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTT
TCGAATGGTTCCAAAACgactgcacaacgacaatgcgggtgaaatggGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAAC
tgataccttattttcatacacaatccactGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgaagtatactacggtta
gagattggctcaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgaaacttcgattagtttgcgtactcgctcaGTTT
TAGAGCTGTGCTGTTTCGAATGGTTCCAAAACAttagtcctgttgttatgatggatgattatGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACacaaacctctaatggataatatagaacaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACac
acaattaaacataaagaagtgcttatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcaggagattgtgtgtact
cacgaatttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACaaccttgataggcgtactgtatggattccaGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACtctcgggaaatgcaactgatgctgcaatcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaat
tgtttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgctagcatggcacaaaaatag
cgttggaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaataactgtccgttttgccatttcttgcGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttggtactgatGTTTTAAAGCTGTGCTGTTATTATG
CTAGGGCACCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagatt
aattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatat
tggcgactaaaaaatatgaacgtcatttagaggtgttctattagtagaaaattggaatttttt

Isolate 232

atatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattac
gaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgactgcacaacgacaatgc
gggtgaaatggGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgataccttattttcatacacaatccactGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACgaagtatactacggttagagattggctcaaGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACgaaacttcgattagtttgcgtactcgctcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACAtta
gtcctgttgttatgatggatgattatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaacctctaatggataata
tagaacaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacacaattaaacataaagaagtgcttatttGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACtcaggagattgtgtgtactcacgaatttttGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaccctt
gataggcgtactgtatggattccaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtctcgggaaatgcaactgatgct
gcaatcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACgctagcatggcacaaaaatagcgttggaatGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACtcaggagattgtgtgtactcacgaatttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtt
tttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttggta
ctgatGTTTTAAAGCTGTGCTGTTATTATGCTAGGGCACCAttgtggtgttctagttttttgttatactgaaataaattttca
gagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaa
tttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaat
tcgaattgtttttcggattcaaagagaatgtgtcgataaaagagatatgaaaggctataattccaaccttatggttaaagggc
aagatt

Woman 46

Isolate 249

tttcttgataggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgaga
cgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctt
gacgttgttcttgacgatgattgccgtcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaactgcaacatacttaat
catttttttctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggaactgaatatcactcgctttttacatcGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACaatcgtcttcattgttgcgatgattttgatGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACaacattagccttttctaactcttcagctgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatg
tcttctaacagttgcttcttgtgcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccgtcaaacaagagcgacagc

gaaacaagcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaacagtttcaagtctgtcttgttacttaGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACACTCTAAATGATAGTTATGAGTTAAATGTTGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACCaaattacagtttcgactgattatggaaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatatgt
tccactctatgaatttaggctcatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtttttaccaatgcttccatatcg
cttatatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtacttgacgaattgaagatgacggaatttaGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCTGTGCTGTTATTATGCT
AGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaa
ttcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattg
gcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtg
tcaataaaagagatatgaaaggctataattccaaccttatggttaaagggctaggtgtcta

Isolate 254

ttgaaggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaat
ggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcttgacgt
tgttcttgacgatgattgccgtcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaactgcaacatacttaatcattt
ttttctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggaactgaatatcactcgctttttacatcGTTTTAGAGCT
GTGCTGTTTCGAATGGTTCCAAAACaatcgtcttcattgttgcgatgattttgatGTTTTAGAGCTGTGCTGTTTCGAATGGT
TCCAAAACtatgtcttctaacagttgcttcttgtgcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccgtcaaac
aagagcgacagcgaaacaagcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaacagtttcaagtctgtcttgtta
cttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACactctaaatgatagttATGAGTTAAATGTTGTTTTAGAGCTGT
GCTGTTTCGAATGGTTCCAAAACcaaattacagtttcgactgattatggaaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTC
CAAAACAtatgttccactctatgaatttaggctcatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtttttaccaat
gcttccatatcgcttatatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtacttgacgaattgaagatgacggaatt
taGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCTGTGC
TGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcgg
taattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaa
acttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattc
aaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggttaaagggctaggt

Woman 51

Isolate 286

tgaaggcgaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatgg
cgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcatttgaaac
gctcttccaacgtcgcttctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaatttaaaggtagttttgcaggtgttt
tatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaaggagctggttttgactatttccaatcgGTTTTAGAGCTGTG
CTGTTTCGAATGGTTCCAAAACagctaagaaaaaaataatttaggaatttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCC
AAAACttccattttatttactcctcgtctggtattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggttaaataag
caatcaaagcaggcacacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatccttttctgagggcgtcggtatttgt
tGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatgccaaccctcaatttacttggtcacaaaGTTTTAGAGCTGTGCT
GTTTCGAATGGTTCCAAAACttctatcttctgaagatatttcacaagtgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAA
AACatcttcttttgacctaacaaaaggatatgtGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctag
ttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttta
gttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtca
tttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggc
tataattccaaccttatggttaaagggctaggtgtcta

Isolate 287

attgcttgataggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgag
acgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaa
tttaaaggtagttttgcaggtgttttatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaaggagctggttttgact
atttccaatcgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagctaagaaaaaaataatttaggaatttaGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACttccattttatttactcctcgtctggtattGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACtggttaaataagcaatcaaagcaggcacacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatc
cttttctgagggcgtcggtatttgttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatgccaaccctcaatttactt
ggtcacaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttctatcttctgaagatatttcacaagtgaGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACatcttcttttgacctaacaaaaggatatgtGTTTTAGAGCTGTGCTGTTATTATG
CTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagatt
aattcagacgtaatttcagaacttagttggccagctaacgaatcacccatttatcagttgatggaac

Woman 63

Isolate 357

Caaagggatagggttattgctttgataggcgaaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaa
actgaagtcctgctgagacgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACagtcaatgctgttaagacagcatttcaaggGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACct
ggagttgctaaagtacctcgtccttggGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaagggctgaaaaacaagcg
atatatcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacctaaatagggttcataaaagttattattGTTTTAG

AGCTGTGCTGTTTCGAATGGTTCCAAAACttagaagcttattttacaggcgaaaaaaccGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACagctaagaaaaaaataatttttaggaatttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttttc
accgcttcagcaacgtcataaggctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattgacaggtgataccttaaca
ataaaaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttctatcttctgaagatatttcacaagtgaGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACtcctctttctatgtttcaattgccaattttGTTTTAGAGCTGTGCTGTTATTATGC
TAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagatta
attcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatatt
ggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgt
gtcaataaaagagatatgaaaggctataattccaaccttatggttaaagggctaggttgttcta

Isolate 358

tatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattacg
aaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctggagttgctaaagtacct
cgtccttggGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaagggctgaaaaacaagcgatatatcattGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACacctaaataggttcataaaagttattattGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACttagaagcttattttacaggcgaaaaaaccGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagctaa
gaaaaaaataatttttaggaatttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttttcaccgcttcagcaacgtca
taaggctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattgacaggtgataccttaacaataaaaaaGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACttctatcttctgaagatatttcacaagtgaGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACtcctctttctatgtttcaattgccaattttGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtg
ttctagtttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcag
aacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaaatatga
acgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatg
aaaggctataattccaaccttatggtaaagggctaggtgtctac

Woman 67

Isolate 380

gtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattacgaaag
ctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatcggaacgagtggagtggtaaa
acatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaaaatactgtaaaatcgtaatctacataaGTTTTAGAGCT
GTGCTGTTTCGAATGGTTCCAAAACaaatgttattgatgagacaacattaaaattGTTTTAGAGCTGTGCTGTTTCGAATGGT
TCCAAAACtgtgcgtataaatgatagcttgctttccttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatagtcctt
gtgatagaagttggtgtaagaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaatttctttgctcttgatgatgacg
gtgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcgttctgctttagtcatatgtgtccttttctGTTTTAGAGCTGT
GCTGTTTCGAATGGTTCCAAAACctactgtattatctgaaatagtatcgttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTC
CAAAACtctcgggaaatgcaactgatgctgcaatcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgttttt
gttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaataactgtccgttttgccatttctt
gcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagaggggaaaatatcaatgccgaatgctgaGTTTTAGAGCTGTGC
TGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttggtactgatGTTTTAAAGCTGTGCTGTTATTATGCTAGGGC
ACCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaa
gacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgac
taaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcgat
aaaagagatatgaaaggctataattccaaccttatggttaaagggcaa

Isolate 381

tattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcgattacgaaagctcaaa
agaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACTatcggaacgagtggagtggtaaaacattt
GTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaaaatactgtaaaatcgtaatctacataaGTTTTAGAGCTGTGCTG
TTTCGAATGGTTCCAAAACtgtgcgtataaatgatagcttgctttccttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAA
ACtgtgcgtataaatgatagcttgctttccttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatagtccttgtgata
gaagttggtgtaagaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaatttctttgctcttgatgatgacggtgtGT
TTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcgttctgctttagtcatatgtgtcctttctGTTTTAGAGCTGTGCTGTT
TCGAATGGTTCCAAAACctactgtattatctgaaatagtatcgttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAAC
tctcgggaaatgcaactgatgctgcaatcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgta
atataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaataactgtccgttttgccatttcttgcGTTT
TAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagaggggaaaatatcaatgccgaatgctgaGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACgatggtacaaaatcatttgttggtactgatGTTTTAAAGCTGTGCTGTTATTATGCTAGGGCACCAtt
gtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgta
attcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaa
atatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcgataaaaga
gatatgaaaggctataattccaaccttatggttaaagggcaa

Isolate 382

Gaaagggaaaggattaatttcttgataggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaa
ctgaagtcctgctgagacgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCG
AATGGTTCCAAAACtatcggaacgagtggagtggtaaaacatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaa
aatactgtaaaatcgtaatctacataaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaatgttattgatgagacaa

cattaaaattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgtgcgtataaatgatagcttgctttccttGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACatagtccttgtgatagaagttggtgtaagaGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACcaatttctttgctctttgatgatgacggtgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcgttc
tgctttagtcatatgtgtcctttctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctactgtattatctgaaatagt
atcgttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtctcgggaaatgcaactgatgctgcaatcaGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACttaataactgtccgttttgccatttcttgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagagggg
aaaatatcaatgccgaatgctgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttggt
actgatGTTTTAAAGCTGTGCTGTTATTATGCTAGGGCACCAttgtggtgttctagttttttgttatactgaaataaattttc
agagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcacccca
atttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaa
ttcgaattgtttttcggattcaaagagaatgtgtcgataaaaaagatatgaaaggctataattccaaccttatggttaaagg

Woman 75

Isolate 427

gcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcg
attacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaactaaagatatt
gactccctgcaggataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaacattagccttttctaactcttcagctgtG
TTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatgtcttctaacagttgcttcttgtgcttGTTTTAGAGCTGTGCTGT
TTCGAATGGTTCCAAAACccgtcaaacaagagcgacagcgaaacaagcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAA
CttttattggttttctaagtgctcgaccatcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagttacttctgcttgg
gtttgataagggtcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgaaatgtggagtcattcaggttgatgatggGTT
TTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaaaaataaatgactttaaagcacttggagGTTTTAGAGCTGTGCTGTTT
CGAATGGTTCCAAAACtagtagccattattattatggcttttatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACa
agctaattctcatctcaccgagatggataGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagtttt
ttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttg
gccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcattta
gaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctata
attccaaccttatggtaaagggctaggtgtctacgct

Isolate 428

tttcttgataggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgaga
cgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaat
atattcgagcaatcaacaaagggattgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttactcatttattactcctc
tgtgctatgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtccttatcagataccgaaaatacaaagttaGTTTTAG
AGCTGTGCTGTTTCGAATGGTTCCAAAACaaactaaagatattgactccctgcaggataGTTTTAGAGCTGTGCTGTTTCGAA
TGGTTCCAAAACaacattagccttttctaactcttcagctgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtatgt
cttctaacagttgcttcttgtgcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccgtcaaacaagagcgacagcg
aaacaagcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACTtttattggttttctaagtgctcgaccatcGTTTTAGAG
CTGTGCTGTTTCGAATGGTTCCAAAACagttacttctgcttgggtttgataagggtcGTTTTAGAGCTGTGCTGTTTCGAATG
GTTCCAAAACgaaatgtggagtcattcaggttgatgatggGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaaaaat
aaatgactttaaagcacttggagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtagtagccattattattatggctt
ttatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCT
GTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaag
gcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattg
ggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcgg
attcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggttaaagggctaggtgtctacgct

Woman 76

Isolate 431

gaaatatggtataatattagtaaaagcacagtaataacaaggaattatcgaaactgaagtcctgctgagacgaatggcgcgat
tacgaaagctcaaaagaaaattttctacgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctgcattattagctttt
gcagagctttataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgaattccacgccaccaagtaaacctgtgaGTTT
TAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaagaattttagatactctacttgaatgctGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACgggtccaaatcagcaatctatgctaaaaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACca
accctatgtttgataatatatttagacgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagtgaagttgaatttttat
ttgagatactaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaatactttacaatatgtgttttcactacGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACaagtgccacagtttgtggctgattggattgGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACcattcaaggactaccctcaacagtaactctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACataa
tatattatataataaatatagaaataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttctatcttctgaagatattt
cacaagtgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatcttcttttgacctaacaaaaggatatgtGTTTTAGA
GCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtggggg
aaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttga
ttgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgttttt
cggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggttaaaggg

Isolate 432

ttgataggcgaatatggtataatattagtaaaagcacagtaataacaaggaattatcgaaactgaagtcctgctgagacgaat
ggcgcgattacgaaagctcaaaagaaaattttctacgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACctgcattat
tagcttttgcagagctttataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgaattccacgccaccaagtaaacct
gtgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaagaattttagatactctacttgaatgctGTTTTAGAGCTGT
GCTGTTTCGAATGGTTCCAAAACgggtccaaatcagcaatctatgctaaaaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTC
CAAAACcaaccctatgtttgataatattttagacgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagtgaagttg
aattttatttgagatactaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaatacttttacaatatgtgttttcact
acGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagtgccacagtttgtggctgattggattgGTTTTAGAGCTGTGC
TGTTTCGAATGGTTCCAAAACaagtgccacagtttgtggctgattggattgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCA
AAACcattcaaggactaccctcaacagtaactctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACataatatattata
taataaatatagaaaataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttctatcttctgaagatatttcacaagtga
GTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatcttcttttgacctaacaaaaggatatgtGTTTTAGAGCTGTGCTG
TTATTATGCTAGGACATCAttgtggtgttctagtttttttgttatactgaaataaattttcagagaatgtgggggaaggcggta
attagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaac
ttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaa
agagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggttaaa

Woman 77

Isolate 433

gcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcg
attacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcattcatcttctc
aagttgccagatataaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaattaactttaataacctactataattctG
TTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaatttctatggaagaaaccgcagcagcaatGTTTTAGAGCTGTGCTGT
TTCGAATGGTTCCAAAACatacaagaatatcactgaactactcaaaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAA
CgaaacttcgattagtttgcgtactcgctcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattagtcctgttgtta
tgatggatgattatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaacctctaatggataatatagaacaaaGTT
TTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacacaattaaacataaagaagtgcttatttGTTTTAGAGCTGTGCTGTTT
CGAATGGTTCCAAAACttttcagctgaggaagtaatcttttcttctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACt
caggagattgtgtgtactcacgaattttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgaatgaagcaggtatcaa
ggattatacttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTT
AGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaccttgataggcgtactgtatggattccaGTTTTAGAGCTGTGCTGTTTCG
AATGGTTCCAAAACtctcgggaaatgcaactgatgctgcaatcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtta
ataactgtccgttttgccatttcttgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagaggggaaaatatcaatgc
cgaatgctgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttggtactgatGTTTTAA
AGCTGTGCTGTTATTATGCTAGGGCACCAttgtggtgttctagtttttgttatactgaaataaattttcagagaatgtgggg
gaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttg
attgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttt
tcggattcaaagagaatgtgtcgataaaatagatatgaaaggctataattccaaccttatggtaaagggcaagatgtctacgc
gggcccaaataaactgaaac

Isolate 433

gcgaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgcga
ttacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcattcatcttctca
agttgccagatataaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaattaactttaataacctactataattctGT
TTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaatttctatggaagaaaccgcagcagcaatGTTTTAGAGCTGTGCTGTT
TCGAATGGTTCCAAAACatacaagaatatcactgaactactcaaaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAAC
gaaacttcgattagtttgcgtactcgctcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattagtcctgttgttat
gatggatgattatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacaaacctctaatggataatatagaacaaaGTTT
TAGAGCTGTGCTGTTTCGAATGGTTCCAAAACacacaattaaacataaagaagtgcttatttGTTTTAGAGCTGTGCTGTTTC
GAATGGTTCCAAAACttttcagctgaggaagtaatcttttcttctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtc
aggagattgtgtgtactcacgaattttttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgaatgaagcaggtatcaag
gattatacttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaattgtttttgttgtaatataaagtcgacGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACaaccttgataggcgtactgtatggattccaGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACtctcgggaaatgcaactgatgctgcaatcaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaat
tgtttttgttgtaatataaagtcgacGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttaataactgtccgttttgcc
atttcttgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagaggggaaaatatcaatgccgaatgctgaGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACgatggtacaaaatcatttgttggtactgatGTTTTAAAGCTGTGCTGTTATTATG
CTAGGGCACCAttgtggtgttctagtttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagatt
aattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatat
tggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatg
tgtcgataaaagagatatgaaaggctataattccaaccttatggttaaagggcaagatgtctacgctggaccttaataatagg
aacgc

Woman 79

Isolate 439

ctttcttgataggcgaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgaga
cgaatggcgcgattacgaaagctcaaaagaaaattttctacgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgg
aaagatggcagaggggttaacgcagcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttttgagttttcaacttatag
agtgtatttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatctggaagcggtcaagaaagcccagtatgGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACacggcacttttaaccgtctctccgcttcttGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACatggtcagaaatcgaagaaaacgaagtcgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaactaa
acaaacgcattgtgtttttagcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcctctatcgaggtttcttacggg
tagggtcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgacgaatttagataacaaaaaagccactgtGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACatggtcagaaatcgaagaaaacgaagtcgtGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACCttatgggagttcttgcaatattagttgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgtaacccgt
cttttccatcagtcccatcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttggaaaaaacacgaaagtgatatta
ctttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtaagaattttagatactctacttgaatgctGTTTTAGAGCTGT
GCTGTTTCGAATGGTTCCAAAACaagacttaaaatcgattagaattgattttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTC
CAAAACtcaagatgataatattttaacgcttctaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccagtttagaa
acagaacaaaactaacgaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagacaaagaagatggcaagtctatcaac
aaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagtagaaacataacgataattccatgaataGTTTTAGAGCTGTGC
TGTTTCGAATGGTTCCAAAACatgttgtagaatcatatcgaccatataaccGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCA
AAACcaattgattgccgttaaaaccgatagaggaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatcttcttttgac
ctaacaaaaggatatgtGTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactga
aataaattttcagagaatgtgggggaaggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacg
aaatcaccccaatttatcagtttgattgggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtcta
ttagtagaaaattcgaattgtttttcggattcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccctta
tggttaaagggctaggtgtctacgccg

Isolate 440

aggcgaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatggcgc
gattacgaaagctcaaaagaaaattttctacgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttggaaagatggca
gaggggttaacgcagcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttttgagttttcaacttatagagtgtatttG
TTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatctggaagcggtcaagaaagcccagtatgGTTTTAGAGCTGTGCTGT
TTCGAATGGTTCCAAAACacggcacttttaaccgtctctccgcttcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAA
CatggtcagaaatcgaagaaaacgaagtcgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaactaaacaaacgcat
tgtgtttttagcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcctctatcgaggtttcttacgggtagggtcGTT
TTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgacgaatttagataacaaaaaagccactgtGTTTTAGAGCTGTGCTGTTT
CGAATGGTTCCAAAACatggtcagaaatcgaagaaaacgaagtcgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACc
gctatctcttgtactggtttcggttctgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcttatgggagttcttgca
atattagttgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgtaacccgtcttttccatcagtcccatcttGTTTTA
GAGCTGTGCTGTTTCGAATGGTTCCAAAACttggaaaaaacacgaaagtgatattactttGTTTTAGAGCTGTGCTGTTTCGA
ATGGTTCCAAAACtaagaattttagatactctacttgaatgctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaaga
cttaaaatcgattagaattgattttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcaagatgataatattttaac
gcttctaaaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccagtttagaaacagaacaaaactaacgaaGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACagacaaagaagatggcaagtctatcaacaaGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACagtagaaacataacgataattccatgaataGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatgttg
tagaatcatatcgaccatataaccGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaattgattgccgttaaaaccga
tagaggaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACatcttcttttgacctaacaaaaggatatgtGTTTTAGAGC
TGTGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaa
ggcggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgatt
gggaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcg
gattcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggtaaagggctaggtgtctacgc

Woman 88

Isolate 476

gaatagtaagctaatttcttgataggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactg
aagtcctgctgagacgaatggcgcgattacgaaagctcaaaagaaaattttctacgagGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACtatctacgagacgaaaacaattgccaatgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagttat
caatgctatatggtttctcatcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgtttaattttgaataatatcatg
tttactcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACggcgaacttgtaacgtttgacggtaattttGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACtaatgtcttttcctagaatgtacatttctaGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACattactttcaagatgtctatgactatatgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtacttgac
gaattgaagatgacggaatttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggttatacatttactaatccatca
gcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCTG
TGCTGTTATTATGCTAGGACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaagg
cggtaattagattaattcaagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgg
gaaacttatatattggcgactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcgga
ttcaaagagaatgtgtcaataaaagagatatgaaaggctataattccaaccttatggttaaagggctaggtgtctacgcttac
taatattttggaccc

Isolate 477

gttaggataagctttttttcttgaaggcgaaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactg
aagtcctgctgagacgaatggcgcgattacgaaagctcaaaagaaaatttttctacgagGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACtatctacgagacgaaaacaattgccaatgtGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACagttat
caatgctatatggtttctcatcttGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgtttaattttgaataatatcatg
tttactcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACgtttaattttgaataatatcatgtttactcGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACggcgaacttgtaacgtttgacggtaattttGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACtaatgtcttttcctagaatgtacatttctaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattacttt
caagatgtctatgactatatgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtacttgacgaattgaagatgacgga
atttaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggttatacatttactaatccatcagcattGTTTTAGAGCTG
TGCTGTTTCGAATGGTTCCAAAACaagctaattctcatctcaccgagatggataGTTTTAGAGCTGTGCTGTTATTATGCTAG
GACATCAttgtggtgttctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaatt
caagacgtaattcagaacttagttggccaagctaacgaaatcaccccaatttatcagtttgattgggaaacttatatattggc
gactaaaaaatatgaacgtcatttagaggtgtgtctattagtagaaaattcgaattgtttttcggattcaaagagaatgtgtc
aataaaagagatatgaaaggctataattccaaccttatggttaaagggctaggtgtctacgctttacttaatttttttgaacc

Woman 97

Isolate 536

gataggcgaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagacgaatgg
cgcgattacgaaagctcaaaagaaaatttttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggaattcaa
agcaaaagcgcttatgaaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgctagacttgaaaaagacagttcaga
catGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattaaaacctcccattatagctctacccgcGTTTTAGAGCTGTG
CTGTTTCGAATGGTTCCAAAACacaacaatattgtcaataagtaccaacccgGTTTTAGAGCTGTGCTGTTTCGAATGGTTCC
AAAACtaataactaaaaattccacgcttgatttagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcaaaggctatt
gataaagctgtatttaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccaaattttgaagagaaaaaactaaaaac
tGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaaagaactttgtgtctcaatttgtgagctGTTTTAGAGCTGTGCT
GTTTCGAATGGTTCCAAAACatcacttctttcgcttcttcaatattcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAA
AACtcttgataatgaatttgagtcagcaatacaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcgctaaagcttgcc
ctaaagttactgttgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACTagatttctacgaacaagtggattgagaaaG
TTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcagtcagttaaagacaccgttattaataaGTTTTAGAGCTGTGCTGT
TTCGAATGGTTCCAAAACttgacctttacgttaacagtccattttgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAA
CaacctcattattaaatttcactactactgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtctttaaagcagttg
atatcaagacatctGTTTTAGAGCTGTGCTGTTATTATGCTAGGGCACCAttgtggtgttctagtttttgttatactgaaat
aaattttcagagaatgtgggggaaggcggtaattagattaattcaagatgtaattcagaacttagttggccaagctaacgaaa
tcaccccagtttatcagtttgattggaaacttatatattggcgactaaaaaatatggacgtcatttagaggtgtgtctattag
tagaaaattcgaattatttttcggattcaaagagaatgtgtcgatagaagagatacgaaagactataattccaaccttatggt
taaagggctaggtgtctacgcttggccaattatttggaacgc

Isolate 537

tttcttgataggcgaatatggtataatattagtaaaagcacagtaataacaaggaatcatcgaaactgaagtcctgctgagac
gaatggcgcgattacgaaagctcaaaagaaaatttttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgga
attcaaagcaaaagcgcttatgaaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgctagacttgaaaaagacag
ttcagacatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattaaaacctcccattatagctctacccgcGTTTTAGA
GCTGTGCTGTTTCGAATGGTTCCAAAACacaacaatattgtcaataagtaccaacccgGTTTTAGAGCTGTGCTGTTTCGAAT
GGTTCCAAAACtaataactaaaaattccacgcttgatttagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcaaag
gctattgataaagctgtatttaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccaaattttgaagagaaaaaact
aaaaactGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaaagaactttgtgtctcaatttgtgagctGTTTTAGAGC
TGTGCTGTTTCGAATGGTTCCAAAACatcacttctttcgcttcttcaatattcattGTTTTAGAGCTGTGCTGTTTCGAATGG
TTCCAAAACtcttgataatgaatttgagtcagcaatacaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcgctaaag
cttgccctaaagttactgttgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcagtcagttaaagacaccgttatt
aataaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgacctttacgttaacagtccattttgaGTTTTAGAGCTG
TGCTGTTTCGAATGGTTCCAAAACaacctcattattaaatttcactactactgcGTTTTAGAGCTGTGCTGTTTCGAATGGTT
CCAAAACtcttttaaagcagttgatatcaagacatctGTTTTAGAGCTGTGCTGTTATTATGCTAGGGCACCAttgtggtgtt
ctagttttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagatgtaattcagaa
cttagttggccaagctaacgaaatcaccccagtttatcagtttgattggaaacttatatattggcgactaaaaaatatggacg
tcatttagaggtgtgtctattagtagaaaattcgaattatttttcggattcaaagagaatgtgtcgatagaagagatacgaaa
gactataattccaaccttatggttaaagggctaggtgtctacgcttagctaatatt

Isolate 538

gataggcgaatatggtataatattagtaaaagcacagtaataacaaggaatcacccgaaactgaagtcctgctgagacgaatg
gcgcgattacgaaagctcaaaagaaaatttttctacgagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtggaattca
aagcaaaagcgcttatgaaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtgctagacttgaaaaagacagttcag
acatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACattaaaacctcccattatagctctacccgcGTTTTAGAGCTGT
GCTGTTTCGAATGGTTCCAAAACacaacaatattgtcaataagtaccaacccgGTTTTAGAGCTGTGCTGTTTCGAATGGTTC
CAAAACtaataactaaaaattccacgcttgatttagGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcaaaggctat
tgataaagctgtatttaatGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACccaaattttgaagagaaaaaactaaaaa
ctGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcaaagaactttgtgtctcaatttgtgagctgttttagagctgtgc
tgtttcgAATGGTTCCAAAACatcacttctttcgcttcttcaatattcattGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCA

AAACTcttgataatgaatttgagtcagcaatacaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACcgctaaagcttgc
cctaaagttactgttgaGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACtcagtcagttaaagacaccgttattaataa
GTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAACttgacctttacgttaacagtccattttgaGTTTTAGAGCTGTGCTG
TTTCGAATGGTTCCAAAACaacctcattattaaatttcactactactgcGTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAA
ACtcttttaaagcagttgatatcaagacatctGTTTTAGAGCTGTGCTGTTATTATGCTAGGGCACCAttgtggtgttctagt
tttttgttatactgaaataaattttcagagaatgtgggggaaggcggtaattagattaattcaagatgtaattcagaacttag
ttggccaagctaacgaaatcaccccagtttatcagtttgattggaaacttatatattggcgactaaaaaatatggacgtcatt
tagaggtgtgtctattagtagaaaattcgaattatttttcggattcaaagagaatgtgtcgatagaagagatacgaaagacta
taattccaaccttatggttaaagggctaggtgtctacgc

II. Etude de la diversité génétique des souches de *S. agalactiae* au sein d'une population de portage vaginal.

A. Etude de la diversité génétique par CRISPR/RFLP et séquençage du locus CRISPR1

Publication n°2 : Clémence Beauruelle, Adeline Pastuszka, Laurent Mereghetti and Philippe Lanotte. **Group B Streptococcus vaginal carriage in pregnant women deciphered by CRISPR analysis.** (2018) Journal of Clinical Microbiology (IF: 3,712) doi: 10.1128/JCM.01949-17.

L'étude précédente a été menée à partir de quelques colonies conservées au moment de l'ensemencement du prélèvement. Lors de l'isolement des souches à partir d'un prélèvement, seulement quelques colonies sont conservées. Cette approche ne nous permet pas d'avoir une vision exhaustive de la population bactérienne à un temps donné. La présence d'une double population ne peut donc pas être observée par cette approche. La répétition des prélèvements dans cette première étude a permis dans certains cas de mettre en évidence l'existence de deux souches différentes pour une même femme. La souche étudiée étant possiblement la souche majoritaire à un temps donné, la première souche isolée pouvait être présente lors du deuxième prélèvement mais n'aurait pas été étudiée. D'autre part, la diversité clonale au sein des isolats provenant d'un même prélèvement n'est pas étudiée dans les laboratoires de bactériologie médicale en routine. Pourtant, les analyses bactériologiques, comme l'antibiogramme ou le typage des souches, ne sont réalisées qu'à partir d'une ou de quelques colonies qui ne reflètent peut-être pas la population globale. Quelques travaux se sont précédemment intéressés à la diversité du portage de SGB mais ne permettaient l'analyse que de quelques souches par prélèvement. Suite à ce constat, nous avons voulu estimer la

diversité de la population de SGB présente en portage pour une patiente donnée et à un temps donné.

Cette étude avait pour objectifs :

i) D'obtenir un aperçu d'une population de *S. agalactiae*, de la diversité des clones et de la cohabitation des différentes lignées dans un environnement défini à un même temps de portage.

ii) D'évaluer une approche de typage basée sur le locus CRISPR1, rapide et adaptée à l'étude d'un nombre important de souches : l'amplification et la restriction enzymatique du locus CRISPR1 (CRISPR-RFLP).

iii) de confirmer l'intérêt du locus CRISPR1 comme outil de typage, notamment par comparaison avec des méthodes classiquement utilisées pour le typage des souches de SGB : typage capsulaire et MLST.

Dans cette étude, toutes les colonies isolées en primo-culture sur une gélose sélective (Granada®, bioMerieux) après ensemencement d'un écouvillon vaginal étaient analysées.

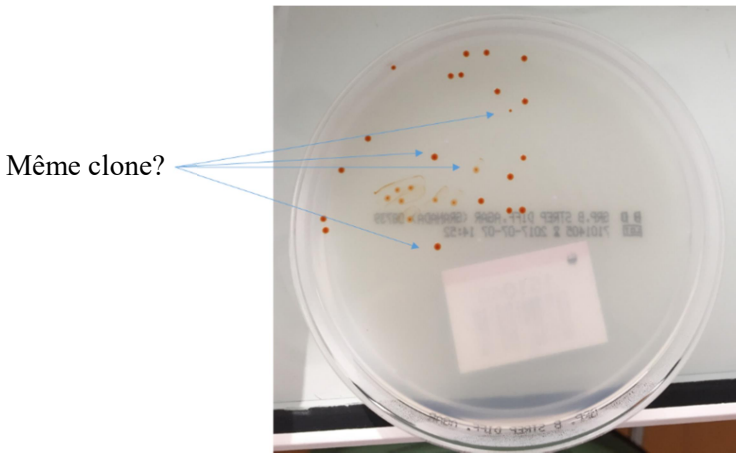


Figure 31. Culure d'un prélèvement vaginal de dépistage de SGB sur gélose Granada®.
Toutes les colonies présentes sur la gélose ont été réisolées et analysées.

Les prélèvements correspondaient à des écouvillons vaginaux prélevés chez dix femmes dans le cadre de leur dépistage systématique de SGB. Un total de 970 isolats étaient analysés. Afin de caractériser ce nombre important d'isolats, nous avons effectué une étape de « screening » par amplification et restriction enzymatique du locus CRISPR1 (CRISPR-

RFLP). Dans un 1er temps, pour chacun des isolats, les loci CRISPR1 étaient amplifiés. Une restriction était ensuite effectuée par deux enzymes sélectionnées en fonction des séquences des loci CRISPR1 préalablement obtenues. L'objectif était d'obtenir un équilibre entre un nombre suffisant de coupures pour permettre une discrimination entre les isolats et un nombre pas trop important pour permettre une interprétation aisée des profils (Tableau 6)

Tableau 6. Sélection des enzymes de restriction.
Exemple de la restriction du locus CRISPR1 par l'enzyme DdeI *pour une souche isolée chez la patiente W2.*

Site de coupure	Taille des fragments théorique	Profil théorique	Profil obtenu
303/306 293 TGAAGTCCTG C TGA▲G ACGAATGGCG 724/727 714 AAAACATTAT C TTA▲G AAATACTTCA 1076/1079 1066 TAATTCAGAA C TTA▲G TTGGCCAAGC	# / Coordinates / Length (bp) 1 / 1077-1526 / 450 2 / 304-724 / 421 3 / 725-1076 / 352 4 / 1-303 / 303	10000 5000 1000 500 100	DdEI 100pb A1 P2 A2 //

Les loci CRISPR1 présentant un profil de restriction différent étaient séquencés et les souches correspondantes étaient typées par les méthodes usuelles : le typage capsulaire et le MLST.

Ces données nous ont permis :

i) d'obtenir des informations sur une population vaginale de** S. agalactiae **chez les femmes en fin de grossesse.

ii) d'évoquer un rôle du microbiote dans la sélection des lignées phylogénétiques.

En effet, nous avons pu mettre en évidence :

- Un portage relativement uniforme des clones de *S. agalactiae* pour une patiente / un microbiote vaginal donné. Un seul clone était retrouvé pour 8 femmes sur 10 avec un maximum de 2 clones portés simultanément. Les clones présents simultanément présentaient un fond génétique commun (même spacers ancestraux au sein du locus CRISPR1, même sérotype capsulaire voire même ST (un cas sur deux), faisant suggérer un impact du microbiote vaginal sur la sélection des souches de portage. Cependant, compte-tenu de la taille limitée de l'échantillon, ce point reste à vérifier sur des populations plus grandes.

iii) de confirmer l'intérêt du locus CRISPR1 comme outil de typage pour l'étude de la diversité clonale des souches de portage, notamment par une approche de restriction enzymatique.

En effet,

- Bien que la diversité des clones soit limitée pour chaque individu, elle est élevée au sein de la population de l'étude ; chaque patiente présentant un à plusieurs profils de restriction qui lui est (sont) propre(s).

- La méthode de typage développée ici permet une comparaison rapide de la composition des loci CRISPR1, particulièrement adaptée à l'étude d'un nombre important de souches. Cette méthode de screening avant séquençage du locus CRISPR1 paraît adaptée pour distinguer les isolats, notamment ceux appartenant à un même ST.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

EPIDEMIOLOGY

Group B *Streptococcus* Vaginal Carriage in Pregnant Women as Deciphered by Clustered Regularly Interspaced Short Palindromic Repeat Analysis

Clemence Beauruelle,[a,b,c] Adeline Pastuszka,[a,b,c] Laurent Mereghetti,[a,b,c] Philippe Lanotte[a,b,c]

[a]Université François Rabelais de Tours, UMR1282 Infectiologie et Santé Publique, Tours, France
[b]INRA, UMR1282 Infectiologie et Santé Publique, Nouzilly, France
[c]CHRU de Tours, Service de Bactériologie-Virologie, Tours, France

ABSTRACT We evaluated the diversity of group B *Streptococcus* (GBS) vaginal carriage populations in pregnant women. For this purpose, we studied each isolate present in a primary culture of a vaginal swab using a new approach based on clustered regularly interspaced short palindromic repeats (CRISPR) locus analysis. To evaluate the CRISPR array composition rapidly, a restriction fragment length polymorphism (RFLP) analysis was performed. For each different pattern observed, the CRISPR array was sequenced and capsular typing and multilocus sequence typing (MLST) were performed. A total of 970 isolates from 10 women were analyzed by CRISPR-RFLP. Each woman carrying GBS isolates presented one to five specific "personal" patterns. Five women showed similar isolates with specific and unique restriction patterns, suggesting the carriage of a single GBS clone. Different patterns were observed among isolates from the other five women. For three of these, CRISPR locus sequencing highlighted low levels of internal modifications in the locus backbone, whereas there were high levels of modifications for the last two women, suggesting the carriage of two different clones. These two clones were closely related, having the same ancestral spacer(s), the same capsular type and, in one case, the same ST, but showed different antibiotic resistance patterns in pairs. Eight of 10 women were colonized by a single GBS clone, while two of them were colonized by two strains, leading to a risk of selection of more-virulent and/or more-resistant clones during antibiotic prophylaxis. This CRISPR analysis made it possible to separate isolates belonging to a single capsular type and sequence type, highlighting the greater discriminating power of this approach.

KEYWORDS CRISPR. *Streptococcus agalactiae*, carriage, genotyping, diversity

Streptococcus agalactiae, or group B *Streptococcus* (GBS), is a major pathogen in human medicine, particularly in neonatology, where this bacterium is still the leading cause of invasive infections in industrialized countries (1). Prevention of neonatal infection associated with GBS requires the detection of vaginal carriage in pregnant women, followed by intrapartum antibiotic prophylaxis for those colonized (2, 3). Indeed, GBS colonizes the gastrointestinal and genital tracts of 10 to 30% of healthy women (4–6).

GBS isolates were initially distinguished on the basis of differences in capsular polysaccharides (CPS), leading to the definition of 10 different serotypes (Ia, Ib, and II to IX) (7, 8). While this typing method highlighted pathogenicity differences among serotypes (serotype III GBS strains are strongly associated with neonatal invasive infections [9, 10]), it was suboptimal in its discriminatory capacity for studying the epidemiology of GBS infection. Thus, several higher-throughput and higher-resolution subtyping methods were developed, such as multilocus sequence typing (MLST) and

Received 13 December 2017 **Returned for modification** 28 January 2018 **Accepted** 28 March 2018

Accepted manuscript posted online 4 April 2018

Citation Beauruelle C, Pastuszka A, Mereghetti L, Lanotte P. 2018. Group B *Streptococcus* vaginal carriage in pregnant women as deciphered by clustered regularly interspaced short palindromic repeat analysis. J Clin Microbiol 56:e01949-17. https://doi.org/10.1128/JCM.01949-17.

Editor Alexander Mellmann, University Hospital Münster

Copyright © 2018 American Society for Microbiology. All Rights Reserved.

Address correspondence to Philippe Lanotte, philippe.lanotte@univ-tours.fr.

multilocus variable-number tandem repeat (VNTR) analysis (MLVA) (11–13). Based on the combination of alleles for seven housekeeping genes, MLST defines sequence types (ST), which can be clustered in clonal complexes (CC) when six of the seven alleles are common. Based on the number of repeats at each of the 6 VNTR loci, MLVA defines a profile consisting of a series of allele numbers.

Clustered regularly interspaced short palindromic repeats (CRISPR) loci consist of short (25- to 40-bp) direct repeats (DR) that are highly conserved within a given CRISPR array, interspaced by various sequences of similar size, called spacers, which are derived from foreign genetic elements. CRISPR and CRISPR-associated (Cas) proteins form the CRISPR-Cas system, which provides defense against mobile genetic elements (MGE) such as phages or plasmids (14). This immune system is adaptive and confers specific and acquired immunity based on short nucleic acid sequence integration from the invaders. A core feature of the CRISPR-Cas system is the ability to acquire a novel spacer in a specific manner at the leader end of the locus, indicating exposure to invasive elements over time. Spacers at the leader end were recently acquired and provide insight into recent MGE exposure. In contrast, ancestral spacers at the trailer end of the CRISPR array provide information about the phylogenetic lineage and are correlated with phylogenetic groups defined by other typing methods. For bacterial phylogenetic analysis, CRISPR-based typing appears to be a practical and powerful alternative (15–17). Subtyping using CRISPR has been employed for other bacterial species such as *Yersinia* and *Salmonella* (18, 19). In GBS, a type II-A system, associated with the CRISPR1 locus, is ubiquitous and functional. The similarity between CRISPR1 spacer sequences and MGE sequences has been previously emphasized (20, 21). Comparative sequence analysis across numerous isolates has revealed that this locus is extremely diverse and evolves *in vivo*, reflecting the dynamics of the system (20–22).

The aim of this study was to determine the homogeneity of GBS populations in pregnant women at the time of their GBS screening. For this purpose, we used an original approach based on a CRISPR array study to analyze all the isolates present in a primary culture from a vaginal swab. To rapidly evaluate the composition of the spacers, we screened isolates by restriction fragment length polymorphism (RFLP) of the CRISPR1 array and, when different restriction patterns were obtained, CRISPR1 loci were sequenced and isolates typed using the most common GBS typing methods (capsular typing and MLST). This method enabled us to compare numerous isolates and to target potentially different clones with the high discriminatory power of CRISPR analysis.

MATERIALS AND METHODS

Bacterial isolates. Isolates were collected between July and September 2016 from the department of obstetrics, gynecology, and fetal medicine of the Tours University Hospital, Tours, France. We randomly selected vaginal swabs from 10 pregnant women. GBS specimen collection was conducted during the late third trimester (35 to 37 weeks' gestation), observing the French recommendations on screening for GBS carriage (2). The protocol was reviewed and approved by the Ethics Committee in Human Research of the University Hospital Center of Tours (Tours, France). We selected plates with at least 20 and up to 200 isolates for practical reasons. Each swab was spread on a selective agar plate (group B *Streptococcus* differential agar, Granada medium; BD) and incubated for 18 to 48 h at 35 ± 2°C under anaerobic conditions according to the manufacturer's recommendations. Bacterial identification was based on growth and orange pigment formation on Granada agar and confirmed by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS; Microflex; Bruker Daltonics). Each isolate present on the agar plate was reisolated and stored at −80°C for the duration of the study and grown secondarily on TSH agar (Trypticase soy agar supplemented with 5% horse blood; bioMérieux).

DNA extraction. Genomic DNA was extracted following enzymatic lysis with mutanolysin (Sigma). A bacterial suspension of 1.5 McFarland standard was prepared in 250 μl water containing 25 U of mutanolysin. The suspension was incubated at 56°C for 1 h, followed by 10 min at 100°C, leading to cell lysis. Lysates were centrifuged at 1,500 × *g* for 3 min, and the DNA-containing supernatants were collected.

CRISPR1 locus amplification. CRISPR1 locus amplification was performed in a T3000 thermocycler (Biometra) using CRISPR1 SEQ-F and CRISPR1 SEQ-R primers targeting the CRISPR1-flanking regions, as previously described (20). PCR amplification was performed in a total volume of 25 μl, containing 0.5 μM each primer, 0.2 μM deoxynucleoside triphosphate (dNTP), 1.5 mM $MgCl_2$, 0.02 U/μl Go *Taq* polymerase (Promega), 1× PCR buffer, and 1 μl extracted DNA. The PCR mixtures were heated to 94°C for 5 min,

TABLE 1 Primers used for amplification and sequencing

Oligonucleotide name	Used for: (patient ID[a]/isolate)	Sequence (5′→3′)
CRISPR1 seq-F	All patients and isolates	GAAGACTCTATGATTTACCGC
CRISPR1 seq-R	All patients and isolates	CAGCAATCACTAAAAGAACCAAC
Sp803F	W1/All isolates	TCTAATGTATACTGTAACTGAGGCAG
Sp504F	W4/All isolates	TTTATCCCTAGTAGGTT
Sp539F	W8/A2	TCTAAGTGCTCGACCATC
Sp8R	W8/A2	TATCCATCTCGGTGAGATGAGAATTAGCTT
Sp31R	W1/All isolates	ATCAGTACCAACAAATGATTTTGTACCATC
Sp36R	W3/All isolates	ACATATCCTTTTGTTAGGTCAAAAGAAGAT
Sp177R	W9/All isolates	AGTTACTGTTGAGGGTAGTCC
275F	W3/All isolates	ACAGACAAAGAAGATGGCAAG

[a]Patients are identified by "W" followed by a number, i.e., W1 is woman 1.

followed by 40 cycles of a denaturation step at 94°C for 30 s, an annealing step at 58°C for 30 s, and an elongation step at 72°C for 3 min, ending with a final extension step at 72°C for 10 min. PCR amplification was verified by electrophoretic migration in a 1% agarose gel. If amplification did not enable us to restrict the PCR product due to low signal intensity, isolates were amplified using internal primers designed after sequencing and covering the entire CRISPR1 locus (Table 1). Indeed, for these CRISPR arrays, two amplifications were necessary, one targeting the 5′ end and one targeting the 3′ end of the locus.

RFLP. Restriction enzymes were selected using NEBcutter V2.0 (http://nc2.neb.com/NEBcutter2/) based on previously sequenced CRISPR1 arrays. We selected two enzymes with three or four cutters for most of the CRISPR1 arrays tested: DdeI and MnlI. Restriction was performed separately for each enzyme in a T3000 thermocycler (Biometra) according to the manufacturer's recommendations. Restriction products were analyzed by electrophoretic migration in a 1.5% agarose gel. For isolates with CRISPR1 arrays showing different sizes and/or different patterns by RFLP, the CRISPR1 locus was sequenced; MLST and capsular typing were performed.

CRISPR1 locus sequencing and analysis. CRISPR1 locus sequencing was performed as previously described (20). Primers targeting internal spacers were required to complete the sequencing of the PCR products for CRISPR1 regions exceeding 1.3 kb (Table 1).

DNA sequences were analyzed and assembled using the BioEdit sequence alignment editor version 7.2.5 (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Spacers, repeats, and flanking regions for each sequence were determined using a macro-enabled Excel tool (kindly provided by P. Horvath, DuPont). This application was used to identify and extract CRISPR features from nucleotide sequences and provided graphic representations of spacers as colored cells in Excel spreadsheets. Spacer sequences were compared to the previously established dictionary of spacers (20). New spacers identified in this study expanded the dictionary and were numbered incrementally. The spacer dictionary is available online (http://crispr.i2bc.paris-saclay.fr/CRISPRcompar/Dict/Dict.php).

MLST and capsular typing. MLST was carried out as previously described (12). Allelic profiles and ST were assigned using the international MLST database (http://pubmlst.org/sagalactiae/). Clonal complexes were defined using eBurst analysis (http://eburst.mlst.net/v3/mlst_datasets/) (23). Capsular type identification was performed using PCR multiplex assays as described previously (24).

Antibiotic susceptibility testing. Susceptibility to antibiotics was tested and interpreted according to the 2016 EUCAST guidelines using a standardized disk diffusion method (http://www.eucast.org/clinical_breakpoints/). We tested the following 14 antibiotics in an MHF medium (Mueller-Hinton agar, 5% horse blood, 20 mg/liter β-NAD): benzylpenicillin (1 unit), ampicillin (2 μg), norfloxacin (10 μg), levofloxacin (5 μg), erythromycin (15 μg), lincomycin (15 μg), pristinamycin (15 μg), gentamicin (500 μg), tetracycline (30 μg), rifampin (5 μg), trimethoprim-sulfamethoxazole (1.25 to 23.75 μg), vancomycin (5 μg), teicoplanin (30 μg), nitrofurantoin (100 μg). To confirm the differences in antibiotic susceptibility observed by disk diffusion, MICs were determined for some isolates by Etest (bioMérieux) for erythromycin and clindamycin.

RESULTS

Population studied. A total of 970 isolates were recovered from 10 GBS-colonized women during their last month of pregnancy. The average age of the cohort was 31 years at sampling time (median, 31 years; minimum, 25 years; maximum, 38 years). Vaginal swabs were taken from each woman and rapidly seeded to agar plates. All GBS isolates present on these plates were then analyzed. An average of 97 isolates per woman was considered (median, 108; minimum, 20; maximum, 164) (Table 2), illustrating differences in bacterial carriage density.

Culture morphology. No phenotypic differences were observed between isolates from any given woman except for one (woman 10), from whom 125 isolates were present on the plate. Among them, 33 isolates were more pigmented on Granada selective agar and more hemolytic on blood agar than the other 92 isolates. These

TABLE 2 Characteristics of the cohort study and results of CRISPR-RFLP using DdeI[a]

Woman ID (age in yr)	No. of isolates	No. of isolates (%) with pattern no.:				
		1	2	3	4	5
W1 (35)	164	162 (99)	2 (1)			
W2 (34)	130	130 (100)				
W3 (32)	48	47 (98)	1 (2)			
W4 (27)	64	64 (100)				
W5 (25)	144	144 (100)				
W6 (26)	121	121 (100)				
W7 (28)	60	60 (100)				
W8 (32)	20	13 (65)	7 (35)			
W9 (38)	97	92 (95)	2 (2)	1 (1)	1 (1)	1 (1)
W10 (29)	125	92 (74)	33 (26)			

[a]A total of 970 isolates were studied. Each pattern was specific for a given woman. Five women carried GBS isolates with a single pattern. The other five carried GBS isolates with several patterns (up to five different patterns). The same distribution was observed with MnlI.

observed colony pigmentation differences may be explained by the level of production of the orange carotenoid pigment, which is tightly regulated.

CRISPR1 amplification and RFLP analysis. For 922 isolates, amplification was performed using CRISPR1 seq-F and CRISPR1 seq-R primers. For 48 isolates from one woman (woman 3), DNA was amplified using internal primers due to low signal intensity. Indeed, two amplifications were necessary for these isolates, and restriction was performed for the two amplicons (96 PCR products, 192 restrictions). CRISPR1 locus sizes were evaluated after PCR amplification and compared for each isolate from the same woman. For 5 of the 10 women, all isolates presented a CRISPR1 locus of the same size (women 2, 4, 5, 6, and 7). For the other five women, a difference in CRISPR1 locus size was observed. Differences were observed for 1% (two isolates, woman 1) to 35% (seven isolates, woman 8) of isolates.

For each isolate, we performed a CRISPR-RFLP with the restriction enzymes DdeI and MnlI. The two restrictions proceeded and were analyzed separately. The resulting restriction fragments were separated according to length by gel electrophoresis. Among all the GBS isolates, a total of 18 different gel band patterns were detected by CRISPR-RFLP. All women studied carried one to five GBS isolates with specific gel band patterns. Each gel band pattern was specific to the isolates from a given woman.

For five women, no differences were observed after CRISPR enzymatic restriction (Table 2). In contrast, different restriction patterns were observed for isolates from the other five women: minor patterns represented 1% (2/164 isolates, woman 1) to 35% (7/20 isolates, woman 8) of all isolates. The two enzymes used detected the same differences. For four of these five women, only two patterns were observed, suggesting the presence of two different clones. For the last woman (woman 9), five patterns were observed, suggesting the presence of five different clones. Nonetheless, for all women, one clone was predominant, representing at least 65% of isolates.

Sequencing. All CRISPR1 loci showing different restriction patterns were sequenced and analyzed (18 isolates, 10 women) (Fig. 1). Each woman carried GBS, characterized by a specific CRISPR array. For three of the five women whose isolates showed CRISPR variations by CRISPR-RFLP (women 1, 3, and 9), variations concerned a single CRISPR type. Indeed, spacers were acquired or deleted in the middle part of the locus, whereas ancestral and more recently acquired spacers were conserved. For woman 1, isolates presented a CRISPR1 locus with one repetition of seven spacers in 99% of isolates (162/164) and two repetitions in 1% of isolates (2/164, e.g., isolates W1 E2). For woman 3, the differences in restriction patterns were due to a middle locus deletion of three spacers for one isolate (isolate W3 C6). For woman 9, four isolates presented a middle locus deletion of one (isolate W9 F1), two (isolates W9 A2), or 14 (isolates W9 A4 and W9 A5) spacers (Fig. 1). For two of the five women with different CRISPR-RFLP, isolates had two different CRISPR types (women 8 and 10). The CRISPR1 locus, however,

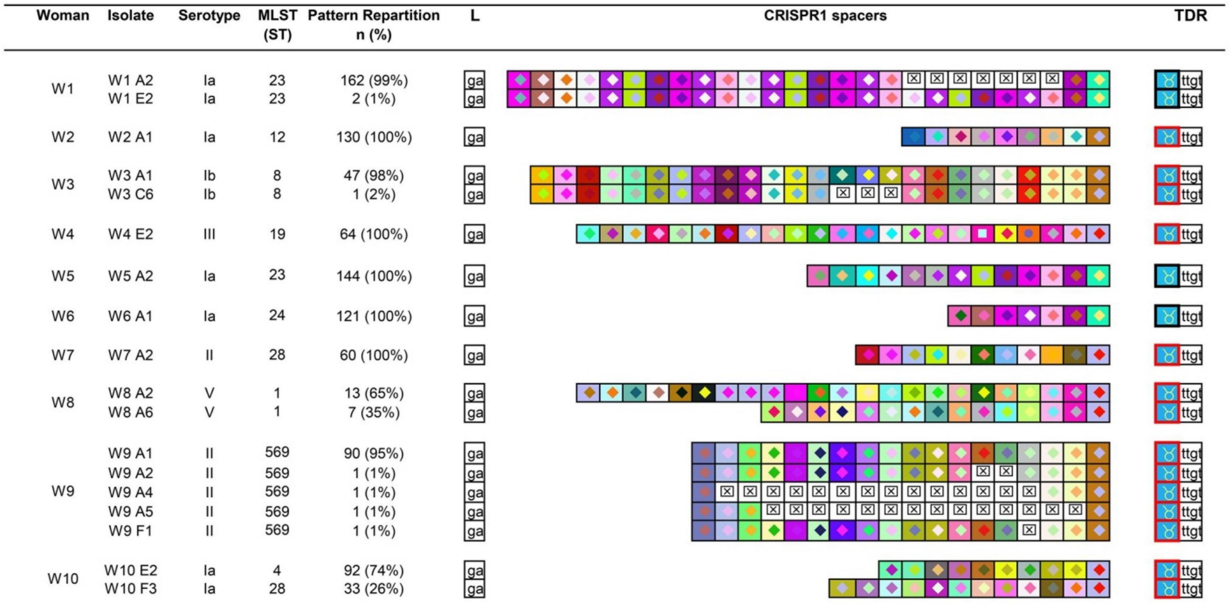


FIG 1 Spacer composition of CRISPR1 arrays for GBS isolates from 10 women. Only one array is represented by different patterns observed after RFLP. The CRISPR1 arrays are represented using a macro-enabled Excel tool whereby spacers are converted into two-color symbols based on the spacer sequence. Gaps (indicating missing spacers) are shown with a boxed "×" symbol after alignment of identical spacers between strains of the same group. Terminal direct repeats (TDRs) are represented by different color borders according to their sequence. The leader sequence (L) is represented by its last two nucleotides. For each woman, the main CRISPR array was represented first. Each woman presented a specific CRISPR array. Three of the 10 women carried isolates belonging to a single CRISPR type with internal modification of the CRISPR1 array (women 1, 3, and 9). Two of the 10 women carried isolates with different CRISPR1 arrays, indicating the presence of two clones (women 8 and 10).

presented common ancestral spacer(s) among all isolates (four [woman 8] and one [woman 10]).

Antibiotic susceptibility testing. Antibiotic susceptibility profiles were determined for the two women with different CRISPR type isolates as a supplementary phenotypic character. Various susceptibilities were observed for the two clones. Isolates from woman 8 were resistant to erythromycin and susceptible to lincomycin (inhibition zone diameters, 17 mm and 29 mm, respectively; erythromycin MIC, 1 mg/liter; clindamycin MIC, 0.125 mg/liter) for the major clone, whereas the minor clone was resistant to both erythromycin and lincomycin (inhibition zone diameter, 7 mm for both; erythromycin MIC, >256 mg/liter; clindamycin MIC, >256 mg/liter). Concerning woman 10, isolates were resistant to both erythromycin and lincomycin (inhibition zone diameters, 17 mm and 15 mm, respectively; erythromycin MIC, 4 mg/liter; clindamycin MIC, 0.125 mg/liter) for the major clone and susceptible to both erythromycin and lincomycin (inhibition zone diameters, 28 mm and 30 mm, respectively; erythromycin MIC, 0.064 mg/liter; clindamycin MIC, 0.064 mg/liter) for the minor clone.

Capsular typing and MLST. Capsular typing and MLST were performed for each isolate showing a different CRISPR-RFLP pattern (Fig. 1). Among isolates belonging to the 12 different CRISPR types, 10 could not be distinguished by capsular typing (isolates from women 1, 2, 5, and 6 and the two CRISPR types isolated from woman 10 [type Ia]; from women 7 and 9 [type II] and the two CRISPR types isolated from woman 8 [type V]). Six could not be distinguished by MLST (isolates from women 1 and 5, which shared the same ST, ST23; isolates from woman 7 and from the minor clone from woman 10, sharing ST28, and the two CRISPR types isolated from woman 8 with the same ST, ST1). All isolates from a given woman presented the same capsular type and the same ST, except those from woman 10, which presented two different STs among all isolates tested (Fig. 1).

Most isolates belonged to capsular types Ia (five women, corresponding to 684 isolates) and II (two women, corresponding to 154 isolates). Capsular types Ib, III, and

V, respectively, were isolated from the three other women, corresponding to 132 isolates. Concerning MLST analysis, isolates belonged to CC23 for three women (women 1 and 5 [ST23] and woman 6 [ST24], corresponding to 429 isolates), to CC10 for three women (women 2 [ST12], 3 [ST8], and 9 [ST569], corresponding to 272 isolates), to CC19 for three women (woman 7 and a minor clone from woman 10 [ST28] and woman 4 [ST19], corresponding to 157 isolates), and CC1 for two women (woman 8 [ST1] and a major clone from woman 10 [ST4], corresponding to 112 isolates).

DISCUSSION

The objective of our study was to explore GBS phylogenetic diversity in a vaginal population taken from pregnant women at the time of their screening, specifically at the end of pregnancy. Given the previous data on CRISPR typing analysis (21, 22), the strategy used was based on CRISPR array analysis. We developed an original approach to compare CRISPR composition based on a CRISPR-RFLP analysis to screen GBS isolates, followed by sequencing and analysis of the CRISPR array.

In our study, GBS carriage in pregnant women was divided into three categories based on CRISPR1 locus diversity. For one-half of the women (women 2, 4, 5, 6, and 7), all isolates shared the same CRISPR1 locus, demonstrating homogenous carriage.

In the second category, isolates shared a CRISPR1 locus with the same backbone (women 1, 3, and 9), but some isolates had a CRISPR array with some minor variations. These internal CRISPR1 locus modifications were spacer deletions (women 3 and 9) or duplications (woman 1). This phenomenon appeared rarely and was probably the result of slipped misalignment during DNA replication or homologous recombination (20, 22, 25). This phenomenon concerned up to 5% of isolates (woman 9), with various numbers of spacers (1 to 14 spacers) but never the spacers located at the CRISPR1 array extremities. Thus, the CRISPR locus appears to evolve *in vivo* mainly through the loss and acquisition of internal spacers. These first two categories consisted of the same type of CRISPR isolates, which probably belonged to the same clone. The same capsular type and the same ST also point to the fact that these were a single clone.

In the latter category, GBS isolates for women 8 and 10 displayed two different CRISPR types. Two clones coexisted with proportions of 65% to 35% (woman 8) and 74% to 26% (woman 10). In this group, the two clones from each woman shared similarities among ancestral spacers: four for woman 8 and one for woman 10. Moreover, for woman 8, the two clones detected shared the same serotype (serotype V) and the same ST (ST1), suggesting a common ancestor. For woman 10, the two clones belonged to two different STs (ST4 and ST28), though both were from serotype Ia. Although we analyzed isolates from a small number of women, we observed the favoring presence of a phylogenetic lineage between GBS clones carried simultaneously, as previously suggested by minor differences in pulsed-field gel electrophoresis (PFGE) patterns ($\leq$2 bands) (26). We can hypothesize that a specific vaginal environment promoted the development of these lineages. To our knowledge, no link between phylogenetic group and vaginal environment has been described. However, the "co-carriage" of several clones might represent an advantage for improving adaptation and virulence by genetic exchange, even more so with closely related clones, which could promote horizontal gene transfer (27–30). Moreover, bacteriophages may modify the equilibrium between several clones, eventually leading to the elimination of some of them according to the *S. agalactiae* CRISPR-Cas system present in the phage-infected isolate. For women 8 and 10, the two clones presented phenotypic differences in hemolytic activity and antimicrobial susceptibilities (GBS isolates from woman 10) and in antimicrobial susceptibilities (GBS isolates from woman 8). We note that genetic diversity was not linked to richness of carriage. For woman 8, who carried two clones simultaneously, only 20 isolates were obtained on the agar plate, whereas 125 isolates were present for woman 10.

Although clonal diversity was limited for each individual person, it was high for the entire cohort. Each woman presented a "personal" profile of GBS carriage strains. Nevertheless, the study was conducted on just 10 women. In contrast, the previously

described PFGE and randomly amplified polymorphic DNA (RAPD) analyses failed to distinguish profiles for some isolates from different women (26, 31, 32). In our study, the CRISPR-RFLP approach appears to be more discriminant than serotyping and MLST. Among the 12 different CRISPR types, 10 and 6 isolates could not be distinguished by capsular typing and MLST, respectively (Fig. 1).

The isolates studied belonged to five different capsular types (Ia, Ib, II, III, V) and to four CCs (CC23, CC10, CC19, and CC1). These serotypes are widely represented and account for more than 90% of GBS isolates worldwide (12, 33–36). Serotype Ia was particularly frequent (present in one-half of the women), whereas serotype III, associated with CC17, which is one of the most prevalent serotypes in carriage among pregnant women, was isolated for just one woman (37–39). This could be explained by the limited number of women studied and the selection of plates with intermediate to heavy carriage (at least 20 isolates), while serotype III isolates appear to be involved in "light" vaginal carriage mainly (37). Concerning MLST analysis and CC clustering, the isolates belonged to four CCs (CC23, CC10, CC19, and CC1), which are the predominant CCs among colonizing strains during pregnancy (35, 36). No ST17 strains (associated with serotype III) were isolated (12, 33, 34).

The three typing methods used in our study (capsular typing, MLST, and CRISPR typing) showed that the GBS population was relatively uniform in all the women examined, with the same capsular type for each carrier, the same ST for all but one, the specific CRISPR type for 8 of the 10 carriers, and no more than two clones per carrier. Previous typing approaches performed to evaluate GBS diversity in carriage have given similar results, with more than two-thirds of GBS carriers colonized by a single clone (same serotype [96%] and the same PFGE profile [86 to 90% in the same sample] with no more than three clones per carrier [26, 31, 40, 41]). Using RAPD analysis, a higher genotypic diversity per individual was reported (same genotype for 39% of carriers with up to four clones per carrier) (32). While the use of broth could explain part of this diversity, RAPD is now notoriously technically dependent and difficult to reproduce, which might explain part of this variability. These studies compared only a few selected isolates on agar plates from vaginal and rectal swabs (5 to 26 isolates) (26, 31, 40, 41). Our study reinforces previous results on the diversity of GBS carriage with the analysis of a large number of isolates per sample. Here, we considered all the isolates present on the agar plates; thus, we analyzed a higher number of colonies per woman (about 100 on average) than did previous studies. To more precisely characterize the genetic links between isolates, however, the whole-genome approach appears to be suitable. Although this approach remains expensive for large studies, it could be helpful in studying clones, especially those from women 8 and 10.

One of the difficulties of our approach was the choice of restriction enzymes because of the high polymorphism of the spacer sequences. The enzymes selected had to provide sufficient restriction to separate isolates but not so much as to be detectable. Here, the two enzymes selected separated the considered CRISPR array into three to six fragments. However, two restriction sites were located in conserved CRISPR flanking regions and did not provide supplemental information about locus composition. The use of two different enzymes improved the discriminatory power of this approach while remaining useful and affordable.

In routine clinical laboratory practice, bacteriological analyses such as antibiotic sensitivity testing or typing are often performed on a single or a very small number of isolates. The isolates selected, however, may not represent the entire population, as several clones may coexist. For two women studied, two clones coexisted and shared paired differences in antibiotic susceptibility. Although the two clones remained susceptible to beta-lactam antibiotics, which are the first-line antibiotics, they had different susceptibilities to the macrolides and lincosamides tested, which are used as alternative drugs for patients allergic to penicillin.

Enzymatic restriction makes it possible to perform rapid comparisons of locus CRISPR compositions for large-scale studies. Applied to CRISPR loci, this is a good screening method before CRISPR sequencing for comparing isolates and can be used

to separate isolates belonging to the same ST. Vaginal carriage from pregnant women is relatively homogenous, with a single clone for 8 of the 10 women and at least two closely related clones simultaneously. Along with possibly being a source of selection for resistant strains, cocarriage of GBS clones may contribute to increasing bacterial virulence by horizontal transfer. Further studies are needed to expand our results to a larger number of pregnant women. Moreover, long-term follow-up of GBS carriage and the interaction between different clones or between very close strains is also an interesting question for understanding the coexistence of bacterial populations *in vivo*. Monitoring based on CRISPR array analysis may be a useful tool for deciphering these interactions.

ACKNOWLEDGMENTS

We are most grateful to P. Horvath for providing the macro-enabled Excel tool for CRISPR analysis and for fruitful discussions. We are most grateful to P. Glaser for providing the initial spacer dictionary. We also thank the Institut de Biologie Intégrative de la Cellule (I2BC) for maintaining the CRISPR's web server (http://crispr.i2bc.paris-saclay.fr) and especially Christine Pourcel.

REFERENCES

1. Edmond KM, Kortsalioudaki C, Scott S, Schrag SJ, Zaidi AK, Cousens S, Heath PT. 2012. Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. Lancet 379:547–556. https://doi.org/10.1016/S0140-6736(11)61651-6.
2. Agence Nationale d'Accréditation et d'Évaluation en Santé. 2001. Antenatal prevention of early neonatal bacterial infection (September 2001). https://www.has-sante.fr/portail/upload/docs/application/pdf/Antenatal_prevention.pdf.
3. Schrag S, Gorwitz R, Fultz-Butts K, Schuchat A. 2002. Prevention of perinatal group B streptococcal disease. Revised guidelines from CDC. MMWR Recomm Rep 51(RR-11):1–22.
4. Regan JA, Klebanoff MA, Nugent RP. 1991. The epidemiology of group B streptococcal colonization in pregnancy. Vaginal Infections and Prematurity Study Group. Obstet Gynecol 77:604–610.
5. Stoll BJ, Schuchat A. 1998. Maternal carriage of group B *Streptococci* in developing countries. Pediatr Infect Dis J 17:499–503. https://doi.org/10.1097/00006454-199806000-00013.
6. Campbell JR, Hillier SL, Krohn MA, Ferrieri P, Zaleznik DF, Baker CJ. 2000. Group B streptococcal colonization and serotype-specific immunity in pregnant women at delivery. Obstet Gynecol 96:498–503.
7. Cieslewicz MJ, Chaffin D, Glusman G, Kasper D, Madan A, Rodrigues S, Fahey J, Wessels MR, Rubens CE. 2005. Structural and genetic diversity of group B *Streptococcus* capsular polysaccharides. Infect Immun 73:3096–3103. https://doi.org/10.1128/IAI.73.5.3096-3103.2005.
8. Slotved H-C, Kong F, Lambertsen L, Sauer S, Gilbert GL. 2007. Serotype IX, a proposed new *Streptococcus agalactiae* serotype. J Clin Microbiol 45:2929–2936. https://doi.org/10.1128/JCM.00117-07.
9. Schuchat A. 1998. Epidemiology of group B streptococcal disease in the United States: shifting paradigms. Clin Microbiol Rev 11:497–513.
10. Poyart C, Réglier-Poupet H, Tazi A, Billoët A, Dmytruk N, Bidet P, Bingen E, Raymond J, Trieu-Cuot P. 2008. Invasive group B streptococcal infections in infants, France. Emerg Infect Dis 14:1647–1649. https://doi.org/10.3201/eid1410.080185.
11. Haguenoer E, Baty G, Pourcel C, Lartigue M-F, Domelier A-S, Rosenau A, Quentin R, Mereghetti L, Lanotte P. 2011. A multi locus variable number of tandem repeat analysis (MLVA) scheme for *Streptococcus agalactiae* genotyping. BMC Microbiol 11:171. https://doi.org/10.1186/1471-2180-11-171.
12. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan M-S, Kunst F, Glaser P, Rusniok C, Crook DWM, Harding RM, Bisharat N, Spratt BG. 2003. Multilocus sequence typing system for group B *Streptococcus*. J Clin Microbiol 41:2530–2536. https://doi.org/10.1128/JCM.41.6.2530-2536.2003.
13. Radtke A, Lindstedt B-A, Afset JE, Bergh K. 2010. Rapid multiple-locus variant-repeat assay (MLVA) for genotyping of *Streptococcus agalactiae*. J Clin Microbiol 48:2502–2508. https://doi.org/10.1128/JCM.00234-10.
14. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712. https://doi.org/10.1126/science.1138140.
15. Grissa I, Vergnaud G, Pourcel C. 2009. Clustered regularly interspaced short palindromic repeats (CRISPRs) for the genotyping of bacterial pathogens. Methods Mol Biol 551:105–116. https://doi.org/10.1007/978-1-60327-999-4_9.
16. Shariat N, Dudley EG. 2014. CRISPRs: molecular signatures used for pathogen subtyping. Appl Environ Microbiol 80:430–439. https://doi.org/10.1128/AEM.02790-13.
17. Barrangou R, Dudley EG. 2016. CRISPR-based typing and next-generation tracking technologies. Annu Rev Food Sci Technol 7:395–411. https://doi.org/10.1146/annurev-food-022814-015729.
18. Pourcel C, Salvignol G, Vergnaud G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151:653–663. https://doi.org/10.1099/mic.0.27437-0.
19. Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C, Weill F-X. 2012. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One 7:e36995. https://doi.org/10.1371/journal.pone.0036995.
20. Lopez-Sanchez M-J, Sauvage E, Da Cunha V, Clermont D, Ratsima Hariniaina E, Gonzalez-Zorn B, Poyart C, Rosinski-Chupin I, Glaser P. 2012. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. Mol Microbiol 85:1057–1071. https://doi.org/10.1111/j.1365-2958.2012.08172.x.
21. Lier C, Baticle E, Horvath P, Haguenoer E, Valentin A-S, Glaser P, Mereghetti L, Lanotte P. 2015. Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. Front Genet 6:214. https://doi.org/10.3389/fgene.2015.00214.
22. Beauruelle C, Pastuszka A, Horvath P, Perrotin F, Mereghetti L, Lanotte P. 2017. CRISPR: a useful genetic feature to follow vaginal carriage of group B *Streptococcus*. Front Microbiol 8:1981. https://doi.org/10.3389/fmicb.2017.01981.
23. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186:1518–1530. https://doi.org/10.1128/JB.186.5.1518-1530.2004.
24. Imperi M, Pataracchia M, Alfarone G, Baldassarri L, Orefici G, Creti R. 2010. A multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of *Streptococcus agalactiae*. J Microbiol Methods 80:212–214. https://doi.org/10.1016/j.mimet.2009.11.010.
25. Brochet M, Rusniok C, Couvé E, Dramsi S, Poyart C, Trieu-Cuot P, Kunst F, Glaser P. 2008. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. Proc

Natl Acad Sci U S A 105:15961–15966. https://doi.org/10.1073/pnas.0803654105.
26. Perez-Ruiz M, Rodríguez-Granger JM, Bautista-Marín MF, Romero-Noguera J, Rosa-Fraile M. 2004. Genetic diversity of *Streptococcus agalactiae* strains colonizing the same pregnant woman. Epidemiol Infect 132:375–378. https://doi.org/10.1017/S0950268803001845.
27. Shen P, Huang HV. 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics 112:441–457.
28. Majewski J, Zawadzki P, Pickerill P, Cohan FM, Dowson CG. 2000. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. J Bacteriol 182:1016–1023. https://doi.org/10.1128/JB.182.4.1016-1023.2000.
29. Salloum M, van der Mee-Marquet N, Valentin-Domelier A-S, Quentin R. 2011. Diversity of prophage DNA regions of *Streptococcus agalactiae* clonal lineages from adults and neonates with invasive infectious disease. PLoS One 6:e20256. https://doi.org/10.1371/journal.pone.0020256.
30. Domelier A-S, van der Mee-Marquet N, Sizaret P-Y, Hery-Arnaud G, Lartigue M-F, Mereghetti L, Quentin R. 2009. Molecular characterization and lytic activities of *Streptococcus agalactiae* bacteriophages and determination of lysogenic-strain features. J Bacteriol 191:4776–4785. https://doi.org/10.1128/JB.00426-09.
31. Hansen SM, Uldbjerg N, Kilian M, Sorensen UBS. 2004. Dynamics of *Streptococcus agalactiae* colonization in women during and after pregnancy and in their infants. J Clin Microbiol 42:83–89. https://doi.org/10.1128/JCM.42.1.83-89.2004.
32. El Aila N, Tency I, Claeys G, Saerens B, De Backer E, Temmerman M, Verhelst R, Vaneechoutte M. 2009. Genotyping of *Streptococcus agalactiae* (group B *Streptococci*) isolated from vaginal and rectal swabs of women at 35-37 weeks of pregnancy. BMC Infect Dis 9:153. https://doi.org/10.1186/1471-2334-9-153.
33. Bohnsack JF, Whiting A, Gottschalk M, Dunn DM, Weiss R, Azimi PH, Philips JB, Weisman LE, Rhoads GG, Lin F-YC. 2008. Population structure of invasive and colonizing strains of *Streptococcus agalactiae* from neonates of six U.S. academic centers from 1995 to 1999. J Clin Microbiol 46:1285–1291. https://doi.org/10.1128/JCM.02105-07.
34. Huber CA, McOdimba F, Pflueger V, Daubenberger CA, Revathi G. 2011. Characterization of invasive and colonizing isolates of *Streptococcus agalactiae* in East African adults. J Clin Microbiol 49:3652–3655. https://doi.org/10.1128/JCM.01288-11.
35. Manning SD, Lewis MA, Springman AC, Lehotzky E, Whittam TS, Davies HD. 2008. Genotypic diversity and serotype distribution of group B *Streptococcus* isolated from women before and after delivery. Clin Infect Dis 46:1829–1837. https://doi.org/10.1086/588296.
36. Jones N, Oliver KA, Barry J, Harding RM, Bisharat N, Spratt BG, Peto T, Crook DW, Consortium OGBS. 2006. Enhanced invasiveness of bovine-derived neonatal sequence type 17 group B *Streptococcus* is independent of capsular serotype. Clin Infect Dis 42:915–924. https://doi.org/10.1086/500324.
37. van der Mee-Marquet N, Jouannet C, Domelier A-S, Arnault L, Lartigue M-F, Quentin R. 2009. Genetic diversity of *Streptococcus agalactiae* strains and density of vaginal carriage. J Med Microbiol 58:169–173. https://doi.org/10.1099/jmm.0.005827-0.
38. Jones N. 2006. Carriage of group B *Streptococcus* in pregnant women from Oxford, UK. J Clin Pathol 59:363–366. https://doi.org/10.1136/jcp.2005.029058.
39. Kunze M, Ziegler A, Fluegge K, Hentschel R, Proempeler H, Berner R. 2011. Colonization, serotypes and transmission rates of group B *Streptococci* in pregnant women and their infants born at a single University Center in Germany. J Perinat Med 39:417–422. https://doi.org/10.1515/JPM.2011.037.
40. Anthony BF, Eisenstadt R, Carter J, Kim KS, Hobel CJ. 1981. Genital and intestinal carriage of group B streptococci during pregnancy. J Infect Dis 143:761–766. https://doi.org/10.1093/infdis/143.6.761.
41. Brzychczy-Włoch M, Pabian W, Majewska E, Zuk MG, Kielbik J, Gosiewski T, Bulanda MG. 2014. Dynamics of colonization with group B streptococci in relation to normal flora in women during subsequent trimesters of pregnancy. New Microbiol 37:307–319.

B. Comparaison des génomes des clones présents en co-culture.

Publication N°3 : Beauruelle Clémence, Branger Maxime, Cochard Thierry, Pastuszka Adeline, Biet Franck, Lanotte Philippe. **Whole-Genome Sequencing confirms co-existence of different colonizing Group B *Streptococcus* isolates underscored by CRISPR typing.** Microbiology Resource Announcements. Manuscrit en préparation. 2019

Dans l'étude précédente, deux femmes (W8 et W10) étaient porteuses de deux types de souches présentant des similarités au sein de leur locus CRISPR1. Concernant la patiente W8, les loci CRISPR1 des deux types de souches présentaient un TDR et quatre spacers ancestraux communs. Elles appartenaient d'autre part au même type capsulaire et au même ST défini par MLST. Concernant la patiente W10, le locus CRISPR1 des deux types de souches présentait un TDR et un spacer ancestral commun, un type capsulaire identique mais un ST différent.

D'autre part, des différences phénotypiques étaient observées pour les deux types de souches et concernaient :

- La sensibilité aux macrolides: les souches de la patientes W8 étaient résistantes à l'érythromycine (CMI > 256 mg/L) et à la clindamycine (CMI >256 mg/L) pour une souche et présentaient une CMI plus faible à l'érythromycine (CMI = 1 mg/L) et une sensibilité à la clarithromycine (CMI = 0,125 mg/L) pour l'autre souche. Les souches de la patiente W10 étaient résistantes à l'érythromycine (CMI = 4 mg/L) et sensibles à la clindamycine (CMI = 0,125 mg/L) pour une souche et sensible à l'érythromycine (CMI = 0,064 mg/L) et à la clindamycine (CMI= 0,064 mg/L) pour l'autre souche.
- L'hémolyse sur gélose au sang et la pigmentation sur gélose Granada® : un type de colonies de la patiente W10 présentait des colonies moins hémolytiques et moins pigmentées que l'autre type (Figure 32. Différence de pigmentation observée sur gélose Granada® entre les deux types de colonies de la patiente W10.

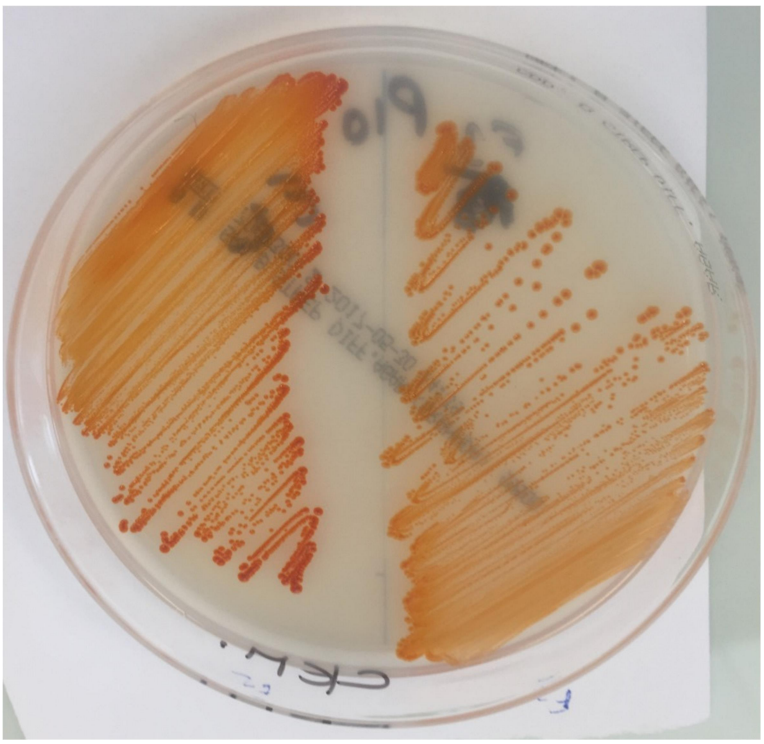


Figure 32. Différence de pigmentation observée sur gélose Granada® entre les deux types de colonies de la patiente W10.

Compte-tenu de ces données, le séquençage total de ces quatre souches apparaissait comme nécessaire pour conclure sur le lien entre ces isolats. Avec l'aide de Franck Biet et de son équipe (« infection bactérienne et immunité des ruminants » INRA Val de Loire, UMR ISP), nous avons ainsi réalisé un séquençage total de ces quatre souches en utilisant la technologie illumina (HiSeq)®. L'analyse des SNP (Single Nucleotide Polymorphism) a permis d'exclure un lien phylogénétique entre les souches isolées de la patiente W10. En revanche, les souches provenant de la patiente W8 étaient très proches et dérivent probablement d'un ancêtre commun. D'autre part, les différences phénotypiques observées ont pu être en partie confirmées par l'analyse des génomes complets.

- Concernant l'hémolyse, la production de l'hémolysine est codée par l'opéron *cyl* régulé par le système à deux composants CovS/CovR, lui-même codé par les gènes *csrS* et *csrR* (Lamy et *al.* 2004). Les souches isolées de la patiente W10, qui présentaient des différences d'hémolyse, étaient porteuses de mutations dans le gene *csrS* (deux mutations pour la souche plus hémolytique et une mutation pour la souche moins hémolytique).

- Concernant la résistance aux macrolides, le gène *ermB*, codant pour la résistance par méthylation de la cible de l'antibiotique, était retrouvé pour la souche de la patiente W8 présentant un haut niveau de résistance aux macrolides.

Au total, ce séquençage de génome complet nous a permis de confirmer en partie les différences phénotypiques observées et de confirmer les données obtenues par l'analyse du locus CRISPR1 sur le lien phylogénétique entre les souches.

Whole-Genome Sequencing confirms co-existence of different colonizing Group B *Streptococcus* isolates underscored by CRISPR typing.

Clémence Beauruelle,[a,b,c] Maxime Branger,[b] Thierry Cochard,[b] Adeline Pastuszka,[a,b,c] Franck Biet,[b] Philippe Lanotte[a,b,c]#

[a] Université François Rabelais de Tours, UMR1282 Infectiologie et Santé Publique, F-37032 Tours, France;
[b] INRA, UMR1282 Infectiologie et Santé Publique, F-37380 Nouzilly, France;
[c] CHRU de Tours, Service de Bactériologie-Virologie, F-37044 Tours, France
Clémence Beauruelle and Maxime Branger contributed equally to this work

Abstract.
Streptococcus agalactiae is a major pathogen in humans and is the leading cause of neonatal infections in industrialized countries. Diversity of strains isolated from two pregnant women was investigated. Here, we present the draft genome sequences of strains (W8A2, W8A6, W10E2 and W10F3) obtained in order to ascertain their phylogenetical link.

Streptococcus agalactiae, or Group B *Streptococcus* (GBS), is a major pathogen in humans and is the leading cause of neonatal infections in industrialized countries (1). GBS belongs to the commensal microbiota that colonizes the gastrointestinal and genital tracts of 10 to 30% of healthy women (2, 3). In pregnant women, GBS can be transmitted to the newborn during delivery leading to an infection in 1-2% of them (4).
In a recent study we explored the diversity of GBS in vaginal carriage population in pregnant women at the screening time (i.e. at 35-37 weeks' gestation by vaginal swabbing according to French clinical guidelines) using a CRISPR typing approach (5). For two women (woman 8 and woman 10), two types of isolates (W8-A2, W8-A6 and W10-E2, W10-F3 respectively) were identified with similarities among the CRISPR array between the 2 types of isolates. Whole Genome Sequencing (WGS) was chosen to affirm or reject a filiation between isolates.

For woman 8, among 20 colonies of GBS, 13 were represented by the W8-A2 type and the 7 others by the W8-A6 type. The Terminal Direct Repeat (TDR) and the ancestral spacer were identical as well as four of the older acquired spacers ("ancestral spacers") within the CRISPR1 array (5). Both types of isolates belong to the capsular serotype V and Sequence Type 1 defined by MLST. Regarding antibiotics susceptibility, W8-A6 type isolates were highly resistant to erythromycin (MIC > 256 mg/L) and clindamycin (MIC > 256 mg/L) whereas W8-A2 type isolates presented a low resistance to erythromycin (MIC = 1 mg/L) and were fully susceptible to clindamycin (MIC = 0.125 mg/L).

For woman 10, among 125 colonies of GBS, 92 were represented by the W10-E2 type and the 33 others by the W10-F3 type. The Terminal Direct Repeat (TDR) and the ancestral spacer were identical within the CRISPR1 array, as well as one of the older spacers. Both types of isolates belonged to capsular serotype Ia but belonged to different ST by MLST (ST4 [W10-E2 type] and ST28 [W10-F3 type]). Concerning antibiotics susceptibility, W10-E2 type isolates were resistant to erythromycin (MIC = 4 mg/L) and susceptible to clindamycin (MIC = 0.125 mg/L) whereas isolates W10-F3 were fully susceptible to erythromycin (MIC = 0.064 mg/L) and to clindamycin (MIC = 0.064 mg/L). Moreover, W10-F3 type isolates were more pigmented on Granada medium and more hemolytic on blood agar than W10-E2 type isolates.

Whole genome sequencing was performed by GenoScreen (Lille, France). Libraries from purified genomic DNA were prepared for Illumina sequencing with Nextera XT sample prep

kit (Illumina, San Diego, USA). Sequencing was realized using Illumina HiSeq in paired-ends of 250 bp. At least 2.8 millions of reads were obtained for each strain for a mean coverage of more than 339X for each, based on 2.1 Mb genome size.

Quality check of sequencing data was perfomed using FastQC (6), trimming of reads realized using sickle (7) before assembly using SPAdes (8). Contigs were annotated using the NCBI prokaryotic genome annotation pipeline (9).

All strains were assembled into less than 80 contigs (from 50 to 69) with a total size of assembly of 2.1 Mb. About 2,000 coding genes were annotated (2,039 to 2,137). Metrics for each sample are shown in Table 1. All contigs and annotations used in this study are available on NCBI (accession no. SSWT00000000 [strain W8A2], SSWU00000000 ([train W8A6], SSWV00000000 [strain W10E2] and SSWW00000000 [strain W10F3]).

WGS analysis confirmed the presence and structure of CRISPR-*cas* genes previously described for these isolates (5). Using Phaster™ (10), remnant phages were discovered on the 4 genomes and an intact prophage sequence (79% identity with *Streptococcus* prophage 315.2, NC_004585.1) was found on W10-F3 isolates (contig 4).

In GBS, hemolysin production is encoded by the *cyl* operon containing 12 genes including the *cylE* gene (11). Transcriptional regulation of the *cyl* operon is mainly controlled by the two-component system CovS/CovR (*csrS* and *csrR* genes) (12). The *cylE* gene was found strictly conserved in the four different isolates. Even if there was no difference in *csrS* and *csrR* genes between both types of woman 8 isolates, two mutations were observed in *csrS* genes of W10-E2 isolate and one for W10-F3 isolate, which may explain the highly hemolytic activity of W10-F3.

Regarding resistance to macrolides, in GBS, erythromycin resistance is commonly due to target modification by an rRNA methyltransferase enzyme encoded by the *ermB* gene. An *ermB* gene was found only in isolate W8-A6, which presented a high level of resistance to both erythromycin and clindamycin.

One of the major outcomes of this study was the measure of the degree of relatedness between isolates originating from each woman. The minimum spanning tree based on Single Nucleotide Polymorphism analysis (Figure 1) allowed to exclude a phylogenetical link between W10-E2 type and W10-F3 type isolates, despite a same serotype and common Terminal Direct repeat and ancestral spacer. This result confirms the difference in Sequence Type found (respectively ST4 and ST28). Isolates from W8-A2 type and W8-A6 type were

genetically closely related. These results show the *in vivo* co-existence of carriage of 2 types of closely-related strains with a probable common ancestor.

Data availability.

The draft genome sequences of strains W8A2, W8A6, W10E2 and W10F3 have been deposited in DDBJ/ENA/GenBank under the accession no. SSWT00000000, SSWU00000000, SSWV00000000 and SSWW00000000, respectively. The version described in this paper is the first version, SSWT01000000, SSWU01000000, SSWV01000000 and SSWW01000000, respectively.

Acknowledgments

Authors are very grateful to Aileen Bonade Bottino for technical assistance.

References

1. **Stoll BJ, Schuchat A.** 1998. Maternal carriage of group B *Streptococci* in developing countries. Pediatr Infect Dis J **17:**499-503.
2. **Regan JA, Klebanoff MA, Nugent RP.** 1991. The epidemiology of group B streptococcal colonization in pregnancy. Vaginal Infections and Prematurity Study Group. Obstet Gynecol **77:**604-610.
3. **Yancey MK, Duff P, Clark P, Kurtzer T, Frentzen BH, Kubilis P.** 1994. Peripartum infection associated with vaginal group B streptococcal colonization. Obstet Gynecol **84:**816-819.
4. **Melin P.** 2011. Neonatal group B streptococcal disease: from pathogenesis to preventive strategies. Clin Microbiol Infect **17:**1294-1303.
5. **Beauruelle C, Pastuszka A, Mereghetti L, Lanotte P.** 2018. Group B *Streptococcus* vaginal carriage in pregnant women as deciphered by clustered regularly interspaced short palindromic repeat analysis. J Clin Microbiol **56**.
6. **Adrews S.** 2010. FastQC : A Quality Control tool for High Throughput Sequence Data, [Software]. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
7. **Joshi N.** 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files., [Software]. Available at https://github.com/najoshi/sickle.

8. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.
9. **Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J.** 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res **44:**6614-6624.
10. **Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS.** 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res **44:**W16-21.
11. **Whidbey C, Harrell MI, Burnside K, Ngo L, Becraft AK, Iyer LM, Aravind L, Hitti J, Adams Waldorf KM, Rajagopal L.** 2013. A hemolytic pigment of Group B *Streptococcus* allows bacterial penetration of human placenta. J Exp Med **210:**1265-1281.
12. **Lamy MC, Zouine M, Fert J, Vergassola M, Couve E, Pellegrini E, Glaser P, Kunst F, Msadek T, Trieu-Cuot P, Poyart C.** 2004. CovS/CovR of group B *Streptococcus*: a two-component global regulatory system involved in virulence. Mol Microbiol **54:**1250-1268.

Table 1. Sequencing, assembly and annotation metrics for each strain.

Strain	Bio Sample Accession	No of reads	Avg coverage (X)	No. of contigs	N_{50} (kb)	Assembly length (bp)	G+C content (%)	No. of protein CDSs
W8A2	SAMN11415932	2 851 690	339	50	139262	2 078 349	35.26	2,059
W8A6	SAMN11415933	2 926 588	348	69	99289	2 086 162	35.29	2,076
W10E2	SAMN11415934	2 851 102	339	55	155056	2 158 522	35.38	2,137
W10F3	SAMN11415935	4 173 810	497	55	149658	2 085 005	35.34	2,097

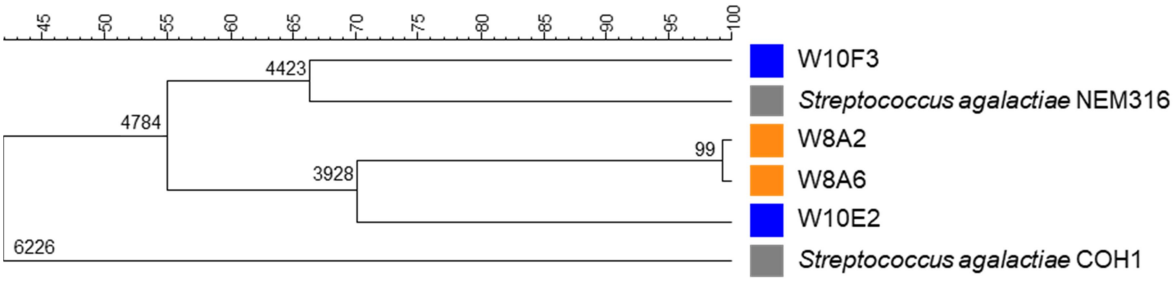


Figure 1. Phylogeny of the clinical isolates of *Streptococcus agalactiae* with reference genomes, resolved with SNPs. UPGMA phylogenetic tree was inferred from the 19,460 SNPs detected in the whole genome sequences of clinical isolates in comparison to the reference strain *Streptococcus agalactiae* NEM316. The number of SNP identified between the strains is indicated.

III. L'approche CRISPR pour le typage des souches de *S. agalactiae* : comparaison aux méthodes usuelles et pouvoir discriminant

Publication N°4 : Clémence Beauruelle, Ludovic Treluyer, Clément Lier, Elodie Baticle, Thierry Cochard, Laurent Mereghetti, Philippe Glaser, Claire Poyart, Philippe Lanotte. **CRISPR is an optimal target for the typing of *Streptococcus agalactiae.*** Manuscrit en préparation. Journal of Clinical Microbiology (IF: 3,712) 2019.

Cette étude a été réalisée en la collaboration avec le Pr Claire Poyart (Hôpital Cochin, APHP), le Pr Philippe Glaser (Institut Pasteur de Paris) et le Dr Thierry Cochard (INRA, Tours-Nouzilly).

Les études précédentes nous ont permis de confirmer l'intérêt du locus CRISPR1 comme outil de typage pour l'étude des souches de portage. L'objectif de ce travail était d'évaluer la pertinence du « typage CRISPR1 » en le comparant aux approches usuelles : par sa capacité à obtenir des données sur l'appartenance phylogénétique des souches d'une part, par son pouvoir discriminant d'autre part.

Pour cela, nous avons sélectionné 250 souches de différentes origines (souches de portage et souches invasives, isolées chez l'adulte ou le nouveau-né au CHRU de Tours ou provenant du CNR des streptocoques de Paris) ainsi que 5 souches de références, soit un total de 255 souches. Pour chacune de ces souches, 3 techniques de typage ont été appliquées : le typage capsulaire, le MLST et le typage CRISPR. En fonction de la composition en spacers et des DRT du locus CRISPR1, les souches étaient réparties en différents groupes. Les spacers et DRT permettaient une première répartition des souches. L'analyse des spacers internes permettait de former des sous-groupes au sein des groupes précédemment définis. Deux sélections de marqueurs (spacers et DRT) discriminants au sein de chaque groupe et sous-groupe ont ensuite été effectuées afin de mettre en évidence des spacers spécifiques de lignée. Une corrélation entre les marqueurs sélectionnés et les séquences-types définies en MLST

était recherchée. D'autre part, le pouvoir discriminant du typage CRISPR était comparé à celui du MLST.

Cette analyse nous a permis de confirmer :

i) la corrélation entre le typage CRISPR et le MLST, notamment par les deux sélections de marqueurs.

ii) le fort pouvoir discriminant de l'approche CRISPR pour le typage des souches de* S. agalactiae *y compris par l'utilisation d'une sélection réduite de marqueurs.

En effet,

- Les DRT et le spacers ancestraux étaient très conservés parmi les 255 souches étudiées, permettant d'obtenir des informations sur l'origine phylogénétique des souches. Par l'analyse des DRT et des spacers ancestraux, les isolats pouvaient être regroupés en clusters présentant une corrélation avec les CC définis par MLST et en sous-groupes. Les groupes et sous-groupes obtenus pour les CC-17 et CC-23 présentaient une corrélation avec l'analyse phylogénétique par comparaison de génome préalablement publié (Da Cunha et *al.* 2014).

- Les spacers plus récents présentaient une grande variabilité, permettant d'obtenir une méthode de typage très discriminante. Le typage CRISPR1 permettait de séparer des isolats au sein d'un même type capsulaire et d'un même ST défini par MLST. Parmi les 255 souches, 229 profils CRISPR1 uniques étaient observés, soit un profil CRISPR1 unique pour 90% des isolats. A l'opposé, 6 types capsulaires et 48 ST définis en MLST étaient représentés, témoignant du fort pouvoir discriminant du typage CRISPR1. Outre l'analyse du locus CRISPR1 dans sa globalité, les deux sélections de spacers (comprenant 94 et 25 marqueurs) offraient un fort index de diversité (0,9947 et 0,9284 respectivement), supérieur à celui du MLST (0,9017).

iii) l'existence de différences de variabilité du locus CRISPR1 en fonction de l'appartenance phylogénétique des souches bactériennes.

Un nombre restreint et une plus faible diversité des spacers étaient observés pour les loci CRISPR1 des souches appartenant au CC-17. Cette faible diversité est probablement le reflet de la clonalité des souches de CC-17, groupe décrit comme très homogène (Sørensen et *al.*

2010; Da Cunha et *al.* 2014). Le suivi des souches de portage développé précédemment a mis en évidence des modifications internes des loci CRISPR1 possiblement dues à un phénomène de recombinaison homologue. Les souches de CC17 ont par ailleurs été caractérisées par un faible taux de recombinaison par rapport aux autres CC, pouvant expliquer, au moins en partie la faible variabilité des loci CRISPR1 des souches de CC17 (Sørensen et *al.* 2010; Da Cunha et *al.* 2014).

A l'opposé, les loci CRISPR1 des souches appartenant au CC-1, CC-4 et CC-19 présentaient une forte variabililté et étaient difficiles à distinguer entre ces CC et à regrouper. Les isolats appartenant à ces CC étaient distribués dans différents groupes définis par MLST et eBURST, suggérant une grande hétérogénéité de ces populations. Cette diversité des loci CRISPR1 pourrait être due i) à une évolution plus rapide de ces isolats ; ii) à un système CRISPR-Cas plus actif ; iii) à une plus grande diversité des EGMs pour ces CC responsables d'une immunisation de ces lignées phylogénétiques.

CRISPR is an optimal target for the typing of *Streptococcus agalactiae*.

Clémence Beauruelle[1, 2, 3], Ludovic Treluyer[4], Adeline Pastuszka[1, 2, 3], Thierry Cochard, Clément Lier[1, 2, 3], Laurent Mereghetti[1, 2, 3], Philippe Glaser[4, 5, 6], Claire Poyart [4, 5, 6, 7] and Philippe Lanotte[1, 2, 3].

[1] Université François Rabelais de Tours, UMR1282 Infectiologie et Santé Publique, F-37032 Tours, France;

[2] INRA, UMR1282 Infectiologie et Santé Publique, F-37380 Nouzilly, France;

[3] CHRU de Tours, Service de Bactériologie-Virologie, F-37044 Tours, France

[4]Department of Bacteriology, University Hospitals Paris Centre-Cochin, French National Center for Streptococci, Assistance Publique - Hôpitaux de Paris (AP-HP).

[5]Département Hospitalo-Universitaire Risks and Pregnancy.

[6]Team Barriers and Pathogens of Cochin Institute, Institut National de la Santé et de la Recherche Médicale (INSERM) U1016 - Unité Mixte de Recherche Centre National de la Recherche Scientifique (UMR CNRS) 8104.

[7]Paris Descartes University.

Running title: CRISPR for GBS typing

Key words: CRISPR, GBS, Typing, MLST

Abstract

We explored the relevance of the CRISPR-based genotyping tool for *Streptococcus agalactiae* typing and we compared this method to current molecular standards (MLST and capsular typing). To this effect, we developed two CRISPR marker schemes (94 and 25 markers). Among the 255 isolates, 229 CRISPR profiles were observed. The 94 and 25 markers made it possible to efficiently separate isolates with a high diversity index (0.9947 and 0.9284, respectively), highlighting a high discriminatory power, superior to that of both capsular typing and MLST (diversity index of 0.9017 for MLST). This method has the advantage of i) being linked to the genetic lineage of the isolate through analysis of the terminal direct repeat and ancestral spacers, and ii) sharing a high discriminatory power through analysis of the leader-end spacers. On the basis of these data, we propose that this method could become a reference method for phylogenetic GBS typing.

Introduction

Streptococcus agalactiae or group B *Streptococcus* is the leading cause of neonatal infections and an emerging pathogen in adults, particularly in elderly and immunocompromized patients (1, 2). It is also found in commensal microbiota that colonize the gastrointestinal and genitourinary tracts of 10 to 30% of healthy women (3–5). Prevention of GBS-related neonatal infection involves the detection of vaginal carriage in pregnant women, followed by intrapartum antibiotic prophylaxis for those colonized (3, 6).

To characterize GBS isolates, two mains typing methods are widely used: serotyping and MLST (Multi Locus Sequence Typing). Serotyping methods, which are based on differences in capsular polysaccharides (CPS), was first evaluated by phenotypic methods and more recently, by genotypic methods (7). To date, 10 serotypes (or capsular types) have been identified (Ia, Ib, II-IX) (8, 9). The prevalence and distribution of serotypes is known to differ between geographical regions, ethnic populations and clinical presentations (10, 11). MLST, a sequence-based method, is now widely used to investigate the population structure and genetic lineage of GBS, and is currently the reference method for GBS typing. This approach is based on the combination of alleles for seven housekeeping genes (Jones et al, 2003). Unique combinations of the alleles at each locus define the allelic profiles, or sequence types (ST). ST can be clustered in clonal complexes (CC) when six of the seven alleles are in common (12). These methods have highlighted the involvement of serotype III and ST17, in causing more invasive neonate disease (S. D. Manning et al. 2009b). Among adults, serotype V isolates belonging to CC1 have been associated with invasive disease (1, 14). More recently, serotype III, Ia and IV have gained relevance in this context (15–18). Despite the clonality observed within GBS populations causing invasive disease, increasing diversity has appeared, hence the need to investigate the population structure. However, between these two typing methods, while serotyping is relatively easy to perform in the laboratory, it is insufficiently discriminating to compare isolates. MLST presents significant disadvantages: it is costly, time-consuming, labor-intensive and specific equipment is required. Moreover, as the sequences targeted evolve slowly, MLST is not highly discriminant for epidemiological studies and local surveillance, and isolates are not easy to distinguish among ST17 (19–21).

CRISPR-Cas [Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated sequences (Cas)] is an adaptive and vertically-transmitted immune system present in a large proportion of prokaryotic genomes (22). CRISPR arrays are made up of a succession of repeated sequences, called direct repeats or DR, interspaced by sequences of a

similar length, called spacers. Most CRISPR arrays are flanked on one side by the leader sequence and on the other side by a trailer sequence, which is preceded by the terminal direct repeat (TDR) of the CRISPR array, corresponding to a degenerated or truncated DR. DR are highly conserved within a locus, whereas spacers vary widely as they derive from foreign genetic elements. In this acquired immunity system, each spacer acquired in response to mobile genetic element (MGE) invasion is integrated in a linear, time resolved manner. Each new spacer is incorporated at the leader end of the CRISPR array, concomitantly to the duplication of a DR at the leader end. Because of the polarized acquisition of spacers deriving from encounters with various MGE over time, CRISPR arrays constitute a chronological archive of past encounters. Thus, spacers located at the leader-end extremity are recently acquired and represent recent contacts with MGE, whereas spacers located at the trailer end represent ancestral spacers. Leader-end spacers enable the differentiation of closely-related strains and trailer-end spacers reflect broader phylogenetic relationships. Indeed, several studies have shown potential for CRISPR-based typing (23).

In GBS, two CRISPR-Cas systems have been characterized, a type I-C system named CRISPR2, which is rare and most often incomplete, suggesting little or no activity; and a type II-A system, named CRISPR1, which is ubiquitous and functional (24). The CRISPR1 array is made up of highly conserved DR of 36 bp, separated by spacers of approximately 30 bp which are extremely diverse in sequence and in number across GBS strains studied. Similarities between CRISPR1 spacer sequences and MGEs have been previously reported (24, 25). Comparative sequence analysis across numerous GBS isolates emphasized that CRISPR1 array is extremely diverse and evolves *in vivo*, demonstrating the dynamics of the system (Lopez-Sanchez et al. 2012; Lier et al. 2015; Clemence Beauruelle et al. 2018; Clémence Beauruelle et al. 2017). A typing method based on CRISPR array analysis appeared to be a useful tool for comparing GBS isolates. This method presents the advantage of being linked to the genetic lineage of the isolate through analysis of the TDR and ancestral spacers. Moreover, exploration of the recent evolution of the isolate, especially encounters with MGE, is possible through the analysis of the leader-end spacers. Based on the spacers' variability it seems to be in congruence with MLST with a greater discriminating power, however these advantages still have to be demonstrated (Lier et al. 2015; Clémence Beauruelle et al. 2017; Clemence Beauruelle et al. 2018). The aim of our study was to evaluate CRISPR1 analysis as a high-resolution *S. agalactiae* typing method and to compare this method to serotyping and MLST.

MATERIALS AND METHODS

Bacterial isolates. 255 *S. agalactiae* isolates were analyzed. 224 isolates were collected at the Tours teaching hospital, France between 2002 and 2016. 26 were collected by the National Reference Center for *Streptococci* in 2012 and five are reference strains. Among the clinical isolates, 198 were isolated from adults and 52 from neonates. Among them, 164 were isolated in a non-invasive setting from a cutaneous swab (n= 2), urinary tract (n= 5), gastric aspirate (n= 7), respiratory tract (n= 9), genital (n= 119) or intestinal tract (n= 23) and 86 were isolated in an invasive setting from blood (n= 49), CSF (n= 33) or joint fluid cultures (n=2). The 5 reference strains considered were 2603V/R (capsular type V, ST110), A909 (capsular type Ia, ST7), NEM316 (capsular type III, ST23), COH1 (capsular type III, ST17) and BM110 (capsular type III, ST17) (Table S1).

DNA extraction. Genomic DNA was extracted following enzymatic lysis with mutanolysin (Sigma). A bacterial suspension of 1.5 McFarland was prepared in 500 µL water containing 50 U of mutanolysin. The suspension was incubated for 1 hr at 56°C, followed by 10 min at 100°C, leading to cell lysis. Lysates were centrifuged for 3 min at 1500 × g and the supernatants containing DNA collected.

Capsular typing. Capsular typing was performed by genotypic methods based on previously described methods (7). The assay identified each serotype (Ia to IX) by analysis of the band pattern on agarose gel.

Multilocus sequence typing (MLST). MLST was carried out as previously described (28). Allelic profiles and sequence type (ST) were assigned using the international MLST database (http://pubmlst.org/sagalactiae/). Clonal complexes (CC) were defined using the stringent group definition (6/7 shared alleles) and eBurst analysis (http://eburst.mlst.net/v3/mlst_datasets/).

CRISPR1 array amplification and sequencing. CRISPR1 array amplification was performed using CRISPR1 PCR-F and CRISPR1 PCR-R primers which target the CRISPR1-flanking regions as previously described (Lopez-Sanchez et al. 2012; Lier et al. 2015; Clémence Beauruelle et al. 2017). PCR products were sequenced using the internal sequencing primers CRISPR1 SEQ-F and CRISPR1 SEQ-R. For CRISPR1 regions exceeding 1.3 kb, primers targeting internal spacers had to be used to complete the sequencing (Primers sequences are presented in supplementary material table S1).

CRISPR1 array analysis. Spacers, repeats, and flanking regions for each sequence were identified using a macro-enabled Excel tool (P. Horvath, DuPont). This tool allows the

identification and extraction of CRISPR features from nucleotide sequences, and the graphic representation of spacers as colored cells in Excel spreadsheets. Spacer sequences were compared to the dictionary of spacers established earlier and expanded previously (Lopez-Sanchez et al. 2012; Clémence Beauruelle et al. 2017; Clemence Beauruelle et al. 2018) New spacers identified in this study expanded the dictionary and were numbered incrementally. The spacer dictionary is available on (http://crispr.i2bc.paris-saclay.fr/CRISPRcompar/Dict/Dict.php).

Spacers and DRT selection. Graphic representation of CRISPR1 arrays made it possible to separate isolates according to their CRISPR1 array composition. Isolates were clustered according to their ancestral spacers and DRT composition. Among each cluster, specific markers, corresponding to spacers and DRT, were selected in view of separating isolates according to their CRISPR1 array similarity. We selected two groups of markers (spacers and DRT), first including 94 markers and a smaller group corresponding to 25 markers.

Data analysis.

The discriminatory index (DI) described by Hunter and Gaston was used as a numerical index for the discriminatory power of each typing method (29). The categorical coefficient, unweighted pair group with arithmetic mean (UPGMA) and the minimum spanning trees (MST) were run using BioNumerics 7.6.2 software (Applied Maths, Sint-Martens-Latem, Belgium). In MST, each circle represents a CRISPR genotype and its size is proportional to the number of strains. The thicker branches link the genotypes differing by only one spacer, the thinner branches link genotypes differing by more than one spacer.

Results

Capsular typing and Multi Locus Sequence Typing

Capsular typing**.** Among the 255 *S. agalactiae* isolates analyzed, 6 capsular types were represented. The main capsular type was type III (n = 93), followed by type Ia (n = 56), V (n = 54), Ib (n = 23), II (n = 17) and IV (n = 12). Non-capsular types VI, VII, VIII, IX and X were also found among our isolates (Figure 1 and table S3).

MLST. A total of 48 different STs were found among the 255 isolates. The most common was ST17 (n = 56), followed by ST23 (n = 37), ST1 (n = 32), ST19 (n = 26), ST8 (n = 12) and ST28 (n = 10). Other STs were represented by fewer isolates (< 10) (Table 1). Invasive isolates belonged mainly to ST17 followed by ST23 and ST1, while non-invasive isolates were distributed mainly among ST23, ST17, ST1 and ST19 isolates (Table S4). eBURST

analysis clustered the ST into 7 clonal complexes (CC17 [grouping ST17, ST291, ST1004], CC23 [grouping ST23, ST220, ST1002, ST481, ST385, ST88, ST144, ST366, ST391], CC1 [grouping ST1, ST196, ST2, ST1005, ST459, ST136, ST370, ST297, ST173], CC19 [grouping ST19, ST28, ST27, ST389, ST182, ST386], CC8 [grouping ST8, ST12, ST10, ST390], CC6 [grouping ST6, ST7, ST255, ST41] and CC4 [grouping ST4, ST3, ST243]), one group with only two ST (ST26 and ST388), and 5 singletons (ST22, ST24, ST130, ST327 and ST569) (Table S4). All CC except CC17 contained mainly non-invasive isolates (CC19 (n=31, 78%), CC1 (n=38, 75%), CC23 (n=34, 71%), CC6 (n=4, 67%), CC8 (n=17, 63%), CC4 (n=3, 60%) and CC17 (n=22, 39%) (Table S5). Using UPGMA and MST, the 48 different profiles (corresponding to the 48 ST) were grouped into 21 clusters with more than one isolate (up to 54 isolates by cluster).

CRISPR1 array analysis.

We generated a complete CRISPR1 sequence for all the isolates. The number of spacers present ranged from 3 to 29 per isolate, corresponding to a CRISPR array size of 266 to 1,800 bp. Among the 255 isolates, a unique CRISPR1 array was observed for 90% of isolates (n= 229).

Using the macro-enabled Excel tool, isolates could be distributed according to their CRISPR1 array homology (supplementary material Figure S1). First, isolates were grouped into 6 clusters according to their DRT and ancestral spacers. These 6 clusters showed a correlation between CC defined by MLST and eBURST analysis: a CC23 cluster, a cluster grouping CC17 and ST130 isolates, a cluster grouping CC1, CC4 and CC19 isolates, a cluster grouping CC6 and CC8 isolates, a cluster grouping ST26-ST388 isolates, and an ST22 cluster (Figure 1). CRISPR1 clusters were then divided into subgroups according to more recent spacers. In this way, a total of 14 clusters and subgroups were defined, including three subgroups in cluster CC1-CC4-CC19, allowing separation between CC19 (2 subgroups) and the group CC1-CC4 (2 subgroups), 5 subgroups in the CC17 cluster (COH1 type, BM110 type, ST130 type and 2 others subgroups), and 2 subgroups in the CC23 cluster (Figure 1 and Supdata Figure S1). This CRISPR genotyping approach can be used to rank some singletons defined by eBURST analysis into clusters: ST24 (CRISPR1 cluster CC23), ST130 (CRISPR1 cluster CC17), ST569 (CRISPR1 cluster CC6-CC8) and ST327 (CRISPR1 cluster CC19) (Table S4, Figure S1).

Among each cluster, spacers and TDR were selected to define phylogenetic lineages. Each spacer selected was cluster- and subgroup-specific. We selected 2 groups of markers. The first

involved 94 markers, including the five different DRT and 89 spacers (Table S6A). The second involved 25 markers including the five different DRT and 20 spacers (Table S6B). Among the 255 isolates, using the UPGMA algorithm, the 94-marker scheme defined a total of 172 different profiles and 42 groups with more than one isolate (up to 8 isolates) (figure S2 and S3). Likewise, the 25 marker selection defined a total of 42 different profiles and 30 clusters with more than one isolate (up to 40 isolates) (Figure 2 and Figure S4). Using this two-marker selection, groups previously defined were divided in different clusters (Figure 2, Figure S2 and S4). Thus, CC17 isolates were divided into 30 groups and 5 groups, respectively, using these two-marker selections (94 and 25 markers, respectively).

Difference among clusters

All isolates shared a high level of correlation between spacer content and MLST type. Nonetheless, some groups displayed lower spacer composition diversity. Among the 26 isolates sharing a mutual CRISPR1 profile, 17 belonged to CC17, and the 9 others belonged to CC8 (n = 2), CC23 (n = 3), CC19 (n = 2), and ST388 (n = 2). Indeed, the majority of CRISPR1 arrays sharing a mutual profile belonged to CC17 isolates and represented 31% (17/54) of them. Among these CC17 isolates with common CRISPR1 array, the main isolates (76%, 13/17) belonged to the first CRISPR1 subgroup (COH1 type) and represented 41% of isolates of the COH1 type subgroup (13/32). By contrast, isolates belonging to CC1 and CC19 shared a high degree of polymorphism and could not be easily clustered into their specific CC defined by MLST and Eburst (supplementary material Figure S1).

Discrimination of isolates by CRISPR analysis

CRISPR1 array analysis makes it possible to separate isolates within a same ST or capsular type. Among the 255 isolates, 6 capsular types and 48 different ST were observed, while CRISPR1 analysis separated them into 229 different CRISPR1 profiles, and the 2-marker scheme into 172 (94-marker scheme) and 42 (25-marker scheme) profiles. The diversity index was compared for the two CRISPR1 specific marker schemes as well as for MLST. The diversity index was 0.9947 for theCRISPR1-specific 94-marker scheme, 0.9284 for the CRISPR1-specific 25-marker scheme, and 0.9017 for the MLST method (Table 2).

Discussion

The aim of our study was to evaluate CRISPR1 analysis as a high-resolution *S. agalactiae* typing method. Methods currently in use present great disadvantages as they may provide inadequate resolution, or are labor-intensive. CRISPR1 array analysis is easy to perform with just one array to be analyzed (unlike MLST which requires 7 loci to be sequenced) and generates data (spacers dictionary) which enables comparison of isolates between different laboratories. In this work, we explored the relevance of the CRISPR1-based genotyping tool and we compared this method to current molecular standards (capsular typing and MLST). In this aim, we developed two schemes of marker selection (spacers and TDR). This analysis was based on 255 isolates from different origins and we simultaneously characterized these isolates by MLST and capsular typing. In this study, we used a representative library of the species. We selected a wide variety of isolates which differed by their anatomic and geographic origin, as well as by their phylogenetic origin.

Among the 255 isolates, 229 CRISPR1 profiles were observed, corresponding to a unique CRISPR array for 90% of isolates. By contrast, 6 capsular types and 48 STs were represented, demonstrating the high discriminatory power of the CRISPR1 sequence-based approach. Selection of two marker groups exhibited 172 and 42 profiles, respectively, with a high diversity index, 0.9947 and 0.9284, respectively, for the 94-marker and the 25-marker selection. The discriminatory power of this CRISPR1 approach is superior to that of both capsular typing and MLST, which shows a diversity index of 0.9017 for MLST. Indeed, the CRISPR1-typing approach and the two CRISPR1 marker group selections, including the simplest, allow to efficiently separate isolates and to successfully discriminate isolates that were considered indistinguishable by MLST.

Typing with CRISPR1 array showed a correlation with the clustering observed with MLST. The two marker group selections showed clonal distribution of the population similar to that obtained by MLST, and so, were able to reflect the phylogenetic structure of the *S. agalactiae* population. The high discriminatory index of the CRISPR1 approach made it possible to distinguish between isolates within a same ST or CC defined by MLST and Eburst analysis. Concerning markers selections, CC17 isolates could be divided into 29 groups by CRISPR1 clustering based on 94 markers, and into 5 groups by CRISPR1 clustering based on 25 markers. Considering CRISPR1 array, CC17 isolates were divided into 40 different profiles. According to DRT and ancestral spacers, CC17 isolates were clustered with ST130 and could be divided into 5 subgroups; one of them concerning ST130 isolates which did not belong to

CC17 defined by MLST (Figures 1). Indeed, CC17 could be divided into 4 subgroups corresponding to different lineages previously highlighted by SNP analysis (Da Cunha et al. 2014). Similarly, CC23 isolates were divided into two subgroups as in SNP analysis (Da Cunha et al. 2014). Nonetheless, among clusters, CC17 isolates shared a low degree of polymorphism compared to other CC, as noted previously (25, 27). This moderate CRISPR1 diversity could be explain by the slow rate of evolution of ST17 isolates. Indeed, isolates belonging to CC17 differ from the other CC by a low rate of recombination, which is known for contribute to CRISPR array diversity (30, 31). Moreover, CC17 isolates encode a specific repertoire of surface proteins, suggesting a specific colonization site, and therefore a genetic isolation (Da Cunha et al. 2014). Conversely, some CC shared a high degree of polymorphism and were difficult to cluster using CRISPR1 array. This relates in particular to isolates belonging to CC1, CC4 and CC19 which appear to be mixed in different clusters, suggesting a greater heterogeneity of this population. Interestingly, CC1 isolates are increasing in adult invasive infections since the 1990's, suggesting evolution of this CC, especially ST1 isolates (1, 14). We can hypothesize that this CRISPR1 diversity could be due to i) a more rapid evolution of these CC isolates, ii) more active CRISPR-Cas system or iii) higher diversity of GME for this CC responsible for immunization of this phylogenetic lineage.

The CRISPR1 clustering approach generated major clusters that corresponded well to the main CC obtained by MLST, offering information about the genetic lineage. Moreover, this approach presents the major advantage of being highly discriminant thanks to the variability of the CRISPR1 array. This approach delivers a unique DNA fingerprint and makes it possible to separate isolates, even clonal bacteria such as *S. agalactiae*. Indeed, PCR-based amplification and sequencing of a single array of limited length (266 to 1,800 bp) offered the same amount of or more information than that obtained by labor intensive methods such as MLST and capsular typing combined. Given these data, we assume that this method could become a reference method for phylogenetic GBS typing. Although whole genome sequencing is increasingly accessible, it is still an expensive technology requiring experience and skill to be used, including bio informatic analysis. Nowadays, whole-genome sequencing still does not appear to be a routine method for genotyping, thus rendering CRISPR-based typing technology useful during the transition from the current molecular typing method to the omic-level.

Acknowledgment

We are most grateful to P. Horvath for providing the macro-enabled Excel tool for CRISPR analysis and to the Institut de Biologie Integrative de la cellule (I2BC) for maintaining the CRISPR's web server (http://crispr.i2bc.paris-saclay.fr) and especially to C. Pourcel.

References

Al Nakib, Malik, Magalie Longo, Asmaa Tazi, Annick Billoet, Josette Raymond, Patrick Trieu-Cuot, and Claire Poyart. 2011. "Comparison of the Diversilab® System with Multi-Locus Sequence Typing and Pulsed-Field Gel Electrophoresis for the Characterization of Streptococcus Agalactiae Invasive Strains." *Journal of Microbiological Methods* 85 (2): 137–42. https://doi.org/10.1016/j.mimet.2011.02.010.

Al Safadi, Rim, Souheila Amor, Geneviève Hery-Arnaud, Barbara Spellerberg, Philippe Lanotte, Laurent Mereghetti, François Gannier, Roland Quentin, and Agnès Rosenau. 2010. "Enhanced Expression of Lmb Gene Encoding Laminin-Binding Protein in Streptococcus Agalactiae Strains Harboring IS1548 in ScpB-Lmb Intergenic Region." Edited by Frank R. DeLeo. *PLoS ONE* 5 (5): e10794. https://doi.org/10.1371/journal.pone.0010794.

Al-Attar, Sinan, Edze R. Westra, John van der Oost, and Stan J.J. Brouns. 2011. "Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs): The Hallmark of an Ingenious Antiviral Defense Mechanism in Prokaryotes." *Biological Chemistry* 392 (4). https://doi.org/10.1515/bc.2011.042.

Alhhazmi, Areej, Donna Hurteau, and Gregory J. Tyrrell. 2016. "Epidemiology of Invasive Group B Streptococcal Disease in Alberta, Canada, from 2003 to 2013." Edited by K. C. Carroll. *Journal of Clinical Microbiology* 54 (7): 1774–81. https://doi.org/10.1128/JCM.00355-16.

Al-Maani, Amal, Laurie Streitenberger Streitenberger, Megan Clarke, Yvonne C. W. Yau, Danuta Kovach, Rick Wray, and Anne Matlow. 2014. "Nosocomial Transmission of Group B Streptococci Proven by Positive Environmental Culture." *Oman Medical Journal* 29 (5): 376–79. https://doi.org/10.5001/omj.2014.100.

Amin, A., Y.M. Abdulrazzaq, and S. Uduman. 2002. "Group B Streptococcal Serotype Distribution of Isolates from Colonized Pregnant Women at the Time of Delivery in United Arab Emirates." *Journal of Infection* 45 (1): 42–46. https://doi.org/10.1053/jinf.2001.0990.

Amitai, Gil, and Rotem Sorek. 2016. "CRISPR-Cas Adaptation: Insights into the Mechanism of Action." *Nature Reviews. Microbiology* 14 (2): 67–76. https://doi.org/10.1038/nrmicro.2015.14.

ANAES. 2001. "[Antenatal prevention of early neonatal bacterial infection (September 2001)]." *Journal De Gynécologie, Obstétrique et Biologie de la Reproduction* 32 (1): 68–74.

Andersson, Anders F., and Jillian F. Banfield. 2008. "Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities." *Science (New York, N.Y.)* 320 (5879): 1047–50. https://doi.org/10.1126/science.1157358.

Anthony, Bascom F., Donald M. Okada, and Calvin J. Hobel. 1978. "Epidemiology of Group B Streptococcus: Longitudinal Observations during Pregnancy." *Journal of Infectious Diseases* 137 (5): 524–530.

Avci, Fikri Y., and Dennis L. Kasper. 2010. "How Bacterial Carbohydrates Influence the Adaptive Immune System." *Annual Review of Immunology* 28 (1): 107–30. https://doi.org/10.1146/annurev-immunol-030409-101159.

Badri, M. S., S. Zawaneh, A. C. Cruz, G. Mantilla, H. Baer, W. N. Spellacy, and E. M. Ayoub. 1977. "Rectal Colonization with Group B Streptococcus: Relation to Vaginal

Colonization of Pregnant Women." *The Journal of Infectious Diseases* 135 (2): 308–12.

Baker, C. J. 1978. "Early Onset Group B Streptococcal Disease." *The Journal of Pediatrics* 93 (1): 124–25.

Baker, C. J. 1997. "Group B Streptococcal Infections." *Clinics in Perinatology* 24 (1): 59–70.

Baker, C.J., and F. Barrett. 1974. "Group B Streptococcal Infections in Infants: The Importance of the Various Serotypes." *JAMA* 230 (8): 1158–60. https://doi.org/10.1001/jama.1974.03240080040025.

Bakker, H. C. den, M. W. Allard, D. Bopp, E. W. Brown, J. Fontana, Z. Iqbal, A. Kinney, et *al*. 2014. "Rapid Whole-genome sequencing for surveillance of *Salmonella enterica* Serovar Enteritidis." *Emerging Infectious Diseases* 20 (8): 1306–14. https://doi.org/10.3201/eid2008.131399.

Ballard, Mark S., Henrik C. Schønheyder, Jenny Dahl Knudsen, Outi Lyytikäinen, Matthew Dryden, Karina J. Kennedy, Louis Valiquette, et al. 2016. "The Changing epidemiology of group B *Streptococcus* bloodstream infection: a multi-national population-based assessment." *Infectious Diseases* 48 (5): 386–91. https://doi.org/10.3109/23744235.2015.1131330.

Barcaite, Egle, A. Bartusevicius, R. Tameliene, M. Kliucinskas, L. Maleckiene, and R. Nadisauskiene. 2008. "Prevalence of maternal group B streptococcal colonisation in european countries." *Acta Obstetricia Et Gynecologica Scandinavica* 87 (3): 260–71. https://doi.org/10.1080/00016340801908759.

Barrangou, Rodolphe. 2013. "CRISPR-Cas systems and RNA-guided interference." *Wiley Interdisciplinary Reviews: RNA* 4 (3): 267–78. https://doi.org/10.1002/wrna.1159.

Barrangou, Rodolphe, and Edward G. Dudley. 2016. "CRISPR-Based typing and next-generation tracking technologies." *Annual Review of Food Science and Technology* 7 (1): 395–411. https://doi.org/10.1146/annurev-food-022814-015729.

Barrangou, Rodolphe, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A. Romero, and Philippe Horvath. 2007. "CRISPR provides acquired resistance against viruses in prokaryotes." *Science (New York, N.Y.)* 315 (5819): 1709–12. https://doi.org/10.1126/science.1138140.

Barrangou, Rodolphe, and Philippe Horvath. 2017. "A decade of discovery: CRISPR functions and applications." *Nature Microbiology* 2 (June): 17092. https://doi.org/10.1038/nmicrobiol.2017.92.

Barrangou, Rodolphe, and Luciano A. Marraffini. 2014. "CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity." *Molecular Cell* 54 (2): 234–44. https://doi.org/10.1016/j.molcel.2014.03.011.

Beauruelle, Clémence, Adeline Pastuszka, Philippe Horvath, Franck Perrotin, Laurent Mereghetti, and Philippe Lanotte. 2017. "CRISPR: a useful genetic feature to follow vaginal carriage of group B *Streptococcus*." *Frontiers in Microbiology* 8. https://doi.org/10.3389/fmicb.2017.01981.

Beauruelle, Clemence, Adeline Pastuszka, Laurent Mereghetti, and Philippe Lanotte. 2018. "Group B *Streptococcus* vaginal carriage in pregnant women deciphered by CRISPR analysis." *Journal of Clinical Microbiology*, April. https://doi.org/10.1128/JCM.01949-17.

Bekker, Vincent, Merijn W. Bijlsma, Diederik van de Beek, Taco W. Kuijpers, and Arie van der Ende. 2014. "Incidence of invasive group B streptococcal disease and pathogen

genotype distribution in newborn babies in the netherlands over 25 years: a nationwide surveillance study." *The Lancet Infectious Diseases* 14 (11): 1083–1089.

Bergseng, H., J.E. Afset, A. Radtke, K. Loeseth, R.V. Lyng, M. Rygg, and K. Bergh. 2009. "Molecular and phenotypic characterization of invasive group B *Streptococcus* strains from infants in Norway 2006–2007." *Clinical Microbiology and Infection* 15 (12): 1182–85. https://doi.org/10.1111/j.1469-0691.2009.02789.x.

Bergseng, H., M. Rygg, L. Bevanger, and K. Bergh. 2008. "Invasive Group B *Streptococcus* (GBS) disease in Norway 1996–2006." *European Journal of Clinical Microbiology & Infectious Diseases* 27 (12): 1193–99. https://doi.org/10.1007/s10096-008-0565-8.

Bidet, Philippe, Naima Brahimi, Céline Chalas, Yannick Aujard, and Edouard Bingen. 2003. "Molecular characterization of serotype III Group B *Streptococcus* isolates causing neonatal meningitis." *The Journal of Infectious Diseases* 188 (8): 1132–1137.

Bingen, E., E. Denamur, N. Lambert-Zechovsky, Y. Aujard, N. Brahimi, P. Geslin, and J. Elion. 1992. "Analysis of DNA restriction fragment length polymorphism extends the evidence for breast milk transmission in *Streptococcus agalactiae* late-onset neonatal infection." *The Journal of Infectious Diseases* 165 (3): 569–73.

Bisharat, N. 2005. "Population structure of Group B *Streptococcus* from a low-incidence region for invasive neonatal disease." *Microbiology* 151 (6): 1875–81. https://doi.org/10.1099/mic.0.27826-0.

Björnsdóttir, E.S., E.R. Martins, H. Erlendsdóttir, G. Haraldsson, J. Melo-Cristino, K.G. Kristinsson, and M. Ramirez. 2016. "Changing epidemiology of Group B streptococcal infections among adults in Iceland: 1975–2014." *Clinical Microbiology and Infection* 22 (4): 379.e9-379.e16. https://doi.org/10.1016/j.cmi.2015.11.020.

Blancas, D., M. Santin, M. Olmo, F. Alcaide, J. Carratala, and F. Gudiol. 2004. "Group B streptococcal disease in nonpregnant adults: incidence, clinical characteristics, and outcome." *European Journal of Clinical Microbiology & Infectious Diseases* 23 (3): 168–73. https://doi.org/10.1007/s10096-003-1098-9.

Bliss, Sandra J., Shannon D. Manning, Patricia Tallman, Carol J. Baker, Mark D. Pearlman, Carl F. Marrs, and Betsy Foxman. 2002. "Group B *Streptococcus* colonization in male and nonpregnant female university students: a cross-sectional prevalence study." *Clinical Infectious Diseases* 34 (2): 184–190.

Bohnsack, J. F. 2002. "Long-Range mapping of the *Streptococcus agalactiae* phylogenetic lineage restriction digest pattern type III-3 reveals clustering of virulence genes." *Infection and Immunity* 70 (1): 134–39. https://doi.org/10.1128/IAI.70.1.134-139.2002.

Bohnsack, John F., April Whiting, Marcelo Gottschalk, Diane Marie Dunn, Robert Weiss, Parvin H. Azimi, Joseph B. Philips, Leonard E. Weisman, George G. Rhoads, and Feng-Ying C. Lin. 2008. "Population structure of invasive and colonizing strains of *Streptococcus agalactiae* from neonates of six U.S. academic centers from 1995 to 1999." *Journal of Clinical Microbiology* 46 (4): 1285–91. https://doi.org/10.1128/JCM.02105-07.

Bolotin, Alexander, Benoit Quinquis, Alexei Sorokin, and S. Dusko Ehrlich. 2005. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) have spacers of extrachromosomal origin." *Microbiology (Reading, England)* 151 (Pt 8): 2551–61. https://doi.org/10.1099/mic.0.28048-0.

Both, Ulrich von, Andrea John, Kirsten Fluegge, Anette Siedler, and Reinhard Berner. 2008. "Molecular epidemiology of invasive neonatal *Streptococcus agalactiae* isolates in Germany." *The Pediatric Infectious Disease Journal* 27 (10): 903–6. https://doi.org/10.1097/INF.0b013e318178d1ff.

Boyer, K. M., and S. P. Gotoff. 1986. "Prevention of early-onset neonatal group B streptococcal disease with selective intrapartum chemoprophylaxis." *The New England Journal of Medicine* 314 (26): 1665–69. https://doi.org/10.1056/NEJM198606263142603.

Brochet, Mathieu, Elisabeth Couvé, Mohamed Zouine, Tatiana Vallaeys, Christophe Rusniok, Marie-Cécile Lamy, Carmen Buchrieser, et al. 2006. "Genomic diversity and evolution within the species *Streptococcus agalactiae*." *Microbes and Infection* 8 (5): 1227–43. https://doi.org/10.1016/j.micinf.2005.11.010.

Brussow, H., C. Canchaya, and W.-D. Hardt. 2004. "Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion." *Microbiology and Molecular Biology Reviews* 68 (3): 560–602. https://doi.org/10.1128/MMBR.68.3.560-602.2004.

Bulle, Béatrice, Laurence Millon, Jean-Mathieu Bart, Montserrat Gállego, Françoise Gambarelli, Montserrat Portús, Lee Schnur, et al. 2002. "Practical approach for typing strains of *Leishmania infantum* by microsatellite analysis." *Journal of Clinical Microbiology* 40 (9): 3391–97. https://doi.org/10.1128/JCM.40.9.3391-3397.2002.

Burstein, David, Christine L. Sun, Christopher T. Brown, Itai Sharon, Karthik Anantharaman, Alexander J. Probst, Brian C. Thomas, and Jillian F. Banfield. 2016. "Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems." *Nature Communications* 7 (February): 10613. https://doi.org/10.1038/ncomms10613.

Campbell, J. R., S. L. Hillier, M. A. Krohn, P. Ferrieri, D. F. Zaleznik, and C. J. Baker. 2000. "Group B streptococcal colonization and serotype-specific immunity in pregnant women at delivery." *Obstetrics and Gynecology* 96 (4): 498–503.

Charpentier, Emmanuelle, Hagen Richter, John van der Oost, and Malcolm F. White. 2015. "Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity." *FEMS Microbiology Reviews* 39 (3): 428–41. https://doi.org/10.1093/femsre/fuv023.

Chatellier, Sonia, Helene Huet, Samia Kenzi, Agnes Rosenau, Pierre Geslin, and Roland Quentin. 1996. "Genetic diversity of RRNA operons of unrelated *Streptococcus agalactiae* strains isolated from cerebrospinal fluid of neonates suffering from meningitis." *Journal of Clinical Microbiology* 34 (11): 2741–2747.

Chatellier, Sonia, Celine Ramanantsoa, Patrick Harriau, Karine Rolland, Agnes Rosenau, and Roland Quentin. 1997. "Characterization of *Streptococcus agalactiae* strains by randomly amplified polymorphic DNA analysis." *Journal of Clinical Microbiology* 35 (10): 2573–2579.

Chau, Man L., Swaine L. Chen, Min Yap, Sri H. P. Hartantyo, Paul K. T. Chiew, Charlene J. Fernandez, Wai K. Wong, et al. 2017. "Group B *Streptococcus* infections caused by improper sourcing and handling of fish for raw consumption, Singapore, 2015-2016." *Emerging Infectious Diseases* 23 (12). https://doi.org/10.3201/eid2312.170596.

Chin, Chen-Shan, Jon Sorenson, Jason B. Harris, William P. Robins, Richelle C. Charles, Roger R. Jean-Charles, James Bullard, et al. 2011. "The Origin of the haitian Cholera

Outbreak Strain." *New England Journal of Medicine* 364 (1): 33–42. https://doi.org/10.1056/NEJMoa1012928.

Christensen, K. K., and P. Christensen. 1978. "Typing of Group B *Streptococci* from the throat and urogenital tract of females." *Scandinavian Journal of Infectious Diseases* 10 (3): 209–12.

Chylinski, K., K. S. Makarova, E. Charpentier, and E. V. Koonin. 2014. "Classification and evolution of Type II CRISPR-Cas systems." *Nucleic Acids Research* 42 (10): 6091–6105. https://doi.org/10.1093/nar/gku241.

Cieslewicz, M. J., D. Chaffin, G. Glusman, D. Kasper, A. Madan, S. Rodrigues, J. Fahey, M. R. Wessels, and C. E. Rubens. 2005. "Structural and genetic diversity of group B *Streptococcus* capsular polysaccharides." *Infection and Immunity* 73 (5): 3096–3103. https://doi.org/10.1128/IAI.73.5.3096-3103.2005.

CNGOF. 1997. "CNGOF - Recommandations Pour La Pratique Médicale - Infection Cervico-Vaginales et Grossesse (1997)." 1997. http://www.cngof.asso.fr/D_PAGES/PURPC_02.HTM.

CNR des Streptocoques. 2014. "Rapport d'activité Du CNR Des Streptocoques." 2014. https://cnr-strep.fr/images/CNR-STREP/rapport/Rapport_activite_CNR_Strep_2014.pdf.

CNR des Streptocoques. 2016. "Rapport d'activité Du CNR Des Streptocoques."

CNR des Streptocoques. "CNR-Strep. Expertise." Accessed June 11, 2018. https://cnr-strep.fr/index.php/expertises/recherche-de-lien-de-clonalite.

Cui, Yujun, Yanjun Li, Olivier Gorgé, Mikhail E. Platonov, Yanfeng Yan, Zhaobiao Guo, Christine Pourcel, et al. 2008. "Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats." Edited by Niyaz Ahmed. *PLoS ONE* 3 (7): e2652. https://doi.org/10.1371/journal.pone.0002652.

Da Cunha, Violette, Mark R. Davies, Pierre-Emmanuel Douarre, Isabelle Rosinski-Chupin, Immaculada Margarit, Sebastien Spinali, Tim Perkins, et al. 2014. "*Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline." *Nature Communications* 5 (August). https://doi.org/10.1038/ncomms5544.

Dallman, Timothy J., Lisa Byrne, Philip M. Ashton, Lauren A. Cowley, Neil T. Perry, Goutam Adak, Liljana Petrovska, et al. 2015. "Whole-genome sequencing for national surveillance of shiga toxin–producing *Escherichia Coli* O157." *Clinical Infectious Diseases* 61 (3): 305–12. https://doi.org/10.1093/cid/civ318.

Deltcheva, Elitza, Krzysztof Chylinski, Cynthia M. Sharma, Karine Gonzales, Yanjie Chao, Zaid A. Pirzada, Maria R. Eckert, Jörg Vogel, and Emmanuelle Charpentier. 2011. "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." *Nature* 471 (7340): 602–7. https://doi.org/10.1038/nature09886.

Denning, D. W., C. J. Baker, N. J. Troup, and L. S. Tompkins. 1989. "Restriction endonuclease analysis of human and bovine group B *Streptococci* for epidemiologic study." *Journal of Clinical Microbiology* 27 (6): 1352–1356.

Deplano, Ariane, Annette Schuermans, Johan Van Eldere, Wolfgang Witte, Hèléne Meugnier, Jerome Etienne, Hajo Grundmann, et al. 2000. "Multicenter evaluation of epidemiological typing of methicillin-resistant *Staphylococcus aureus* Strains by Repetitive-Element PCR *a*nalysis." *Journal of Clinical Microbiology* 38 (10): 3527–3533.

Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath, and S. Moineau. 2008. "Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*." *Journal of Bacteriology* 190 (4): 1390–1400. https://doi.org/10.1128/JB.01412-07.

Diedrick, M. J., A. E. Flores, S. L. Hillier, R. Creti, and P. Ferrieri. 2010. "Clonal analysis of colonizing group B *streptococcus,* serotype IV, an emerging pathogen in the United States." *Journal of Clinical Microbiology* 48 (9): 3100–3104. https://doi.org/10.1128/JCM.00277-10.

Dillon, H. C., E. Gray, M. A. Pass, and B. M. Gray. 1982. "Anorectal and vaginal carriage of group B *Streptococci* during pregnancy." *The Journal of Infectious Diseases* 145 (6): 794–99.

Domelier, A.-S., N. van der Mee-Marquet, L. Arnault, L. Mereghetti, P. Lanotte, A. Rosenau, M.-F. Lartigue, and R. Quentin. 2008. "Molecular characterization of erythromycin-resistant *streptococcus agalactiae* strains." *Journal of Antimicrobial Chemotherapy* 62 (6): 1227–33. https://doi.org/10.1093/jac/dkn388.

Edmond, Karen M, Christina Kortsalioudaki, Susana Scott, Stephanie J Schrag, Anita KM Zaidi, Simon Cousens, and Paul T Heath. 2012. "Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis." *The Lancet* 379 (9815): 547–56. https://doi.org/10.1016/S0140-6736(11)61651-6.

Eickhoff, Theodore C., Jerome O. Klein, A. Kathleen Daly, David Ingall, and Maxwell Finland. 1964. "Neonatal sepsis and other infections due to group B beta-hemolytic *Streptococci*." *New England Journal of Medicine* 271 (24): 1221–28. https://doi.org/10.1056/NEJM196412102712401.

Elliott, J. A., K. D. Farmer, and R. R. Facklam. 1998. "Sudden Increase in isolation of group B *Streptococci*, serotype V, is not due to emergence of a new pulsed-field gel electrophoresis type." *Journal of Clinical Microbiology* 36 (7): 2115–2116.

Eyre, David W, Tanya Golubchik, N Claire Gordon, Rory Bowden, Paolo Piazza, Elizabeth M Batty, Camilla L C Ip, et al. 2012. "A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium Difficile* for outbreak detection and surveillance." *BMJ Open* 2 (3): e001124. https://doi.org/10.1136/bmjopen-2012-001124.

Fabre, Laëtitia, Jian Zhang, Ghislaine Guigon, Simon Le Hello, Véronique Guibert, Marie Accou-Demartin, Saïana de Romans, et al. 2012. "CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections." *PloS One* 7 (5): e36995. https://doi.org/10.1371/journal.pone.0036995.

Farley, M. M. 2001. "Group B streptococcal disease in nonpregnant adults." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 33 (4): 556–61. https://doi.org/10.1086/322696.

Fasola, ELIZABETH, CAROL Livdahl, and PATRICIA Ferrieri. 1993. "Molecular analysis of multiple isolates of the major serotypes of group B *Streptococci*." *Journal of Clinical Microbiology* 31 (10): 2616–2620.

Feil, Edward J., Bao C. Li, David M. Aanensen, William P. Hanage, and Brian G. Spratt. 2004. "EBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data." *Journal of Bacteriology* 186 (5): 1518–30. https://doi.org/10.1128/JB.186.5.1518-1530.2004.

Ferrieri, Patricia, Ruth Lynfield, Roberta Creti, and Aurea E. Flores. 2013. "Serotype IV and invasive group B *Streptococcus* disease in neonates, Minnesota, USA, 2000–20101." *Emerging Infectious Diseases* 19 (4): 551–58. https://doi.org/10.3201/eid1904.121572.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, and J. M. Merrick. 1995. "Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd." *Science (New York, N.Y.)* 269 (5223): 496–512.

Florindo, C., V. Damiao, I. Silvestre, C. Farinha, F. Rodrigues, F. Nogueira, F. Martins-Pereira, R. Castro, and M. J. Borrego. 2014. "Epidemiological surveillance of colonising group B *Streptococcus* epidemiology in the lisbon and Tagus Valley regions, Portugal (2005 to 2012): Emergence of a New Epidemic Type IV/Clonal Complex 17 Clone." Methods 10 (16): 20.

Foxman, B. 2006. "Incidence and duration of group B *Streptococcus* by serotype among male and female college students living in a single dormitory." *American Journal of Epidemiology* 163 (6): 544–51. https://doi.org/10.1093/aje/kwj075.

Fricke, W. F., M. K. Mammel, P. F. McDermott, C. Tartera, D. G. White, J. E. LeClerc, J. Ravel, and T. A. Cebula. 2011. "Comparative genomics of 28 *Salmonella Enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution." *Journal of Bacteriology* 193 (14): 3556–68. https://doi.org/10.1128/JB.00297-11.

Frothingham, Richard, and Winifred A. Meeker-O'Connell. 1998. "Genetic diversity in the *Mycobacterium Tuberculosis* complex based on variable numbers of tandem DNA repeats." *Microbiology* 144 (5): 1189–1196.

Fry, R.M. 1938. "Fatal infections by haemolytic *Streptococcus* group B." *The Lancet* 231 (5969): 198–199.

Gardy, Jennifer L., James C. Johnston, Shannan J. Ho Sui, Victoria J. Cook, Lena Shah, Elizabeth Brodkin, Shirley Rempel, et al. 2011. "Whole-genome sequencing and social-network analysis of a tuberculosis outbreak." *New England Journal of Medicine* 364 (8): 730–739.

Gasiunas, G., R. Barrangou, P. Horvath, and V. Siksnys. 2012. "Cas9-CrRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." *Proceedings of the National Academy of Sciences* 109 (39): E2579–86. https://doi.org/10.1073/pnas.1208507109.

Ginevra, C., N. Jacotin, L. Diancourt, G. Guigon, R. Arquilliere, H. Meugnier, G. Descours, et al. 2012. "*Legionella Pneumophila* sequence type 1/Paris pulsotype subtyping by spoligotyping." *Journal of Clinical Microbiology* 50 (3): 696–701. https://doi.org/10.1128/JCM.06180-11.

Glaser, Philippe, Christophe Rusniok, Carmen Buchrieser, Fabien Chevalier, Lionel Frangeul, Tarek Msadek, Mohamed Zouine, et al. 2002. "genome Sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease." *Molecular Microbiology* 45 (6): 1499–1513.

Gordillo, M. E., K. V. Singh, C. J. Baker, and B. E. Murray. 1993. "Typing of group B *Streptococci*: comparison of pulsed-field gel electrophoresis and conventional electrophoresis." *Journal of Clinical Microbiology* 31 (6): 1430–1434.

Granlund, M., L. Oberg, M. Sellin, and M. Norgren. 1998. "Identification of a novel insertion element, IS1548, in group B *Streptococci*, predominantly in strains causing endocarditis." *The Journal of Infectious Diseases* 177 (4): 967–76.

Grissa, Ibtissem, Gilles Vergnaud, and Christine Pourcel. 2007. "The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats." *BMC Bioinformatics* 8 (1): 172. https://doi.org/10.1186/1471-2105-8-172.

Groenen, Peter, Annelies E. Bunschoten, Dick van Soolingen, and Jan DA van Errtbden. 1993. "Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis;* application for strain differentiation by a novel typing method." *Molecular Microbiology* 10 (5): 1057–1065.

Haguenoer, Eve, Gaelle Baty, Christine Pourcel, Marie-Frédérique Lartigue, Anne-Sophie Domelier, Agnès Rosenau, Roland Quentin, Laurent Mereghetti, and Philippe Lanotte. 2011. "A Multi Locus Variable Number of Tandem Repeat Analysis (MLVA) scheme for *Streptococcus agalactiae* genotyping." *BMC Microbiology* 11 (1): 171.

Håkansson, Stellan, Pia Axemo, Katarina Bremme, Anna-Lena Bryngelsson, Marie Carlsson Wallin, Carl-Magnus Ekström, Margareta Granlund, et al. 2008. "Group B streptococcal carriage in sweden: a national study on risk factors for mother and infant colonisation." *Acta Obstetricia Et Gynecologica Scandinavica* 87 (1): 50–58. https://doi.org/10.1080/00016340701802888.

Hale, Caryn, Kyle Kleppe, Rebecca M. Terns, and Michael P. Terns. 2008. "Prokaryotic silencing (Psi)RNAs in *Pyrococcus furiosus*." *RNA (New York, N.Y.)* 14 (12): 2572–79. https://doi.org/10.1261/rna.1246808.

Hansen, Søren Mose, Niels Uldbjerg, Mogens Kilian, and Uffe B. Skov Sørensen. 2004. "Dynamics of *Streptococcus agalactiae* colonization in women during and after pregnancy and in their infants." *Journal of Clinical Microbiology* 42 (1): 83–89.

Harrison, Lee H., Diane M. Dwyer, and Judith A. Johnson. 1995. "Emergence of Serotype V Group B streptococcal infection among infants and adults." *The Journal of Infectious Diseases* 171 (2): 511–513.

Hauge, Majbritt, Christina Jespersgaard, Knud Poulsen, and Mogens Kilian. 1996. "Population structure of *Streptococcus agalactiae* reveals an association between specific evolutionary lineages and putative virulence factors but not disease." *Infection and Immunity* 64 (3): 919–925.

Hedegaard, Jakob, Majbritt Hauge, Jeppe Fage-Larsen, Kim Kusk Mortensen, Mogens Kilian, Hans Uffe Sperling-Petersen, and Knud Poulsen. 2000. "Investigation of the translation-initiation factor IF2 Gene, *InfB*, as a tool to study the population structure of *Streptococcus agalactiae*." *Microbiology* 146 (7): 1661–1670.

Heidelberg, John F., William C. Nelson, Thomas Schoenfeld, and Devaki Bhaya. 2009. "Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes." Edited by Niyaz Ahmed. *PLoS ONE* 4 (1): e4169. https://doi.org/10.1371/journal.pone.0004169.

Held, Nicole L., Alfa Herrera, Hinsby Cadillo-Quiroz, and Rachel J. Whitaker. 2010. "CRISPR associated diversity within a population of *Sulfolobus islandicus*." Edited by Paul J. Planet. *PLoS ONE* 5 (9): e12988. https://doi.org/10.1371/journal.pone.0012988.

Held, Nicole L., and Rachel J. Whitaker. 2009. "Viral biogeography revealed by signatures in *Sulfolobus Islandicus* genomes." *Environmental Microbiology* 11 (2): 457–66. https://doi.org/10.1111/j.1462-2920.2008.01784.x.

Heler, Robert, Poulami Samai, Joshua W. Modell, Catherine Weiner, Gregory W. Goldberg, David Bikard, and Luciano A. Marraffini. 2015. "Cas9 specifies functional viral targets during CRISPR–Cas adaptation." *Nature* 519 (7542): 199–202. https://doi.org/10.1038/nature14245.

Helmig, Rikke, Niels Uldbjerg, Jane Boris, and Mogens Kilian. 1993. "Clonal analysis of *Streptococcus agalactiae* isolated from infants with neonatal sepsis or meningitis and their mothers and from healthy pregnant women." *Journal of Infectious Diseases* 168 (4): 904–909.

Herbert, Mark A., Catriona JE Beveridge, David McCormick, Emmelien Aten, Nicola Jones, Lori AS Snyder, and Nigel J. Saunders. 2005. "Genetic islands of *Streptococcus agalactiae* strains NEM316 and 2603VR and their presence in other group B streptococcal strains." *BMC Microbiology* 5 (1): 31.

Hery-Arnaud, G., G. Bruant, P. Lanotte, S. Brun, B. Picard, A. Rosenau, N. van der Mee-Marquet, P. Rainard, R. Quentin, and L. Mereghetti. 2007. "Mobile genetic elements provide evidence for a bovine origin of clonal complex 17 of *Streptococcus agalactiae*." *Applied and Environmental Microbiology* 73 (14): 4668–72. https://doi.org/10.1128/AEM.02604-06.

Hery-Arnaud, G., G. Bruant, P. Lanotte, S. Brun, A. Rosenau, N. van der Mee-Marquet, R. Quentin, and L. Mereghetti. 2005. "Acquisition of insertion sequences and the *GBSi1* intron by *Streptococcus agalactiae* isolates correlates with the evolution of the species." *Journal of Bacteriology* 187 (17): 6248–52. https://doi.org/10.1128/JB.187.17.6248-6252.2005.

Hidalgo-Cantabrana, Claudio, Rosemary Sanozky-Dawes, and Rodolphe Barrangou. 2018. "Insights into the human virome using CRISPR spacers from microbiomes." *Viruses* 10 (9). https://doi.org/10.3390/v10090479.

Hoe, Nancy, Kazumitsu Nakashima, Diana Grigsby, Xi Pan, Shu Jun Dou, Steven Naidich, Marianne Garcia, Emily Kahn, David Bergmire-Sweat, and James M. Musser. 1999. "Rapid molecular genetic subtyping of serotype M1 Group A *Streptococcus* strains." *Emerging Infectious Diseases* 5 (2): 254.

Horvath, Philippe, and Rodolphe Barrangou. 2010. "CRISPR/Cas, the immune system of Bacteria and Archaea." *Science (New York, N.Y.)* 327 (5962): 167–70. https://doi.org/10.1126/science.1179555.

Horvath, Philippe, Dennis A. Romero, Anne-Claire Coûté-Monvoisin, Melissa Richards, Hélène Deveau, Sylvain Moineau, Patrick Boyaval, Christophe Fremaux, and Rodolphe Barrangou. 2008. "Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*." *Journal of Bacteriology* 190 (4): 1401–12. https://doi.org/10.1128/JB.01415-07.

Huber, C. A., F. McOdimba, V. Pflueger, C. A. Daubenberger, and G. Revathi. 2011. "Characterization of invasive and colonizing isolates of *Streptococcus agalactiae* in east african adults." *Journal of Clinical Microbiology* 49 (10): 3652–55. https://doi.org/10.1128/JCM.01288-11.

Huet, H., C. Martin, P. Geslin, F. Grimont, and R. Quentin. 1993. "Ribotyping of *Streptococcus agalactiae* strains isolated from vaginas of asymptomatic women." *Research in Microbiology* 144 (6): 457–65. https://doi.org/10.1016/0923-2508(93)90053-5.

Hunter, P. R., and M. A. Gaston. 1988. “Numerical index of the discriminatory ability of typing systems: an application of Simpson’s index of diversity.” *Journal of Clinical Microbiology* 26 (11): 2465–66.

Imperi, Monica, Marco Pataracchia, Giovanna Alfarone, Lucilla Baldassarri, Graziella Orefici, and Roberta Creti. 2010. “A Multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of *Streptococcus agalactiae*.” *Journal of Microbiological Methods* 80 (2): 212–14. https://doi.org/10.1016/j.mimet.2009.11.010.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakata. 1987. “Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in E*scherichia coli,* and identification of the gene product.” *Journal of Bacteriology* 169 (12): 5429–33.

Jackson, Lisa A., Roberta Hilsdon, Monica M. Farley, Lee H. Harrison, Arthur L. Reingold, Brian D. Plikaytis, Jay D. Wenger, and Anne Schuchat. 1995. “Risk factors for group B streptococcal disease in adults.” *Annals of Internal Medicine* 123 (6): 415–420.

Jackson, Paul J., Eliza A. Walthers, Abdullahi S. Kalif, Kara L. Richmond, Debra M. Adair, Karen K. Hill, Cheryl R. Kuske, et al. 1997. “Characterization of the variable-number tandem repeats in VrrA from different *Bacillus Anthracis* isolates.” *Applied and Environmental Microbiology* 63 (4): 1400–1405.

Jansen, Rund, Jam D. A. van Embden, Wim Gaastra, and Leo M. Schouls. 2002. “Identification of a novel family of sequence repeats among prokaryotes.” *Omics: A Journal of Integrative Biology* 6 (1): 23–33. https://doi.org/10.1089/15362310252780816.

Jansen, Ruud, Jan D. A. van Embden, Wim Gaastra, and Leo M. Schouls. 2002. “Identification of genes that are associated with DNA repeats in prokaryotes.” *Molecular Microbiology* 43 (6): 1565–75.

Jauréguy, F., M. Carton, J. Teboul, M.-J. Butel, P. Panel, J.-C. Ghnassia, and F. Doucet-Populaire. 2003. “[Risk factors and screening strategy for group B streptococcal colonization in pregnant women: results of a prospective study].” *Journal De Gynécologie, Obstétrique Et Biologie De La Reproduction* 32 (2): 132–38.

Jelínková, JARMILA, and JITKA Motlová. 1985. “MLVA.” *Journal of Clinical Microbiology* 21 (3): 361–362.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. “A Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” *Science (New York, N.Y.)* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Joensen, K. G., F. Scheutz, O. Lund, H. Hasman, R. S. Kaas, E. M. Nielsen, and F. M. Aarestrup. 2014. “Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*.” *Journal of Clinical Microbiology* 52 (5): 1501–10. https://doi.org/10.1128/JCM.03617-13.

Jones, Nicola, John F. Bohnsack, Shinji Takahashi, Karen A. Oliver, Man-Suen Chan, Frank Kunst, Philippe Glaser, et al. 2003. “Multilocus sequence typing system for group B *Streptococcus*.” *Journal of Clinical Microbiology* 41 (6): 2530–36.

Jones, Nicola, Karen A. Oliver, Joanne Barry, Rosalind M. Harding, Naiel Bisharat, Brian G. Spratt, Tim Peto, Derrick W. Crook, and Oxford Group B. Streptococcus Consortium. 2006. “Enhanced invasiveness of bovine-derived neonatal sequence type 17 Group B *Streptococcus* is independent of capsular serotype.” *Clinical Infectious Diseases* 42 (7): 915–924.

Jourdan-Da Silva, Nathalie J, Denise Antona, Caroline Six, scarlett Georges, Véronique Goulet, Philippe Judlin, and Daniel Lévy-Bruhl. 2008. "Infections néonatales à Streptocoque B En France : Données d'incidence de 1997 à 2006 et Pr Atiques de Prévention En Maternité," 2008. http://opac.invs.sante.fr/doc_num.php?explnum_id=3286.

Kamerbeek, Judith, L. E. O. Schouls, Arend Kolk, Miranda Van Agterveld, Dick Van Soolingen, Sjoukje Kuijper, Annelies Bunschoten, et al. 1997. "Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology." *Journal of Clinical Microbiology* 35 (4): 907–914.

Keefe, Gregory P. 1997. "*Streptococcus agalactiae* mastitis: a review." *The Canadian Veterinary Journal* 38 (7): 429.

Kexel, G., and S. Schönbohm. 1965. "S*treptococcus agalactiae* als Erreger von Säuglingsmeningitiden." *DMW - Deutsche Medizinische Wochenschrift* 90 (06): 258–61. https://doi.org/10.1055/s-0028-1111328.

Kiely, R. A., L. Cotter, A. M. Mollaghan, B. Cryan, A. Coffey, and B. Lucey. 2011. "Emergence of Group B *Streptococcus* serotype IV in women of child-bearing age in Ireland." *Epidemiology and Infection* 139 (02): 236–38. https://doi.org/10.1017/S0950268810001275.

Kieper, Sebastian N., Cristóbal Almendros, Juliane Behler, Rebecca E. McKenzie, Franklin L. Nobrega, Anna C. Haagsma, Jochem N.A. Vink, Wolfgang R. Hess, and Stan J.J. Brouns. 2018. "Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation." *Cell Reports* 22 (13): 3377–84. https://doi.org/10.1016/j.celrep.2018.02.103.

Ko, W. C., H. C. Lee, L. R. Wang, C. T. Lee, A. J. Liu, and J. J. Wu. 2001. "Serotyping and antimicrobial susceptibility of Group B *Streptococcus* over an eight-year period in southern Taiwan." *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 20 (5): 334–39.

Koeck, Jean-Louis, Berthe-Marie Njanpop-Lafourcade, Sonia Cade, Emmanuelle Varon, Lassana Sangare, Samina Valjevac, Gilles Vergnaud, and Christine Pourcel. 2005. "Evaluation and selection of tandem repeat loci for S*treptococcus pneumoniae* MLVA strain typing." *BMC Microbiology* 5 (November): 66. https://doi.org/10.1186/1471-2180-5-66.

Koonin, Eugene V, Kira S Makarova, and Feng Zhang. 2017. "Diversity, classification and evolution of CRISPR-Cas systems." *Current Opinion in Microbiology* 37 (June): 67–78. https://doi.org/10.1016/j.mib.2017.05.008.

Köser, Claudio U., Matthew TG Holden, Matthew J. Ellington, Edward JP Cartwright, Nicholas M. Brown, Amanda L. Ogilvy-Stuart, Li Yang Hsu, et al. 2012. "Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak." *New England Journal of Medicine* 366 (24): 2267–2275.

Krohn, Marijane A., Sharon L. Hillier, and Carol J. Baker. 1999. "Maternal peripartum complications associated with vaginal group B *streptococci* colonization." *Journal of Infectious Diseases* 179 (6): 1410–1415.

Kunin, Victor, Rotem Sorek, and Philip Hugenholtz. 2007. "Evolutionary conservation of sequence and secondary structures in CRISPR Repeats." *Genome Biol* 8 (4): R61.

Kwatra, Gaurav, Marianne C Cunnington, Elizabeth Merrall, Peter V Adrian, Margaret Ip, Keith P Klugman, Wing Hung Tam, and Shabir A Madhi. 2016. "Prevalence of

maternal colonisation with group B *Streptococcus*: A Systematic Review and Meta-analysis." *The Lancet Infectious Diseases* 16 (9): 1076–84. https://doi.org/10.1016/S1473-3099(16)30055-X.

Laboratoire de microbiologie du réseau EPIBAC, InVS. 2016. "Bulletin Du Réseau de Surveillance Des Infections Invasives Bactériennes - Reseau EPIBAC." 2016. http://invs.santepubliquefrance.fr/Dossiers-thematiques/Maladies-infectieuses/Maladies-a-prevention-vaccinale/Infections-invasives-d-origine-bacterienne-Reseau-EPIBAC/Bulletin-du-reseau-de-surveillance-des-infections-invasives-bacteriennes.

Labrie, Simon J., Julie E. Samson, and Sylvain Moineau. 2010. "Bacteriophage resistance mechanisms." *Nature Reviews. Microbiology* 8 (5): 317–27. https://doi.org/10.1038/nrmicro2315.

Lachenauer, C. S., D. L. Kasper, J. Shimada, Y. Ichiman, H. Ohtsuka, M. Kaku, L. C. Paoletti, P. Ferrieri, and L. C. Madoff. 1999. "Serotypes VI and VIII predominate among Group B *Streptococci* isolated from pregnant japanese women." *The Journal of Infectious Diseases* 179 (4): 1030–1033.

Lamagni, T. L., C. Keshishian, A. Efstratiou, R. Guy, K. L. Henderson, K. Broughton, and E. Sheridan. 2013. "Emerging trends in the epidemiology of invasive group B streptococcal disease in England and Wales, 1991-2010." *Clinical Infectious Diseases* 57 (5): 682–88. https://doi.org/10.1093/cid/cit337.

Lambertsen, L., K. Ekelund, I. C. Skovsted, A. Liboriussen, and H.-C. Slotved. 2010. "Characterisation of invasive group B *Streptococci* from adults in Denmark 1999 to 2004." *European Journal of Clinical Microbiology & Infectious Diseases* 29 (9): 1071–77. https://doi.org/10.1007/s10096-010-0941-z.

Lamy, Marie-Cécile, Shaynoor Dramsi, Annick Billoët, Hélène Réglier-Poupet, Asmaa Tazi, Josette Raymond, François Guérin, et al. 2006. "Rapid detection of the 'highly virulent' group B *Streptococcus* ST-17 Clone." *Microbes and Infection* 8 (7): 1714–22. https://doi.org/10.1016/j.micinf.2006.02.008.

Lamy, Marie-Cécile, Mohammed Zouine, Juliette Fert, Massimo Vergassola, Elisabeth Couve, Elisabeth Pellegrini, Philippe Glaser, et al. 2004. "*CovS/CovR* of Group B *Streptococcus:* a two-component global regulatory system involved in virulence." *Molecular Microbiology* 54 (5): 1250–68. https://doi.org/10.1111/j.1365-2958.2004.04365.x.

Lancefield, Rebecca. 1934. "A serological differentiation of specific types of bovine hemolytic *Streptococci* (GROUP B)."http://jem.rupress.org/content/jem/59/4/441.full.pdf

Le Doare, Kirsty, and Paul T. Heath. 2013. "An overview of global GBS epidemiology." *Vaccine* 31 (August): D7–12. https://doi.org/10.1016/j.vaccine.2013.01.009.

Le Thomas-Bories, I., F. Fitoussi, P. Mariani-Kurkdjian, J. Raymond, N. Brahimi, P. Bidet, V. Lefranc, and E. Bingen. 2001. "Clonal relationship between U.S. and french serotype V group B *Streptococcus* isolates." *Journal of Clinical Microbiology* 39 (12): 4526–28. https://doi.org/10.1128/JCM.39.12.4526-4528.2001.

Leal, Carlos A. G., Guilherme A. Queiroz, Felipe L. Pereira, Guilherme C. Tavares, and Henrique C. P. Figueiredo. 2019. "*Streptococcus agalactiae* sequence type 283 in farmed fish, Brazil." *Emerging Infectious Diseases* 25 (4): 776–79. https://doi.org/10.3201/eid2504.180543.

Lee, B. K., Y. R. Song, M. Y. Kim, J. H. Yang, J. H. Shin, Y. S. Seo, K. Y. Oh, et al. 2010. "Epidemiology of Group B S*treptococcus* in korean pregnant women." *Epidemiology and Infection* 138 (02): 292. https://doi.org/10.1017/S0950268809990859.

Leenay, Ryan T., and Chase L. Beisel. 2017. "Deciphering, communicating, and engineering the CRISPR PAM." *Journal of Molecular Biology* 429 (2): 177–91. https://doi.org/10.1016/j.jmb.2016.11.024.

Lewin, E. B., and M. S. Amstey. 1981. "Natural history of group B *Streptococcus* colonization and its therapy during pregnancy." *American Journal of Obstetrics and Gynecology* 139 (5): 512–15.

Li, Shunming, Jingya Huang, Zhiyao Chen, Dan Guo, Zhenjiang Yao, and Xiaohua Ye. 2017. "Antibiotic prevention for maternal group B streptococcal colonization on neonatal gbs-related adverse outcomes: a meta-analysis." *Frontiers in Microbiology* 8 (March). https://doi.org/10.3389/fmicb.2017.00374.

Lier, Clément, Elodie Baticle, Philippe Horvath, Eve Haguenoer, Anne-Sophie Valentin, Philippe Glaser, Laurent Mereghetti, and Philippe Lanotte. 2015. "Analysis of the Type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages." *Frontiers in Genetics* 6. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466440/.

Lillestøl, Reidun K., Peter Redder, Roger A. Garrett, and Kim Brügger. 2006. "A putative viral defence mechanism in archaeal cells." *Archaea (Vancouver, B.C.)* 2 (1): 59–72.

Lin, F.-Y. C., A. Whiting, E. Adderson, S. Takahashi, D. M. Dunn, R. Weiss, P. H. Azimi, et al. 2006. "Phylogenetic lineages of invasive and colonizing strains of serotype III Group B *Streptococci* from neonates: A Multicenter Prospective Study." *Journal of Clinical Microbiology* 44 (4): 1257–61. https://doi.org/10.1128/JCM.44.4.1257-1261.2006.

Lindstedt, Bjørn-Arne. 2005. "Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria." *ELECTROPHORESIS* 26 (13): 2567–82. https://doi.org/10.1002/elps.200500096.

Liu, F., R. Barrangou, P. Gerner-Smidt, E. M. Ribot, S. J. Knabel, and E. G. Dudley. 2011. "Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* Subsp. Enterica." *Applied and Environmental Microbiology* 77 (6): 1946–56. https://doi.org/10.1128/AEM.02625-10.

Liu, F., S. Kariyawasam, B. M. Jayarao, R. Barrangou, P. Gerner-Smidt, E. M. Ribot, S. J. Knabel, and E. G. Dudley. 2011. "Subtyping *Salmonella enterica* Serovar Enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs)." *Applied and Environmental Microbiology* 77 (13): 4520–26. https://doi.org/10.1128/AEM.00468-11.

Liu, G, J Zhu, K Chen, T Gao, H Yao, Y Liu, W Zhang, and C Lu. 2016. "Development of *Streptococcus agalactiae* vaccines for Tilapia." *Diseases of Aquatic Organisms* 122 (2): 163–70. https://doi.org/10.3354/dao03084.

Lopez-Sanchez, Maria-José, Elisabeth Sauvage, Violette Da Cunha, Dominique Clermont, Elisoa Ratsima Hariniaina, Bruno Gonzalez-Zorn, Claire Poyart, Isabelle Rosinski-Chupin, and Philippe Glaser. 2012. "The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome." *Molecular Microbiology* 85 (6): 1057–71. https://doi.org/10.1111/j.1365-2958.2012.08172.x.

Luan, S.-L., M. Granlund, M. Sellin, T. Lagergard, B. G. Spratt, and M. Norgren. 2005. "Multilocus sequence typing of swedish invasive group B *Streptococcus* isolates indicates a neonatally associated genetic lineage and capsule switching." *Journal of Clinical Microbiology* 43 (8): 3727–33. https://doi.org/10.1128/JCM.43.8.3727-3733.2005.

Lunel, Frans Verduyn, Lidia Licciardello, Stefania Stefani, Henri A. Verbrugh, Willem J. G. Melchers, Jacques F. G. M. Meis, Stewart Scherer, and Alex van Belkum. 1998. "Lack of consistent short sequence repeat polymorphisms in genetically homologous colonizing and invasive *Candida albicans* strains." *Journal of Bacteriology* 180 (15): 3771–78.

Maiden, Martin CJ, Jane A. Bygraves, Edward Feil, Giovanna Morelli, Joanne E. Russell, Rachel Urwin, Qing Zhang, et al. 1998. "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms." *Proceedings of the National Academy of Sciences* 95 (6): 3140–3145.

Makarova, Kira S, Nick V Grishin, Svetlana A Shabalina, Yuri I Wolf, and Eugene V Koonin. 2006. "A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action." *Biology Direct* 1 (March): 7. https://doi.org/10.1186/1745-6150-1-7.

Makarova, Kira S., Daniel H. Haft, Rodolphe Barrangou, Stan J. J. Brouns, Emmanuelle Charpentier, Philippe Horvath, Sylvain Moineau, et al. 2011. "Evolution and classification of the CRISPR–Cas systems." *Nature Reviews Microbiology* 9 (6): 467–77. https://doi.org/10.1038/nrmicro2577.

Makarova, Kira S., Yuri I. Wolf, Omer S. Alkhnbashi, Fabrizio Costa, Shiraz A. Shah, Sita J. Saunders, Rodolphe Barrangou, et al. 2015. "An updated evolutionary classification of CRISPR-Cas systems." *Nature Reviews Microbiology* 13 (11): 722–36. https://doi.org/10.1038/nrmicro3569.

Manning, S. D., A. C. Springman, E. Lehotzky, M. A. Lewis, T. S. Whittam, and H. D. Davies. 2009a. "Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada." *Journal of Clinical Microbiology* 47 (4): 1143–48. https://doi.org/10.1128/JCM.01424-08.

Manning, Shannon D., Maggi A. Lewis, A. Cody Springman, Erica Lehotzky, Thomas S. Whittam, and H. Dele Davies. 2008. "Genotypic diversity and serotype distribution of group B *Streptococcus* isolated from women before and after delivery." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 46 (12): 1829–37. https://doi.org/10.1086/588296.

Martins, E. R., J. Melo-Cristino, M. Ramirez, and the portuguese group for the study of streptococcal infections. 2012. "Dominance of serotype Ia among group B *Streptococci* Causing invasive infections in nonpregnant adults in Portugal." *Journal of Clinical Microbiology* 50 (4): 1219–27. https://doi.org/10.1128/JCM.05488-11.

Martins, E. R., M. A. Pessanha, M. Ramirez, J. Melo-Cristino, and and the Portuguese Group for the Study of Streptococcal Infections. 2007. "Analysis of group B *streptococcal* isolates from infants and pregnant women in Portugal revealing two lineages with enhanced invasiveness." *Journal of Clinical Microbiology* 45 (10): 3224–29. https://doi.org/10.1128/JCM.01182-07.

Matsubara, Kousaku, Kiyoshi Hoshina, and Yoko Suzuki. 2013. "Early-onset and late-onset Group B streptococcal disease in Japan: a nationwide surveillance study, 2004–2010." *International Journal of Infectious Diseases* 17 (6): e379–84. https://doi.org/10.1016/j.ijid.2012.11.027.

Meehan, Mary, Robert Cunney, and Mary Cafferkey. 2014. "Molecular epidemiology of group B *Streptococci* in Ireland reveals a diverse population with evidence of capsular switching." *European Journal of Clinical Microbiology & Infectious Diseases* 33 (7): 1155–62. https://doi.org/10.1007/s10096-014-2055-5.

Mee-Marquet, N. van der, A.-S. Domelier, L. Mereghetti, P. Lanotte, A. Rosenau, W. van Leeuwen, and R. Quentin. 2006. "Prophagic DNA fragments in *Streptococcus agalactiae* strains and association with neonatal meningitis." *Journal of Clinical Microbiology* 44 (3): 1049–58. https://doi.org/10.1128/JCM.44.3.1049-1058.2006.

Mee-Marquet, N. van der, L. Fourny, L. Arnault, A.-S. Domelier, M. Salloum, M.-F. Lartigue, and R. Quentin. 2008. "Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites." *Journal of Clinical Microbiology* 46 (9): 2906–11. https://doi.org/10.1128/JCM.00421-08.

Mee-Marquet, N. van der, C. Jouannet, A.-S. Domelier, L. Arnault, M.-F. Lartigue, and R. Quentin. 2009. "genetic diversity of *Streptococcus agalactiae* strains and density of vaginal carriage." *Journal of Medical Microbiology* 58 (2): 169–73. https://doi.org/10.1099/jmm.0.005827-0.

Mee-Marquet, Nathalie van der, Anne-Sophie Domelier, Laurent Mereghetti, Philippe Lanotte, Agnès Rosenau, Willem van Leeuwen, and Roland Quentin. 2006. "Prophagic DNA fragments in *Streptococcus agalactiae* strains and association with neonatal meningitis." *Journal of Clinical Microbiology* 44 (3): 1049–58. https://doi.org/10.1128/JCM.44.3.1049-1058.2006.

Mojica, F. J., C. Díez-Villaseñor, E. Soria, and G. Juez. 2000. "Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria." *Molecular Microbiology* 36 (1): 244–46.

Mojica, F. J. M., C. Diez-Villasenor, J. Garcia-Martinez, and C. Almendros. 2009. "Short motif sequences determine the targets of the prokaryotic CRISPR defence system." *Microbiology* 155 (3): 733–40. https://doi.org/10.1099/mic.0.023960-0.

Mojica, Francisco J. M., César Díez-Villaseñor, Jesús García-Martínez, and Elena Soria. 2005. "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements." *Journal of Molecular Evolution* 60 (2): 174–82. https://doi.org/10.1007/s00239-004-0046-3.

Mokrousov, I., A. Vyazovaya, V. Kolodkina, E. Limeschenko, L. Titov, and O. Narvskaya. 2009. "Novel macroarray-based method of *Corynebacterium diphtheriae* genotyping: evaluation in a field study in Belarus." *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 28 (6): 701–3. https://doi.org/10.1007/s10096-008-0674-4.

Mokrousov, Igor, Elena Limeschenko, Anna Vyazovaya, and Olga Narvskaya. 2007. "*Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci." *Biotechnology Journal* 2 (7): 901–6. https://doi.org/10.1002/biot.200700035.

Moore, Matthew R., Stephanie J. Schrag, and Anne Schuchat. 2003. "Effects of intrapartum antimicrobial prophylaxis for prevention of group-B-streptococcal disease on the

incidence and ecology of early-onset neonatal sepsis." *The Lancet Infectious Diseases* 3 (4): 201–213.

Morozumi, Miyuki, Takeaki Wajima, Misako Takata, Satoshi Iwata, and Kimiko Ubukata. 2016. "Molecular characteristics of group B *Streptococci* isolated from adults with invasive infections in Japan." Edited by B. A. Forbes. *Journal of Clinical Microbiology* 54 (11): 2695–2700. https://doi.org/10.1128/JCM.01183-16.

Moyo, SYLVESTER R., JOHAN A. Maeland, and KÅRE Bergh. 2002. "Typing of human isolates of *Streptococcus agalactiae* (group B *Streptococcus*, GBS) strains from Zimbabwe." *Journal of Medical Microbiology* 51 (7): 595–662. https://doi.org/10.1099/0022-1317-51-7-595.

Musser, James M., Stephen J. Mattingly, Roland Quentin, Alain Goudeau, and Robert K. Selander. 1989. "Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B *Streptococcus*) causing invasive neonatal disease." *Proceedings of the National Academy of Sciences* 86 (12): 4731–4735.

Nagano, N., Y. Nagano, M. Toyama, K. Kimura, T. Tamura, K. Shibayama, and Y. Arakawa. 2012. "Nosocomial spread of multidrug-resistant group B *Streptococci* with reduced penicillin susceptibility belonging to Clonal Complex 1." *Journal of Antimicrobial Chemotherapy* 67 (4): 849–56. https://doi.org/10.1093/jac/dkr546.

Nagano, YUKIKO, N. Nagano, S. Takahashi, K. Murono, K. Fujita, F. Taguchi, and Y. Okuwaki. 1991. "restriction endonuclease digest patterns of chromosomal DNA from group B β-haemolytic *Streptococci*." *Journal of Medical Microbiology* 35 (5): 297–303.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, and E. Kumlin. 1987. "Variable Number of Tandem Repeat (VNTR) markers for human gene mapping." *Science (New York, N.Y.)* 235 (4796): 1616–22.

Nitschke, H., P. Slickers, E. Muller, R. Ehricht, and S. Monecke. 2014. "DNA microarray-based typing of *Streptococcus agalactiae* isolates." *Journal of Clinical Microbiology* 52 (11): 3933–43. https://doi.org/10.1128/JCM.02411-14.

Noya, F. J., M. A. Rench, T. G. Metzger, G. Colman, J. Naidoo, and C. J. Baker. 1987. "Unusual occurrence of an epidemic of Type Ib/c group B streptococcal sepsis in a neonatal intensive care unit." *The Journal of Infectious Diseases* 155 (6): 1135–44.

Oost, John van der, Matthijs M. Jore, Edze R. Westra, Magnus Lundgren, and Stan J.J. Brouns. 2009. "CRISPR-based adaptive and heritable immunity in prokaryotes." *Trends in Biochemical Sciences* 34 (8): 401–7. https://doi.org/10.1016/j.tibs.2009.05.002.

Oost, John van der, Edze R. Westra, Ryan N. Jackson, and Blake Wiedenheft. 2014. "Unravelling the structural and mechanistic basis of CRISPR–Cas systems." *Nature Reviews Microbiology* 12 (7): 479–92. https://doi.org/10.1038/nrmicro3279.

Patras, Kathryn A., and Kelly S. Doran. 2016. "A murine model of group B *Streptococcus* vaginal colonization." *Journal of Visualized Experiments*, no. 117 (November). https://doi.org/10.3791/54708.

Persson, E., S. Berg, B. Trollfors, P. Larsson, E. Ek, E. Backhaus, B.E.B. Claesson, et al. 2004. "Serotypes and clinical manifestations of invasive group B streptococcal infections in western Sweden 1998-2001." *Clinical Microbiology and Infection* 10 (9): 791–96. https://doi.org/10.1111/j.1469-0691.2004.00931.x.

Pettengill, James B., Ruth E. Timme, Rodolphe Barrangou, Magaly Toro, Marc W. Allard, Errol Strain, Steven M. Musser, and Eric W. Brown. 2014. "The evolutionary history and

diagnostic utility of the CRISPR-Cas system within *Salmonella enterica* Ssp. *Enterica*." *PeerJ* 2 (April): e340. https://doi.org/10.7717/peerj.340.

Phares, Christina R., Ruth Lynfield, Monica M. Farley, Janet Mohle-Boetani, Lee H. Harrison, Susan Petit, Allen S. Craig, et al. 2008. "Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005." *JAMA* 299 (17): 2056–65. https://doi.org/10.1001/jama.299.17.2056.

Pourcel, C., K. Hormigos, L. Onteniente, O. Sakwinska, R. H. Deurenberg, and G. Vergnaud. 2009. "Improved Multiple-Locus variable-number tandem-repeat assay for *Staphylococcus aureus* genotyping, providing a highly informative technique together with strong phylogenetic value." *Journal of Clinical Microbiology* 47 (10): 3121–28. https://doi.org/10.1128/JCM.00267-09.

Pourcel, C., G. Salvignol, and G. Vergnaud. 2005. "CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies." *Microbiology (Reading, England)* 151 (Pt 3): 653–63. https://doi.org/10.1099/mic.0.27437-0.

Pourcel, C, Paolo Visca, Baharak Afshar, Silvia D'Arezzo, Gilles Vergnaud, and Norman K. Fry. 2007. "Identification of Variable-Number Tandem-Repeat (VNTR) sequences in *Legionella pneumophila* and development of an optimized multiple-locus VNTR analysis typing scheme." *Journal of Clinical Microbiology* 45 (4): 1190–99. https://doi.org/10.1128/JCM.02078-06.

Poyart, C, Hélène Réglier-Poupet, Asmaa Tazi, Annick Billoët, Nicolas Dmytruk, Philippe Bidet, Edouard Bingen, Josette Raymond, and Patrick Trieu-Cuot. 2008. "Invasive group B streptococcal infections in infants, France." *Emerging Infectious Diseases* 14 (10): 1647–49. https://doi.org/10.3201/eid1410.080185.

Poyart, C., A. Tazi, H. Reglier-Poupet, A. Billoet, N. Tavares, J. Raymond, and P. Trieu-Cuot. 2007. "Multiplex PCR Assay for Rapid and Accurate Capsular Typing of Group B Streptococci." *Journal of Clinical Microbiology* 45 (6): 1985–88. https://doi.org/10.1128/JCM.00159-07.

Poyart, Claire, Hélène Réglier-Poupet, Asmaa Tazi, Annick Billoët, Nicolas Dmytruk, Philippe Bidet, Edouard Bingen, Josette Raymond, and Patrick Trieu-Cuot. 2008. "Invasive Group B streptococcal infections in infants, France." *Emerging Infectious Diseases* 14 (10): 1647–49. https://doi.org/10.3201/eid1410.080185.

Pride, D. T., C. L. Sun, J. Salzman, N. Rao, P. Loomer, G. C. Armitage, J. F. Banfield, and D. A. Relman. 2011. "Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time." *Genome Research* 21 (1): 126–36. https://doi.org/10.1101/gr.111732.110.

Quentin, Roland, Helene Huet, Fu-Sheng Wang, Pierre Geslin, Alain Goudeau, and Robert K. Selander. 1995. "Characterization of *Streptococcus agalactiae* strains by multilocus enzyme genotype and serotype: identification of multiple virulent clone families that cause invasive neonatal disease." *Journal of Clinical Microbiology* 33 (10): 2576–2581.

Radtke, A., B.-A. Lindstedt, J. E. Afset, and K. Bergh. 2010. "Rapid Multiple-Locus Variant-Repeat Assay (MLVA) for genotyping of *Streptococcus agalactiae*." *Journal of Clinical Microbiology* 48 (7): 2502–8. https://doi.org/10.1128/JCM.00234-10.

Regan, J. A., M. A. Klebanoff, and R. P. Nugent. 1991. "The epidemiology of group B streptococcal colonization in pregnancy. vaginal infections and prematurity study group." *Obstetrics and Gynecology* 77 (4): 604–10.

Richards, Vincent P., Ping Lang, Paulina D. Pavinski Bitar, Tristan Lefébure, Ynte H. Schukken, Ruth N. Zadoks, and Michael J. Stanhope. 2011. "Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*." *Infection, Genetics and Evolution* 11 (6): 1263–75. https://doi.org/10.1016/j.meegid.2011.04.019.

Richter, Corinna, James T. Chang, and Peter C. Fineran. 2012. "Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems." *Viruses* 4 (10): 2291–2311. https://doi.org/10.3390/v4102291.

Rolland, Karine, Corinne Marois, Veronique Siquier, Blandine Cattier, and Roland Quentin. 1999. "Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *hylb* gene analysis." *Journal of Clinical Microbiology* 37 (6): 1892–1898.

Rolland, Karine, Laurent Mereghetti, Stephane Watt, Sonia Chatellier, and Roland Quentin. 2002. "TRNA gene clusters at the 3' end of RRNA operons are specific to virulent subgroups of *Streptococcus agalactiae* strains, as demonstrated by molecular differential analysis at the population level." *Microbiology* 148 (5): 1493–1499.

Rosen, Geoffrey H, Tara M Randis, Purnahamsi V Desai, Katherine J Sapra, Bing Ma, Pawel Gajer, Michael S Humphrys, Jacques Ravel, Shari E Gelber, and Adam J Ratner. 2017. "Group B *Streptococcus* and the vaginal microbiota." *The Journal of Infectious Diseases* 216 (6): 744–51. https://doi.org/10.1093/infdis/jix395.

Rosenau, A., K. Martins, S. Amor, F. Gannier, P. Lanotte, N. van der Mee-Marquet, L. Mereghetti, and R. Quentin. 2007. "Evaluation of the ability of *Streptococcus agalactiae* strains isolated from genital and neonatal specimens to bind to human fibrinogen and correlation with characteristics of the *FbsA* and *FbsB* Genes." *Infection and Immunity* 75 (3): 1310–17. https://doi.org/10.1128/IAI.00996-06.

Sabat, A. J., A. Budimir, D. Nashev, R. Sá-Leão, J. M. Van Dijl, F. Laurent, H. Grundmann, A. W. Friedrich, ESCMID Study Group of Epidemiological Markers (ESGEM), and others. 2013. "Overview of molecular typing methods for outbreak detection and epidemiological surveillance." *Euro Surveill* 18 (4): 20380.

Salloum, Mazen, Nathalie van der Mee-Marquet, Anne-Sophie Valentin-Domelier, and Roland Quentin. 2011. "Diversity of prophage DNA regions of *Streptococcus agalactiae* clonal lineages from adults and neonates with invasive infectious disease." *PloS One* 6 (5): e20256. https://doi.org/10.1371/journal.pone.0020256.

Schrag, Stephanie, Rachel Gorwitz, Kristi Fultz-Butts, and Anne Schuchat. 2002. "Prevention of perinatal group B streptococcal disease. Revised Guidelines from CDC." *MMWR. Recommendations and Reports: Morbidity and Mortality Weekly Report. Recommendations and Reports / Centers for Disease Control* 51 (RR-11): 1–22.

Schuchat, A. 1995. "Group B streptococcal disease in newborns: a global perspective on prevention." *Biomedicine & Pharmacotherapy = Biomédecine & Pharmacothérapie* 49 (1): 19–25. https://doi.org/10.1016/0753-3322(96)82573-X.

Schuchat, A., K. Deaver-Robinson, B. D. Plikaytis, K. M. Zangwill, J. Mohle-Boetani, and J. D. Wenger. 1994. "Multistate case-control study of maternal risk factors for neonatal group B streptococcal disease. The Active Surveillance Study Group." *The Pediatric Infectious Disease Journal* 13 (7): 623–29.

Schuchat, Anne. 1998. “Epidemiology of group B streptococcal disease in the United States: shifting paradigms.” *Clinical Microbiology Reviews* 11 (3): 497–513.

Schurch, A. C., K. Kremer, O. Daviena, A. Kiers, M. J. Boeree, R. J. Siezen, and D. van Soolingen. 2010. “High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster.” *Journal of Clinical Microbiology* 48 (9): 3403–6. https://doi.org/10.1128/JCM.00370-10.

Schwartz, B., A. Schuchat, M. J. Oxtoby, S. L. Cochi, A. Hightower, and C. V. Broome. 1991. “Invasive Group B streptococcal disease in adults. A population-based study in metropolitan Atlanta.” *JAMA* 266 (8): 1112–14.

Schwartz, D. C., and C. R. Cantor. 1984. “Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis.” *Cell* 37 (1): 67–75.

Selander, Robert K., Dominique A. Caugant, Howerd Ochman, James M. Musser, Marion N. Gilmour, and Thomas S. Whittam. 1986. “Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics.” *Applied and Environmental Microbiology* 51 (5): 873.

Shah, Shiraz A., Susanne Erdmann, Francisco J.M. Mojica, and Roger A. Garrett. 2013. “Protospacer recognition motifs: mixed identities and functional diversity.” *RNA Biology* 10 (5): 891–99. https://doi.org/10.4161/rna.23764.

Shariat, N., and E. G. Dudley. 2014. “CRISPRs: molecular signatures used for pathogen subtyping.” *Applied and Environmental Microbiology* 80 (2): 430–39. https://doi.org/10.1128/AEM.02790-13.

Shariat, N., M. K. Kirchner, C. H. Sandt, E. Trees, R. Barrangou, and E. G. Dudley. 2013. “Subtyping of *Salmonella enterica* serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results.” *Journal of Clinical Microbiology* 51 (7): 2328–36. https://doi.org/10.1128/JCM.00608-13.

Shariat, Nikki, Michael J. DiMarzio, Shuang Yin, Lisa Dettinger, Carol H. Sandt, James R. Lute, Rodolphe Barrangou, and Edward G. Dudley. 2013. “The combination of CRISPR-MVLST and PFGE provides increased discriminatory power for differentiating human clinical isolates of *Salmonella enterica* Subsp. *Enterica* Serovar Enteritidis.” *Food Microbiology* 34 (1): 164–73. https://doi.org/10.1016/j.fm.2012.11.012.

Shariat, Nikki, Carol H. Sandt, Michael J. DiMarzio, Rodolphe Barrangou, and Edward G. Dudley. 2013. “CRISPR-MVLST subtyping of *Salmonella enterica* Subsp. Enterica Serovars Typhimurium and Heidelberg and application in identifying outbreak isolates.” *BMC Microbiology* 13 (1): 254.

Shmakov, Sergey, Aaron Smargon, David Scott, David Cox, Neena Pyzocha, Winston Yan, Omar O. Abudayyeh, et al. 2017. “Diversity and evolution of class 2 CRISPR–Cas systems.” *Nature Reviews Microbiology* 15 (3): 169–82. https://doi.org/10.1038/nrmicro.2016.184.

Siegel, J. D., G. H. McCracken, N. Threlkeld, B. Milvenan, and C. R. Rosenfeld. 1980. “Single-Dose Penicillin Prophylaxis against Neonatal Group B streptococcal infections. a controlled trial in 18,738 newborn infants.” *The New England Journal of Medicine* 303 (14): 769–75. https://doi.org/10.1056/NEJM198010023031401.

Six, Anne, Caroline Joubrel, Asmaa Tazi, and Claire Poyart. 2014. “Infections materno-fœtales à *Streptococcus agalactiae.*” *La Presse Médicale* 43 (6): 706–14. https://doi.org/10.1016/j.lpm.2014.04.008.

Skoff, Tami H., Monica M. Farley, Susan Petit, Allen S. Craig, William Schaffner, Ken Gershman, Lee H. Harrison, et al. 2009. "Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990–2007." *Clinical Infectious Diseases* 49 (1): 85–92. https://doi.org/10.1086/599369.

Slotved, H.-C., F. Kong, L. Lambertsen, S. Sauer, and G. L. Gilbert. 2007. "Serotype IX, a proposed New *Streptococcus agalactiae* serotype." *Journal of Clinical Microbiology* 45 (9): 2929–36. https://doi.org/10.1128/JCM.00117-07.

Société Française de Néonatologie, and Société Française de Pédiatrie. 2017. "Recommandations - Version Sept 2017. Recommandations de bonnes pratiques. prise en charge du nouveau - ne a risque d'infection neonatale bacterienne precoce (≥ 34 SA)." September 2017. http://www.societe-francaise-neonatalogie.fr/wp-content/uploads/2017/02/LABEL-HAS-RECOMMANDATIONS-INBP.09.2017.pdf.

Sørensen, Uffe B. Skov, Knud Poulsen, Claudia Ghezzo, Immaculada Margarit, and Mogens Kilian. 2010. "Emergence and global dissemination of host-specific *Streptococcus agalactiae* clones." *MBio* 1 (3). https://doi.org/10.1128/mBio.00178-10.

Springman, A. C., D. W. Lacher, G. Wu, N. Milton, T. S. Whittam, H. D. Davies, and S. D. Manning. 2009. "Selection, recombination, and virulence gene diversity among group B streptococcal genotypes." *Journal of Bacteriology* 191 (17): 5419–27. https://doi.org/10.1128/JB.00369-09.

Stringer, Jacqueline, and W.R Maxted. 1979. "Phage typing of group B *Streptococci*." *The Lancet* 313 (8111): 327–328.

Suara, RO. 1994. "Carriage of group B *Streptococci* in pregnant gambian mothers and their infants. - PubMed - NCBI." 1994. https://www.ncbi.nlm.nih.gov/pubmed/7963736.

Sun, Christine L., Brian C. Thomas, Rodolphe Barrangou, and Jillian F. Banfield. 2016. "Metagenomic reconstructions of bacterial CRISPR loci constrain population histories." *The ISME Journal* 10 (4): 858–70. https://doi.org/10.1038/ismej.2015.162.

Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht. 2001. "Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units." *Journal of Clinical Microbiology* 39 (10): 3563–71. https://doi.org/10.1128/JCM.39.10.3563-3571.2001.

Takahashi, S., E. E. Adderson, Y. Nagano, N. Nagano, M. R. Briesacher, and J. F. Bohnsack. 1998. "Identification of a highly encapsulated, genetically related group of invasive type III Group B *Streptococci*." *The Journal of Infectious Diseases* 177 (4): 1116–19.

Takahashi, Shinji, Shauna Detrick, April A. Whiting, Anne J. Blaschke-Bonkowksy, Youko Aoyagi, Elisabeth E. Adderson, and John F. Bohnsack. 2002a. "Correlation of phylogenetic lineages of group B *Streptococci,* identified by analysis of restriction-digestion patterns of genomic DNA, with *InfB* alleles and mobile genetic elements." *The Journal of Infectious Diseases* 186 (7): 1034–1038. https://doi.org/10.1086/342950.

Tazi, A, Olivier Disson, Samuel Bellais, Abdelouhab Bouaboud, Isabelle Tardieux, Patrick Trieu-Cuot, Marc Lecuit, and Claire Poyart. 2011. "[A breakthrough in the understanding of neonatal group B streptococcus meningitis]." *Médecine Sciences: M/S* 27 (4): 362–64. https://doi.org/10.1051/medsci/2011274010.

Tazi, A., P.C. Morand, H. Réglier-Poupet, N. Dmytruk, A. Billoët, D. Antona, P. Trieu-Cuot, and C. Poyart. 2011. "Invasive group B streptococcal infections in adults, France (2007–

2010)." *Clinical Microbiology and Infection* 17 (10): 1587–89. https://doi.org/10.1111/j.1469-0691.2011.03628.x.

Tazi, Asmaa, Olivier Disson, Samuel Bellais, Abdelouhab Bouaboud, Nicolas Dmytruk, Shaynoor Dramsi, Michel-Yves Mistou, et al. 2010. "The surface protein HvgA mediates Group B *Streptococcus* hypervirulence and meningeal tropism in neonates." *The Journal of Experimental Medicine* 207 (11): 2313–22. https://doi.org/10.1084/jem.20092594.

Teatero, S., A. McGeer, D. E. Low, A. Li, W. Demczuk, I. Martin, and N. Fittipaldi. 2014. "Characterization of invasive group B *Streptococcus* strains from the greater Toronto area, Canada." *Journal of Clinical Microbiology* 52 (5): 1441–47. https://doi.org/10.1128/JCM.03554-13.

Teatero, Sarah, Taryn B. T. Athey, Paul Van Caeseele, Greg Horsman, David C. Alexander, Roberto G. Melano, Aimin Li, et al. 2015. "Emergence of serotype IV group B *Streptococcus* adult invasive disease in Manitoba and Saskatchewan, Canada, is driven by clonal sequence type 459 strains." Edited by N. A. Ledeboer. *Journal of Clinical Microbiology* 53 (9): 2919–26. https://doi.org/10.1128/JCM.01128-15.

Tettelin, Hervé, Vega Masignani, Michael J. Cieslewicz, Claudio Donati, Duccio Medini, Naomi L. Ward, Samuel V. Angiuoli, et al. 2005. "Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome.'" *Proceedings of the National Academy of Sciences of the United States of America* 102 (39): 13950–13955.

Tettelin, Hervé, Vega Masignani, Michael J. Cieslewicz, Jonathan A. Eisen, Scott Peterson, Michael R. Wessels, Ian T. Paulsen, et al. 2002. "complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*." *Proceedings of the National Academy of Sciences* 99 (19): 12391–12396.

Timme, Ruth E., James B. Pettengill, Marc W. Allard, Errol Strain, Rodolphe Barrangou, Chris Wehnes, Joann S. Van Kessel, Jeffrey S. Karns, Steven M. Musser, and Eric W. Brown. 2013. "Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. Enterica inferred from genome-wide reference-free SNP characters." *Genome Biology and Evolution* 5 (11): 2109–23. https://doi.org/10.1093/gbe/evt159.

Tompkins, Lucy. 1992. "The Use of Molecular Methods in Infectious Diseases." *The New England Journal of Medicine* 127 (3): 339–347.

Trivalle, C., E. Martin, P. Martel, B. Jacque, J. F. Menard, and J. F. Lemeland. 1998. "Group B streptococcal bacteraemia in the elderly." *Journal of Medical Microbiology* 47 (7): 649–52.

Tsolia, M., M. Psoma, S. Gavrili, V. Petrochilou, S. Michalas, N. Legakis, and Th Karpathios. 2003. "Group B Streptococcus Colonization of Greek Pregnant Women and Neonates: Prevalence, Risk Factors and Serotypes." *Clinical Microbiology and Infection* 9 (8): 832–838.

Tyler, K. D., G. Wang, S. D. Tyler, and W. M. Johnson. 1997. "Factors Affecting Reliability and Reproducibility of Amplification-Based DNA Fingerprinting of Representative Bacterial Pathogens." *Journal of Clinical Microbiology* 35 (2): 339.

Tyrrell, Gregory J., Linda D. Senzilet, John S. Spika, Daniel A. Kertesz, Merrilyn Alagaratnam, Marguerite Lovgren, and James A. Talbot. 2000. "Invasive disease due to group B streptococcal infection in adults: results from a Canadian, population-based, active

laboratory surveillance study—1996." *The Journal of Infectious Diseases* 182 (1): 168–173.

Tyson, Gene W., and Jillian F. Banfield. 2007. "Rapidly Evolving CRISPRs implicated in acquired resistance of microorganisms to viruses." *Environmental Microbiology* 10 (1): 200–207. https://doi.org/10.1111/j.1462-2920.2007.01444.x.

Ueno, Hiroyuki, Yoshihiko Yamamoto, Akiko Yamamichi, Koji Kikuchi, Sumie Kobori, and Motonobu Miyazaki. 2012. "Characterization of Group B *Streptococcus* isolated from women in Saitama City, Japan." *Japanese Journal of Infectious Diseases* 65 (6): 516–521.

Valkenburg-van den Berg, Arijaan W., Arwen J. Sprij, Paul M. Oostvogel, Johan A.E.M. Mutsaers, Wouter B. Renes, Frits R. Rosendaal, and P. Joep Dörr. 2006. "Prevalence of colonisation with Group B *Streptococci* in pregnant women of a multi-ethnic population in the Netherlands." *European Journal of Obstetrics & Gynecology and Reproductive Biology* 124 (2): 178–83. https://doi.org/10.1016/j.ejogrb.2005.06.007.

Van Belkum, A, S. Scherer, W. Van Leeuwen, D. Willemse, L. Van Alphen, and H. Verbrugh. 1997. "Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*." *Infection and Immunity* 65 (12): 5017–5027.

Van Belkum, A. 1999. "Short sequence repeats in microbial pathogenesis and evolution." *Cellular and Molecular Life Sciences: CMLS* 56 (9–10): 729–34.

Van Orden, Mason J., Peter Klein, Kesavan Babu, Fares Z. Najar, and Rakhi Rajan. 2017. "Conserved DNA motifs in the type II-A CRISPR leader region." *PeerJ* 5 (April): e3161. https://doi.org/10.7717/peerj.3161.

Verani, Jennifer R., Lesley McGee, Stephanie J. Schrag, and Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention (CDC). 2010. "Prevention of perinatal group B streptococcal disease--revised guidelines from CDC, 2010." *MMWR. Recommendations and Reports: Morbidity and Mortality Weekly Report. Recommendations and Reports / Centers for Disease Control* 59 (RR-10): 1–36.

Vu-Thien, Hoang, Gaëlle Corbineau, Katia Hormigos, Brigitte Fauroux, Harriet Corvol, Annick Clément, Gilles Vergnaud, and Christine Pourcel. 2007. "Multiple-locus variable-number tandem-repeat analysis for longitudinal survey of sources of *Pseudomonas aeruginosa* infection in cystic fibrosis patients." *Journal of Clinical Microbiology* 45 (10): 3175–83. https://doi.org/10.1128/JCM.00702-07.

Wang, Rui, Ming Li, Luyao Gong, Songnian Hu, and Hua Xiang. 2016. "DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula Hispanica*." *Nucleic Acids Research* 44 (9): 4266–77. https://doi.org/10.1093/nar/gkw260.

Wehnes, C. A., T. G. Rehberger, R. Barrangou, and A. H. Smith. 2014. "Short Communication: Determination of *Salmonella* clustered regularly interspaced short palindromic repeats (CRISPR) diversity on dairy farms in Wisconsin and Minnesota." *Journal of Dairy Science* 97 (10): 6370–77. https://doi.org/10.3168/jds.2013-7595.

Wei, Yunzhou, Megan T. Chesne, Rebecca M. Terns, and Michael P. Terns. 2015. "Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*." *Nucleic Acids Research* 43 (3): 1749–58. https://doi.org/10.1093/nar/gku1407.

Welsh, John, and Michael McClelland. 1990. “Fingerprinting genomes using PCR with arbitrary primers.” *Nucleic Acids Research* 18 (24): 7213–7218.

Westra, Edze R., Angus Buckling, and Peter C. Fineran. 2014. “CRISPR-Cas systems: beyond adaptive immunity.” *Nature Reviews. Microbiology* 12 (5): 317–26. https://doi.org/10.1038/nrmicro3241.

Whitney, Cg, S Daly, S Limpongsanurak, Mr Festin, Kk Thinn, T Chipato, P Lumbiganon, et al. 2004. “The international infections in pregnancy study: group B streptococcal colonization in pregnant women.” *The Journal of Maternal-Fetal & Neonatal Medicine* 15 (4): 267–74. https://doi.org/10.1080/14767050410001668617.

Williams, John GK, Anne R. Kubelik, Kenneth J. Livak, J. Antoni Rafalski, and Scott V. Tingey. 1990. “DNA polymorphisms amplified by arbitrary primers are useful as genetic markers.” *Nucleic Acids Research* 18 (22): 6531–6535.

Wright, Addison V, and Jennifer A Doudna. 2016. “protecting genome integrity during CRISPR immune adaptation.” *Nature Structural & Molecular Biology* 23 (10): 876–83. https://doi.org/10.1038/nsmb.3289.

Yancey, M. K., P. Duff, P. Clark, T. Kurtzer, B. H. Frentzen, and P. Kubilis. 1994. “Peripartum infection associated with vaginal group B streptococcal colonization.” *Obstetrics and Gynecology* 84 (5): 816–19.

Yosef, I., M. G. Goren, and U. Qimron. 2012. “Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*.” *Nucleic Acids Research* 40 (12): 5569–76. https://doi.org/10.1093/nar/gks216.

Zhang, G. W., M. Kotiw, and G. Daggard. 2002. “A RAPD-PCR genotyping assay which correlates with serotypes of group B *Streptococci*.” *Letters in Applied Microbiology* 35 (3): 247–250.

Zhang, Shaokang, Yanlong Yin, Marcus B. Jones, Zhenzhen Zhang, Brooke L. Deatherage Kaiser, Blake A. Dinsmore, Collette Fitzgerald, Patricia I. Fields, and Xiangyu Deng. 2015. “*Salmonella* serotype determination utilizing high-throughput genome sequencing data.” Edited by N. A. Ledeboer. *Journal of Clinical Microbiology* 53 (5): 1685–92. https://doi.org/10.1128/JCM.00323-15.

FIGURES

Table 1. Characteristics of the 255 GBS isolates according to their ST.

ST	N° of isolates in ST	Serotype (n° of isolates)	Source (n° of isolates)*
1	32	II (2), V (30)	AC (22), AI (9), NI (1)
2	5	Ia (1), Ib (1), II (1), IV (1), V (1)	AC (5)
3	1	IV (1)	AC (1)
4	3	Ia (2), II(1)	AC (1), AI (2)
6	3	Ib (3)	AC (2), NI (1)
7	2	Ia (1), V (1)	AC (1), Ref. (1)
8	12	Ib (11), V (1)	AC (8), AI (4)
10	7	Ib (1), II (2), IV (1), V (3)	AC (4), AI (2), NI (1)
12	7	Ia (1), Ib (4), II (1), III (1)	AC (4), AI (3),
17	55	III (56)	AC (18), AI (2), NC (3), NI (31), Ref. (2)
19	26	II (1), III (22), V (3)	AC (18), AI (5), NI (3)
22	2	II (2)	AC (2)
23	37	Ia (34), III (3)	AC (22), AI (7), NC (2), NI (5), Ref. (1)
24	4	Ia (4)	AC (3), NI (1)
26	3	V (3)	AC (2), AI (1)
27	1	III (1)	AC (1)
28	10	Ia (1), II(5), III(2), V (2)	AC (8), NC (2)
41	1	III (1)	AC (1)
88	1	Ia (1)	AC (1)
110	1	V (1)	Ref. (2603 V/R)
130	3	Ia (2), V (1)	AC (3)
136	1	IV (1)	AC (1)
144	2	Ia (2)	AC (2)
173	1	V (1)	AI (1)
182	1	III (1)	AC (1)
196	8	Ia (1), Ib (1), IV (6)	AC (7), AI (1)
220	2	Ia (2)	AC (2)
223	1	Ia (1)	AC (1)
243	1	Ia (1)	AC (1)
255	1	Ib (1)	AI (1)
291	1	IV (1)	AI (1)
297	1	V (1)	AC (1)
305	1	Ia (1)	AC (1)
327	2	V (2)	AC (2)
366	1	III (1)	AC (1)
370	1	V	AC (1)
385	1	Ia (1)	AI (1)
386	1	II (1)	AI (1)
388	2	V (2)	AC (2)
389	1	III (1)	AC (1)
390	1	Ib (1)	AC (1)
391	1	III (1)	AC (1)

459	1	IV (1)	AI (1)
481	1	III (1)	AI (1)
569	1	II (1)	AC (1)
1002	1	III (1)	AC (1)
1004	1	III (1)	AC (1)
1005	1	V (1)	AC (1)

*A: Adult, N: Neonate, C: Non-invasive isolates, I: invasive isolates

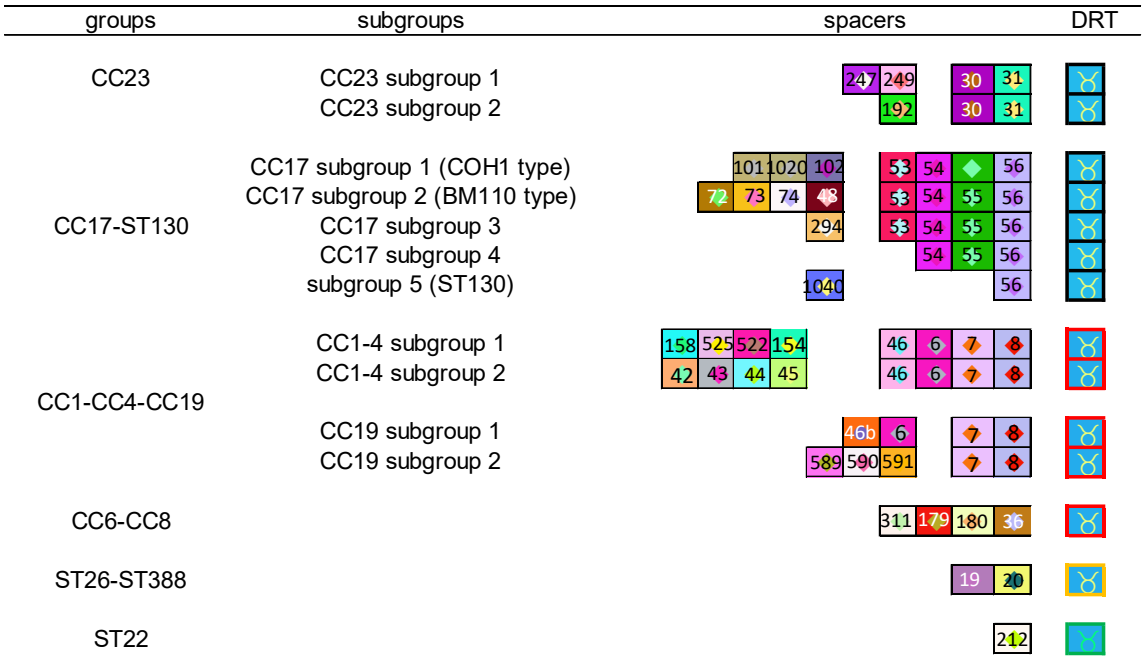


Figure 1. CRISPR clustering according to DRT and ancestral spacers. A total of 14 clusters were obtained. The six groups were defined by DRT and ancestral spacers. These groups were then divided into subgroups, based on more recent spacers.

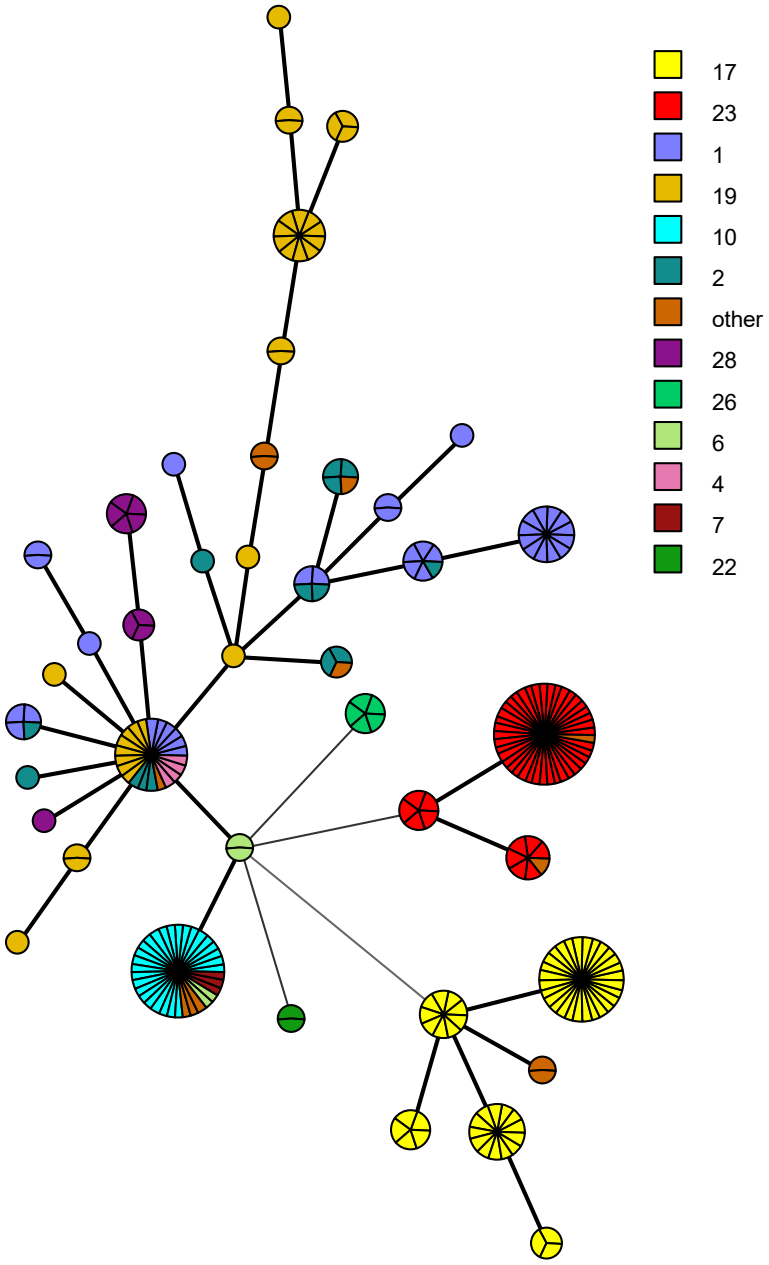


Figure 2. Minimum spanning tree (MST) representation of the 25 CRISPR1 marker scheme clustering. Each circle represents a CRISPR1 genotype and its size is proportional to the number of strains.

Table 2. Diversity index of the three typing methods used in this study (CRISPR typing, MLST and capsular typing). The two-marker group selection makes it possible to separate isolates with a high diversity index, superior to both MLST and capsular typing, highlighting the high discriminatory power of this CRISPR typing approach.

Typing method		Diversity index
CRISPR typing	94 markers	0.9939
	25 markers	0.9295
MLST		0.9017

Supplementary material

Table S1. Characteristics of GBS isolates of the study.

strains	Adult (A) / Neonate (N)	Invasive (I) / Non-invasive (C)	Source	ST	CC	serotype
Tours-S313	A	I	Blood	17	17	III
Tours-S322	A	I	Blood	23	23	Ia
Tours-S148	A	I	Blood	23	23	Ia
Tours-S125	A	I	Blood	23	23	Ia
Tours-E198	A	I	Blood	385	23	Ia
Tours-S264	A	I	Blood	23	23	Ia
Tours-S151	A	I	Blood	23	23	III
Tours-S180	A	I	Blood	8	8	Ib
Tours-E237	A	I	Blood	10	8	Ib
Tours-E207	A	I	Blood	10	8	V
Tours-S210	A	I	Blood	12	8	Ib
Tours-S402	A	I	Blood	12	8	Ib
Tours-S378	A	I	Blood	12	8	Ib
Tours-S274	A	I	Blood	255	6	Ib
Tours-S182	A	I	Blood	1	1	V
Tours-S247	A	I	Blood	1	1	V
Tours-S293	A	I	Blood	173	1	V
Tours-S280	A	I	Blood	1	1	V
Tours-S299	A	I	Blood	1	1	V
Tours-S147	A	I	Blood	1	1	V
Tours-S257	A	I	Blood	1	1	V
Tours-S352	A	I	Blood	19	19	III
Tours-S234	A	I	Blood	19	19	III
Tours-S146	A	I	Blood	4	4	Ia
Tours-E546	A	I	Blood	386	19	II
Tours-S291	A	I	CSF	23	23	Ia
Tours-S318	A	I	CSF	23	23	Ia
Tours-S105	A	I	CSF	19	19	III
Tours-C8	A	C	Cutaneous	17	17	III
Tours-C13	A	C	Cutaneous	1	1	V
Tours-C75	A	C	Urinary tract	305	23	Ia
Tours-C81	A	C	Urinary tract	12	8	III
Tours-C39	A	C	Urinary tract	2	1	Ia
Tours-C96	A	C	Urinary tract	196	1	Ib
Tours-C106	A	C	Urinary tract	19	19	III
Tours-Sag249	A	C	Genital tract	1	1	V
Tours-Sag397	A	C	Genital tract	19	19	III
Tours-Sag152	N	C	Gastric aspirate	23	23	Ia
Tours-Sag434	N	C	Gastric aspirate	23	23	Ia
Tours-S1	N	I	Blood	17	17	III
Tours-S14	N	I	Blood	17	17	III

Tours-S128	N	I	Blood	17	17	III
Tours-S94	N	I	Blood	17	17	III
Tours-S143	N	I	Blood	17	17	III
Tours-S129	N	I	Blood	17	17	III
Tours-S103	N	I	Blood	17	17	III
Tours-S12	N	I	Blood	23	23	Ia
Tours-S158	N	I	Blood	23	23	Ia
Tours-S165	N	I	Blood	481	23	III
Tours-S120	N	I	CSF	17	17	III
Tours-S202	N	I	CSF	17	17	III
Tours-S130	N	I	CSF	17	17	III
Tours-S135	N	I	CSF	17	17	III
Tours-S107	N	I	CSF	17	17	III
Tours-S7	N	I	CSF	17	17	III
Tours-S195	N	I	CSF	17	17	III
Tours-S192	N	I	CSF	17	17	III
Tours-S39	N	I	CSF	17	17	III
Tours-S191	N	I	CSF	17	17	III
Tours-S184	N	I	CSF	17	17	III
Tours-S194	N	I	CSF	17	17	III
Tours-S173	N	I	CSF	17	17	III
Tours-S34	N	I	CSF	17	17	III
Tours-S179	N	I	CSF	17	17	III
Tours-S118	N	I	CSF	17	17	III
Tours-S2	N	I	CSF	17	17	III
Tours-S24	N	I	CSF	17	17	III
Tours-S102	N	I	CSF	23	23	Ia
Tours-S108	N	I	CSF	23	23	Ia
Tours-S101	N	I	CSF	23	23	Ia
Tours-S167	N	I	CSF	24	singleton	Ia
Tours-S92	N	I	CSF	8	8	Ib
Tours-S47	N	I	CSF	6	6	Ib
Tours-S163	N	I	CSF	1	1	V
Tours-S127	N	I	CSF	19	19	III
Tours-S193	N	I	CSF	19	19	III
Tours-S152	N	I	CSF	19	19	III
Tours-C214	A	C	Respiratory tract	17	17	III
Tours-C52	A	C	Respiratory tract	223	23	Ia
Tours-C20	A	C	Respiratory tract	23	23	Ia
Tours-C12	A	C	Respiratory tract	391	23	III
Tours-C165	A	C	Respiratory tract	8	8	Ib
Tours-C115	A	C	Respiratory tract	8	8	Ib
Tours-C23	A	C	Respiratory tract	1	1	V
Tours-C66	A	C	Respiratory tract	1	1	V
Tours-Sag1	A	C	Genital tract	19	19	III
Tours-Sag4	A	C	Genital tract	23	23	Ia
Tours-Sag8	A	C	Genital tract	23	23	Ia

Tours-Sag10	A	C	Genital tract	182	19	III
Tours-Sag14	A	C	Genital tract	23	23	Ia
Tours-Sag17	A	C	Genital tract	28	19	II
Tours-Sag20	A	C	Genital tract	27	19	III
Tours-Sag22	A	C	Genital tract	23	23	Ia
Tours-Sag25	A	C	Genital tract	19	19	III
Tours-Sag30	A	C	Genital tract	19	19	III
Tours-Sag37	A	C	Genital tract	1	1	V
Tours-Sag38	A	C	Genital tract	24	singleton	Ia
Tours-Sag42	A	C	Genital tract	1	1	V
Tours-Sag43	A	C	Genital tract	22	singleton	II
Tours-Sag45	A	C	Genital tract	136	1	IV
Tours-Sag90	A	C	Genital tract	28	19	Ia
Tours-Sag94	A	C	Genital tract	23	23	Ia
Tours-Sag96	A	C	Genital tract	28	19	II
Tours-Sag98	A	C	Genital tract	28	19	V
Tours-Sag106	A	C	Genital tract	28	19	III
Tours-Sag109	A	C	Genital tract	17	17	III
Tours-Sag113	A	C	Genital tract	366	23	III
Tours-Sag117	A	C	Genital tract	130	singleton	Ia
Tours-Sag121	A	C	Genital tract	130	singleton	Ia
Tours-Sag128	A	C	Genital tract	17	17	III
Tours-Sag136	A	C	Genital tract	196	1	IV
Tours-Sag138	A	C	Genital tract	196	1	IV
Tours-Sag140	A	C	Genital tract	2	1	IV
Tours-Sag142	A	C	Genital tract	10	8	V
Tours-Sag148	A	C	Genital tract	1	1	V
Tours-Sag153	A	C	Genital tract	1	1	V
Tours-Sag161	A	C	Genital tract	1	1	V
Tours-Sag169	A	C	Genital tract	2	1	V
Tours-Sag171	A	C	Genital tract	1002	23	III
Tours-Sag183	A	C	Genital tract	144	23	Ia
Tours-Sag195	A	C	Genital tract	3	4	IV
Tours-Sag200	A	C	Genital tract	19	19	V
Tours-Sag209	A	C	Genital tract	17	17	III
Tours-Sag211	A	C	Genital tract	12	8	II
Tours-Sag228	A	C	Genital tract	1	1	V
Tours-Sag232	A	C	Genital tract	23	23	Ia
Tours-Sag234	A	C	Genital tract	17	17	III
Tours-Sag236	A	C	Genital tract	17	17	III
Tours-Sag247	A	C	Genital tract	12	8	Ib
Tours-Sag259	A	C	Genital tract	19	19	II
Tours-Sag274	A	C	Genital tract	26	26/388	V
Tours-Sag278	A	C	Genital tract	19	19	III
Tours-Sag280	A	C	Genital tract	17	17	III
Tours-Sag286	A	C	Genital tract	6	6	Ib
Tours-Sag288	A	C	Genital tract	23	23	Ia

Tours-Sag293	A	C	Genital tract	23	23	Ia
Tours-Sag299	A	C	Genital tract	17	17	III
Tours-Sag305	A	C	Genital tract	243	4	Ia
Tours-Sag313	A	C	Genital tract	23	23	Ia
Tours-Sag324	A	C	Genital tract	17	17	III
Tours-Sag328	A	C	Genital tract	23	23	Ia
Tours-Sag330	A	C	Genital tract	196	1	IV
Tours-Sag337	A	C	Genital tract	17	17	III
Tours-Sag343	A	C	Genital tract	19	19	III
Tours-Sag346	A	C	Genital tract	327	singleton	V
Tours-Sag357	A	C	Genital tract	6	6	Ib
Tours-Sag362	A	C	Genital tract	28	19	II
Tours-Sag365	A	C	Genital tract	1	1	V
Tours-Sag371	A	C	Genital tract	196	1	IV
Tours-Sag382	A	C	Genital tract	23	23	Ia
Tours-Sag387	A	C	Genital tract	23	23	Ia
Tours-Sag393	A	C	Genital tract	1	1	V
Tours-Sag396	A	C	Genital tract	19	19	III
Tours-Sag400	A	C	Genital tract	19	19	III
Tours-Sag409	A	C	Genital tract	1	1	V
Tours-Sag416	A	C	Genital tract	23	23	Ia
Tours-Sag420	A	C	Genital tract	22	singleton	II
Tours-Sag428	A	C	Genital tract	1	1	V
Tours-Sag432	A	C	Genital tract	7	6	V
Tours-Sag435	A	C	Genital tract	19	19	III
Tours-Sag440	A	C	Genital tract	8	8	V
Tours-Sag443	A	C	Genital tract	1	1	V
Tours-Sag446	A	C	Genital tract	327	singleton	V
Tours-Sag456	A	C	Genital tract	88	23	Ia
Tours-Sag459	A	C	Genital tract	1	1	V
Tours-Sag463	A	C	Genital tract	17	17	III
Tours-Sag468	A	C	Genital tract	23	23	Ia
Tours-Sag471	A	C	Genital tract	1	1	V
Tours-Sag475	A	C	Genital tract	17	17	III
Tours-Sag477	A	C	Genital tract	19	19	III
Tours-Sag485	A	C	Genital tract	19	19	V
Tours-Sag488	A	C	Genital tract	10	8	V
Tours-Sag492	A	C	Genital tract	1004	17	III
Tours-Sag498	A	C	Genital tract	8	8	Ib
Tours-Sag499	A	C	Genital tract	10	8	II
Tours-Sag505	A	C	Genital tract	23	23	Ia
Tours-Sag507	A	C	Genital tract	24	singleton	Ia
Tours-Sag521	A	C	Genital tract	17	17	III
Tours-Sag536	A	C	Genital tract	26	26/388	V
Tours-Sag544	A	C	Genital tract	1005	1	V
Tours-Sag545	A	C	Genital tract	17	17	III
Tours-Sag564	A	C	Genital tract	297	1	V

Tours-P1E4	A	C	Genital tract	23	23	Ia
Tours-P2A1	A	C	Genital tract	12	8	Ia
Tours-P3A1	A	C	Genital tract	8	8	Ib
Tours-P4E2	A	C	Genital tract	19	19	III
Tours-P5A2	A	C	Genital tract	23	23	Ia
Tours-P6A1	A	C	Genital tract	24	singleton	Ia
Tours-P7A2	A	C	Genital tract	28	19	II
Tours-P8A2	A	C	Genital tract	1	1	V
Tours-P9A1	A	C	Genital tract	569	singleton	II
Tours-P10E2	A	C	Genital tract	4	4	II
Tours-C191	A	C	Genital tract	17	17	III
Tours-C133	A	C	Genital tract	220	23	Ia
Tours-C194	A	C	Genital tract	220	23	Ia
Tours-C188	A	C	Genital tract	196	1	IV
Tours-C221	A	C	Genital tract	1	1	II
Tours-C197	A	C	Genital tract	2	1	Ib
Tours-C1	A	C	Genital tract	19	19	III
Tours-C104	A	C	Genital tract	19	19	III
Tours-C213	A	C	Intestinal tract	17	17	III
Tours-C38	A	C	Intestinal tract	17	17	III
Tours-C87	A	C	Intestinal tract	130	singleton	V
Tours-C230	A	C	Intestinal tract	144	23	Ia
Tours-C196	A	C	Intestinal tract	23	23	Ia
Tours-C57	A	C	Intestinal tract	23	23	Ia
Tours-C18	A	C	Intestinal tract	23	23	Ia
Tours-C226	A	C	Intestinal tract	23	23	III
Tours-C175	A	C	Intestinal tract	390	8	Ib
Tours-C40	A	C	Intestinal tract	8	8	Ib
Tours-C234	A	C	Intestinal tract	8	8	Ib
Tours-C112	A	C	Intestinal tract	8	8	Ib
Tours-C193	A	C	Intestinal tract	10	8	IV
Tours-C166	A	C	Intestinal tract	41	6	III
Tours-C15	A	C	Intestinal tract	2	1	II
Tours-C219	A	C	Intestinal tract	196	1	IV
Tours-C120	A	C	Intestinal tract	1	1	V
Tours-C124	A	C	Intestinal tract	1	1	V
Tours-C50	A	C	Intestinal tract	1	1	II
Tours-C99	A	C	Intestinal tract	370	1	V
Tours-C113	A	C	Intestinal tract	389	19	III
Tours-C97	A	C	Intestinal tract	28	19	V
Tours-C94	A	C	Genital tract	388	26/388	V
Tours-C82	A	C	Intestinal tract	388	26/388	V
Paris-620	A	I	Blood	291	17	IV
Paris-576	N	I	joint fluid	17	17	III
Paris-569	A	I	Blood	17	17	III
Paris-567	N	C	Gastric aspirate	17	17	III
Paris-562	N	I	Blood	17	17	III

Paris-533	N	c	Gastric aspirate	17	17	III
Paris-531	A	I	Blood	4	4	Ia
Paris-509	N	I	CSF	17	17	III
Paris-503	N	C	Gastric aspirate	28	19	III
Paris-498	N	I	Blood	17	17	III
Paris-493	N	I	autopsy	10	8	II
Paris-483	N	C	Gastric aspirate	28	19	II
Paris-475	A	I	Blood	8	8	Ib
Paris-447	N	I	Respiratory tract	17	17	III
Paris-442	A	I	Blood	1	1	V
Paris-437	A	I	CSF	196	1	Ia
Paris-429	N	C	Gastric aspirate	17	17	III
Paris-361	A	I	Blood	26	26/388	V
Paris-614	A	I	Blood	459	1	IV
Paris-606	A	I	Blood	1	1	V
Paris-598	N	I	Blood	17	17	III
Paris-548	A	I	joint fluid	1	1	V
Paris-524	A	I	Blood	8	8	Ib
Paris-476	A	C	Genital tract	19	19	III
Paris-472	A	I	Blood	19	19	V
Paris-470	A	I	Blood	19	19	III
2603 V/R	Ref strain	Ref strain	Ref strain	110	19	V
A909	Ref strain	Ref strain	Ref strain	7	4	Ia
NEM316	Ref strain	Ref strain	Ref strain	23	23	III
COH1	Ref strain	Ref strain	Ref strain	17	17	III
BM110	Ref strain	Ref strain	Ref strain	17	17	III

Table S2. Primers used for CRISPR1 locus amplification and sequencing

Primer name	Sequence (5' → 3')
CRISPR1 PCR-F	GAGACACAGAGCGACACTATC
CRISPR1 PCR-R	CATTTCTTTCTCCACTATTATAAC
CRISPR1 SEQ-F	GAAGACTCTATGATTTACCGC
CRISPR1 SEQ-R	CAGCAATCACTAAAAGAACCAAC
Sp39F	GCCTTTTCTAACTCTTCAG
Sp40F	TATGTCTTCTAACAGTTGCTTCTTGTGCTT
Sp75F	AGAGTGTGTCCAAGACCAGAGTTACTGTTT
Sp107F	ACCATCAAGGCTCTTATCTGCAGATTGTTA
Sp109F	ACAAAATCTTTTGTTGCTCCTGGACGTATT
Sp154F	GATATTTGGAAAGATTCTGATTTTGGTAAG
Sp177R	AGTTACTGTTGAGGGTAGTCC
Sp246F	GATTAGTTTGCGTACTCGCTC
Sp263F	AAGTGAAGTTGAATTTTATTTGAGATACTA
Sp275F	ACAGACAAAGAAGATGGCAAG
Sp295F	TTGGAAAGATGGCAGAGGGGTTAACGCAGC
Sp298F	CTTTTAACCGTCTCTCCGCTTC
Sp299F	ATGGTCAGAAATCGAAGAAAACGAAGTCGT
Sp305bisF	TTGGAAAAAACACGAAAGTGATATTACTTT
Sp325F	GGATGATTTCGATTATGCGGCGGTGGTTGA
Sp328F	ATTACTTTCAAGATGTCTATGACTATATGC
Sp335F	AAGGGTGTTAGATGATAATACCTTTTTTAA
Sp348F	CGTTCTGCTTTAGTCATATGTGTCCTTTCT
Sp351F	ACAAACCTCTAATGGATAATATAGAACAAA
Sp379F	CATGCCTTTGAAAATAAATTCCGAGCCATT
Sp436F	CCATCCAACCATTTGATTTTGTGTAAATAT
Sp504F	TTTATCCCTAGTAGGTT
Sp511F	GCTTGGGTTTGATAAGGG
Sp531F	AGTTTAGAGTTAACGTCTGTATTTTAGATA
Sp539F	TCTAAGTGCTCGACCATC
Sp650F	ATGCTTGTCAAAGGTAATAAATCTGGAGAT
Sp803F	TCTAATGTATACTGTAACTGAGGCAG
Sp823F	AACACAGCTTCCTCGAAAGGGATATATCTA
Sp830F	TCAAAAGTTTCCACTAATAGCGTTT
Sp856F	TATCAAGTCTGCTTTGTCAACAAAGTCAGC
Sp904bisF	GATATGGTGGACAGGTTCAGCAGGTCATGT
Sp2069F	CCGTGCAAGAGTTAAGGAACTGACACAGCA
Sp2132F	ATGGCTAATTATTGCGTGATGTTAGCGGTT
Sp2166F	TGTTTTTTCTTTCCATATGTGGTGGCTTTC
Sp2191F	AACCTAAAAAGATGTATGAGAAGTTGCCTT
Sp2199F	TCAATATAAATCTGTCCAGTCTTTTTCTCA
Sp2201F	GAAAAATTAAATACTTTCCAATAATTATAA
Sp2211F	AAAATCTAAAAATACATTTTACCGTTTGAA
Sp2228F	ACGTTGAGGGGTTAGTTGCTGCTGGTCTGG

Sp2234F	AGTTTGTTTATGATGAACACCAAAATTTAA
Sp2259F	TAATTTTGTTATATCTATATTTAAGTCTTT
Sp2293F	TTTTCTACTTTATTTCCTACTTCATTTGTT
Sp2312F	CTGTTATATCAACAATTACCAAGTCTGCTG
Sp2325F	ATCGGCAATTTCTTGAAGACGTTCAGCGTT
Sp2359F	AAGGATTTTACTTCAAAGATTTTTTTGTTG
Sp8R	TATCCATCTCGGTGAGATGAGAATTAGCTT
Sp20R	AGATGTCTTGATATCAACTGCTTTAAAAGA
Sp31R	ATCAGTACCAACAAATGATTTTGTACCATC
Sp36R	ACATATCCTTTTGTTAGGTCAAAAGAAGAT
Sp42R	TAAGTAACAAGACAGACTTGAA
Sp52R	TTTATTTTTTCTCAGTTCCTTGATTTTAGA
Sp87R	TATGCAAAGCAGTCACCAGC

Table S3. Capsular type repartition of the isolates according to their origin

Capsular type	Carriage n	%	Invasive strains	%	Reference strains	tot
Ia	39	24	16	19	1	56
Ib	13	8	10	12	0	24
II	15	9	2	2	0	17
III	48	29	42	48	3	92
IV	10	6	2	2	0	12
V	39	24	14	16	1	54
VI	0	0	0	0	0	0
VII	0	0	0	0	0	0
VIII	0	0	0	0	0	0
IX	0	0	0	0	0	0
X	0	0	0	0	0	0
total	164		86		5	255

Table S4. Sequence type repartition of the isolates according to their origin.

ST	CC (MLST/Eburst)	CC (CRISPR)	Carriage		Invasive		Ref. strains	total
			n	%	n	%		
17	17	17*	21	13	33	39	2	54
23	23	23	24	15	12	14	1	37
1	1	1 - 4	22	13	10	12	0	32
19	19	19	18	11	8	9	0	26
8	8	8	8	5	4	5	0	12
28	19	19	10	6	0	0	0	10
196	1	1 - 4	7	4	1	1	0	8
10	8	8	4	2	3	4	0	7
12	8	8	4	2	3	4	0	7
2	1	1 - 4	5	3	0	0	0	5
24	Singleton	23	3	2	1	1	0	4
4	4	1 - 4	1	1	2	2	0	3
6	6	6	2	1	1	1	0	3
26	None	ST388-ST26	2	1	1	1	0	3
130	Singleton	ST 130	3	2	0	0	0	3
22	Singleton	ST22	2	1	0	0	0	2
144	23	23	2	1	0	0	0	2
220	23	23	2	1	0	0	0	2
327	Singleton	19	2	1	0	0	0	2
388	None	ST388-ST26	2	1	0	0	0	2
7	6	6	1	1	0	0	1	2
3	4	1 - 4	1	1	0	0	0	1
27	19	19	1	1	0	0	0	1
41	6	6	1	1	0	0	0	1
88	23	23	1	1	0	0	0	1
110	19	19	0	0	0	0	1	1
136	1	1 - 4	1	1	0	0	0	1

173	1	1 - 4	0	0	1	1	0	1
182	19	19	1	1	0	0	0	1
223	23	23	1	1	0	0	0	1
243	4	1 - 4	1	1	0	0	0	1
255	6	6	0	0	1	1	0	1
291	17	17*	0	0	1	1	0	1
297	1	1 - 4	1	1	0	0	0	1
305	23	23	1	1	0	0	0	1
366	23	23	1	1	0	0	0	1
370	1	1 - 4	1	1	0	0	0	1
385	23	23	0	0	1	1	0	1
386	19	19	0	0	1	1	0	1
389	19	19	1	1	0	0	0	1
390	8	8	1	1	0	0	0	1
391	23	23	1	1	0	0	0	1
459	1	1 - 4	0	0	1	1	0	1
481	23	23	0	0	1	1	0	1
569	Singleton	8	1	1	0	0	0	1
1002	23	23	1	1	0	0	0	1
1004	17	17*	1	1	0	0	0	1
1005	1	1 - 4	1	1	0	0	0	1
Total			165		85		5	255

Table S5. Clonal complex repartition of the isolates according to their origin**.** Percentage corresponding to percentage of CC isolates according to their origin

CC	non invasive		invasive		total
	n	%	n	%	
23	34	71	14	29	48
17	22	39	34	61	56
1	38	75	13	25	51
19	31	78	9	23	40
6	4	67	2	33	6
8	17	63	10	37	27
4	3	60	2	40	5
26/388	4	80	1	20	5
singleton	11	92	1	8	12
TOTAL	164		86		250


Isolates
CC
ST
capsular type
leader
spacers
DR1
Tours-S291 23 23 Ia
Tours-C230 23 144 Ia
Tours-C75 23 305 Ia
Tours-C133 23 220 Ia
Tours-S12 23 23 Ia
Tours-S322 23 23 Ia
Tours-S102 23 23 Ia
Tours-C52 23 223 Ia
Tours-S148 23 23 Ia
Tours-S125 23 23 Ia
Tours-S158 23 23 Ia
Tours-S.ag14 23 23 Ia
ours-S.ag 505 23 23 Ia
Tours-S108 23 23 Ia
Tours-S101 23 23 Ia
fours-S.ag232 23 23 Ia
fours-S.ag313 23 23 Ia
Tours-S.ag94 23 23 Ia
Tours-C57 23 23 Ia
Tours-C18 23 23 Ia
Tours-P1E4 23 23 Ia
ours-S.ag 387 23 23 Ia
Paris-464 23 23 Ia
fours-S.ag288 23 23 Ia
ours-S.ag 468 23 23 Ia
fours-S.ag434 23 23 Ia
Tours-S318 23 23 Ia
Tours-S264 23 23 Ia
Tours-P5A2 23 23 Ia
Tours-C196 23 23 Ia
Tours-E198 23 385 Ia
ours-S.ag 328 23 23 Ia
Tours-S.ag8 23 23 Ia
ours-S.ag 416 23 23 Ia
fours-S.ag152 23 23 Ia
Tours-C20 23 23 Ia
fours-S.ag183 23 144 Ia
S.ag22 23 23 Ia
S.ag4 Singleton 24 Ia
S167 Singleton 24 Ia
S.ag38 Singleton 24 Ia
S.ag 507 Singleton 24 Ia
S.ag293 23 23 Ia
P6A1 Singleton 24 Ia
S.ag 382 23 23 Ia
C194 23 220 Ia
Tours-C226 23 23 III
Tours-S165 23 481 III
NEM 316 23 23 III
Tours-S151 23 23 III
Tours-C12 23 391 III
fours-S.ag113 23 366 III
fours-S.ag171 23 1002 III
ours-S.ag 456 23 88 Ia
Tours-S14 17 17 III
Tours-S120 17 17 III
Tours-S128 17 17 III
Paris-562 17 17 III
Tours-S.ag128 17 17 III
Tours-S94 17 17 III
Tours-S313 17 17 III
Tours-C191 17 17 III
Tours-S202 17 17 III
Tours-S.ag299 17 17 III
Tours-S.ag 46 17 17 III
Paris-509 17 17 III
Tours-S.ag 54 17 17 III
Tours-S143 17 17 III
Tours-S179 17 17 III
Tours-S2 17 17 III
Paris-498 17 17 III
Tours-S129 17 17 III
Tours-S130 17 17 III
Tours-S135 17 17 III
Tours-S107 17 17 III
Tours-S.ag 47 17 17 III
Tours-S.ag109 17 17 III
Paris-569 17 17 III
Tours-C213 17 17 III
Tours-C214 17 17 III
Tours-S7 17 17 III
Paris-567 17 17 III
Tours-S195 17 17 III
Tours-S192 17 17 III
Paris-533 17 17 III
COH1 17 17 III
BM110 17 17 III
Tours-S.ag 52 17 17 III
Tours-S103 17 17 III
Tours-S.ag492 17 1004 III
Tours-S39 17 17 III
Tours-S191 17 17 III
Tours-C8 17 17 III
Tours-S.ag23 17 17 III
Tours-S.ag280 17 17 III
Tours-C38 17 17 III
Paris-598 17 17 III
Paris-447 17 17 III
Tours-S.ag234 17 17 III
Paris-429 17 17 III
Tours-S.ag 33 17 17 III
Tours-S184 17 17 III
Tours-S194 17 17 III
Tours-S173 17 17 III
Tours-S34 17 17 III
Tours-S.ag209 17 17 III
Tours-S.ag 32 17 17 III
Paris-620 17 291 IV
Paris-576 17 III
Tours-S118 17 17 III
Tours-S24 17 17 III
Tours-C87 singleton 130 V
Tours-S.ag117 singleton 130 Ia
Tours-S.ag121 singleton 130 Ia
S180 8 8 Ib
S92 8 8 Ib
C165 8 8 Ib
C115 8 8 Ib
P3A1 8 8 Ib
C175 8 390 Ib
S.ag211 8 12 II
Paris-524 8 8 Ib
C40 8 8 Ib
C168 4 5 Ib
S.ag 498 8 8 Ib
S.ag 440 8 8 V
Paris-475 8 8 Ib
C234 8 8 Ib
C112 8 8 Ib
S210 8 12 Ib
C81 8 12 III
S.ag 488 8 10 V
P9A1 singleton 569 II
E207 8 10 V
S.ag 499 8 10 II
S.ag142 8 10 IV
E237 8 10 Ib
C193 8 10 IV
Paris-493 8 10 II
P2A1 8 12 Ia
S.ag247 8 12 Ib
S402 8 12 Ib
S378 8 12 Ib
S.ag432 6 7 V
A909 6 7 Ia
C166 6 41 III
S274 6 255 Ib

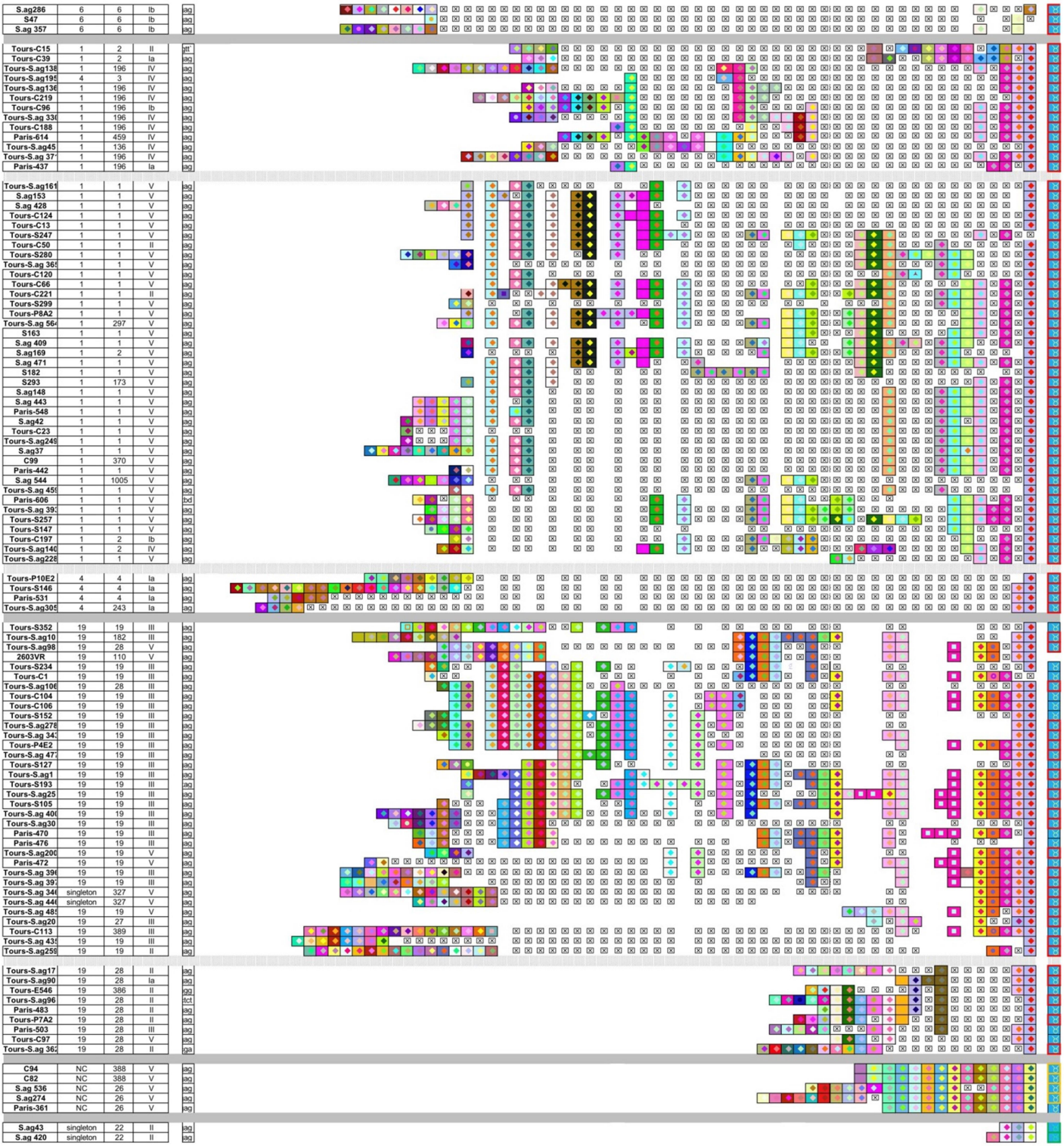

S.ag286 6 6 Ib ag
S47 6 6 Ib ag
S.ag 357 6 6 Ib ag
Tours-C15 1 2 II gtt
Tours-C39 1 2 Ia ag
Tours-S.ag138 1 196 IV ag
Tours-S.ag195 4 3 IV ag
Tours-S.ag136 1 196 IV ag
Tours-C219 1 196 IV ag
Tours-C96 1 196 Ib ag
Tours-S.ag 330 1 196 IV ag
Tours-C188 1 196 IV ag
Paris-614 1 459 IV ag
Tours-S.ag45 1 136 IV ag
Tours-S.ag 37 1 196 IV ag
Paris-437 1 196 Ia ag
Tours-S.ag161 1 1 V ag
S.ag153 1 1 V ag
S.ag 428 1 1 V ag
Tours-C124 1 1 V ag
Tours-C13 1 1 V ag
Tours-S247 1 1 V ag
Tours-C50 1 1 II ag
Tours-S280 1 1 V ag
Tours-S.ag 365 1 1 V ag
Tours-C120 1 1 V ag
Tours-C66 1 1 V ag
Tours-C221 1 1 II ag
Tours-S299 1 1 V ag
Tours-P8A2 1 1 V ag
Tours-S.ag 564 1 297 V ag
S163 1 1 V ag
S.ag 409 1 1 V ag
S.ag169 1 2 V ag
S.ag 471 1 1 V ag
S182 1 1 V ag
S293 1 173 V ag
S.ag148 1 1 V ag
S.ag 443 1 1 V ag
Paris-548 1 1 V ag
S.ag42 1 1 V ag
Tours-C23 1 1 V ag
Tours-S.ag249 1 1 V ag
S.ag37 1 1 V ag
C99 1 370 V ag
Paris-442 1 1 V ag
S.ag 544 1 1005 V ag
Tours-S.ag 459 1 1 V ag
Paris-606 1 1 V bd
Tours-S.ag 393 1 1 V ag
Tours-S257 1 1 V ag
Tours-S147 1 1 V ag
Tours-C197 1 2 Ib ag
Tours-S.ag140 1 2 IV ag
Tours-S.ag228 1 1 V ag
Tours-P10E2 4 4 Ia ag
Tours-S146 4 4 Ia ag
Paris-531 4 4 Ia ag
Tours-S.ag305 4 243 Ia ag
Tours-S352 19 19 III ag
Tours-S.ag10 19 182 III ag
Tours-S.ag98 19 28 V ag
2603VR 19 110 V ag
Tours-S234 19 19 III ag
Tours-C1 19 19 III ag
Tours-S.ag106 19 28 III ag
Tours-C104 19 19 III ag
Tours-C106 19 19 III ag
Tours-S152 19 19 III ag
Tours-S.ag278 19 19 III ag
Tours-S.ag 342 19 19 III ag
Tours-P4E2 19 19 III ag
Tours-S.ag 472 19 19 III ag
Tours-S127 19 19 III ag
Tours-S.ag1 19 19 III ag
Tours-S193 19 19 III ag
Tours-S.ag25 19 19 III ag
Tours-S105 19 19 III ag
Tours-S.ag 400 19 19 III ag
Tours-S.ag30 19 19 III ag
Paris-470 19 19 III ag
Paris-476 19 19 III ag
Tours-S.ag200 19 19 V ag
Paris-472 19 19 V ag
Tours-S.ag 396 19 19 III ag
Tours-S.ag 397 19 19 III ag
Tours-S.ag 346 singleton 327 V ag
Tours-S.ag 446 singleton 327 V ag
Tours-S.ag 485 19 19 V ag
Tours-S.ag20 19 27 III ag
Tours-C113 19 389 III ag
Tours-S.ag 435 19 19 III ag
Tours-S.ag259 19 19 II ag
Tours-S.ag17 19 28 II ag
Tours-S.ag90 19 28 Ia ag
Tours-E546 19 386 II gg
Tours-S.ag96 19 28 II tct
Paris-483 19 28 II ag
Tours-P7A2 19 28 II ag
Paris-503 19 28 III ag
Tours-C97 19 28 V ag
Tours-S.ag 362 19 28 II gga
C94 NC 388 V ag
C82 NC 388 V ag
S.ag 536 NC 26 V ag
S.ag274 NC 26 V ag
Paris-361 NC 26 V ag
S.ag43 singleton 22 II ag
S.ag 420 singleton 22 II ag

Figure S1. CRISPR1 spacer content of all (255) isolates. The CRISPR1 arrays are represented using a macro-enabled Excel tool, whereby spacers are converted into two-color symbols based on spacer sequence. Gaps (= missing spacers) are shown with a checked boxed symbol ☒ after alignment of identical spacers between strains of the same group. MLVA allelic profiles correspond to the number of repeats of each VNTR (SAG2, SAG3, SAG4, SAG7, SAG21, and SAG22). Terminal Direct Repeats (TDRs) are represented using a different color outline according to their sequence. Isolates were distributed according to their CRISPR1 array homology. Isolates were clustered according to their ancestral spacers and DRT composition into 6 clusters (CC23, CC17 – ST130, CC1-CC4**,** CC19, CC6-CC8, ST26-ST28 and ST22.

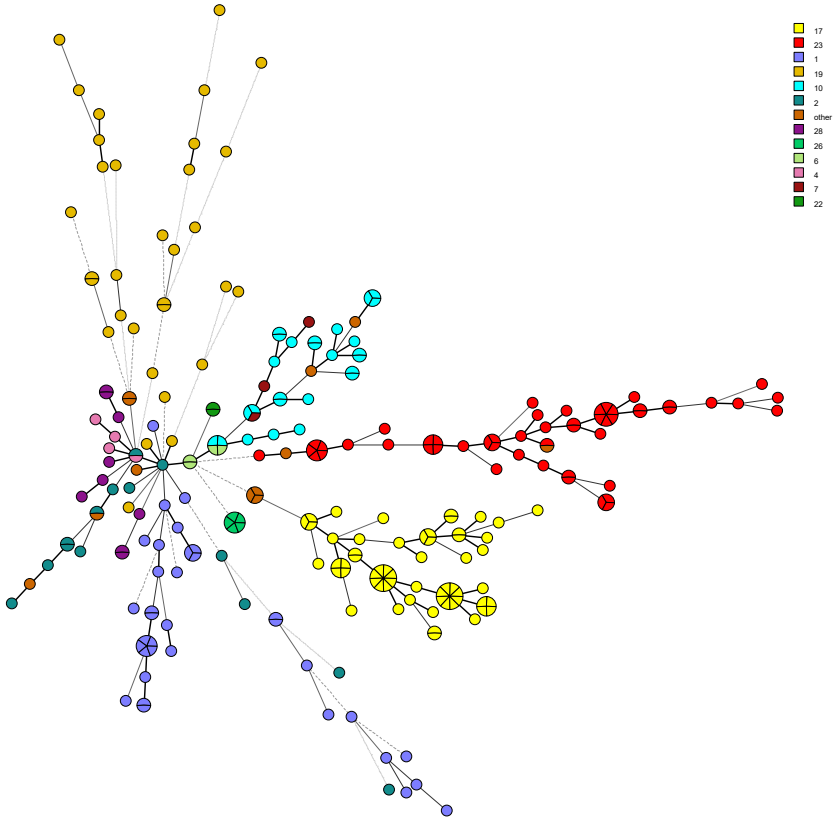


Figure S2. Minimum spanning tree (MST) representation of the 94 CRISPR1 markers clustering scheme. Each circle represents a CRISPR1 genotype and its size is proportional to the number of strains.

A

a	b	c	a	b	c
	DRT	53		44	22
	31	53		43	19
	30	46		42	22
	249	38		315	14
	247	41		314	14
	355	36		41	24
	353	24		39	27
	354	24		325	16
	246	27		326	15
	352	21		830	11
	402	26		542	11
	192	7		543	11
	466	22		544	12
	24	7		545	11
	DRTCC17	61		46 bis	19
	56	58		89	21
	55	48		88	18
	54	45		86	15
	53	47		87	19
	102	29		85	13
	1020	15		84	12
	101	16		83	8
	243	4		82	15
	48	12		120	12
	74	13		119	12
	73	14		506	14
	72	16		327	17
	1040	3		343	13
	294	5		342	11
	DRT CC1-19-10	134		341	17
	36	33		340	5
	180	25		502	10
	179	17		822	7
	178	14		335	10
	262	11		532	2
	295	13		373	8
	275	15		592	4
	307	13		591	4
	8	97		589	5
	7	45		590	6
	6	53		417	2
	46	29		427	2
	154	8		DRT ST26-388	5
	522	6		20	5
	525	5		19	5
	158	3		DRT ST22	2
	45	23		212	2

B

a	b	c
	DRT	53
	247	41
	192	7
	DRT	60
	56	57
	102	28
	72	16
	1040	3
	294	5
	DRT	135
	36	34
	8	97
	6	53
	46	29
	154	8
	45	23
	42	22
	46 bis	19
	89	21
	88	18
	87	19
	373	8
	590	6
	DRT ST26-388	5
	DRT ST22	2

Table S6. CRISPR1 marker selection. A. The 94-markers selection included 88 spacers and the five different TDR. B The 25-marker selection included the five different DRT and 20 spacers. a: spacers and TDR graphical representation / b: spacer number according to CRISPtionary or DRT c: number of representation among isolates studied.

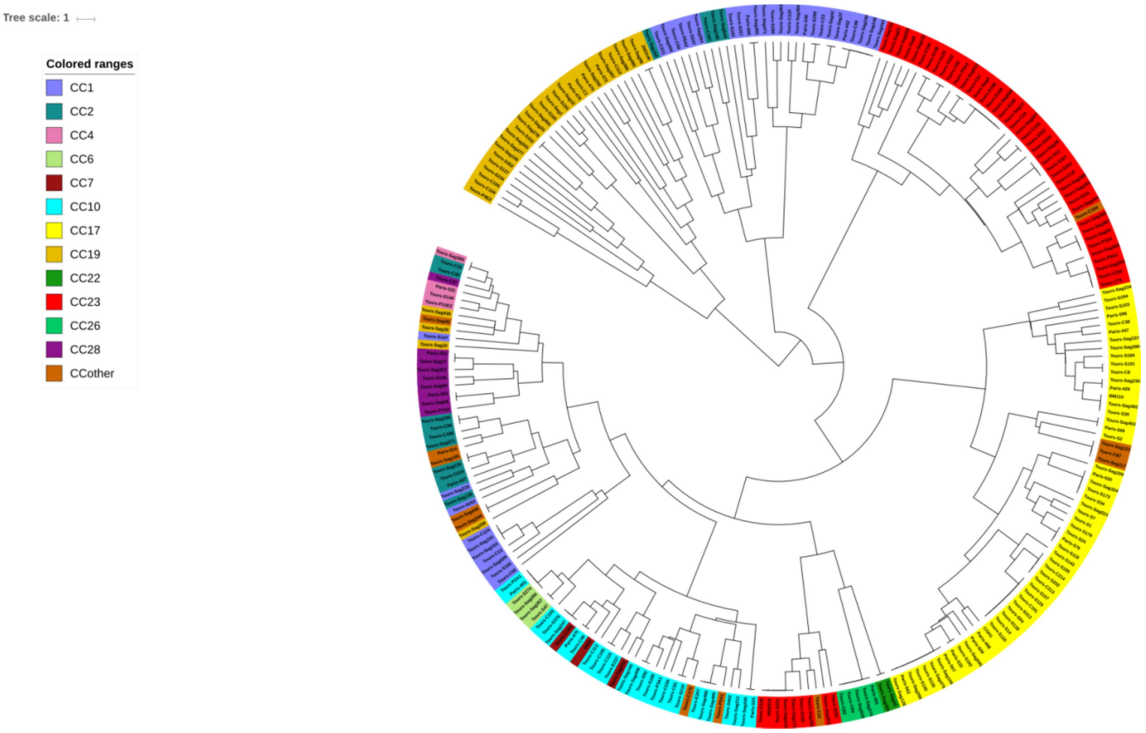


Figure S3. Unweighted pair group (UPGMA) circular tree representation of the 94 CRISPR1 markers clustering scheme.

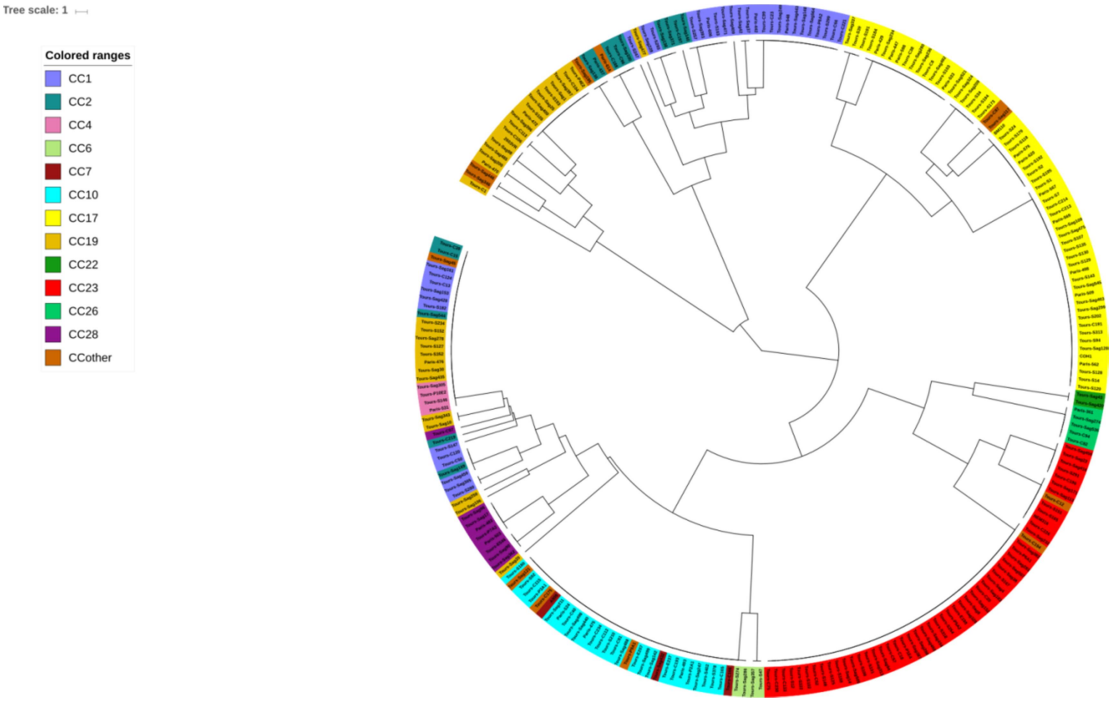


Figure S4. Unweighted pair group (UPGMA) circular tree representation of the 25 CRISPR1 markers clustering scheme.

Discussion et perspectives

L'objectif principal de ce travail de thèse était d'évaluer le système CRISPR-Cas de type II-A, et plus particulièrement le locus CRISPR1, comme marqueur épidémiologique chez *S. agalactiae*. L'utilisation du locus CRISPR1 comme outil de typage ou « typage CRISPR » a été évalué dans différents contextes au cours de ce travail.

Dans un premier temps, nous nous sommes intéressés aux souches de portage génital et ce, dans différents contextes environnementaux : au sein d'un microbiote vaginal en évolution (même patiente, différents temps d'échantillonnage), au sein d'un même microbiote vaginal (même patiente, même situation clinique [femmes enceintes, 35 à 38 semaines d'aménorrhée]) à un temps donné. Le suivi longitudinal des souches de portage nous a permis de mettre en évidence une activité du système *in vivo* par l'acquisition polarisée de spacers et de constater que les loci CRISPR1 étaient relativement stables dans le temps, confirmant leur intérêt comme outil de typage. Ces premières études ont permis de mettre en évidence l'intérêt majeur de cette technique pour l'étude des souches de portage génital de SGB ; pour le suivi longitudinal puis pour l'étude d'une population. D'autre part, une très bonne corrélation était observée entre le typage CRISPR et les autres méthodes de typage : le MLST et le MLVA.

Cet outil a ensuite été évalué à partir d'un souchier représentatif de la diversité de l'espèce (souches provenant à la fois de contextes cliniques différents (souches de portage et souches invasives, isolées de l'adulte ou de l'enfant) et de zones géographiques différentes (différents centres hospitaliers français)). Nous avons ainsi pu confirmer le fort pouvoir discriminant du typage CRISPR (un locus CRISPR1 unique pour 90% des souches et un indice de discrimination de 0,9947 et 0,9284 pour les deux selections de marqueurs) ainsi que la forte corrélation entre le typage CRISPR et les divisions phylogénétiques (définies par MLST et eBURST) et sous-divisons (définies par analyse des génomes totaux pour les souches de ST-17). Cette méthode de typage possède l'avantage

i) d'être liée à la lignée génétique des souches par l'analyse du DRT et des spacers ancestraux. Le typage CRISPR permet ainsi d'obtenir des données sur l'origine phylogénétique des souches en lien avec les CC définis par MLST et analyse eBURST (les souches appartenant au CC-1, CC-4 et CC-19 sont cependant plus difficilement distinguables).

ii) d'explorer l'évolution récente des souches, particulièrement les rencontres avec les EGMs, par l'analyse des spacers au niveau de l'extrémité leader. Par la diversité de ces spacers plus récents, le typage CRISPR offre une approche plus discriminante permettant, à l'instar du PFGE, de séparer les isolats de souches proches, notamment au sein du même ST.

D'autre part, la comparaison des quatre génomes séquencés nous permettait de confirmer les données sur les relations phylogénétiques entre ces souches obtenues par typage CRISPR. De manière similaire aux résultats obtenus par le typage CRISPR, l'analyse des SNPs mettaient en évidence pour une des deux patientes des souches très proches, dérivant probablement d'un même ancêtre commun. Cette technique de séquençage total de génome n'apportait ici pas de plus-value sur la relation entre ces souches par rapport au typage CRISPR. Compte-tenu de ces éléments, à l'issue de ce travail de thèse, nous pensons que le typage CRISPR pourrait être proposé comme nouvelle méthode de référence pour le typage des souches de SGB.

Ce travail de thèse nous a également permis de démontrer des différences en termes d'évolution et de polymorphisme des loci CRISPR1 en fonction de l'origine phylogénétique des souches. Les études précédentes avaient mis en évidence un locus CRISPR1 moins polymorphe chez les souches de CC17 avec un nombre de spacers significativement moins important (Lier et *al.* 2015). Dans notre travail, nous avons pu constater que les souches appartenant à la lignée hyper-virulente de CC-17 présentaient un locus CRISPR1 moins polymorphe et une habilité à acquérir de nouveaux spacers qui semblait plus faible que pour les autres lignées. En effet, malgré un nombre relativement important de souches de CC-17 sélectionnées pour le suivi des souches de portage, aucune modification du locus CRISPR1 n'était observée. D'autre part, les 2/3 des souches présentant un locus CRISPR1 non unique au sein des 255 souches d'origines différentes appartenaient au CC-17. Ainsi, 30% des souches de CC-17 de cette étude présentaient un locus CRISPR1 non unique. Nous avons mis en évidence que les variations des loci CRISPR1 étaient en partie dues à des modifications internes du locus (et non polarisées), possiblement en lien avec un phénomène de recombinaison homologue. Les souches de CC-17 ont par ailleurs été décrites comme appartenant à un groupe très homogène caractérisé par un faible taux de recombinaison par rapport aux autres CC, pouvant expliquer, au moins en partie, la faible variabilité des loci

CRISPR1 des souches de CC-17 (Sørensen et *al.* 2010; Da Cunha et *al.* 2014). Cette faible variabilité pourrait également être due à une plus faible ou une absence d'activité du système CRISPR-Cas de type II-A chez SGB. Cette dernière hypothèse fait l'objet d'un travail de thèse en cours actuellement au sein de notre laboratoire (Thèse d'université d'Adeline Pastuszka).

Une des limites principales de ce travail de thèse concernait l'approche culturale utilisée. En effet, seules les souches présentes après culture étaient analysées. Les souches plus « fragiles » aux conditions de transport (absence de milieu de transport, sensibilité à la dessiccation ou à la température de transport) ou plus exigeantes en facteurs de croissance (gélose utilisée pour les souches de portage : gélose sélective de type Granada®) pourraient ne pas avoir été analysées. Pour s'affranchir de ce biais, une approche indépendante de la culture semble plus appropriée. Ainsi, une des perspectives de ce travail repose sur l'étude d'une population de SGB par séquençage haut débit des loci CRISPR1 directement à partir du prélèvement (« deep sequencing »). Les échantillons seront placés immédiatement après le prélèvement dans un milieu de transport approprié pour la conservation des micro-organismes (RNAlatter), puis les loci CRISPR1 seront amplifiés puis séquencés afin d'obtenir une vision globale, indépendante de la culture de la diversité de ces populations. Cette approche pourra également être utilisée afin d'évaluer la diversité des souches de SGB dans d'autres écosystèmes (digestif, urinaire, pulmonaire…). Une autre limite de notre travail repose sur les souches sélectionnées. Bien que nous ayons sélectionné des souches de contextes cliniques différents et de centres hospitaliers différents, toutes les souches analysées étaient d'une part des souches humaines, et d'autre part des souches françaises. Ainsi, il serait intéressant d'étendre cette analyse à des souches animales (souches bovines ou souches isolées de poisson) et des souches isolées d'autres continents.

Ce travail de thèse nous a permis d'obtenir un aperçu de la cohabitation des différentes souches de SGB dans leur environnement naturel. Par le typage CRISPR nous avons mis en évidence un portage relativement homogène des souches de SGB, à la fois au cours du temps et au sein d'une population ; 83% des femmes étaient porteuses d'une même souche (même type CRISPR) au cours du temps et 80% des femmes présentaient un portage uniforme à un temps donné. De plus, les souches de SGB cohabitant au sein d'un même microbiote présentaient un fond génétique commun, laissant suggérer la possibilité d'un impact du microbiote vaginal sur la sélection des souches de portage. Du fait du faible effectif de la

cohorte sélectionnée, ces données seraient néanmoins à confirmer sur une cohorte plus importante. Le rôle du microbiote vaginal sur la sélection de souches de SGB reste aujourd'hui à évaluer. Rosen et al ont étudié l'impact du microbiote vaginal sur le portage de SGB (Rosen et *al.* 2017). Cette étude mettait en évidence que, bien que le portage de SGB ne soit pas relié à un « community state types », certains taxons majoritaires étaient associés à la colonisation par SGB, suggérant un rôle potentiel du microbiote dans la colonisation à SGB. Cette étude a été réalisée chez des femmes non enceintes et ne distinguait pas les différentes lignées de SGB. Ainsi, une perspective de ce travail serait d'étudier le rôle du microbiote dans la colonisation par les différentes lignées de SGB, notamment de la lignée ST-17, particulièrement chez les femmes enceintes (analyse par métagénomique ciblée (ou métagénétique)). D'autre part, outre un aperçu sur la cohabitation des différents clones de SGB, cette thèse nous a permis d'entrevoir la cohabitation entre ces bactéries et les EGMs présents dans cet environnement. Ainsi, une analyse de métagénomique globale (ou shot gun sequencing) nous permettrait d'obtenir une vision plus globale du microbiote, y compris viral (notamment les bactériophages).

Une autre approche à plus long terme consisterait à analyser l'aptitude des différentes lignées de *S. agalactiae* à la colonisation dans un modèle murin (Patras & Doran 2016). Des souches de *S. agalactiae* de différentes lignées phylogénétiques seront administrées individuellement puis collectivement afin d'évaluer la capacité des différentes lignées à coloniser leur hôte. Un suivi des souches de colonisation sera ensuite effectué à différents temps. Ceci nous permettra de mettre en évidence une éventuelle différence de colonisation des souches en fonction de leur appartenance phylogénétique. Cette étude nous permettra d'avoir un premier élément de compréhension de la physiopathologie des infections à *S. agalactiae*.

Par ce travail, nous avons pu démontrer une acquisition polarisée de spacers *in vivo* en réponse à des EGMs (plasmides et bactériophages). Cette acquisition n'a, à ce jour, jamais été démontrée *in vitro* chez *S. agalactiae.* D'autre part, la fonctionnalité du système CRISPR-Cas de type II-A a été mise en évidence précédemment *in vitro* uniquement en réponse à un élements integratif conjugatif (réduction du transfert d'un plasmide en présence d'un spacer correspondant) (Lopez-Sanchez et *al.* 2012). Une réponse spécifique à un bactériophage n'a pas encore été obtenue *in vitro* chez *S. agalactiae* à ce jour. D'autre part, à notre connaissance, aucun bactériophage lytique de *S. agalactiae* n'a été isolé. Une des perspectives

de ce travail porte sur l'analyse fonctionnelle du système CRISPR-Cas chez *S. agalactiae* et l'étude de son interaction avec des bactériophages. Ceci reposera sur la détection, l'isolement et la multiplication de phages lytiques à partir d'échantillons biologiques de différentes origines (prélèvements génitaux et digestifs). Ces phages seront ensuite caractérisés afin d'étudier leur impact *in vitro* sur les souches de *S. agalactiae* appartenant à différentes lignées phylogénétiques. Ces souches seront soumises à des infections par des bactériophages préalablement caractérisés et la composition du locus CRISPR1 des souches bactériennes infectées sera analysée par PCR et séquençage. Cette même analyse pourra ensuite être répétée *in vivo* dans le modèle rongeur précédemment évoqué.

Conclusion

Préalablement à ce travail, le locus CRISPR1 de *S. agalactiae* était décrit comme ubiquitaire et relativement polymorphe, faisant de ce locus un marqueur phylogénétique séduisant pour le génotypage et l'analyse phylogénétique des souches. L'objectif principal de ce travail de thèse était d'évaluer le locus CRISPR1 de *S. agalactiae* comme marqueur épidémiologique.

Pour cela, la 1e étape de notre travail a consisté à mettre en évidence une activité du système *in vivo* par l'acquisition polarisé de spacers chez des souches de portage. Ce suivi longitudinal de souches nous a permis par ailleurs de constater que les loci CRISPR1 étaient relativement stables dans le temps et pouvaient ainsi être utilisés comme outil de typage pour le suivi épidémiologique des souches. La 2e étape a consisté à développer une approche de typage basée sur le locus CRISPR1 adaptée à l'analyse d'un nombre important de souches (amplification et restriction enzymatique du locus CRISPR1 (CRISPR-RFLP)). L'outil développé nous a permis d'étudier par approche culturale une population de portage vaginal, permettant de mettre en évidence une relative homogénéité des souches au sein de cette niche. La 3e étape consistait à étendre ces données de typage à un nombre relativement important de souches humaines de différentes origines, permettant de i) confirmer le fort polymorphisme du locus CRISPR1 ; ii) mettre en évidence des différences de polymorphisme en fonction de l'origine phylogénétique des souches ; iii) confirmer le lien entre la composition du locus CRISPR1 et l'origine phylogénétique des souches ainsi que la forte corrélation entre cette méthode de typage et notamment la méthode de référence actuelle : le MLST ; iv) mettre en évidence un fort pouvoir discriminant du typage CRISPR, supérieur au MLST, permettant de séparer les souches au sein d'un même ST.

L'analyse de ce locus permet d'obtenir des données sur l'origine phylogénétique des souches et son appartenance aux différentes lignées en lien avec les CC et les ST définis par MLST, mais également d'avoir une approche plus discriminante permettant de séparer les isolats au sein des mêmes CC et ST. Les liens phylogénétiques entre des souches présentant des loci CRISPR1 relativement proches ont par ailleurs été confirmés par l'analyse comparative des génomes complets. Outre son intérêt comme marqueur de la phylogénie de l'espèce, l'analyse de ce locus nous donne également des informations sur l'évolution récente des isolats, notamment sur leurs contacts récents avec des EGMs. La méthode de typage basée sur le locus CRISPR est une méthode génotypique simple, reproductible, facile à interpréter,

facilement exportable entre les laboratoires et apparaît comme un outil pratique pour typer les isolats de SGB. Toutes ces données font qu'à l'issue de ce travail de thèse nous suggérons que le typage CRISPR puisse devenir la méthode de référence pour le typage des souches de *S. agalactiae*.

Références bibliographiques

Al Nakib, Malik, Magalie Longo, Asmaa Tazi, Annick Billoet, Josette Raymond, Patrick Trieu-Cuot, and Claire Poyart. 2011. "Comparison of the Diversilab® System with Multi-Locus Sequence Typing and Pulsed-Field Gel Electrophoresis for the Characterization of *Streptococcus agalactiae* Invasive Strains." *Journal of Microbiological Methods* 85 (2): 137–42. https://doi.org/10.1016/j.mimet.2011.02.010.

Al Safadi, Rim, Souheila Amor, Geneviève Hery-Arnaud, Barbara Spellerberg, Philippe Lanotte, Laurent Mereghetti, François Gannier, Roland Quentin, and Agnès Rosenau. 2010. "Enhanced expression of *lmb* gene encoding laminin-binding protein in *Streptococcus agalactiae* strains harboring IS1548 in *scpB-lmb* intergenic region." Edited by Frank R. DeLeo. *PLoS ONE* 5 (5): e10794. https://doi.org/10.1371/journal.pone.0010794.

Al-Attar, Sinan, Edze R. Westra, John van der Oost, and Stan J.J. Brouns. 2011. "Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs): The hallmark of an ingenious antiviral defense mechanism in prokaryotes." *Biological Chemistry* 392 (4). https://doi.org/10.1515/bc.2011.042.

Alhhazmi, Areej, Donna Hurteau, and Gregory J. Tyrrell. 2016. "Epidemiology of invasive group B streptococcal disease in Alberta, Canada, from 2003 to 2013." Edited by K. C. Carroll. *Journal of Clinical Microbiology* 54 (7): 1774–81. https://doi.org/10.1128/JCM.00355-16.

Al-Maani, Amal, Laurie Streitenberger Streitenberger, Megan Clarke, Yvonne C. W. Yau, Danuta Kovach, Rick Wray, and Anne Matlow. 2014. "Nosocomial transmission of group B *Streptococci* proven by positive environmental culture." *Oman Medical Journal* 29 (5): 376–79. https://doi.org/10.5001/omj.2014.100.

Amin, A., Y.M. Abdulrazzaq, and S. Uduman. 2002a. "Group B streptococcal serotype distribution of isolates from colonized pregnant women at the time of delivery in United Arab Emirates." *Journal of Infection* 45 (1): 42–46. https://doi.org/10.1053/jinf.2001.0990.

Amitai, Gil, and Rotem Sorek. 2016. "CRISPR-Cas adaptation: insights into the mechanism of action." *Nature Reviews. Microbiology* 14 (2): 67–76. https://doi.org/10.1038/nrmicro.2015.14.

ANAES. 2001. "[Antenatal prevention of early neonatal bacterial infection (September 2001)]." *Journal De Gynécologie, Obstétrique et Biologie de la Reproduction* 32 (1): 68–74.

Andersson, Anders F., and Jillian F. Banfield. 2008. "Virus population dynamics and acquired virus resistance in natural microbial communities." *Science (New York, N.Y.)* 320 (5879): 1047–50. https://doi.org/10.1126/science.1157358.

Anthony, Bascom F., Donald M. Okada, and Calvin J. Hobel. 1978. "Epidemiology of Group B *Streptococcu*s: longitudinal observations during pregnancy." *Journal of Infectious Diseases* 137 (5): 524–530.

Avci, Fikri Y., and Dennis L. Kasper. 2010. "How Bacterial Carbohydrates Influence the Adaptive Immune System." *Annual Review of Immunology* 28 (1): 107–30. https://doi.org/10.1146/annurev-immunol-030409-101159.

Badri, M. S., S. Zawaneh, A. C. Cruz, G. Mantilla, H. Baer, W. N. Spellacy, and E. M. Ayoub. 1977. "Rectal colonization with Group B *Streptococcus*: relation to vaginal colonization of pregnant women." *The Journal of Infectious Diseases* 135 (2): 308–12.

Baker, C. J. 1978. "Early onset Group B streptococcal disease." *The Journal of Pediatrics* 93 (1): 124–25.

Baker, C. J. 1997. “Group B streptococcal infections.” *Clinics in Perinatology* 24 (1): 59–70.

Baker, Carol J., and Fred F. Barrett. 1974. “Group B streptococcal infections in infants: the importance of the various serotypes.” *JAMA* 230 (8): 1158–60. https://doi.org/10.1001/jama.1974.03240080040025.

Bakker, Henk C. den, Marc W. Allard, Dianna Bopp, Eric W. Brown, John Fontana, Zamin Iqbal, Aristea Kinney, et *al.* 2014. “Rapid whole-genome sequencing for surveillance of *Salmonella Enterica* Serovar Enteritidis.” *Emerging Infectious Diseases* 20 (8): 1306–14. https://doi.org/10.3201/eid2008.131399.

Ballard, Mark S., Henrik C. Schønheyder, Jenny Dahl Knudsen, Outi Lyytikäinen, Matthew Dryden, Karina J. Kennedy, Louis Valiquette, et *al.* 2016. “The Changing epidemiology of Group B *Streptococcus* bloodstream infection: a multi-national population-based assessment.” *Infectious Diseases* 48 (5): 386–91. https://doi.org/10.3109/23744235.2015.1131330.

Barbera, L, L. lafont, A. ouachee, S. Dos Santos, A-S. Valentin, M. Decalonne, C. Decreux, C. Hombrock, O. Lehiani, V. Morange, G. Petit, V. Vialette, N. Van der mee-marquet. 2016. Etude multicentrique des bactériémies associées à Streptococcus agalactiae (2007-2015). RICAI (Réunion Interdisciplinaire de chimiothéparie anti-infectieuse) : Communication iorale CO-039.

Barcaite, Egle, Arnoldas Bartusevicius, Rasa Tameliene, Mindaugas Kliucinskas, Laima Maleckiene, and Ruta Nadisauskiene. 2008. “Prevalence of maternal group B streptococcal colonisation in european countries.” *Acta Obstetricia Et Gynecologica Scandinavica* 87 (3): 260–71. https://doi.org/10.1080/00016340801908759.

Barrangou, Rodolphe. 2013. “CRISPR-Cas Systems and RNA-Guided Interference.” *Wiley Interdisciplinary Reviews: RNA* 4 (3): 267–78. https://doi.org/10.1002/wrna.1159.

Barrangou, Rodolphe, and Edward G. Dudley. 2016. “CRISPR-Based typing and next-generation tracking technologies.” *Annual Review of Food Science and Technology* 7 (1): 395–411. https://doi.org/10.1146/annurev-food-022814-015729.

Barrangou, Rodolphe, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A. Romero, and Philippe Horvath. 2007. “CRISPR provides acquired resistance against viruses in prokaryotes.” *Science (New York, N.Y.)* 315 (5819): 1709–12. https://doi.org/10.1126/science.1138140.

Barrangou, Rodolphe, and Philippe Horvath. 2017. “A decade of discovery: CRISPR functions and applications.” *Nature Microbiology* 2 (June): 17092. https://doi.org/10.1038/nmicrobiol.2017.92.

Barrangou, Rodolphe, and Luciano A. Marraffini. 2014. “CRISPR-Cas Systems: prokaryotes upgrade to adaptive immunity.” *Molecular Cell* 54 (2): 234–44. https://doi.org/10.1016/j.molcel.2014.03.011.

Bekker, Vincent, Merijn W. Bijlsma, Diederik van de Beek, Taco W. Kuijpers, and Arie van der Ende. 2014. “Incidence of invasive group B streptococcal disease and pathogen genotype distribution in newborn babies in the netherlands over 25 years: A Nationwide Surveillance Study.” *The Lancet Infectious Diseases* 14 (11): 1083–1089.

Belkum, A. van. 1999. “Short sequence repeats in microbial pathogenesis and evolution.” *Cellular and Molecular Life Sciences: CMLS* 56 (9–10): 729–34.

Bergseng, H., J.E. Afset, A. Radtke, K. Loeseth, R.V. Lyng, M. Rygg, and K. Bergh. 2009. “Molecular and phenotypic characterization of invasive group B *Streptococcus* strains from infants in Norway 2006–2007.” *Clinical Microbiology and Infection* 15 (12): 1182–85. https://doi.org/10.1111/j.1469-0691.2009.02789.x.

Bergseng, H., M. Rygg, L. Bevanger, and K. Bergh. 2008. “Invasive Group B *Streptococcus* (GBS) disease in Norway 1996–2006.” *European Journal of Clinical Microbiology & Infectious Diseases* 27 (12): 1193–99. https://doi.org/10.1007/s10096-008-0565-8.

Bidet, Philippe, Naima Brahimi, Céline Chalas, Yannick Aujard, and Edouard Bingen. 2003. "Molecular characterization of serotype III Group B–*Streptococcus* isolates causing neonatal meningitis." *The Journal of Infectious Diseases* 188 (8): 1132–1137.

Bingen, E., E. Denamur, N. Lambert-Zechovsky, Y. Aujard, N. Brahimi, P. Geslin, and J. Elion. 1992. "Analysis of DNA restriction fragment length polymorphism extends the evidence for breast milk transmission in *Streptococcus agalactiae* late-onset neonatal infection." *The Journal of Infectious Diseases* 165 (3): 569–73.

Bisharat, N. 2005. "Population structure of Group B *Streptococcus* from a low-incidence region for invasive neonatal disease." *Microbiology* 151 (6): 1875–81. https://doi.org/10.1099/mic.0.27826-0.

Björnsdóttir, E.S., E.R. Martins, H. Erlendsdóttir, G. Haraldsson, J. Melo-Cristino, K.G. Kristinsson, and M. Ramirez. 2016. "Changing epidemiology of group B streptococcal infections among adults in Iceland: 1975–2014." *Clinical Microbiology and Infection* 22 (4): 379.e9-379.e16. https://doi.org/10.1016/j.cmi.2015.11.020.

Blancas, D., M. Santin, M. Olmo, F. Alcaide, J. Carratala, and F. Gudiol. 2004. "Group B streptococcal disease in nonpregnant adults: incidence, clinical characteristics, and outcome." *European Journal of Clinical Microbiology & Infectious Diseases* 23 (3): 168–73. https://doi.org/10.1007/s10096-003-1098-9.

Bliss, Sandra J., Shannon D. Manning, Patricia Tallman, Carol J. Baker, Mark D. Pearlman, Carl F. Marrs, and Betsy Foxman. 2002. "Group B *Streptococcus* colonization in male and nonpregnant female university students: a cross-sectional prevalence study." *Clinical Infectious Diseases* 34 (2): 184–190.

Bohnsack, J. F. 2002. "Long-Range Mapping of the *Streptococcus agalactiae* phylogenetic lineage restriction digest pattern type III-3 reveals clustering of virulence genes." *Infection and Immunity* 70 (1): 134–39. https://doi.org/10.1128/IAI.70.1.134-139.2002.

Bohnsack, John F., April Whiting, Marcelo Gottschalk, Diane Marie Dunn, Robert Weiss, Parvin H. Azimi, Joseph B. Philips, Leonard E. Weisman, George G. Rhoads, and Feng-Ying C. Lin. 2008. "Population Structure of invasive and colonizing strains of *Streptococcus agalactiae* from neonates of six u.s. academic centers from 1995 to 1999." *Journal of Clinical Microbiology* 46 (4): 1285–91. https://doi.org/10.1128/JCM.02105-07.

Bolotin, Alexander, Benoit Quinquis, Alexei Sorokin, and S. Dusko Ehrlich. 2005. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) have spacers of extrachromosomal origin." *Microbiology (Reading, England)* 151 (Pt 8): 2551–61. https://doi.org/10.1099/mic.0.28048-0.

Both, Ulrich von, Andrea John, Kirsten Fluegge, Anette Siedler, and Reinhard Berner. 2008. "molecular epidemiology of invasive neonatal *Streptococcus agalactiae* isolates in Germany." *The Pediatric Infectious Disease Journal* 27 (10): 903–6. https://doi.org/10.1097/INF.0b013e318178d1ff.

Boyer, K. M., and S. P. Gotoff. 1986. "prevention of early-onset neonatal group B streptococcal disease with selective intrapartum chemoprophylaxis." *The New England Journal of Medicine* 314 (26): 1665–69. https://doi.org/10.1056/NEJM198606263142603.

Brochet, Mathieu, Elisabeth Couvé, Mohamed Zouine, Tatiana Vallaeys, Christophe Rusniok, Marie-Cécile Lamy, Carmen Buchrieser, et *al.* 2006. "genomic diversity and evolution within the species *Streptococcus agalactiae*." *Microbes and Infection* 8 (5): 1227–43. https://doi.org/10.1016/j.micinf.2005.11.010.

Brussow, H., C. Canchaya, and W.-D. Hardt. 2004. "Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion." *Microbiology and*

Molecular Biology Reviews 68 (3): 560–602. https://doi.org/10.1128/MMBR.68.3.560-602.2004.

Bulle, Béatrice, Laurence Millon, Jean-Mathieu Bart, Montserrat Gállego, Françoise Gambarelli, Montserrat Portús, Lee Schnur, et *al.* 2002. "Practical approach for typing strains of leishmania infantum by microsatellite analysis." *Journal of Clinical Microbiology* 40 (9): 3391–97. https://doi.org/10.1128/JCM.40.9.3391-3397.2002.

Burstein, David, Christine L. Sun, Christopher T. Brown, Itai Sharon, Karthik Anantharaman, Alexander J. Probst, Brian C. Thomas, and Jillian F. Banfield. 2016. "Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems." *Nature Communications* 7 (February): 10613. https://doi.org/10.1038/ncomms10613.

Charpentier, Emmanuelle, Hagen Richter, John van der Oost, and Malcolm F. White. 2015. "Biogenesis pathways of rna guides in archaeal and bacterial CRISPR-Cas adaptive immunity." *FEMS Microbiology Reviews* 39 (3): 428–41. https://doi.org/10.1093/femsre/fuv023.

Chatellier, Sonia, Helene Huet, Samia Kenzi, Agnes Rosenau, Pierre Geslin, and Roland Quentin. 1996. "Genetic diversity of RRNA operons of unrelated streptococcus agalactiae strains isolated from cerebrospinal fluid of neonates suffering from meningitis." *Journal of Clinical Microbiology* 34 (11): 2741–2747.

Chatellier, Sonia, Celine Ramanantsoa, Patrick Harriau, Karine Rolland, Agnes Rosenau, and Roland Quentin. 1997. "Characterization of *Streptococcus agalactiae* Strains by randomly amplified polymorphic DNA analysis." *Journal of Clinical Microbiology* 35 (10): 2573–2579.

Chau, Man L., Swaine L. Chen, Min Yap, Sri H. P. Hartantyo, Paul K. T. Chiew, Charlene J. Fernandez, Wai K. Wong, et *al.* 2017. "Group B *Streptococcus* infections caused by improper sourcing and handling of fish for raw consumption, Singapore, 2015-2016." *Emerging Infectious Diseases* 23 (12). https://doi.org/10.3201/eid2312.170596.

Chin, Chen-Shan, Jon Sorenson, Jason B. Harris, William P. Robins, Richelle C. Charles, Roger R. Jean-Charles, James Bullard, et *al.* 2011. "The Origin of the haitian cholera outbreak strain." *New England Journal of Medicine* 364 (1): 33–42. https://doi.org/10.1056/NEJMoa1012928.

Christensen, K. K., and P. Christensen. 1978. "Typing of Group B *Streptococci* from the throat and urogenital tract of females." *Scandinavian Journal of Infectious Diseases* 10 (3): 209–12.

Chylinski, K., K. S. Makarova, E. Charpentier, and E. V. Koonin. 2014. "Classification and evolution of type II CRISPR-Cas systems." *Nucleic Acids Research* 42 (10): 6091–6105. https://doi.org/10.1093/nar/gku241.

CNGOF. 1997. "CNGOF - Recommandations Pour La Pratique Médicale - Infection Cervico-Vaginales et Grossesse (1997)." 1997. http://www.cngof.asso.fr/D_PAGES/PURPC_02.HTM.

CNR des Streptocoques. 2014. "Rapport d'activité du cnr des Streptocoques." 2014. https://cnr-strep.fr/images/CNR-STREP/rapport/Rapport_activite_CNR_Strep_2014.pdf.

CNR des Streptocoques. 2016. "Rapport d'activité Du CNR Des Streptocoques."

CNR des Streptocoques. "CNR-Strep. Expertise." Accessed June 11, 2018. https://cnr-strep.fr/index.php/expertises/recherche-de-lien-de-clonalite.

Cui, Yujun, Yanjun Li, Olivier Gorgé, Mikhail E. Platonov, Yanfeng Yan, Zhaobiao Guo, Christine Pourcel, et *al.* 2008. "Insight into microevolution of *Yersinia pestis* by Clustered Regularly Interspaced Short Palindromic Repeats." Edited by Niyaz Ahmed. *PLoS ONE* 3 (7): e2652. https://doi.org/10.1371/journal.pone.0002652.

Da Cunha, Violette, Mark R. Davies, Pierre-Emmanuel Douarre, Isabelle Rosinski-Chupin, Immaculada Margarit, Sebastien Spinali, Tim Perkins, et *al.* 2014. "*Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline." *Nature Communications* 5 (August). https://doi.org/10.1038/ncomms5544.

Dallman, Timothy J., Lisa Byrne, Philip M. Ashton, Lauren A. Cowley, Neil T. Perry, Goutam Adak, Liljana Petrovska, et *al.* 2015. "Whole-Genome Sequencing for national surveillance of shiga toxin–producing *Escherichia coli* O157." *Clinical Infectious Diseases* 61 (3): 305–12. https://doi.org/10.1093/cid/civ318.

Deltcheva, Elitza, Krzysztof Chylinski, Cynthia M. Sharma, Karine Gonzales, Yanjie Chao, Zaid A. Pirzada, Maria R. Eckert, Jörg Vogel, and Emmanuelle Charpentier. 2011. "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." *Nature* 471 (7340): 602–7. https://doi.org/10.1038/nature09886.

Denning, D. W., C. J. Baker, N. J. Troup, and L. S. Tompkins. 1989. "Restriction endonuclease analysis of human and bovine group B *Streptococci* for epidemiologic study." *Journal of Clinical Microbiology* 27 (6): 1352–1356.

Deplano, Ariane, Annette Schuermans, Johan Van Eldere, Wolfgang Witte, Hèléne Meugnier, Jerome Etienne, Hajo Grundmann, et *al.* 2000. "Multicenter evaluation of epidemiological typing of methicillin-resistant *Staphylococcus aureus* strains by repetitive-element PCR analysis." *Journal of Clinical Microbiology* 38 (10): 3527–3533.

Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath, and S. Moineau. 2008. "Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*." *Journal of Bacteriology* 190 (4): 1390–1400. https://doi.org/10.1128/JB.01412-07.

Diedrick, M. J., A. E. Flores, S. L. Hillier, R. Creti, and P. Ferrieri. 2010. "Clonal analysis of colonizing group B *Streptococcus*, Serotype IV, an emerging pathogen in the United States." *Journal of Clinical Microbiology* 48 (9): 3100–3104. https://doi.org/10.1128/JCM.00277-10.

Dillon, H. C., E. Gray, M. A. Pass, and B. M. Gray. 1982. "Anorectal and vaginal carriage of Group B *Streptococci* during pregnancy." *The Journal of Infectious Diseases* 145 (6): 794–99.

Domelier, A.-S., N. van der Mee-Marquet, L. Arnault, L. Mereghetti, P. Lanotte, A. Rosenau, M.-F. Lartigue, and R. Quentin. 2008. "Molecular characterization of erythromycin-resistant *Streptococcus agalactiae* strains." *Journal of Antimicrobial Chemotherapy* 62 (6): 1227–33. https://doi.org/10.1093/jac/dkn388.

Edmond, Karen M, Christina Kortsalioudaki, Susana Scott, Stephanie J Schrag, Anita KM Zaidi, Simon Cousens, and Paul T Heath. 2012. "Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis." *The Lancet* 379 (9815): 547–56. https://doi.org/10.1016/S0140-6736(11)61651-6.

Eickhoff, Theodore C., Jerome O. Klein, A. Kathleen Daly, David Ingall, and Maxwell Finland. 1964. "Neonatal sepsis and other infections due to group B beta-hemolytic *Streptococci.*" *New England Journal of Medicine* 271 (24): 1221–28. https://doi.org/10.1056/NEJM196412102712401.

Elliott, J. A., K. D. Farmer, and R. R. Facklam. 1998. "Sudden increase in isolation of group B *Streptococci*, serotype V, is not due to emergence of a new pulsed-field gel electrophoresis type." *Journal of Clinical Microbiology* 36 (7): 2115–2116.

Eyre, David W, Tanya Golubchik, N Claire Gordon, Rory Bowden, Paolo Piazza, Elizabeth M Batty, Camilla L C Ip, et *al.* 2012. "A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and

surveillance." *BMJ Open* 2 (3): e001124. https://doi.org/10.1136/bmjopen-2012-001124.

Fabre, Laëtitia, Jian Zhang, Ghislaine Guigon, Simon Le Hello, Véronique Guibert, Marie Accou-Demartin, Saïana de Romans, et *al.* 2012. "CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections." *PloS One* 7 (5): e36995. https://doi.org/10.1371/journal.pone.0036995.

Farley, M. M. 2001. "Group B streptococcal disease in nonpregnant adults." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 33 (4): 556–61. https://doi.org/10.1086/322696.

Fasola, ELIZABETH, CAROL Livdahl, and PATRICIA Ferrieri. 1993. "Molecular analysis of multiple isolates of the major serotypes of group B *Streptococci*." *Journal of Clinical Microbiology* 31 (10): 2616–2620.

Feil, Edward J., Bao C. Li, David M. Aanensen, William P. Hanage, and Brian G. Spratt. 2004. "EBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data." *Journal of Bacteriology* 186 (5): 1518–30. https://doi.org/10.1128/JB.186.5.1518-1530.2004.

Ferrieri, Patricia, Ruth Lynfield, Roberta Creti, and Aurea E. Flores. 2013. "Serotype IV and invasive group B *Streptococcus* disease in neonates, Minnesota, USA, 2000–20101." *Emerging Infectious Diseases* 19 (4): 551–58. https://doi.org/10.3201/eid1904.121572.

Florindo, C., V. Damiao, I. Silvestre, C. Farinha, F. Rodrigues, F. Nogueira, F. Martins-Pereira, R. Castro, and M. J. Borrego. 2014. "Epidemiological surveillance of colonising group B *Streptococcus* epidemiology in the lisbon and tagus valley regions, Portugal (2005 to 2012): Emergence of a New Epidemic Type IV/Clonal Complex 17 Clone." *Methods* 10 (16): 20.

Foxman, B. 2006. "Incidence and duration of Group B *Streptococcus* by serotype among male and female college students living in a single dormitory." *American Journal of Epidemiology* 163 (6): 544–51. https://doi.org/10.1093/aje/kwj075.

Fricke, W. F., M. K. Mammel, P. F. McDermott, C. Tartera, D. G. White, J. E. LeClerc, J. Ravel, and T. A. Cebula. 2011. "Comparative genomics of 28 *Salmonella* Enterica isolates: evidence for CRISPR-mediated adaptive sublineage evolution." *Journal of Bacteriology* 193 (14): 3556–68. https://doi.org/10.1128/JB.00297-11.

Frothingham, Richard, and Winifred A. Meeker-O'Connell. 1998. "Genetic diversity in the *Mycobacterium tuberculosi*s complex based on variable numbers of tandem DNA repeats." *Microbiology* 144 (5): 1189–1196.

Fry, R.M. 1938. "Fatal Infections by haemolytic *Streptococcus* Group B." *The Lancet* 231 (5969): 198–199.

Gardy, Jennifer L., James C. Johnston, Shannan J. Ho Sui, Victoria J. Cook, Lena Shah, Elizabeth Brodkin, Shirley Rempel, et *al.* 2011. "Whole-genome sequencing and social-network analysis of a tuberculosis outbreak." *New England Journal of Medicine* 364 (8): 730–739.

Gasiunas, G., R. Barrangou, P. Horvath, and V. Siksnys. 2012. "Cas9-CrRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." *Proceedings of the National Academy of Sciences* 109 (39): E2579–86. https://doi.org/10.1073/pnas.1208507109.

Ginevra, C., N. Jacotin, L. Diancourt, G. Guigon, R. Arquilliere, H. Meugnier, G. Descours, et *al.* 2012. "*Legionella Pneumophila* Sequence Type 1/Paris pulsotype subtyping by spoligotyping." *Journal of Clinical Microbiology* 50 (3): 696–701. https://doi.org/10.1128/JCM.06180-11.

Glaser, Philippe, Christophe Rusniok, Carmen Buchrieser, Fabien Chevalier, Lionel Frangeul, Tarek Msadek, Mohamed Zouine, et *al.* 2002. "Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease." *Molecular Microbiology* 45 (6): 1499–1513.

Gordillo, M. E., K. V. Singh, C. J. Baker, and B. E. Murray. 1993. "Typing of Group B *Streptococci*: comparison of pulsed-field gel electrophoresis and conventional electrophoresis." *Journal of Clinical Microbiology* 31 (6): 1430–1434.

Granlund, M., L. Oberg, M. Sellin, and M. Norgren. 1998. "Identification of a novel insertion element, IS1548, in Group B *Streptococci*, predominantly in strains causing endocarditis." *The Journal of Infectious Diseases* 177 (4): 967–76.

Grissa, Ibtissem, Gilles Vergnaud, and Christine Pourcel. 2007. "The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats." *BMC Bioinformatics* 8 (1): 172. https://doi.org/10.1186/1471-2105-8-172.

Groenen, Peter, Annelies E. Bunschoten, Dick van Soolingen, and Jan DA van Errtbden. 1993. "Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method." *Molecular Microbiology* 10 (5): 1057–1065.

Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 2005; 1:e60. [PubMed: 16292354]

Haguenoer, Eve, Gaelle Baty, Christine Pourcel, Marie-Frédérique Lartigue, Anne-Sophie Domelier, Agnès Rosenau, Roland Quentin, Laurent Mereghetti, and Philippe Lanotte. 2011. "A Multi Locus Variable Number of Tandem Repeat Analysis (MLVA) scheme for *Streptococcus agalactiae* genotyping." *BMC Microbiology* 11 (1): 171.

Håkansson, Stellan, Pia Axemo, Katarina Bremme, Anna-Lena Bryngelsson, Marie Carlsson Wallin, Carl-Magnus Ekström, Margareta Granlund, et *al.* 2008. "Group B streptococcal carriage in Sweden: a national study on risk factors for mother and infant colonisation." *Acta Obstetricia Et Gynecologica Scandinavica* 87 (1): 50–58. https://doi.org/10.1080/00016340701802888.

Hale, Caryn, Kyle Kleppe, Rebecca M. Terns, and Michael P. Terns. 2008. "Prokaryotic silencing (Psi)RNAs in *Pyrococcus Furiosus*." *RNA (New York, N.Y.)* 14 (12): 2572–79. https://doi.org/10.1261/rna.1246808.

Hansen, Søren Mose, Niels Uldbjerg, Mogens Kilian, and Uffe B. Skov Sørensen. 2004. "Dynamics of *Streptococcus agalactiae* colonization in women during and after pregnancy and in their infants." *Journal of Clinical Microbiology* 42 (1): 83–89.

Harrison, Lee H., Diane M. Dwyer, and Judith A. Johnson. 1995. "Emergence of serotype V Group B streptococcal infection among infants and adults." *The Journal of Infectious Diseases* 171 (2): 511–513.

Hauge, Majbritt, Christina Jespersgaard, Knud Poulsen, and Mogens Kilian. 1996. "Population structure of *Streptococcus agalactiae* reveals an association between specific evolutionary lineages and putative virulence factors but not disease." *Infection and Immunity* 64 (3): 919–925.

Hedegaard, Jakob, Majbritt Hauge, Jeppe Fage-Larsen, Kim Kusk Mortensen, Mogens Kilian, Hans Uffe Sperling-Petersen, and Knud Poulsen. 2000. "Investigation of the Translation-Initiation Factor IF2 Gene, InfB, as a tool to study the population structure of *Streptococcus agalactiae*." *Microbiology* 146 (7): 1661–1670.

Heidelberg, John F., William C. Nelson, Thomas Schoenfeld, and Devaki Bhaya. 2009. "Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes." Edited by Niyaz Ahmed. *PLoS ONE* 4 (1): e4169. https://doi.org/10.1371/journal.pone.0004169.

Held, Nicole L., Alfa Herrera, Hinsby Cadillo-Quiroz, and Rachel J. Whitaker. 2010. "CRISPR associated diversity within a population of *Sulfolobus islandicus*." Edited by Paul J. Planet. *PLoS ONE* 5 (9): e12988. https://doi.org/10.1371/journal.pone.0012988.

Held, Nicole L., and Rachel J. Whitaker. 2009. "viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes." *Environmental Microbiology* 11 (2): 457–66. https://doi.org/10.1111/j.1462-2920.2008.01784.x.

Heler, Robert, Poulami Samai, Joshua W. Modell, Catherine Weiner, Gregory W. Goldberg, David Bikard, and Luciano A. Marraffini. 2015. "Cas9 specifies functional viral targets during CRISPR–Cas adaptation." *Nature* 519 (7542): 199–202. https://doi.org/10.1038/nature14245.

Helmig, Rikke, Niels Uldbjerg, Jane Boris, and Mogens Kilian. 1993. "Clonal analysis of *Streptococcus agalactiae* isolated from infants with neonatal sepsis or meningitis and their mothers and from healthy pregnant women." *Journal of Infectious Diseases* 168 (4): 904–909.

Herbert, Mark A., Catriona JE Beveridge, David McCormick, Emmelien Aten, Nicola Jones, Lori AS Snyder, and Nigel J. Saunders. 2005. "Genetic islands of *streptococcus agalactiae* strains NEM316 and 2603VR and their presence in other group B streptococcal strains." *BMC Microbiology* 5 (1): 31.

Hery-Arnaud, G., G. Bruant, P. Lanotte, S. Brun, B. Picard, A. Rosenau, N. van der Mee-Marquet, P. Rainard, R. Quentin, and L. Mereghetti. 2007. "Mobile genetic elements provide evidence for a bovine origin of clonal complex 17 of *Streptococcus agalactiae*." *Applied and Environmental Microbiology* 73 (14): 4668–72. https://doi.org/10.1128/AEM.02604-06.

Hery-Arnaud, G., G. Bruant, P. Lanotte, S. Brun, A. Rosenau, N. van der Mee-Marquet, R. Quentin, and L. Mereghetti. 2005. "Acquisition of insertion sequences and the GBSi1 Intron by *Streptococcus* agalactiae isolates correlates with the evolution of the species." *Journal of Bacteriology* 187 (17): 6248–52. https://doi.org/10.1128/JB.187.17.6248-6252.2005.

Hidalgo-Cantabrana, Claudio, Rosemary Sanozky-Dawes, and Rodolphe Barrangou. 2018. "Insights into the human virome using CRISPR spacers from microbiomes." *Viruses* 10 (9). https://doi.org/10.3390/v10090479.

Hoe, Nancy, Kazumitsu Nakashima, Diana Grigsby, Xi Pan, Shu Jun Dou, Steven Naidich, Marianne Garcia, Emily Kahn, David Bergmire-Sweat, and James M. Musser. 1999. "Rapid molecular genetic subtyping of serotype M1 Group A *Streptococcus* strains." *Emerging Infectious Diseases* 5 (2): 254.

Horvath, Philippe, and Rodolphe Barrangou. 2010. "CRISPR/Cas, the immune system of bacteria and archaea." *Science (New York, N.Y.)* 327 (5962): 167–70. https://doi.org/10.1126/science.1179555.

Horvath, Philippe, Dennis A. Romero, Anne-Claire Coûté-Monvoisin, Melissa Richards, Hélène Deveau, Sylvain Moineau, Patrick Boyaval, Christophe Fremaux, and Rodolphe Barrangou. 2008. "Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus.*" *Journal of Bacteriology* 190 (4): 1401–12. https://doi.org/10.1128/JB.01415-07.

Huber, C. A., F. McOdimba, V. Pflueger, C. A. Daubenberger, and G. Revathi. 2011. "Characterization of invasive and colonizing isolates of *Streptococcus agalactiae* in east african adults." *Journal of Clinical Microbiology* 49 (10): 3652–55. https://doi.org/10.1128/JCM.01288-11.

Huet, H., C. Martin, P. Geslin, F. Grimont, and R. Quentin. 1993. "Ribotyping of *Streptococcus agalactiae* strains isolated from vaginas of asymptomatic women."

Research in Microbiology 144 (6): 457–65. https://doi.org/10.1016/0923-2508(93)90053-5.

Imperi, Monica, Marco Pataracchia, Giovanna Alfarone, Lucilla Baldassarri, Graziella Orefici, and Roberta Creti. 2010. "A Multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of *Streptococcus agalactiae*." *Journal of Microbiological Methods* 80 (2): 212–14. https://doi.org/10.1016/j.mimet.2009.11.010.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakata. 1987. "Nucleotide sequence of the *Iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia Coli*, and identification of the gene product." *Journal of Bacteriology* 169 (12): 5429–33.

Jackson, Lisa A., Roberta Hilsdon, Monica M. Farley, Lee H. Harrison, Arthur L. Reingold, Brian D. Plikaytis, Jay D. Wenger, and Anne Schuchat. 1995. "Risk factors for group B streptococcal disease in adults." *Annals of Internal Medicine* 123 (6): 415–420.

Jackson, Paul J., Eliza A. Walthers, Abdullahi S. Kalif, Kara L. Richmond, Debra M. Adair, Karen K. Hill, Cheryl R. Kuske, et *al.* 1997. "Characterization of the variable-number tandem repeats in *VrrA* from different *bacillus anthracis* isolates." *Applied and Environmental Microbiology* 63 (4): 1400–1405.

Jansen, Rund, Jam D. A. van Embden, Wim Gaastra, and Leo M. Schouls. 2002a. "Identification of a novel family of sequence repeats among prokaryotes." *Omics: A Journal of Integrative Biology* 6 (1): 23–33. https://doi.org/10.1089/15362310252780816.

Jansen, Ruud, Jan D. A. van Embden, Wim Gaastra, and Leo M. Schouls. 2002b. "Identification of genes that are associated with DNA repeats in prokaryotes." *Molecular Microbiology* 43 (6): 1565–75.

Jauréguy, F., M. Carton, J. Teboul, M.-J. Butel, P. Panel, J.-C. Ghnassia, and F. Doucet-Populaire. 2003. "[Risk factors and screening strategy for group B streptococcal colonization in pregnant women: results of a prospective study]." *Journal De Gynécologie, Obstétrique Et Biologie De La Reproduction* 32 (2): 132–38.

Jelínková, JARMILA, and JITKA Motlová. 1985. "Worldwide distribution of two new serotypes of group B *Streptococci:* Type IV and provisional type V." *Journal of Clinical Microbiology* 21 (3): 361–362.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." *Science (New York, N.Y.)* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Joensen, K. G., F. Scheutz, O. Lund, H. Hasman, R. S. Kaas, E. M. Nielsen, and F. M. Aarestrup. 2014. "Real-Time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli.*" *Journal of Clinical Microbiology* 52 (5): 1501–10. https://doi.org/10.1128/JCM.03617-13.

Jones, Nicola, John F. Bohnsack, Shinji Takahashi, Karen A. Oliver, Man-Suen Chan, Frank Kunst, Philippe Glaser, et *al.* 2003. "Multilocus Sequence Typing System for Group B *Streptococcus.*" *Journal of Clinical Microbiology* 41 (6): 2530–36.

Jones, Nicola, Karen A. Oliver, Joanne Barry, Rosalind M. Harding, Naiel Bisharat, Brian G. Spratt, Tim Peto, Derrick W. Crook, and Oxford Group B. Streptococcus Consortium. 2006. "Enhanced invasiveness of bovine-derived neonatal sequence type 17 Group B *Streptococcus* is independent of capsular serotype." *Clinical Infectious Diseases* 42 (7): 915–924.

Jourdan-Da Silva, Nathalie J, Denise Antona, Caroline Six, scarlett Georges, Véronique Goulet, Philippe Judlin, and Daniel Lévy-Bruhl. 2008. "Infections néonatales à

Streptocoque B en france : Données d'incidence de 1997 à 2006 et Pr Atiques de Prévention En Maternité," 2008. http://opac.invs.sante.fr/doc_num.php?explnum_id=3286.

Kamerbeek, Judith, L. E. O. Schouls, Arend Kolk, Miranda Van Agterveld, Dick Van Soolingen, Sjoukje Kuijper, Annelies Bunschoten, et *al.* 1997. "Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology." *Journal of Clinical Microbiology* 35 (4): 907–914.

Keefe, Gregory P. 1997. "*Streptococcus agalactiae* Mastitis: A Review." *The Canadian Veterinary Journal* 38 (7): 429.

Kexel, G., and S. Schönbohm. 1965. "*Streptococcus agalactiae* als Erreger von Säuglingsmeningitiden." *DMW - Deutsche Medizinische Wochenschrift* 90 (06): 258–61. https://doi.org/10.1055/s-0028-1111328.

Kiely, R. A., L. Cotter, A. M. Mollaghan, B. Cryan, A. Coffey, and B. Lucey. 2011. "Emergence of group B *Streptococcus* serotype iv in women of child-bearing age in Ireland." *Epidemiology and Infection* 139 (02): 236–38. https://doi.org/10.1017/S0950268810001275.

Kieper, Sebastian N., Cristóbal Almendros, Juliane Behler, Rebecca E. McKenzie, Franklin L. Nobrega, Anna C. Haagsma, Jochem N.A. Vink, Wolfgang R. Hess, and Stan J.J. Brouns. 2018. "Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation." *Cell Reports* 22 (13): 3377–84. https://doi.org/10.1016/j.celrep.2018.02.103.

Ko, W. C., H. C. Lee, L. R. Wang, C. T. Lee, A. J. Liu, and J. J. Wu. 2001. "Serotyping and Antimicrobial susceptibility of Group B *Streptococcus* over an eight-year period in southern Taiwan." *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 20 (5): 334–39.

Koeck, Jean-Louis, Berthe-Marie Njanpop-Lafourcade, Sonia Cade, Emmanuelle Varon, Lassana Sangare, Samina Valjevac, Gilles Vergnaud, and Christine Pourcel. 2005. "Evaluation and selection of tandem repeat loci for *Streptococcus Pneumoniae* MLVA strain typing." *BMC Microbiology* 5 (November): 66. https://doi.org/10.1186/1471-2180-5-66.

Koonin, Eugene V, Kira S Makarova, and Feng Zhang. 2017. "Diversity, classification and evolution of CRISPR-Cas systems." *Current Opinion in Microbiology* 37 (June): 67–78. https://doi.org/10.1016/j.mib.2017.05.008.

Köser, Claudio U., Matthew TG Holden, Matthew J. Ellington, Edward JP Cartwright, Nicholas M. Brown, Amanda L. Ogilvy-Stuart, Li Yang Hsu, et *al.* 2012. "Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak." *New England Journal of Medicine* 366 (24): 2267–2275.

Krohn, Marijane A., Sharon L. Hillier, and Carol J. Baker. 1999. "Maternal peripartum complications associated with vaginal group B *Streptococci c*olonization." *Journal of Infectious Diseases* 179 (6): 1410–1415.

Kunin, Victor, Rotem Sorek, and Philip Hugenholtz. 2007. "Evolutionary conservation of sequence and secondary structures in CRISPR Repeats." *Genome Biol* 8 (4): R61.

Kwatra, Gaurav, Marianne C Cunnington, Elizabeth Merrall, Peter V Adrian, Margaret Ip, Keith P Klugman, Wing Hung Tam, and Shabir A Madhi. 2016. "Prevalence of maternal colonisation with group B *Streptococcus:* a systematic review and meta-analysis." *The Lancet Infectious Diseases* 16 (9): 1076–84. https://doi.org/10.1016/S1473-3099(16)30055-X.

Laboratoire de microbiologie du réseau EPIBAC, InVS. 2016. "Bulletin Du Réseau de Surveillance Des Infections Invasives Bactériennes - Reseau EPIBAC." 2016. http://invs.santepubliquefrance.fr/Dossiers-thematiques/Maladies-

infectieuses/Maladies-a-prevention-vaccinale/Infections-invasives-d-origine-bacterienne-Reseau-EPIBAC/Bulletin-du-reseau-de-surveillance-des-infections-invasives-bacteriennes.

Labrie, Simon J., Julie E. Samson, and Sylvain Moineau. 2010. "Bacteriophage resistance mechanisms." *Nature Reviews. Microbiology* 8 (5): 317–27. https://doi.org/10.1038/nrmicro2315.

Lachenauer, C. S., D. L. Kasper, J. Shimada, Y. Ichiman, H. Ohtsuka, M. Kaku, L. C. Paoletti, P. Ferrieri, and L. C. Madoff. 1999. "Serotypes VI and VIII predominate among group B *Streptococci* isolated from pregnant japanese women." *The Journal of Infectious Diseases* 179 (4): 1030–1033.

Lamagni, T. L., C. Keshishian, A. Efstratiou, R. Guy, K. L. Henderson, K. Broughton, and E. Sheridan. 2013. "Emerging trends in the epidemiology of invasive group b streptococcal disease in England and Wales, 1991-2010." *Clinical Infectious Diseases* 57 (5): 682–88. https://doi.org/10.1093/cid/cit337.

Lambertsen, L., K. Ekelund, I. C. Skovsted, A. Liboriussen, and H.-C. Slotved. 2010. "Characterisation of invasive group B *Streptococci* from adults in Denmark 1999 to 2004." *European Journal of Clinical Microbiology & Infectious Diseases* 29 (9): 1071–77. https://doi.org/10.1007/s10096-010-0941-z.

Lamy, Marie-Cécile, Shaynoor Dramsi, Annick Billoët, Hélène Réglier-Poupet, Asmaa Tazi, Josette Raymond, François Guérin, et *al.* 2006. "Rapid detection of the 'highly virulent' group B *Streptococcus* ST-17 clone." *Microbes and Infection* 8 (7): 1714–22. https://doi.org/10.1016/j.micinf.2006.02.008.

Lamy, Marie-Cécile, Mohammed Zouine, Juliette Fert, Massimo Vergassola, Elisabeth Couve, Elisabeth Pellegrini, Philippe Glaser, et *al.* 2004. "*CovS/CovR* of Group B *Streptococcus:* a two-component global regulatory system involved in virulence." *Molecular Microbiology* 54 (5): 1250–68. https://doi.org/10.1111/j.1365-2958.2004.04365.x.

Lancefield, Rebecca. 1934. "A Serological differentiation of specific types of bovine hemolytic *Streptococci* (GROUP B)." 1934. http://jem.rupress.org/content/jem/59/4/441.full.pdf.

Le Doare, Kirsty, and Paul T. Heath. 2013. "An overview of global gbs epidemiology." *Vaccine* 31 (August): D7–12. https://doi.org/10.1016/j.vaccine.2013.01.009.

Le Thomas-Bories, I., F. Fitoussi, P. Mariani-Kurkdjian, J. Raymond, N. Brahimi, P. Bidet, V. Lefranc, and E. Bingen. 2001. "Clonal Relationship between U.S. and French Serotype V Group B *Streptococcu*s isolates." *Journal of Clinical Microbiology* 39 (12): 4526–28. https://doi.org/10.1128/JCM.39.12.4526-4528.2001.

Leal, Carlos A. G., Guilherme A. Queiroz, Felipe L. Pereira, Guilherme C. Tavares, and Henrique C. P. Figueiredo. 2019. "*Streptococcus agalactiae* sequence type 283 in farmed fish, Brazil." *Emerging Infectious Diseases* 25 (4): 776–79. https://doi.org/10.3201/eid2504.180543.

Lee, B. K., Y. R. Song, M. Y. Kim, J. H. Yang, J. H. Shin, Y. S. Seo, K. Y. Oh, et *al.* 2010. "Epidemiology of Group B *Streptococcus* in Korean pregnant women." *Epidemiology and Infection* 138 (02): 292. https://doi.org/10.1017/S0950268809990859.

Leenay, Ryan T., and Chase L. Beisel. 2017. "Deciphering, communicating, and engineering the CRISPR PAM." *Journal of Molecular Biology* 429 (2): 177–91. https://doi.org/10.1016/j.jmb.2016.11.024.

Lewin, E. B., and M. S. Amstey. 1981. "Natural history of Group B *Streptococcus* colonization and its therapy during pregnancy." *American Journal of Obstetrics and Gynecology* 139 (5): 512–15.

Li, Shunming, Jingya Huang, Zhiyao Chen, Dan Guo, Zhenjiang Yao, and Xiaohua Ye. 2017. "Antibiotic prevention for maternal group B streptococcal colonization on neonatal GBS-related adverse outcomes: A meta-analysis." *Frontiers in Microbiology* 8 (March). https://doi.org/10.3389/fmicb.2017.00374.

Lier, Clément, Elodie Baticle, Philippe Horvath, Eve Haguenoer, Anne-Sophie Valentin, Philippe Glaser, Laurent Mereghetti, and Philippe Lanotte. 2015. "Analysis of the Type II-A CRISPR-Cas System of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages." *Frontiers in Genetics* 6. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466440/.

Lillestøl, Reidun K., Peter Redder, Roger A. Garrett, and Kim Brügger. 2006. "A Putative viral defence mechanism in archaeal cells." *Archaea (Vancouver, B.C.)* 2 (1): 59–72.

Lin, F.-Y. C., A. Whiting, E. Adderson, S. Takahashi, D. M. Dunn, R. Weiss, P. H. Azimi, et *al.* 2006. "Phylogenetic lineages of invasive and colonizing strains of serotype III Group B *Streptococci* from neonates: a multicenter prospective study." *Journal of Clinical Microbiology* 44 (4): 1257–61. https://doi.org/10.1128/JCM.44.4.1257-1261.2006.

Lindstedt, Bjørn-Arne. 2005. "Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria." *ELECTROPHORESIS* 26 (13): 2567–82. https://doi.org/10.1002/elps.200500096.

Liu, F., R. Barrangou, P. Gerner-Smidt, E. M. Ribot, S. J. Knabel, and E. G. Dudley. 2011. "Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* Subsp. Enterica." *Applied and Environmental Microbiology* 77 (6): 1946–56. https://doi.org/10.1128/AEM.02625-10.

Liu, F., S. Kariyawasam, B. M. Jayarao, R. Barrangou, P. Gerner-Smidt, E. M. Ribot, S. J. Knabel, and E. G. Dudley. 2011. "Subtyping *Salmonella enterica* Serovar Enteritidis isolates from different sources by using sequence typing based on virulence genes and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)." *Applied and Environmental Microbiology* 77 (13): 4520–26. https://doi.org/10.1128/AEM.00468-11.

Liu, G, J Zhu, K Chen, T Gao, H Yao, Y Liu, W Zhang, and C Lu. 2016. "Development of *Streptococcus Agalactiae* vaccines for Tilapia." *Diseases of Aquatic Organisms* 122 (2): 163–70. https://doi.org/10.3354/dao03084.

Lopez-Sanchez, Maria-José, Elisabeth Sauvage, Violette Da Cunha, Dominique Clermont, Elisoa Ratsima Hariniaina, Bruno Gonzalez-Zorn, Claire Poyart, Isabelle Rosinski-Chupin, and Philippe Glaser. 2012. "The Highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome." *Molecular Microbiology* 85 (6): 1057–71. https://doi.org/10.1111/j.1365-2958.2012.08172.x.

Luan, S.-L., M. Granlund, M. Sellin, T. Lagergard, B. G. Spratt, and M. Norgren. 2005. "Multilocus sequence typing of swedish invasive Group B *Streptococcus* Isolates indicates a neonatally associated genetic lineage and capsule switching." *Journal of Clinical Microbiology* 43 (8): 3727–33. https://doi.org/10.1128/JCM.43.8.3727-3733.2005.

Lunel, Frans Verduyn, Lidia Licciardello, Stefania Stefani, Henri A. Verbrugh, Willem J. G. Melchers, Jacques F. G. M. Meis, Stewart Scherer, and Alex van Belkum. 1998. "Lack of consistent short sequence repeat polymorphisms in genetically homologous colonizing and invasive *candida Albicans* Strains." *Journal of Bacteriology* 180 (15): 3771–78.

Maiden, Martin CJ, Jane A. Bygraves, Edward Feil, Giovanna Morelli, Joanne E. Russell, Rachel Urwin, Qing Zhang, et *al.* 1998. "Multilocus Sequence Typing: a portable

approach to the identification of clones within populations of pathogenic microorganisms." *Proceedings of the National Academy of Sciences* 95 (6): 3140–3145.

Makarova, Kira S, Nick V Grishin, Svetlana A Shabalina, Yuri I Wolf, and Eugene V Koonin. 2006. "A Putative RNA-Interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action." *Biology Direct* 1 (March): 7. https://doi.org/10.1186/1745-6150-1-7.

Makarova, Kira S., Daniel H. Haft, Rodolphe Barrangou, Stan J. J. Brouns, Emmanuelle Charpentier, Philippe Horvath, Sylvain Moineau, et *al.* 2011. "Evolution and classification of the CRISPR–Cas Systems." *Nature Reviews Microbiology* 9 (6): 467–77. https://doi.org/10.1038/nrmicro2577.

Makarova, Kira S., Yuri I. Wolf, Omer S. Alkhnbashi, Fabrizio Costa, Shiraz A. Shah, Sita J. Saunders, Rodolphe Barrangou, et *al.* 2015. "An updated evolutionary classification of CRISPR-Cas systems." *Nature Reviews Microbiology* 13 (11): 722–36. https://doi.org/10.1038/nrmicro3569.

Manning, S. D., A. C. Springman, E. Lehotzky, M. A. Lewis, T. S. Whittam, and H. D. Davies. 2009. "Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada." *Journal of Clinical Microbiology* 47 (4): 1143–48. https://doi.org/10.1128/JCM.01424-08.

Manning, Shannon D., Maggi A. Lewis, A. Cody Springman, Erica Lehotzky, Thomas S. Whittam, and H. Dele Davies. 2008. "Genotypic diversity and serotype distribution of group B *Streptococcus* isolated from women before and after delivery." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 46 (12): 1829–37. https://doi.org/10.1086/588296.

Martins, E. R., J. Melo-Cristino, M. Ramirez, and the Portuguese Group for the Study of streptococcal Infections. 2012. "Dominance of serotype Ia among Group B *Streptococci* causing invasive infections in nonpregnant adults in Portugal." *Journal of Clinical Microbiology* 50 (4): 1219–27. https://doi.org/10.1128/JCM.05488-11.

Martins, E. R., M. A. Pessanha, M. Ramirez, J. Melo-Cristino, and and the Portuguese Group for the Study of streptococcal Infections. 2007. "Analysis of Group B streptococcal isolates from infants and pregnant women in portugal revealing two lineages with enhanced invasiveness." *Journal of Clinical Microbiology* 45 (10): 3224–29. https://doi.org/10.1128/JCM.01182-07.

Matsubara, Kousaku, Kiyoshi Hoshina, and Yoko Suzuki. 2013. "Early-Onset and Late-Onset Group B streptococcal disease in japan: a nationwide surveillance study, 2004–2010." *International Journal of Infectious Diseases* 17 (6): e379–84. https://doi.org/10.1016/j.ijid.2012.11.027.

Meehan, Mary, Robert Cunney, and Mary Cafferkey. 2014. "Molecular Epidemiology of Group B *Streptococci* in ireland reveals a diverse population with evidence of capsular switching." *European Journal of Clinical Microbiology & Infectious Diseases* 33 (7): 1155–62. https://doi.org/10.1007/s10096-014-2055-5.

Mee-Marquet, N. van der, A.-S. Domelier, L. Mereghetti, P. Lanotte, A. Rosenau, W. van Leeuwen, and R. Quentin. 2006. "Prophagic DNA fragments in *Streptococcus Agalactiae* strains and association with neonatal meningitis." *Journal of Clinical Microbiology* 44 (3): 1049–58. https://doi.org/10.1128/JCM.44.3.1049-1058.2006.

Mee-Marquet, N. van der, L. Fourny, L. Arnault, A.-S. Domelier, M. Salloum, M.-F. Lartigue, and R. Quentin. 2008. "Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital

body sites." *Journal of Clinical Microbiology* 46 (9): 2906–11. https://doi.org/10.1128/JCM.00421-08.

Mee-Marquet, N. van der, C. Jouannet, A.-S. Domelier, L. Arnault, M.-F. Lartigue, and R. Quentin. 2009. "Genetic diversity of *Streptococcus agalactiae* strains and density of vaginal carriage." *Journal of Medical Microbiology* 58 (2): 169–73. https://doi.org/10.1099/jmm.0.005827-0.

Mee-Marquet, Nathalie van der, Anne-Sophie Domelier, Laurent Mereghetti, Philippe Lanotte, Agnès Rosenau, Willem van Leeuwen, and Roland Quentin. 2006. "Prophagic DNA fragments in *Streptococcus agalactiae* strains and association with neonatal meningitis." *Journal of Clinical Microbiology* 44 (3): 1049–58. https://doi.org/10.1128/JCM.44.3.1049-1058.2006.

Mojica, F. J., C. Díez-Villaseñor, E. Soria, and G. Juez. 2000. "Biological Significance of a Family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria." *Molecular Microbiology* 36 (1): 244–46.

Mojica, F. J. M., C. Diez-Villasenor, J. Garcia-Martinez, and C. Almendros. 2009. "Short motif sequences determine the targets of the prokaryotic crispr defence system." *Microbiology* 155 (3): 733–40. https://doi.org/10.1099/mic.0.023960-0.

Mojica, Francisco J. M., César Díez-Villaseñor, Jesús García-Martínez, and Elena Soria. 2005. "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements." *Journal of Molecular Evolution* 60 (2): 174–82. https://doi.org/10.1007/s00239-004-0046-3.

Mokrousov, I., A. Vyazovaya, V. Kolodkina, E. Limeschenko, L. Titov, and O. Narvskaya. 2009. "Novel macroarray-based method of *Corynebacterium diphtheriae* genotyping: evaluation in a field study in Belarus." *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 28 (6): 701–3. https://doi.org/10.1007/s10096-008-0674-4.

Mokrousov, Igor, Elena Limeschenko, Anna Vyazovaya, and Olga Narvskaya. 2007. "*Corynebacterium Diphtheriae* spoligotyping based on combined use of two CRISPR Loci." *Biotechnology Journal* 2 (7): 901–6. https://doi.org/10.1002/biot.200700035.

Moore, Matthew R., Stephanie J. Schrag, and Anne Schuchat. 2003. "Effects of intrapartum antimicrobial prophylaxis for prevention of group-B-streptococcal disease on the incidence and ecology of early-onset neonatal sepsis." *The Lancet Infectious Diseases* 3 (4): 201–213.

Morozumi, Miyuki, Takeaki Wajima, Misako Takata, Satoshi Iwata, and Kimiko Ubukata. 2016. "Molecular characteristics of Group B *Streptococci* isolated from adults with invasive infections in Japan." Edited by B. A. Forbes. *Journal of Clinical Microbiology* 54 (11): 2695–2700. https://doi.org/10.1128/JCM.01183-16.

Moyo, Sylvester R., Johan A. Maeland, and Kåre Bergh. 2002. "Typing of human isolates of *Streptococcus agalactiae* (Group B *Streptococcus*, GBS) strains from Zimbabwe." *Journal of Medical Microbiology* 51 (7): 595–662. https://doi.org/10.1099/0022-1317-51-7-595.

Musser, James M., Stephen J. Mattingly, Roland Quentin, Alain Goudeau, and Robert K. Selander. 1989. "Identification of a high-virulence clone of type III *Streptococcus agalactiae* (Group B *Streptococcus*) causing invasive neonatal disease." *Proceedings of the National Academy of Sciences* 86 (12): 4731–4735.

Nagano, N., Y. Nagano, M. Toyama, K. Kimura, T. Tamura, K. Shibayama, and Y. Arakawa. 2012. "Nosocomial spread of multidrug-resistant group B *Streptococci* with reduced penicillin susceptibility belonging to clonal complex 1." *Journal of Antimicrobial Chemotherapy* 67 (4): 849–56. https://doi.org/10.1093/jac/dkr546.

Nagano, YUKIKO, N. Nagano, S. Takahashi, K. Murono, K. Fujita, F. Taguchi, and Y. Okuwaki. 1991. "Restriction endonuclease digest patterns of chromosomal DNA from Group B β-Haemolytic *Streptococci." Journal of Medical Microbiology* 35 (5): 297–303.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, and E. Kumlin. 1987. "Variable Number of Tandem Repeat (VNTR) markers for human gene mapping." *Science (New York, N.Y.)* 235 (4796): 1616–22.

Nitschke, H., P. Slickers, E. Muller, R. Ehricht, and S. Monecke. 2014. "DNA microarray-based typing of *Streptococcus agalactiae* isolates." *Journal of Clinical Microbiology* 52 (11): 3933–43. https://doi.org/10.1128/JCM.02411-14.

Noya, F. J., M. A. Rench, T. G. Metzger, G. Colman, J. Naidoo, and C. J. Baker. 1987. "Unusual occurrence of an epidemic of type ib/c group b streptococcal sepsis in a neonatal intensive care unit." *The Journal of Infectious Diseases* 155 (6): 1135–44.

Oost, John van der, Matthijs M. Jore, Edze R. Westra, Magnus Lundgren, and Stan J.J. Brouns. 2009. "CRISPR-Based Adaptive and heritable immunity in Prokaryotes." *Trends in Biochemical Sciences* 34 (8): 401–7. https://doi.org/10.1016/j.tibs.2009.05.002.

Oost, John van der, Edze R. Westra, Ryan N. Jackson, and Blake Wiedenheft. 2014. "Unravelling the structural and mechanistic basis of CRISPR–Cas systems." *Nature Reviews Microbiology* 12 (7): 479–92. https://doi.org/10.1038/nrmicro3279.

Patras, Kathryn A., and Kelly S. Doran. 2016. "A murine model of Group B *Streptococcus* vaginal colonization." *Journal of Visualized Experiments*, no. 117 (November). https://doi.org/10.3791/54708.

Persson, E., S. Berg, B. Trollfors, P. Larsson, E. Ek, E. Backhaus, B.E.B. Claesson, et *al.* 2004. "Serotypes and clinical manifestations of invasive group b streptococcal infections in western Sweden 1998-2001." *Clinical Microbiology and Infection* 10 (9): 791–96. https://doi.org/10.1111/j.1469-0691.2004.00931.x.

Pettengill, James B., Ruth E. Timme, Rodolphe Barrangou, Magaly Toro, Marc W. Allard, Errol Strain, Steven M. Musser, and Eric W. Brown. 2014. "The evolutionary history and diagnostic utility of the CRISPR-Cas System within *Salmonella enterica* Ssp. Enterica." *PeerJ* 2 (April): e340. https://doi.org/10.7717/peerj.340.

Phares, Christina R., Ruth Lynfield, Monica M. Farley, Janet Mohle-Boetani, Lee H. Harrison, Susan Petit, Allen S. Craig, et *al.* 2008. "Epidemiology of Invasive Group B streptococcal Disease in the United States, 1999-2005." *JAMA* 299 (17): 2056–65. https://doi.org/10.1001/jama.299.17.2056.

Pourcel, C., K. Hormigos, L. Onteniente, O. Sakwinska, R. H. Deurenberg, and G. Vergnaud. 2009. "Improved multiple-locus variable-number tandem-repeat assay for *Staphylococcus aureus* genotyping, providing a highly informative technique together with strong phylogenetic value." *Journal of Clinical Microbiology* 47 (10): 3121–28. https://doi.org/10.1128/JCM.00267-09.

Pourcel, C., G. Salvignol, and G. Vergnaud. 2005. "CRISPR elements in *Yersinia Pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies." *Microbiology (Reading, England)* 151 (Pt 3): 653–63. https://doi.org/10.1099/mic.0.27437-0.

Pourcel, C, Paolo Visca, Baharak Afshar, Silvia D'Arezzo, Gilles Vergnaud, and Norman K. Fry. 2007. "Identification of Variable-Number Tandem-Repeat (VNTR) sequences in *Legionella Pneumophila* and development of an optimized multiple-locus VNTR analysis typing scheme." *Journal of Clinical Microbiology* 45 (4): 1190–99. https://doi.org/10.1128/JCM.02078-06.

Poyart, C, Hélène Réglier-Poupet, Asmaa Tazi, Annick Billoët, Nicolas Dmytruk, Philippe Bidet, Edouard Bingen, Josette Raymond, and Patrick Trieu-Cuot. 2008. "Invasive Group B streptococcal infections in infants, France." *Emerging Infectious Diseases* 14 (10): 1647–49. https://doi.org/10.3201/eid1410.080185.

Poyart, C., A. Tazi, H. Reglier-Poupet, A. Billoet, N. Tavares, J. Raymond, and P. Trieu-Cuot. 2007. "Multiplex PCR assay for rapid and accurate capsular typing of group B *Streptococci*." *Journal of Clinical Microbiology* 45 (6): 1985–88. https://doi.org/10.1128/JCM.00159-07.

Pride, D. T., C. L. Sun, J. Salzman, N. Rao, P. Loomer, G. C. Armitage, J. F. Banfield, and D. A. Relman. 2011. "Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time." *Genome Research* 21 (1): 126–36. https://doi.org/10.1101/gr.111732.110.

Quentin, Roland, Helene Huet, Fu-Sheng Wang, Pierre Geslin, Alain Goudeau, and Robert K. Selander. 1995. "Characterization of *Streptococcus agalactiae* strains by multilocus enzyme genotype and serotype: identification of multiple virulent clone families that cause invasive neonatal disease." *Journal of Clinical Microbiology* 33 (10): 2576–2581.

Radtke, A., B.-A. Lindstedt, J. E. Afset, and K. Bergh. 2010. "Rapid Multiple-Locus Variant-Repeat Assay (MLVA) for genotyping of *Streptococcus agalactiae*." *Journal of Clinical Microbiology* 48 (7): 2502–8. https://doi.org/10.1128/JCM.00234-10.

Regan, J. A., M. A. Klebanoff, and R. P. Nugent. 1991. "The Epidemiology of group B streptococcal colonization in pregnancy. Vaginal infections and prematurity study group." *Obstetrics and Gynecology* 77 (4): 604–10.

Richards, Vincent P., Ping Lang, Paulina D. Pavinski Bitar, Tristan Lefébure, Ynte H. Schukken, Ruth N. Zadoks, and Michael J. Stanhope. 2011. "Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*." *Infection, Genetics and Evolution* 11 (6): 1263–75. https://doi.org/10.1016/j.meegid.2011.04.019.

Richter, Corinna, James T. Chang, and Peter C. Fineran. 2012. "Function and regulation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) / CRISPR Associated (Cas) systems." *Viruses* 4 (10): 2291–2311. https://doi.org/10.3390/v4102291.

Rolland, Karine, Corinne Marois, Veronique Siquier, Blandine Cattier, and Roland Quentin. 1999. "Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *HylB* Gene analysis." *Journal of Clinical Microbiology* 37 (6): 1892–1898.

Rolland, Karine, Laurent Mereghetti, Stephane Watt, Sonia Chatellier, and Roland Quentin. 2002. "TRNA gene clusters at the 3′ end of RRNA operons are specific to virulent subgroups of *Streptococcus agalactiae* strains, as demonstrated by molecular differential analysis at the population level." *Microbiology* 148 (5): 1493–1499.

Rosen, Geoffrey H, Tara M Randis, Purnahamsi V Desai, Katherine J Sapra, Bing Ma, Pawel Gajer, Michael S Humphrys, Jacques Ravel, Shari E Gelber, and Adam J Ratner. 2017. "Group B *Streptococcus* and the vaginal microbiota." *The Journal of Infectious Diseases* 216 (6): 744–51. https://doi.org/10.1093/infdis/jix395.

Rosenau, A., K. Martins, S. Amor, F. Gannier, P. Lanotte, N. van der Mee-Marquet, L. Mereghetti, and R. Quentin. 2007. "Evaluation of the ability of *Streptococcus agalactiae* strains isolated from genital and neonatal specimens to bind to human fibrinogen and correlation with characteristics of the *FbsA* and *FbsB* Genes." *Infection and Immunity* 75 (3): 1310–17. https://doi.org/10.1128/IAI.00996-06.

Salloum, Mazen, Nathalie van der Mee-Marquet, Anne-Sophie Valentin-Domelier, and Roland Quentin. 2011. "Diversity of prophage DNA regions of *Streptococcus agalactiae* clonal lineages from adults and neonates with invasive infectious disease." *PloS One* 6 (5): e20256. https://doi.org/10.1371/journal.pone.0020256.

Schrag, Stephanie, Rachel Gorwitz, Kristi Fultz-Butts, and Anne Schuchat. 2002. "Prevention of perinatal Group B streptococcal disease. Revised guidelines from CDC." *MMWR. Recommendations and Reports: Morbidity and Mortality Weekly Report. Recommendations and Reports / Centers for Disease Control* 51 (RR-11): 1–22.

Schuchat, A. 1995. "Group B streptococcal disease in newborns: A global perspective on prevention." *Biomedicine & Pharmacotherapy = Biomédecine & Pharmacothérapie* 49 (1): 19–25. https://doi.org/10.1016/0753-3322(96)82573-X.

Schuchat, A., K. Deaver-Robinson, B. D. Plikaytis, K. M. Zangwill, J. Mohle-Boetani, and J. D. Wenger. 1994. "Multistate case-control study of maternal risk factors for neonatal group B streptococcal disease. The active surveillance study group." *The Pediatric Infectious Disease Journal* 13 (7): 623–29.

Schuchat, Anne. 1998. "Epidemiology of Group B streptococcal disease in the United States: shifting paradigms." *Clinical Microbiology Reviews* 11 (3): 497–513.

Schurch, A. C., K. Kremer, O. Daviena, A. Kiers, M. J. Boeree, R. J. Siezen, and D. van Soolingen. 2010. "High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster." *Journal of Clinical Microbiology* 48 (9): 3403–6. https://doi.org/10.1128/JCM.00370-10.

Schwartz, B., A. Schuchat, M. J. Oxtoby, S. L. Cochi, A. Hightower, and C. V. Broome. 1991. "Invasive Group B streptococcal disease in adults. A population-based study in metropolitan Atlanta." *JAMA* 266 (8): 1112–14.

Schwartz, D. C., and C. R. Cantor. 1984. "Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis." *Cell* 37 (1): 67–75.

Selander, Robert K., Dominique A. Caugant, Howerd Ochman, James M. Musser, Marion N. Gilmour, and Thomas S. Whittam. 1986. "Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics." *Applied and Environmental Microbiology* 51 (5): 873.

Shah, Shiraz A., Susanne Erdmann, Francisco J.M. Mojica, and Roger A. Garrett. 2013. "Protospacer recognition motifs: mixed identities and functional diversity." *RNA Biology* 10 (5): 891–99. https://doi.org/10.4161/rna.23764.

Shariat, N., and E. G. Dudley. 2014. "CRISPRs: molecular signatures used for pathogen subtyping." *Applied and Environmental Microbiology* 80 (2): 430–39. https://doi.org/10.1128/AEM.02790-13.

Shariat, N., M. K. Kirchner, C. H. Sandt, E. Trees, R. Barrangou, and E. G. Dudley. 2013. "Subtyping of *Salmonella enterica* Serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results." *Journal of Clinical Microbiology* 51 (7): 2328–36. https://doi.org/10.1128/JCM.00608-13.

Shariat, Nikki, Michael J. DiMarzio, Shuang Yin, Lisa Dettinger, Carol H. Sandt, James R. Lute, Rodolphe Barrangou, and Edward G. Dudley. 2013. "The Combination of CRISPR-MVLST and PFGE provides increased discriminatory power for differentiating human clinical isolates of *Salmonella enterica* Subsp. Enterica Serovar Enteritidis." *Food Microbiology* 34 (1): 164–73. https://doi.org/10.1016/j.fm.2012.11.012.

Shariat, Nikki, Carol H. Sandt, Michael J. DiMarzio, Rodolphe Barrangou, and Edward G. Dudley. 2013. "CRISPR-MVLST subtyping of *Salmonella enterica* subsp. Enterica

Serovars Typhimurium and Heidelberg and application in identifying outbreak isolates." *BMC Microbiology* 13 (1): 254.

Shmakov, Sergey, Aaron Smargon, David Scott, David Cox, Neena Pyzocha, Winston Yan, Omar O. Abudayyeh, et *al.* 2017. "Diversity and evolution of class 2 CRISPR–Cas systems." *Nature Reviews Microbiology* 15 (3): 169–82. https://doi.org/10.1038/nrmicro.2016.184.

Siegel, J. D., G. H. McCracken, N. Threlkeld, B. Milvenan, and C. R. Rosenfeld. 1980. "Single-dose penicillin prophylaxis against neonatal group B streptococcal infections. A Controlled Trial in 18,738 Newborn Infants." *The New England Journal of Medicine* 303 (14): 769–75. https://doi.org/10.1056/NEJM198010023031401.

Six, Anne, Caroline Joubrel, Asmaa Tazi, and Claire Poyart. 2014. "Infections materno-fœtales à *Streptococcus agalactiae*." *La Presse Médicale* 43 (6): 706–14. https://doi.org/10.1016/j.lpm.2014.04.008.

Skoff, Tami H., Monica M. Farley, Susan Petit, Allen S. Craig, William Schaffner, Ken Gershman, Lee H. Harrison, et *al.* 2009. "Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990–2007." *Clinical Infectious Diseases* 49 (1): 85–92. https://doi.org/10.1086/599369.

Slotved, H.-C., F. Kong, L. Lambertsen, S. Sauer, and G. L. Gilbert. 2007. "Serotype IX, a proposed new *Streptococcus agalactiae* serotype." *Journal of Clinical Microbiology* 45 (9): 2929–36. https://doi.org/10.1128/JCM.00117-07.

Société Française de Néonatologie, and Société Française de Pédiatrie. 2017. "Recommandations - Version Sept 2017. Recommandations de Bonnes Pratiques. Prise En Charge Du Nouveau - Né à Risque d'infection Néonatale Bactérienne Précoce (≥34 SA)." September 2017. http://www.societe-francaise-neonatalogie.fr/wp-content/uploads/2017/02/LABEL-HAS-RECOMMANDATIONS-INBP.09.2017.pdf.

Sørensen, Uffe B. Skov, Knud Poulsen, Claudia Ghezzo, Immaculada Margarit, and Mogens Kilian. 2010. "Emergence and global dissemination of host-specific *Streptococcus agalactiae* clones." *MBio* 1 (3). https://doi.org/10.1128/mBio.00178-10.

Springman, A. C., D. W. Lacher, G. Wu, N. Milton, T. S. Whittam, H. D. Davies, and S. D. Manning. 2009. "Selection, recombination, and virulence gene diversity among group B streptococcal genotypes." *Journal of Bacteriology* 191 (17): 5419–27. https://doi.org/10.1128/JB.00369-09.

Stringer, Jacqueline, and W.R Maxted. 1979. "Phage typing of group B *Streptococci." The Lancet* 313 (8111): 327–328.

Suara, RO. 1994. "Carriage of Group B *Streptococci* in pregnant gambian mothers and their infants. - PubMed - NCBI." 1994. https://www.ncbi.nlm.nih.gov/pubmed/7963736.

Sun, Christine L., Brian C. Thomas, Rodolphe Barrangou, and Jillian F. Banfield. 2016. "Metagenomic Reconstructions of bacterial CRISPR loci constrain population histories." *The ISME Journal* 10 (4): 858–70. https://doi.org/10.1038/ismej.2015.162.

Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht. 2001. "Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units." *Journal of Clinical Microbiology* 39 (10): 3563–71. https://doi.org/10.1128/JCM.39.10.3563-3571.2001.

Takahashi, S., E. E. Adderson, Y. Nagano, N. Nagano, M. R. Briesacher, and J. F. Bohnsack. 1998. "Identification of a highly encapsulated, genetically related group of invasive type III group B *Streptococci*." *The Journal of Infectious Diseases* 177 (4): 1116–19.

Takahashi, Shinji, Shauna Detrick, April A. Whiting, Anne J. Blaschke-Bonkowksy, Youko Aoyagi, Elisabeth E. Adderson, and John F. Bohnsack. 2002. "Correlation of

phylogenetic lineages of group B *Streptococci*, identified by analysis of restriction-digestion patterns of genomic DNA, with *InfB* alleles and mobile genetic elements." *The Journal of Infectious Diseases* 186 (7): 1034–1038.

Tazi, A, Olivier Disson, Samuel Bellais, Abdelouhab Bouaboud, Isabelle Tardieux, Patrick Trieu-Cuot, Marc Lecuit, and Claire Poyart. 2011a. "[A breakthrough in the understanding of neonatal group B *Streptococcus* meningitis]." *Médecine Sciences: M/S* 27 (4): 362–64. https://doi.org/10.1051/medsci/2011274010.

Tazi, A., P.C. Morand, H. Réglier-Poupet, N. Dmytruk, A. Billoët, D. Antona, P. Trieu-Cuot, and C. Poyart. 2011b. "Invasive group B streptococcal infections in adults, France (2007–2010)." *Clinical Microbiology and Infection* 17 (10): 1587–89. https://doi.org/10.1111/j.1469-0691.2011.03628.x.

Tazi, Asmaa, Olivier Disson, Samuel Bellais, Abdelouhab Bouaboud, Nicolas Dmytruk, Shaynoor Dramsi, Michel-Yves Mistou, et *al.* 2010. "The surface protein HvgA mediates group B *Streptococcus* hypervirulence and meningeal tropism in neonates." *The Journal of Experimental Medicine* 207 (11): 2313–22. https://doi.org/10.1084/jem.20092594.

Teatero, S., A. McGeer, D. E. Low, A. Li, W. Demczuk, I. Martin, and N. Fittipaldi. 2014. "Characterization of invasive group B *Streptococcus* strains from the greater Toronto Area, Canada." *Journal of Clinical Microbiology* 52 (5): 1441–47. https://doi.org/10.1128/JCM.03554-13.

Teatero, Sarah, Taryn B. T. Athey, Paul Van Caeseele, Greg Horsman, David C. Alexander, Roberto G. Melano, Aimin Li, et *al.* 2015. "Emergence of serotype IV group B *Streptococcus* adult invasive disease in Manitoba and Saskatchewan, Canada, is driven by clonal sequence type 459 strains." Edited by N. A. Ledeboer. *Journal of Clinical Microbiology* 53 (9): 2919–26. https://doi.org/10.1128/JCM.01128-15.

Tettelin, Hervé, Vega Masignani, Michael J. Cieslewicz, Claudio Donati, Duccio Medini, Naomi L. Ward, Samuel V. Angiuoli, et *al.* 2005. "Genome analysis of multiple pathogenic isolates of *Streptococcus agalactia*e: implications for the microbial 'pan-genome.'" *Proceedings of the National Academy of Sciences of the United States of America* 102 (39): 13950–13955.

Tettelin, Hervé, Vega Masignani, Michael J. Cieslewicz, Jonathan A. Eisen, Scott Peterson, Michael R. Wessels, Ian T. Paulsen, et *al.* 2002. "Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*." *Proceedings of the National Academy of Sciences* 99 (19): 12391–12396.

Timme, Ruth E., James B. Pettengill, Marc W. Allard, Errol Strain, Rodolphe Barrangou, Chris Wehnes, Joann S. Van Kessel, Jeffrey S. Karns, Steven M. Musser, and Eric W. Brown. 2013. "Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. Enterica inferred from genome-wide reference-free SNP characters." *Genome Biology and Evolution* 5 (11): 2109–23. https://doi.org/10.1093/gbe/evt159.

Tompkins, Lucy. 1992. "The use of molecular methods in infectious diseases." *The New England Journal of Medicine* 127 (3): 339–347.

Trivalle, C., E. Martin, P. Martel, B. Jacque, J. F. Menard, and J. F. Lemeland. 1998. "Group B streptococcal bacteraemia in the elderly." *Journal of Medical Microbiology* 47 (7): 649–52.

Tsolia, M., M. Psoma, S. Gavrili, V. Petrochilou, S. Michalas, N. Legakis, and Th Karpathios. 2003. "Group B *Streptococcus* colonization of greek pregnant women and neonates: prevalence, risk factors and serotypes." *Clinical Microbiology and Infection* 9 (8): 832–838.

Tyler, K. D., G. Wang, S. D. Tyler, and W. M. Johnson. 1997. “Factors affecting reliability and reproducibility of amplification-based DNA fingerprinting of representative bacterial pathogens.” *Journal of Clinical Microbiology* 35 (2): 339.

Tyrrell, Gregory J., Linda D. Senzilet, John S. Spika, Daniel A. Kertesz, Merrilyn Alagaratnam, Marguerite Lovgren, and James A. Talbot. 2000. “Invasive disease due to Group B streptococcal infection in adults: results from a canadian, population-based, active laboratory surveillance study—1996.” *The Journal of Infectious Diseases* 182 (1): 168–173.

Tyson, Gene W., and Jillian F. Banfield. 2007. “Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses.” *Environmental Microbiology* 10 (1): 200–207. https://doi.org/10.1111/j.1462-2920.2007.01444.x.

Ueno, Hiroyuki, Yoshihiko Yamamoto, Akiko Yamamichi, Koji Kikuchi, Sumie Kobori, and Motonobu Miyazaki. 2012. “Characterization of Group B *Streptococcus* isolated from women in Saitama city, Japan.” *Japanese Journal of Infectious Diseases* 65 (6): 516–521.

Valkenburg-van den Berg, Arijaan W., Arwen J. Sprij, Paul M. Oostvogel, Johan A.E.M. Mutsaers, Wouter B. Renes, Frits R. Rosendaal, and P. Joep Dörr. 2006. “Prevalence of colonisation with Group B *Streptococci* in pregnant women of a multi-ethnic population in the Netherlands.” *European Journal of Obstetrics & Gynecology and Reproductive Biology* 124 (2): 178–83. https://doi.org/10.1016/j.ejogrb.2005.06.007.

Van Belkum, Alex, Stewart Scherer, Willem Van Leeuwen, Diana Willemse, Loek Van Alphen, and Henri Verbrugh. 1997. “Variable number of tandem repeats in clinical strains of *Haemophilus Influenzae*.” *Infection and Immunity* 65 (12): 5017–5027.

Van Orden, Mason J., Peter Klein, Kesavan Babu, Fares Z. Najar, and Rakhi Rajan. 2017. “Conserved DNA motifs in the type II-A CRISPR leader region.” *PeerJ* 5 (April): e3161. https://doi.org/10.7717/peerj.3161.

Verani, Jennifer R., Lesley McGee, Stephanie J. Schrag, and Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention (CDC). 2010. “Prevention of perinatal group B streptococcal disease--revised guidelines from CDC, 2010.” *MMWR. Recommendations and Reports: Morbidity and Mortality Weekly Report. Recommendations and Reports / Centers for Disease Control* 59 (RR-10): 1–36.

Vu-Thien, Hoang, Gaëlle Corbineau, Katia Hormigos, Brigitte Fauroux, Harriet Corvol, Annick Clément, Gilles Vergnaud, and Christine Pourcel. 2007. “Multiple-locus variable-number tandem-repeat analysis for longitudinal survey of sources of *Pseudomonas Aeruginosa* infection in cystic fibrosis patients.” *Journal of Clinical Microbiology* 45 (10): 3175–83. https://doi.org/10.1128/JCM.00702-07.

Wang, Rui, Ming Li, Luyao Gong, Songnian Hu, and Hua Xiang. 2016. “DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula Hispanica*.” *Nucleic Acids Research* 44 (9): 4266–77. https://doi.org/10.1093/nar/gkw260.

Wehnes, C. A., T. G. Rehberger, R. Barrangou, and A. H. Smith. 2014. “Short communication: determination of *Salmonella* Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) diversity on dairy farms in Wisconsin and Minnesota.” *Journal of Dairy Science* 97 (10): 6370–77. https://doi.org/10.3168/jds.2013-7595.

Wei, Yunzhou, Megan T. Chesne, Rebecca M. Terns, and Michael P. Terns. 2015. “Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*.” *Nucleic Acids Research* 43 (3): 1749–58. https://doi.org/10.1093/nar/gku1407.

Welsh, John, and Michael McClelland. 1990. “Fingerprinting genomes using PCR with arbitrary primers.” *Nucleic Acids Research* 18 (24): 7213–7218.

Westra, Edze R., Angus Buckling, and Peter C. Fineran. 2014. “CRISPR-Cas systems: beyond adaptive immunity.” *Nature Reviews. Microbiology* 12 (5): 317–26. https://doi.org/10.1038/nrmicro3241.

Whitney, Cg, S Daly, S Limpongsanurak, Mr Festin, Kk Thinn, T Chipato, P Lumbiganon, et *al.* 2004. “The International infections in pregnancy study: group B streptococcal colonization in pregnant women.” *The Journal of Maternal-Fetal & Neonatal Medicine* 15 (4): 267–74. https://doi.org/10.1080/14767050410001668617.

Williams, John GK, Anne R. Kubelik, Kenneth J. Livak, J. Antoni Rafalski, and Scott V. Tingey. 1990. “DNA polymorphisms amplified by arbitrary primers are useful as genetic markers.” *Nucleic Acids Research* 18 (22): 6531–6535.

Wright, Addison V, and Jennifer A Doudna. 2016. “Protecting genome integrity during CRISPR immune adaptation.” *Nature Structural & Molecular Biology* 23 (10): 876–83. https://doi.org/10.1038/nsmb.3289.

Yancey, M. K., P. Duff, P. Clark, T. Kurtzer, B. H. Frentzen, and P. Kubilis. 1994. “Peripartum infection associated with vaginal group B streptococcal colonization.” *Obstetrics and Gynecology* 84 (5): 816–19.

Yosef, I., M. G. Goren, and U. Qimron. 2012. “Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*.” *Nucleic Acids Research* 40 (12): 5569–76. https://doi.org/10.1093/nar/gks216.

Zhang, G. W., M. Kotiw, and G. Daggard. 2002. “A RAPD-PCR genotyping assay which correlates with serotypes of group B *Streptococci.*” *Letters in Applied Microbiology* 35 (3): 247–250.

Zhang, Shaokang, Yanlong Yin, Marcus B. Jones, Zhenzhen Zhang, Brooke L. Deatherage Kaiser, Blake A. Dinsmore, Collette Fitzgerald, Patricia I. Fields, and Xiangyu Deng. 2015. “*Salmonella* serotype determination utilizing high-throughput genome sequencing data.” Edited by N. A. Ledeboer. *Journal of Clinical Microbiology* 53 (5): 1685–92. https://doi.org/10.1128/JCM.00323-15.

Valorisation de la recherche

Liste des communications orales

- **Beauruelle C**, Pastuszka A, Horvath P, Mereghetti L, Lanotte P. Suivi du portage génital des souches de *Streptococcus agalactiae* par l'analyse du locus CRISPR1. **Journée Recherche. Faculté de Médecine – CHRU Tours. 2016.**

- **Beauruelle C**, Pastuszka A, Mereghetti L, Lanotte P. Analyse du typage CRISPR de la diversité clonale de *Streptococcus agalactiae* en portage vaginal. **Journée d'animation scientifique de la Federation de Recherche en Infectiologie Centre – Loire (FéRi). 2017.**

- **Beauruelle C**, Pastuszka A, Horvath P, Mereghetti L, Lanotte P. Etude du portage génital de *Streptococcus agalactiae* par typage CRISPR. **Congrès de la Société Française de Microbiologie (SFM). 2017.**

Clémence BEAURUELLE

Locus CRISPR de *Streptococcus agalactiae* : marqueur génétique de la phylogénie de l'espèce et de l'évolution récente des isolats.

Résumé

Nous avons étudié l'intérêt du locus CRISPR1 (associé au système CRISPR-Cas de type II-A), comme marqueur épidémiologique pour le typage et l'analyse phylogénétique des souches de *Streptococcus agalactiae*, le streptocoque du groupe B (SGB). Par ce travail de thèse, nous avons pu mettre en évidence i) l'activité du système CRISPR-Cas *in vivo* ainsi que sa faible vitesse d'évolution ii) une conservation des marqueurs ancestraux, permettant d'obtenir des informations sur les lignées phylogénétiques (congruence entre le typage CRISPR et le MLST), iii) un polymorphisme du locus CRISPR1 (notamment des spacers d'acquisition récente), offrant une méthode de typage très discriminante (séparation des isolats au sein d'un même ST définit par MLST). L'analyse de ces spacers nous donne également des informations sur l'évolution récente des isolats, notamment de leurs contacts avec les EGMs. Nous avons montré l'intérêt de cet outil pour le suivi de souches de portage ou l'étude d'une population. Ainsi, à l'issu de ce travail de thèse nous proposons le typage CRISPR comme méthode de référence pour le typage et l'analyse phylogénétique des souches de SGB.

Mots clés : *Streptococcus agalactiae*, Streptocoque du groupe B, CRISPR, Typage

Abstract

We studied the relevance of the CRISPR1 array (associated to a CRISPR-Cas II-A type) as an epidemiological marker for genotyping and phylogenetic analyses of *Streptococcus agalactiae* (or Group B Streptococcus (GBS)) isolates. We demonstrate that i) spacer acquisition events occurred *in vivo* which strongly suggest that the CRISPR1-Cas system is functionally active for adaptation ii) ancestral markers (TDR and ancestral spacers) are highly conserved and reflect the phylogenetic structure of the GBS population (in congruence with MLST) iii) CRISPR1 array shared a high degree of polymorphism (especially for leader end spacers) offering a highly discriminatory typing method (allow to separate isolates within a same ST defined by MLST). Leader end analysis also provides specific evidence on isolates recent evolution, especially encounters with MGEs. CRISPR1 array appears as a useful genetic feature to follow vaginal carriage of GBS in women and for evaluate the diversity of GBS vaginal carriage population. On the basis of these data, we assume that this method could pretend to be a reference method for phylogenetic GBS typing.

Keywords: *Streptococcus agalactiae*, Group B Streptococcus, CRISPR, Typing