UNIVERSITÉ FRANÇOIS-RABELAIS DE TOURS

ÉCOLE DOCTORALE : Santé, Sciences, Technologies

E.A. 3854 - BACTERIES ET RISQUE MATERNO-FOETAL

THÈSE

présentée par :

Geneviève HÉRY-ARNAUD

soutenue le : **17 décembre 2009**

pour obtenir le grade de : **Docteur de l'Université François - Rabelais**

Discipline/ Spécialité : SCIENCES DE LA VIE ET DE LA SANTÉ

Infectiologie Cellulaire et Moléculaire, Vaccinologie

RÔLE DES ÉLÉMENTS GÉNÉTIQUES MOBILES DANS L'ÉVOLUTION ET LA VIRULENCE DE *STREPTOCOCCUS AGALACTIAE*

THÈSE dirigée par :

Dr MEREGHETTI Laurent	Maître de Conférence, HDR, Université François – Rabelais, Tours

RAPPORTEURS :

Dr GLASER Philippe	Chef de Laboratoire, HDR, Institut Pasteur, Paris
Dr POURCEL Christine	Chef de Laboratoire, HDR, Université Paris Sud-11, Paris

---

JURY :

Dr GLASER Philippe	Chef de Laboratoire, HDR, Institut Pasteur, Paris
Pr GOUDEAU Alain	Professeur des Universités, Université François – Rabelais, Tours
Dr MEREGHETTI Laurent	Maître de Conférence, HDR, Université François – Rabelais, Tours
Pr PICARD Bertrand	Professeur des Universités, Université Paris Nord-13, Paris
Dr POURCEL Christine	Chef de Laboratoire, HDR, Université Paris Sud-11, Paris
Pr QUENTIN Roland	Professeur des Universités, Université François – Rabelais, Tours

-Remerciements-

Madame le Docteur Christine Pourcel,

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

Monsieur le Docteur Philippe Glaser,

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

Monsieur le Professeur Bertrand Picard,

Merci de m'avoir encouragée dans cette voie. Soyez assuré de mon profond respect et soyez remercié d'avoir accepté de juger mon travail de recherche.

Monsieur le Professeur Alain Goudeau,

Merci de juger aujourd'hui mon travail. Merci pour votre soutien et vos précieux conseils. Soyez assuré de mon profond respect.

Monsieur le Professeur Roland Quentin,

Vous êtes à la source du projet et de la passion scientifique qui anime votre équipe. Y être accueillie fut pour moi une grande chance. Soyez assuré de ma gratitude et de mon profond respect.

Monsieur le Docteur Laurent Mereghetti,

Ce fut un bonheur de travailler sous ton encadrement. Merci pour la richesse de ton enseignement et pour tes conseils toujours judicieux. J'espère avoir fait honneur à la confiance que tu m'as accordée. Sois assuré de mon profond respect.

Merci à toute l'équipe EA3854 "Bactéries et risque materno-foetal", à l'équipe de l'unité U966 INSERM et au personnel du laboratoire de Bactériologie-Virologie de Bretonneau du CHRU de Tours, pour l'accueil et les échanges fructueux qui ont jalonné ce travail.

Philippe Lanotte, merci pour ton enseignement si juste, tes conseils et ton soutien amical.
Stella Debaets, merci pour ta gentillesse, ton aide technique et la qualité du travail accompli.
Merci à Agnès Rosenau et à Frédérique Lartigue pour les échanges scientifiques et amicaux.
Merci à Monique Lavalade et Laurence Arnault pour leur soutien technique.
Merci à Catherine Gaudy-Graffin, Pascal Vaudin, Nadine Gaudin et Alain Moreau pour leur soutien technique, logistique et amical.
Merci à l'ensemble des étudiants avec qui j'ai eu le plaisir de travailler pour l'accomplissement de ce travail expérimental. En particulier, merci à Yann Ferrandez pour sa persévérance, son travail et sa gentillesse.

Merci à toute l'équipe du Département de Bactériologie-Virologie et Parasitologie-Mycologie du CHRU de Brest. Merci pour votre soutien qui a beaucoup compté ces dernières années.

Monsieur le Professeur Christopher Payan, merci pour votre soutien, vos encouragements et la confiance que vous m'accordez. J'espère y faire honneur à travers ce travail.
Madame le Docteur Marie-Louise Abalain-Colloc, merci pour toutes ces années pendant lesquelles tu m'as suivie, conseillée et encouragée avec gentillesse et confiance. J'espère y faire honneur à travers ce travail.
Geneviève Le Lay et Didier Tandé, merci pour avoir partagé avec moi votre grande expérience de la bactériologie, et m'avoir aidée et soutenue dans mon projet.
Merci à Adissa, Dorothée, Elodie, Marie-Christine, Odile, Sophie, Valérie, aux techniciennes, aux secrétaires, et aux internes du service, pour ces années de collaboration fructueuse et conviviale.
Madame le Professeur Anne-Marie Le Flohic, merci pour m'avoir encouragée dans cette voie lorsque j'étais toute jeune interne.
Monsieur le Professeur Gilles Nevez, merci pour vos encouragements.

Merci aux membres de l'équipe EA3882 du "Laboratoire Universitaire de Biodiversité et Ecologie Microbienne".

Monsieur le Professeur Georges Barbier, merci pour vos conseils avisés.
Merci à Jean-Luc Jany pour sa disponibilité et l'aide très précieuse apportée en phylogénie.
Stéphanie Gouriou, merci pour ta collaboration fructueuse, ton soutien permanent, le dynamisme et la bonne humeur que tu insuffles à toute l'équipe.
Aux autres " animateurs" de l'équipe, en particulier Rozenn, Sylvie, Sophie et Michèle, merci pour votre bonne humeur et votre soutien amical.

Merci à tous mes amis, qui m'ont suivie et soutenue dans cette aventure.

En particulier, ma très chère AnSo,

merci pour ton aide dans la relecture minutieuse du manuscrit,

merci pour ton amitié fidèle, si chère à mes yeux.

A tous les miens, merci pour votre soutien si précieux.

A mes parents, merci pour votre amour et vore soutien indéfectible.

A mes beaux-parents, merci pour votre gentillesse et votre aide si précieuse.

A François, Hélène, Sophie et ma petite filleule Clémence,

A mes petites grands-mères tant aimées,

Merci.

A Bertrand, merci pour ton soutien inépuisable.

A Alix, mon immense bonheur.

Résumé

TITRE : ROLE DES ELEMENTS GENETIQUES MOBILES DANS L'EVOLUTION ET LA VIRULENCE DE *STREPTOCOCCUS AGALACTIAE*

Streptococcus agalactiae est une bactérie pathogène opportuniste présente chez diverses espèces animales. Au sein de l'espèce humaine, elle est retrouvée en portage digestif et vaginal, mais elle est également une cause importante d'infections néonatales et d'infections du sujet adulte. *S. agalactiae* présente un squelette génomique conservé parmi les différentes souches et des zones de plasticité localisées dans des îlots génomiques. Cette variabilité génomique est liée en partie à la présence d'éléments génétiques mobiles (EGM) qui participent aux évènements de recombinaison homologue et aux transferts horizontaux. Notre travail a porté sur la compréhension du rôle de certains EGM dans l'évolution et la virulence de *S. agalactiae*.

Dans la première partie du travail, nous avions pour objectif de valider l'approche par multilocus sequence typing (MLST) pour l'étude de la structure génétique des populations de *S. agalactiae*. Cette validation a été réalisée en analysant une collection de souches précédemment caractérisées par multilocus enzyme electrophoresis (MLEE), et en ré-interprétant les résultats du MLEE à la lumière des outils bioinformatiques développés pour le MLST. Notre travail a montré l'existence de la lignée clonale CC23, non différenciée en MLEE, dont le caractère singulier est conforté par diverses particularités biologiques spécifiques de cette lignée. Par ailleurs, nos résultats montrent une dispersion de CC19 dont le fond génétique semble partagé par une large proportion de la population, confortant l'hypothèse évolutive selon laquelle les lignées clonales actuelles dériveraient d'un seul clone génétiquement proche de CC19 (Brochet *et al.*, 2008b).

Nous avons, dans un deuxième temps, déterminé la prévalence de 11 EGM (sept séquences d'insertion IS*1548*, IS*861*, IS*1381*, ISS*a4*, ISS*ag1*, ISS*ag2*, ISS*ag9*, l'intron du groupe II GBSi1 et trois transposases SAG0448, SAG0610 et SAG1241). La distribution de ces EGM a été confrontée à la structure génétique de la population établie par MLST. Cette confrontation a permis i) de démontrer que l'acquisition des EGM corrélait avec l'évolution de l'espèce, ii) de proposer une hypothèse sur la chronologie d'acquisition des EGM selon laquelle l'émergence des principales lignées clonales chez *S. agalactiae* serait corrélée à un événement génétique lié aux EGM, et iii) de conforter l'hypothèse de l'émergence du clone CC17 à partir d'un ancêtre bovin. En effet, au sein de notre collection, 100% des souches bovines portent GBSi1, et seulement 31,7% des souches humaines. Or, les souches humaines possédant l'intron du groupe II GBSi1 sont majoritairement regroupées au sein du CC17, qui est un clone hyperinvasif impliqué dans les infections néonatales. Ainsi, GBSi1 apparaît être un marqueur de l'origine bovine des souches et aurait été conservé au cours de l'évolution par les souches humaines du CC17.

Dans la troisième partie du travail, nous avons débuté l'analyse de l'implication des EGM dans la physiopathologie des infections à *S. agalactiae*. Après avoir déterminé le nombre de copies des EGM présents chez chacune des souches par Southern blot, les sites d'insertion potentiellement associés à des facteurs de virulence ont été recherchés par Ligation Mediated-PCR. Elle a permis de mettre en évidence huit nouveaux sites d'insertion, dont cinq situés dans la séquence codante d'un gène. Les gènes cibles sont impliqués pour la plupart dans des fonctions métaboliques de la bactérie (transport de substrats cationiques ou carbonés, biosynthèse d'acide aminé). L'inactivation de ces gènes ne semble pas corréler au pouvoir pathogène des souches, ce qui pourrait s'expliquer par la redondance des systèmes de transport mis en évidence par le séquençage des génomes complets de *S. agalactiae*. Deux sites d'insertion sont cependant significativement associés à l'origine écologique de la souche.

Mots Clés :
Streptococcus agalactiae, multilocus sequence typing, éléments génétiques mobiles, évolution, virulence.

Résumé en anglais

TITRE : THE ROLE OF MOBILE GENETIC ELEMENTS IN THE EVOLUTION AND VIRULENCE OF *STREPTOCOCCUS AGALACTIAE*

Streptococcus agalactiae is an opportunistic pathogen in numerous animal species. The bacterium is a common inhabitant of the human gastrointestinal and genitourinary tracts, but it is also a major cause of neonatal infections and invasive infections in adults. *S. agalactiae* harbors a conserved genomic skeleton and areas of plasticity localized on genomic islands. The genomic plasticity is partly linked to the presence of mobile genetic elements (MGE), which participate to homologous recombination and horizontal transfers. Our study aimed to explain the role of MGE in the evolution and the virulence of *S. agalactiae*.

The first part of our work aimed to validate the multilocus sequence typing (MLST) for the study of the *S. agalactiae* population structure. We analyzed a collection of strains previously typed by multilocus enzyme electrophoresis (MLEE), and we reinterpreted the results with bioinformatic tools developed for MLST. We showed that MLST, contrarily to MLEE, identified the clonal complex CC23, a lineage which shows specific biological features. In addition, our results showed the dispatching of CC19 whose genetic background is shared by numerous strains, consistent with the hypothesis that the major clones have emerged from a single clone genetically closed to CC19.

We next determined the prevalence of 11 MGE (seven insertion sequence IS*1548*, IS*861*, IS*1381*, ISS*a4*, ISS*ag1*, ISS*ag2*, ISS*ag9*; the group II intron GBSi1; and three transposases SAG0448, SAG0610, SAG1241). The distribution of the MGE was correlated to the genetic structure of the population established by MLST. This allowed us i) to demonstrate that MGE acquisition correlates with the evolution of the species, ii) to propose a hypothesis for the chronology of acquisition of the MGE suggesting that the emergence of the main genetic lineages of *S. agalactiae* was correlated to genetic events related to the MGE, and iii) to support the hypothesis for a bovine origin for CC17. Indeed, 100% of the bovine strains harbor GBSi1, but only 31.7% of the human strains. The human strains harboring GBSi1 are mainly grouped into CC17, a hyperinvasive clone implicated in neonatal infections. Thus, GBSi1 acts as a bovine marker and would have been kept by CC17 the human strains during the *S. agalactiae* evolution.

In the third part of the study, we analyzed the impact of the MGE on the pathophysiology of *S. agalactiae* infections. We firstly determined the copy number of the MGE for each strain by Southern blotting. Then, insertion sites potentially associated with virulence factors were searched by Ligation Mediated-PCR. We found eight new insertion sites, of which five are located in the ORF. The target genes are mostly implicated in metabolic functions of the bacterium (transport of cationic or carbohydrates substrates, biosynthesis of amino acids). The inactivation of these genes does not seem to correlate with the pathogenicity, that could be explain by the redundancy observed in the systems of transport showed by the complete sequencing of the genomes of *S. agalactiae*. However, two insertion sites were significantly associated with the ecological origin of the strains.

Keywords :
Streptococcus agalactiae, multilocus sequence typing, mobile genetic elements, evolution, virulence.

Table des matières

I-INTRODUCTION 1

II-REVUE DE LA LITTERATURE 4

A. EPIDEMIOLOGIE DE *STREPTOCOCCUS AGALACTIAE* 5

1. *S. agalactiae* à travers l'Histoire 5

2. *S. agalactiae*, un pathogène majeur chez l'homme 6

2.1. *S. agalactiae*, un pathogène redouté chez la femme enceinte 6

2.1.1. Portage vaginal 6

2.1.2. Infections maternelles à *S. agalactiae* 7

2.2. *S. agalactiae*, première cause des infections néonatales 8

2.2.1. Colonisation néonatale 8

2.2.1.1. *Mode de transmission* 8

2.2.1.2. *Facteurs favorisant la transmission* 8

2.2.2. Physiopathologie des infections néonatales 8

2.2.3. Stratégie de prévention des infections néonatales 9

2.2.4. Impact de la politique de prévention des infections néonatales 10

2.2.5. Questions en suspens 10

2.3. *S. agalactiae*, pathogène émergent chez les adultes 11

2.3.1. Evolution de l'épidémiologie 11

2.3.2. Formes cliniques 11

2.3.3. Facteurs de risque 12

2.4. Epidémiologie des sérotypes de *S. agalactiae* 12

2.4.1. Sérotypes et origine écologique des souches 13

2.4.2. Sérotypes et origine géographique des souches 13

2.4.3. Sérotypes et pouvoir pathogène 14

2.5. De l'épidémiologie moléculaire à la phylogénie de *S. agalactiae* 14

2.5.1. Historique 14

2.5.2. Structure des populations de *S. agalactiae* établie par Multilocus Enzyme Electrophoresis (MLEE) 14

2.5.2.1. *Généralités* 14

2.5.2.2. *Principe* 15

2.5.2.3. *Mise en évidence de clone(s) hyperinvasif(s)* 15

2.5.3. Structure des populations de *S. agalactiae* établie par les méthodes génotypiques 15

2.5.3.1. *Restriction enzymatique de l'ADN chromosomique : technique de Restriction Fragment Length Polymorphism (RFLP)* 16

2.5.3.2. *Par électrophorèse en champ pulsé : technique de Pulsed Field Gel Electrophoresis (PFGE)* 16

2.5.4. Structure des populations de *S. agalactiae* établie par les méthodes nucléotidiques 17

2.5.4.1. *Amplification aléatoire de différentes parties du génome : technique de Randomly Amplified Polymorphic DNA (RAPD)* 17

2.5.4.2. *Amplification de fragments spécifiques de lignées* 17

2.5.4.3. *Amplification-séquençage pour l'étude du polymorphisme unilocus* 18

2.5.4.4. *Amplification-séquençage pour l'étude du polymorphisme multilocus : technique de Multilocus Sequence Typing (MLST)* 18

2.5.4.4.1. Généralités 18

2.5.4.4.2. L'épidémiologie de *S. agalactiae* à l'échelle mondiale 19

2.5.4.4.3. Apport du MLST à l'étude de la structure des populations de *S. agalactiae* 20

B. LES ELEMENTS GENETIQUES MOBILES 21
1. Généralités sur les éléments génétiques mobiles 21
1.1. Historique 21
1.2. Définition et terminologie 21
1.3. Nomenclature 22
1.4. Impact des éléments génétiques mobiles dans la virulence bactérienne 23
1.4.1. Eléments génétiques mobiles et gènes de virulence pré-existants 23
1.4.2. Eléments génétiques mobiles et acquisition de gènes de virulence 24
2. Les éléments génétiques mobiles dans le génome de *S. agalactiae* 25
2.1. Les séquences d'insertion 25
2.1.1. Généralités 25
2.1.1.1. Définition 25
2.1.1.2. Nomenclature 26
2.1.1.3. Organisation typique d'une séquence d'insertion 26
2.1.1.3.1. Les répétitions inversées terminales (IR) 26
2.1.1.3.2. Les transposases 27
2.1.1.3.3. Les répétitions cibles 27
2.1.1.4. Classification : les familles de séquences d'insertion 27
2.1.1.5. Le mécanisme réactionnel 28
2.1.1.6. La spécificité de cible 28
2.1.2. Les séquences d'insertion chez *S. agalactiae* 29
2.1.2.1. L'IS861 29
2.1.2.1.1. Caractéristiques générales 29
2.1.2.1.2. Localisation génomique et impact sur le phénotype 29
2.1.2.2. L'IS1381 30
2.1.2.2.1. Caractéristiques générales 30
2.1.2.2.2. Localisation génomique et impact sur le phénotype 30
2.1.2.3. L'IS1548 31
2.1.2.3.1. Caractéristiques générales 31
2.1.2.3.2. Localisation génotypique et impact sur le phénotype 31
2.1.2.4. L'ISSa4 32
2.1.2.4.1. Caractéristiques générales 32
2.1.2.4.2. Localisation génomique et impact sur le phénotype 32
2.1.2.5. L'ISSag2 32
2.1.2.5.1. Caractéristiques générales 32
2.1.2.5.2. Localisation génomique et impact sur le phénotype 32
2.1.2.6. L'ISSag1 33
2.1.2.6.1. Caractéristiques générales 33
2.1.2.6.2. Localisation génomique et impact sur le phénotype 33
2.1.2.7. tISSag10, à la frontière entre IS et transposon 33
2.1.2.7.1. Caractéristiques générales des transposons unitaires 33
2.1.2.7.2. Caractéristiques particulières de tIS*Sag10* 34
2.2. Les transposons composites 35
2.2.1. Caractéristiques générales 35
2.2.2. Les transposons composites de *S. agalactiae* *35*
2.3. Les transposons conjugatifs et autres transposons dérivés 36
2.3.1. Caractéristiques générales 36
2.3.1.1. Définition et structure 36
2.3.1.2. Mécanisme de transposition 37
2.3.1.3. Les dérivés des transposons conjugatifs 37
2.3.2. Les transposons conjugatifs et leurs dérivés chez *S. agalactiae* 37
2.4. Les îlots génomiques et les îlots de pathogénicité 38
2.4.1. Caractéristiques générales 38
2.4.2. Les îlots génomiques et les îlots de pathogénicité chez *S. agalactiae* 39
2.5. Les introns du groupe II 39
2.5.1. Généralités 39
2.5.1.1. Définition et nomenclature 40
2.5.1.2. Structure des introns du groupe II 40

2.5.1.3. Mécanisme réactionnel *40*
2.5.1.3.1. Auto-épissage et assemblage du complexe ribonucléoprotéique 40
2.5.1.3.2. Mobilité des introns du groupe II 41
2.5.2.Les introns du groupe II chez *S. agalactiae* 41
2.5.2.1. Caractéristiques générales de GBSi1 *42*
2.5.2.2. Impact sur le génome et le phénotype *42*

III-PARTIE EXPERIMENTALE 43

PARTIE 1. STRUCTURE GENETIQUE DE *STREPTOCOCCUS AGALACTIAE* 44

A-Introduction* *45
B-Matériel et Méthodes* *48
1. Souches bactériennes 49
2. Extraction de l'ADN génomique bactérien 49
2.1. Extraction de l'ADN total 49
2.2. Contrôle de la pureté et de la concentration des ADN 50
3. Sérotypage des souches 50
4. Caractérisation génétique des souches par multilocus sequence typing 50
4.1. Le multilocus sequence typing de *S. agalactiae* *50*
4.2. Amplification des gènes de ménage 51
4.3. Purification des produits de PCR 51
4.4. Séquençage des produits de PCR 52
4.5. Analyse *in silico* *52*
4.5.1.Analyse des séquences nucléotidiques 52
4.5.2.Assignement des allèles et des profils alléliques 53
4.5.3.Analyse phylogénétique 53
4.5.3.1. Analyses à partir des profils alléliques *53*
4.5.3.1.1. Analyse basée sur les complexes clonaux 53
4.5.3.1.2. Analyse basée sur la méthode des distances 54
4.5.3.2. Analyses à partir des séquences nucléotidiques *54*
4.5.3.2.1. Phylogénie basée sur la méthode des distances 54
4.5.3.2.2. Phylogénie basée sur le maximum de vraisemblance 54
4.5.3.2.3. Choix de l'outgroup 55
4.5.4.Evaluation du phénomène de recombinaison homologue dans la population de *Streptococcus agalactiae* à partir des données MLST 55
4.5.4.1. Exploitation des résultats issus de l'analyse eBURST *55*
4.5.4.2. Analyse de la structure de la population sous forme de réseau *55*
4.5.4.3. Evaluation du ratio recombinaison/mutation *56*
5. Caractérisation génétique des souches par multilocus enzyme electrophoresis 56
6. Analyses statistiques 57
6.1. Clonalité de la population 57
6.2. Pouvoir discriminant 57
6.3. Congruence entre les techniques de typage 57
6.3.1.Analyse en classification croisée 57
6.3.2.Analyse des correspondances 58
C-Résultats* *59
1. Structure génétique de la population définie par multilocus enzyme electrophoresis (MLEE) après ré-analyse des données 60
2. Structure génétique de la population définie en multilocus sequence typing (MLST) 61
3. Pouvoir résolutif des deux méthodes de typage 62

4. Congruence entre les lignées phylogénétiques définies par les deux méthodes de typage 62
5. Relations évolutives entre les lignées clonales 63
5.1. Lignées clonales 63
5.2. Positionnement de CC23 64
5.3. Niveau de recombinaison dans la population 64
5.3.1. Evaluation de la pertinence de l'analyse eBURST 64
5.3.2. Analyse de la structure de la population 65
5.3.3. Evaluation du ratio recombinaison/mutation 65
D-Discussion 67

PARTIE 2. CORRELATION ENTRE LA PHYLOGENIE DE *STREPTOCOCCUS AGALACTIAE* ET LA DISTRIBUTION DES EGM 71
A-Introduction 72
B-Matériel et Méthodes 74
1. Souches bactériennes 75
1.1. Souches humaines 75
1.2. Souches bovines 75
2. Extraction de l'ADN génomique bactérien 76
3. Sérotypage des souches 76
3.1. Amplification par PCR des gènes responsables des sérotypes 76
3.2. Migration et révélation des produits amplifiés 77
4. Prévalence des éléments génétiques mobiles (EGM) 77
4.1. Identification des EGM par PCR 78
4.2. Séquençage des fragments amplifiés 79
5. Analyse de la structure génétique de la population par multilocus sequence typing 79
6. Analyses statistiques 79
6.1. Analyse factorielle des correspondances 79
6.2. Autres analyses statistiques 80
C-Résultats 81
1. Distribution des sérotypes 82
2. Etude des éléments génétiques mobiles 82
2.1. Prévalence des éléments génétiques mobiles 82
2.2. Eléments génétiques mobiles et sérotypes 83
2.3. Eléments génétiques mobiles et potentiel invasif 84
2.4. Eléments génétiques mobiles et réservoir écologique des souches 85
3. Structure génétique de la population 85
3.1. Polymorphisme allélique des sept *loci* 85
3.2. Profils alléliques et sérotypes 85
3.3. Profils alléliques et potentiel infectieux 86
3.4. Profils alléliques et réservoir écologique des souches 86
3.5. Liens phylogénétiques entre les souches 86
3.5.1. Au sein de la collection 86
3.5.2. Au sein de la banque de données mondiale 87
3.6. Distribution des souches humaines et bovines au sein des lignées clonales 87
3.7. Distribution des éléments génétiques mobiles au sein des lignées clonales 88
3.8. Confrontation entre complexes clonaux, éléments génétiques mobiles, sérotypes et origines écologiques 88
3.9. Hypothèse sur la chronologie d'acquisition des EGM 90
D-Discussion 92

PARTIE 3. SITES D'INSERTION DES EGM DANS LE GENOME DE *STREPTOCOCCUS AGALACTIAE* 98

***A-Introduction* 99**

***B-Matériel et Méthodes* 101**

1. Souches bactériennes 102

2. Nombre de copies des EGM 102

2.1. Hybridation selon Southern 102

2.1.1.Microrestriction du génome bactérien 102

2.1.2.Migration des fragments de restriction 102

2.1.3.Transfert sur membrane 102

2.1.4.Hybridation des sondes marquées 103

2.1.5.Révélation 103

2.2.Analyse des données *in silico* 103

3. Identification des sites d'insertion des EGM 103

3.1. Extraction des fragments de restriction à partir du gel de migration (Protocole A)

3.1.1.Principe 104

3.1.2.Restriction du génome et PCR de vérification 104

3.1.3.Ligation et clonage des produits de digestion dans le plasmide PUC18 105

3.1.4.Lyse des clones bactériens et extraction plasmidique 105

3.2. PCR inverse (Protocole B) 106

3.2.1.Principe 106

3.2.2.Restriction du génome 106

3.2.3.Réaction d'autoligation 106

3.2.4.PCR primaire 106

3.2.5.PCR nichée 107

3.3. Ligation-Mediated PCR (Protocole C) 108

3.3.1.Principe 108

3.3.2.Restriction du génome 108

3.3.3.Assemblage de l'adaptateur 108

3.3.4.Ligation de l'adaptateur avec les produits de digestion 108

3.3.5.LM-PCR primaire 109

3.3.6.LM-PCR nichée 109

3.4.Ligation des amplicons dans le vecteur Pgem®-T Easy (Promega) 110

3.5.Séquençage 110

3.6.PCR de jonction entre l'EGM et son site d'insertion 110

***C-Résultats* 112**

1. Nombre de copies des EGM et profils d'hybridation 113

1.1.L'IS*1381* 113

1.2.L'IS*861* 113

1.3.GBSi1 113

1.4.L'IS*1548* 114

1.5.L'IS*Sa4* 114

1.6.Bilan du Southern blot 114

2. Comparaison des trois protocoles testés 115

2.1.Extraction des fragments de restriction (Protocole A) 115

2.2.PCR inverse (Protocole B) 115

2.3.Ligation Mediated-PCR (Protocole C) 116

3. Identification des sites d'insertion des EGM par LM-PCR 116

3.1.1.Identification des sites d'insertion de GBSi1 116

3.1.2.Identification des sites d'insertion de l'IS*1548* 117

3.1.3.Identification des sites d'insertion de l'IS*1381* 117

4. Distribution des copies des EGM sur le génome 118

4.1.Sites d'insertion de GBSi1 118

4.2.Sites d'insertion de l'IS*1548* 118

***D-Discussion* 120**

IV-CONCLUSION ET PERSPECTIVES **125**

V-BIBLIOGRAPHIE **129**

Annexes **150**

Publications et communications scientifiques **157**

Liste des tableaux

Tableau I. Caractéristiques majeures des principales familles de séquences d'insertion. **p27**

Tableau II. Les 113 souches de *S. agalactiae* analysées pour l'étude comparée multilocus sequence typing (MLST) *vs* multilocus enzyme electrophoresis (MLEE). **p49**

Tableau III. Amorces utilisées pour l'analyse en multilocus sequence typing. **p51**

Tableau IV. Caractéristiques des *loci* du MLST appliqué à la collection de 113 souches de *S. agalactiae*. **p61**

Tableau V. Tableau croisé des concordances et discordances entre les paires de souches (*n* = 12656) classées en groupes clonaux (GC) par MLEE et en complexes clonaux (CC) par MLST. **p62**

Tableau VI. Tableau de contingence des correspondances entre le typage MLEE et le typage MLST. **p62**

Tableau VII. Répartition des variants simple-locus (SLV) des quatre complexe clonaux (CC) en fonction de la différence du nombre de nucléotides (un ou plus) avec la séquence du sequence type (ST) fondateur du CC. **p66**

Tableau VIII. Origines écologiques des 98 souches de *S. agalactiae* de la collection issues des réservoirs humain et bovin. **p75**

Tableau IX. Amorces utilisées pour le sérotypage par PCR des souches de *S. agalactiae*. **p77**

Tableau X. Amorces utilisées pour l'amplification par PCR des éléments génétiques mobiles (EGM) de *S. agalactiae*. **p78**

Tableau XI. Prévalence des sérotypes au sein des souches humaines et bovines de *S. agalactiae*. **p82**

Tableau XII. Prévalence des éléments génétiques mobiles en fonction des sérotypes étudiés chez les 98 souches de *S. agalactiae*. **p83**

Tableau XIII. Prévalence des éléments génétiques mobiles (EGM) en fonction du potentiel de virulence et de l'origine écologique des souches de *S. agalactiae*. **p84**

Tableau XIV. Distribution des profils alléliques (ST) des souches de *S. agalactiae* en fonction de l'origine écologique et du sérotype.
p86

Tableau XV. Répartition des origines écologiques, des éléments génétiques mobiles (EGM) et des sérotypes au sein des sept complexes clonaux (CC) définis en MLST dans la population de *S. agalactiae*.
p88

Tableau XVI. Origine écologique, sérotype et prévalence de cinq éléments génétiques mobiles (EGM) pour les 52 souches de *S. agalactiae*.
p102

Tableau XVII. Amorces utilisées pour l'amplification par PCR des éléments génétiques mobiles (EGM) de *S. agalactiae*.
p105

Tableau XVIII. Amorces utilisées pour l'amplification par PCR inverse des cinq éléments génétiques mobiles (EGM) de *S. agalactiae*.
p107

Tableau XIX. Amorces utilisées pour l'amplification par Ligation-Mediated PCR des éléments génétiques mobiles (EGM) de *S. agalactiae*.
p110

Tableau XX. Amorces utilisées pour l'amplification des sites d'insertion des éléments génétiques mobiles (EGM) chez *S. agalactiae*.
p111

Tableau XXI. Sites d'insertion des éléments génétiques mobiles (EGM) identifiés par PCR inverse chez *S. agalactiae*.
p115

Tableau XXII. Sites d'insertion de GBSi1 identifiés par Ligation-Mediated PCR pour différentes souches de *S. agalactiae*.
p115

Tableau XXIII. Sites d'insertion de l'IS*1548* identifiés par Ligation-Mediated PCR pour différentes souches de *S. agalactiae*.
p116

Tableau XXIV. Distribution des copies de l'IS*1548* et de GBSi1 au sein de la collection (n=52) de *S. agalactiae*.
p118

Liste des figures

Figure 1. Janus ou les deux faces de *S. agalactiae* (buste romain de Janus, Musée du Vatican).
p6

Figure 2. Structure de la population de *S. agalactiae* établie par Multilocus Enzyme Electrophoresis aux Etats-Unis (Musser *et al.*, 1989).
p15

Figure 3. Structure de la population de *S. agalactiae* établie par Multilocus Enzyme Electrophoresis en France (Quentin *et al.*, 1995).
p15

Figure 4. Principe du multilocus sequence typing (MLST).
p18

Figure 5. Représentation graphique en « snap-shot », obtenue avec le logiciel eBURST, représentant la population de *S. agalactiae* sous forme de complexes clonaux (CC).
p20

Figure 6. Représentation des EGM en modules fonctionnels (Toussaint & Merlin, 2002).
p23

Figure 7. Structure d'une séquence d'insertion (IS).
p26

Figure 8. Site catalytique de la transposase.
p26

Figure 9. Duplication provoquée par l'insertion d'une IS.
p26

Figure 10. Modes de transposition des séquences d'insertion (IS).
p28

Figure 11. Les différentes catégories de transposons rencontrées chez *S. agalactiae*.
p33

Figure 12. Structure du transposon composite *scpB-lmb* de *S. agalactiae*.
p35

Figure 13. Structure d'un intron du groupe II.
p40

Figure 14. Structure de l'ORF d'un intron du groupe II.
p40

Figure 15. Mécanisme réactionnel de l'autoépissage des introns du groupe II.
p41

Figure 16. Mécanisme réactionnel du transfert des introns du groupe II.
p41

Figure 17. Représentation schématique de la répartition des sept loci (*adhP, pheS, atr, glnA, sdhA, glcK,* et *tkt*) étudiés en multilocus sequence typing sur le génome de la souche de *S. agalactiae* NEM 316.
p50

Figure 18. Structure de la population de *S. agalactiae* déterminée par multilocus enzyme electrophoresis (MLEE).
p60

Figure 19. Structure de la population de *S. agalactiae* déterminée par multilocus sequence typing (MLST).
p61

Figure 20. Deux modèles d'évolution de la population de *S. agalactiae* obtenus à l'issue de l'analyse par by multilocus sequence typing (MLST).
p63

Figure 21. Arbre phylogénétique reconstruit à l'aide du logiciel PAUP version 4 beta 10 avec la méthode de maximum de vraisemblance à partir des séquences concaténées générées après la combinaison des sept *l*oci étudiés dans le système MLST (Jones *et al.,* 2003*)*.
p64

Figure 22. Analyse de la structure des populations de *S. agalactiae* déterminée par MLST après analyse avec le logiciel eBURST v3 (http://eburst.mlst.net) (Feil *et al.*, 2004).
p65

Figure 23. Représentation de la structure de la population de *S. agalactiae* sous forme de réseau phylogénétique interconnecté.
p66

Figure 24. Histogramme représentant la prévalence de 11 éléments génétiques mobiles au sein de la collection de souches humaines et bovines de *S. agalactiae*.
p85

Figure 25. Dendrogramme représentant les liens évolutifs entre les 98 souches de *S. agalactiae* obtenu avec la méthode de neighbor-joining avec le logiciel MEGA v4.0 et l'indice de Kimura à partir des séquences nucléotidiques du supergène obtenues la méthode d'analyse multilocus sequence typing décrite par Jones *et al.* (Jones *et al.*, 2003).
p87

Figure 26. Analyse factorielle des correspondances multiples entre les 98 souches de *S. agalactiae* basée sur la répartition des éléments génétiques mobiles (EGM).
p89

Figure 27. Reconstitution du statut EGM (présence ou absence des 11 Eléments Génétiques Mobiles) aux principaux nœuds du dendrogramme obtenu en NJ (voir Figure 25).
p90

Figure 28. Chronologie d'acquisition des EGM.
p91

Figure 29. Principe du protocole A.
p104

Figure 30. Principe du protocole B.
p106

Figure 31. Principe du protocole C.
p108

Figure 32. Image de film autoradiographique révélant l'hybridation de la sonde IS*861* après réalisation d'un Southern blot sur les souches de *S. agalactiae*.
p113

Figure 33. Représentation schématique des régions flanquant GBSi1 chez *S. agalactiae* (d'après Luan *et al.,* 2003).
p116

Figure 34. Représentation schématique des régions flanquant l'IS*1548* chez *S. agalactiae*.
p117

Figure 35. Représentation des différents sites d'insertion d'éléments génétiques mobiles dans le génome de *S. agalactiae* de sérotype III (NEM316) et de sérotype V (2603V/R) (adapté de Herbert *et al.,* 2005).
p124

Liste des annexes

Annexe A. Représentation graphique de la totalité de la base de données multilocus sequence typing (MLST) de *S. agalactiae* obtenue avec l'algorithme eBURST v4 qui détermine les lignées clonales appelées complexes clonaux (CC) en MLST.
p151

Annexe B. Fragments de restriction mis en évidence avec la sonde IS*1381* par Southern blot pour les 37 souches IS*1381*+.
p152

Annexe C. Fragments de restriction mis en évidence avec la sonde IS*861* par Southern blot pour les 40 souches IS*861*+.
p153

Annexe D. Fragments de restriction mis en évidence avec la sonde GBSi1 par Southern blot pour les 16 souches GBSi1+.
p154

Annexe E. Fragments de restriction mis en évidence avec la sonde IS*1548* par Southern blot pour les 19 souches IS*1548*+.
p155

Annexe F. Fragments de restriction mis en évidence avec la sonde IS*Sa4* par Southern blot pour les 4 souches IS*Sa4*+.
p156

Liste des abréviations

ABC	ATP-binding cassette
Abs	absorbance
ADN	acide désoxyribonucléique
AFC	analyse factorielle des correspondances
ANAES	agence nationale d'accréditation et d'évaluation en santé
ARN	acide ribonucléique
ARNr	acide ribonucléique ribosomal
ARNt	acide ribonucléique de transfert
ATP	adénosine triphosphate
BET	bromure d'éthidium
$CaCl_2$	chlorure de calcium
CAMP-test	test de Christie Atkins Munch-Petersen
CC	complexe clonal
CDC	center for disease control and prevention
CGH	hybridation génomique comparative
CIME	*cis* mobilizable element
CIP	collection de l'Institut Pasteur
CO_2	dioxyde de carbone
CTn	transposons conjugatifs
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytosine 5'-triphosphate
DDE	Asparagine Asparagine Glutamine
dGTP	2'-deoxyguanosine 5'-triphosphate
DLV	double locus variant
dNTPs	2'-deoxynucléotidyl 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
DO	densité optique
EDTA	acide éthylènediaminetétraacétique
EGM	élément génétique mobile
ET	électrophorétype
FR	fragment de restriction
GC	groupe clonal
GC%	pourcentage de liaison guanine-cytosine dans la molécule d'ADN
GI	genomic island
HCl	acide chlorhydrique
hLRT	hierarchical likelihood ratio
HPLC-RT	reverse phase-high pressure liquide chromatography
I_A	indice d'association
ICE	integrative conjugative element
IME	integrative and mobilizable element
IPTG	Isopropyl-beta-thio-galactoside
IR	inverted repeat
IRL	inverted repeat left
IRR	inverted repeat right
IS	insertion sequence
IV	intra-veineuse
Kb	kilobase
KCl	chlorure de potassium
LB	Luria-Bertani
LCR	liquide céphalo-rachidien
$MgCl_2$	chlorure de magnésium
ML	maximum likelihood
MLEE	multilocus enzyme electrophoresis
MLST	multilocus sequence typing
MTn	transposons mobilisables
NaCl	chlorure de sodium

NaOH	hydroxyde de sodium
NCBI	national center for biotechnology information
NJ	neighbor-joining
nt	nucléotide
OMS	organisation mondiale de la santé
ORF	open reading frame
pb	paire de bases
PCR	polymerase chain reaction
PFGE	pulse field gel electrophoresis
PI	pathogenicity island
PHI	pairwise homoplasy index
RAPD	random amplified polymorphic DNA
RDP	restriction digest pattern
RFLP	restriction fragment length polymorphism
RPDM	rupture prématurée des membranes
RT	reverse transcriptase
RT-PCR	reverse transcription-polymerase chain reaction
SDS	dodecyl sodium sulfate
SLV	single locus variant
SSC	sodium citrate
ST	sequence type
TAE	tris acétate EDTA
TBE	tris borate EDTA
TE	tris base EDTA
Tn	transposon
Tris	trishydroxyméthylaminométhane
UPGMA	unweighted pair group method with arithmetic mean
UV	ultra-violets

I-INTRODUCTION

Initialement décrit comme agent de mammites bovines à la fin du XIXe siècle, *Streptococcus agalactiae* est longtemps resté une étiologie spécifique à la médecine vétérinaire (Nocard & Mollereau, 1887). De récents travaux ont montré qu'il était également présent en portage ou impliqué en pathologie chez d'autres animaux (Devriese, 1991, Elliott *et al.*, 1990, Lammler *et al.*, 1998, Schenkman *et al.*, 1994). Ce n'est que dans les années 1960 que ce pathogène opportuniste, présent en portage digestif et vaginal au sein de l'espèce humaine, est devenu une cause majeure d'infections néonatales (Eickhoff *et al.*, 1964). La contamination des nouveaux-nés survient principalement au moment de l'accouchement, par contact avec les sécrétions vaginales de la mère. Les recommandations des sociétés savantes, qui préconisent une antibioprophylaxie pour réduire la transmission materno-fœtale de la bactérie (Anonymous, 2001, Anonymous, 2002), ont permis une diminution significative de l'incidence des infections néonatales (Schrag *et al.*, 2000). Toutefois, *S. agalactiae* représente toujours la première étiologie d'infections néonatales dans les pays industrialisés (Phares *et al.*, 2008). Plus récemment, il a aussi émergé en tant qu'important pathogène responsable d'infections chez les sujets adultes débilités (Skoff *et al.*, 2009).

Le degré d'immunodépression de l'hôte est un facteur majeur dans l'expression du pouvoir pathogène de la bactérie (Krause *et al.*, 2009), comme en témoigne l'atteinte préférentielle des adultes immunodéprimés et des nouveaux-nés (Schuchat, 1998). Il est également évident que toutes les souches de *S. agalactiae* ne présentent pas le même risque infectieux pour le nouveau-né. En effet, l'incidence relativement faible des infections néonatales contraste avec la fréquence élevée du portage vaginal asymptomatique des futures parturientes, évaluée entre 5 et 36% selon les études (Barcaite *et al.*, 2008, Loulergue *et al.*, 2004, Puopolo *et al.*, 2005). Il apparaît que les souches de colonisation vaginale présentent une plus grande diversité génétique que les souches impliquées dans les infections invasives. Ainsi, les travaux menés depuis vingt ans pour une meilleure compréhension de la structure génétique de *S. agalactiae* ont montré l'existence de génogroupes à plus haut risque infectieux pour le nouveau-né (Jones *et al.*, 2003, Musser *et al.*, 1989, Quentin *et al.*, 1995, Rolland *et al.*, 1999, Takahashi *et al.*, 1998).

La diversité génomique de *S. agalactiae* montre un squelette génomique ou « core génome » très conservé au sein des différentes souches et des zones de plasticité localisées dans des îlots, représentant parfois des îlots de pathogénicité (Glaser *et al.*, 2002, Herbert *et al.*, 2005, Tettelin *et al.*, 2002). Cette variabilité génomique est liée pour partie à la présence d'éléments génétiques mobiles (EGM), qui en participant aux évènements de recombinaison homologue et aux transferts horizontaux, représentent une source majeure de diversité génétique favorisant l'adaptabilité bactérienne. Par ailleurs, plusieurs EGM sont proches de gènes codant des facteurs de virulence, comme le gène de la hyaluronate lyase, l'opéron codant le transporteur de l'hémolysine, ou le gène codant une protéine d'adhésion à la laminine (Granlund *et al.*, 1998, Luan *et al.*, 2003, Sellin *et al.*, 2000, Spellerberg *et al.*, 2000). Ainsi, en modulant l'expression des gènes, les EGM peuvent avoir un impact direct sur le potentiel infectieux des bactéries (Hacker & Kaper, 2000).

Ce travail débute par une revue de la littérature portant sur les principaux éléments d'épidémiologie des infections à *S. agalactiae*, en particulier les aspects phylogénétiques liés à la bactérie ; nous faisons également une revue complète des EGM présents chez *S. agalactiae*. Nous présentons ensuite notre travail expérimental dont l'objectif général a été de comprendre le rôle des éléments génétiques mobiles dans l'évolution et la virulence de *S. agalactiae*. Dans la première partie expérimentale, nous avons analysé la structure génétique de la population par multilocus sequence typing complété par une ré-analyse des résultats obtenus lors d'une précédente étude par multilocus enzyme electrophoresis grâce à de nouveaux outils bioinformatiques. Dans la seconde partie, nous avons confronté l'analyse phylogénétique à la distribution des EGM, en montrant une corrélation entre la chronologie d'acquisition des EGM et l'évolution de l'espèce. La distribution particulière de certains EGM a également permis de conforter l'hypothèse de l'émergence du clone humain hyperinvasif à partir d'un ancêtre bovin (Bisharat *et al.*, 2004). Enfin, au cours de la troisième partie du travail, nous avons déterminé le nombre de copies des EGM, puis identifié et analysé les sites d'insertion des EGM pour déterminer le rôle potentiel de ces EGM dans la virulence de *S. agalactiae*.

II-REVUE DE LA LITTERATURE

A. ÉPIDÉMIOLOGIE DE *STREPTOCOCCUS AGALACTIAE*

1. *S. agalactiae* à travers l'Histoire

S. agalactiae est une bactérie qui appartient à la famille des *Streptococcaceae* et au genre *Streptococcus*. Le genre *Streptococcus* (*streptus* : flexible et *coccus* : grain) fut attribué pour la première fois en 1874 par Billroth qui observa dans des lésions d'érysipèle un microorganisme en forme de coques assemblées en chaînettes. Au début du XXe siècle, le caractère hémolytique des streptocoques fut décrit par Schottmuller, puis utilisé par Brown comme critère de classification. C'est en 1933 que la première classification moderne des streptocoques fut établie par Rebecca Lancefield. Cette classification, basée sur les propriétés antigéniques d'un composé pariétal des streptocoques, le polyoside C, a permis d'identifier *S. agalactiae* comme un streptocoque du groupe B (Lancefield & Freimer, 1966).

L'espèce *S. agalactiae* a été initialement décrite chez les bovins. Identifiée par Nocard et Mollereau en 1887 comme le «streptocoque des mastites bovines», elle a été baptisée par Lehmann et Neumann en 1896 avec un nom qui y fait allusion, « *agalactiae* » signifiant « absence de lait » (Lehmann & Neumann, 1896, Nocard & Mollereau, 1887). L'infection à *S. agalactiae* est en effet associée à une diminution de la quantité (et de la qualité) des produits laitiers (Keefe *et al.*, 1997). *S. agalactiae* est un pathogène obligatoire de la glande mammaire des bovins et est resté pendant longtemps une des principales étiologies des mammites chroniques ou aiguës (Dobbins, 1977, Jain, 1979). Grâce à l'utilisation des antibiotiques en médecine vétérinaire et aux mesures sanitaires, les mastites bovines à *S. agalactiae* ont considérablement régressé dans la plupart des pays (Keefe *et al.*, 1997). Concomitamment, l'espèce a fait son apparition chez l'homme. Les premiers cas d'infection néonatale à *S. agalactiae* ont été décrits par Eickhoff *et al.* en 1964 (Eickhoff *et al.*, 1964). *S. agalactiae* a été isolée depuis chez d'autres animaux, comme le chien, le chat, le cheval, la chèvre, le singe, le cochon d'Inde, le rat, la souris, l'éléphant, le dauphin, la grenouille, le crocodile, et même les poissons (Devriese, 1991, Elliott *et al.*, 1990, Lammler *et al.*, 1998, Wibawan *et al.*, 1993, Schenkman *et al.*, 1994, Evans *et al.*, 2006, Evans *et al.*, 2009, Bishop *et al.*, 2007). Ces cas sont pour la plupart anecdotiques et le pouvoir pathogène y est généralement peu documenté. Actuellement, *S. agalactiae* compte donc deux réservoirs majeurs : l'homme et le bovin.

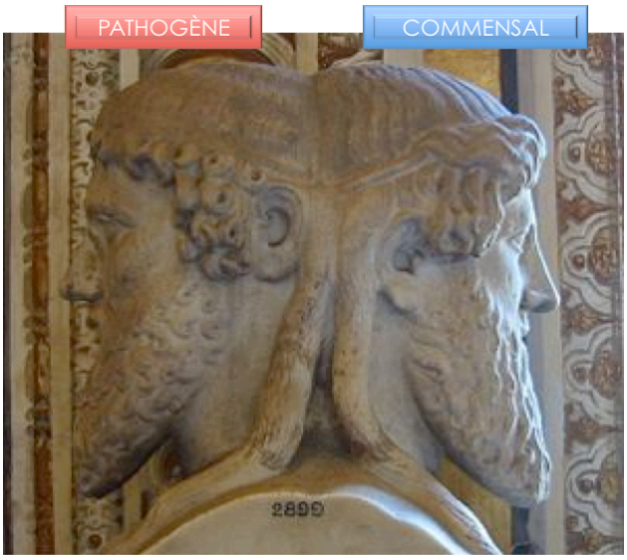


Figure 1. Janus ou les deux faces de *S. agalactiae*
(buste romain de Janus, Musée du Vatican).

2. *S. agalactiae*, un pathogène majeur chez l'homme

Chez l'homme, *S. agalactiae* se présente tel Janus avec ses deux visages tournés en sens contraire (**Figure 1**). Un premier portrait de la bactérie dépeint un commensal retrouvé principalement en portage digestif et vaginal, et plus rarement sur d'autres sites anatomiques. Le deuxième portrait dépeint un pathogène redoutable pour la femme enceinte et surtout les nouveaux-nés, et, depuis une quinzaine d'années, pour les personnes âgées ou ayant un terrain débilité.

2.1. *S. agalactiae,* un pathogène redouté chez la femme enceinte

2.1.1. Portage vaginal

Le tractus gastrointestinal est le réservoir naturel de *S. agalactiae*. Il est aussi la source de la colonisation vaginale. La colonisation vaginale est soit transitoire, soit chronique, soit intermittente (Schrag *et al.*, 2002). Le portage vaginal asymptomatique varie de 5,1 à 36% selon les études internationales. Aux Etats-Unis, les études rapportent des taux allant de 20 à 30%, avec des différences entre les origines ethniques (Puopolo *et al.*, 2005, Schuchat, 1995, Regan *et al.*, 1991). En Europe, la prévalence semble plus importante dans les pays de l'Est et du Nord, et un tiers des études rapporte des taux d'au moins 20% (Barcaite *et al.*, 2008). En France, la prévalence de la colonisation vaginale chez les femmes enceintes a été estimée entre 5,1 et 22%, avec une moyenne proche des 10% (Loulergue *et al.*, 2004). Dans les pays en voie de développement, les taux de portage sont globalement similaires à ceux des pays industrialisés bien que *S. agalactiae* semble être une étiologie plus rare d'infection néonatale (Stoll & Schuchat, 1998).

Les fluctuations constatées entre les pays proviennent avant tout des différences dans les modalités de prélèvement, de transport et de culture : les prévalences sont plus fortes quand le prélèvement est recto-vaginal, quand un milieu de transport est utilisé, quand l'acheminement du prélèvement ne dépasse pas un jour, et quand un milieu d'enrichissement est utilisé (Hakansson *et al.*, 2008, Badri *et al.*, 1977, Chhuy *et al.*, 2005, Jaureguy *et al.*, 2003, Schrag *et al.*, 2002). Ces paramètres, qui amélioreraient la sensibilité de détection du *S. agalactiae,* sont pour la plupart préconisés dans les recommandations nationales de nombreux pays, notamment celles du CDC (Center for Disease Control and prevention) aux Etats-

Unis (Schrag *et al.*, 2002). En France, la stratégie de dépistage de *S. agalactiae* est différente (Anonymous, 2001). Elle s'appuie sur les critères définis par l'OMS pour proposer un dépistage de masse qui est un compromis entre fiabilité, coût modéré, sensibilité, spécificité, prélèvement non traumatisant et applicabilité à une large population cible (Wilson & Jungner, 1968). Les recommandations éditées par l'ANAES (Agence Nationale d'Accréditation et d'Evaluation en Santé) en 2001 précisent l'inutilité du prélèvement ano-rectal et du milieu d'enrichissement dont l'utilisation n'a pas démontré de supériorité dans la prévention des infections materno-fœtales, mais qui par contre engendre un surcoût (Anonymous, 2001).

2.1.2. Infections maternelles à *S. agalactiae*

Si l'impact de la colonisation vaginale maternelle sur les infections néonatales à *S. agalactiae* a été largement étudié, peu d'études ont concerné l'impact de la colonisation vaginale sur les infections maternelles à *S. agalactiae*. Cette bactérie serait la cause de 15% des infections de la femme enceinte en *post-partum* ou en *post-abortum* (Baker, 1997). *S. agalactiae* est 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), celles-ci s'accompagnant parfois d'épisodes bactériémiques (Regan *et al.*, 1996). Le risque de développer une chorioamniotite est lié à la densité de colonisation vaginale (Yancey *et al.*, 1994, Krohn *et al.*, 1999). Par contre, la densité semble sans influence sur le risque de survenue d'endométrite du *post-partum* (Krohn *et al.*, 1999). Quant à l'association avec un risque augmenté de prématurité, d'hypotrophie de l'enfant, ou de rupture prématurée des membranes, ce problème complexe reste discuté : certaines études montrent que ce risque est majoré en cas de colonisation massive à *S. agalactiae* (Regan *et al.*, 1996, Feikin *et al.*, 2001 , Galask *et al.*, 1984, Ferrieri *et al.*, 1977), alors que d'autres études ne retrouvent pas cette relation (Valkenburg-van den Berg *et al.*, 2009, Garland *et al.*, 2000). En revanche, il est incontesté que la colonisation maternelle par *S. agalactiae* est à l'origine de la colonisation néonatale et qu'elle constitue le facteur de risque majeur de l'infection néonatale à *S. agalactiae* (Baker, 1997).

2.2. *S. agalactiae*, première cause des infections néonatales

2.2.1. Colonisation néonatale

2.2.1.1. Mode de transmission

La transmission verticale est le mode principal de transmission de la mère à l'enfant (75% des cas) (Winn, 2007). La transmission horizontale est possible après la naissance, soit à partir d'un portage pharyngé ou cutané de la mère ou d'autres contacts nosocomiaux ou communautaires, soit par l'allaitement maternel (Bingen *et al.*, 1992, Gagneur *et al.*, 2009, Winn, 2007). Les voies de contamination par transmission verticale sont principalement la voie ascendante et l'inhalation des secrétions vaginales au passage de la filière génitale maternelle (Winn, 2007). Le taux de transmission est de l'ordre de 50% à 75% (Quentin *et al.*, 2002). Dans 10 à 20% des cas, la colonisation du nouveau-né évoluera vers une pathologie grave susceptible d'entraîner le décès du nouveau-né ou de graves séquelles neurologiques (méningite, septicémie, infection pulmonaire, leucomalacie périventriculaire notamment) (Brimil *et al.*, 2006, Schrag *et al.*, 2002, Schuchat, 1998). A ce titre, *S. agalactiae* représente la première cause d'infections néonatales précoces dans la plupart des pays industrialisés et serait responsable de la moitié des infections néonatales en France chaque année (Anonymous, 2001, Schuchat, 1999).

2.2.1.2. Facteurs favorisant la transmission

La présence de *S. agalactiae* en portage vaginal maternel et la densité de colonisation sont les éléments majeurs qui interviennent sur le taux de contamination et donc sur le risque pour le nouveau-né de développer une infection (Krohn *et al.*, 1999). La rupture prématurée des membranes (RPDM) supérieures à 12 à 18 heures et l'ouverture du col utérin sont également des facteurs de risque. Les touchers vaginaux répétés (>6), les manœuvres d'extraction instrumentale, les césariennes sur RPDM, majorent le risque d'infections materno-fœtales lorsqu'il y a un portage vaginal (Schuchat, 1998).

2.2.2. Physiopathologie des infections néonatales

Les infections materno-fœtales sont regroupées sous deux entités :

1) une forme précoce qui survient dans les 7 premiers jours de vie du nouveau-né et principalement au cours des 48 premières heures ; elle associe une bactériémie et une atteinte pulmonaire (87% des cas), et plus rarement une méningite (10 % des cas) (Schuchat, 1998). Cette forme est essentiellement due à la contamination du liquide amniotique après rupture des membranes ou de l'inhalation des sécrétions vaginales par le nouveau-né au moment de l'accouchement (Shet & Ferrieri, 2004),

2) une forme tardive qui apparaît entre le 8ème jour et le 3ème mois de vie et associe des manifestations cliniques variées (ostéite, cellulite) à une atteinte méningée dans 50 % des cas (Baker, 1997). La pathogénèse de cette forme est moins bien comprise. Une transmission horizontale, notamment par le biais de l'allaitement maternel, pourrait jouer un rôle (Shet & Ferrieri, 2004).

2.2.3. Stratégie de prévention des infections néonatales

Dans le cadre de la prévention anténatale de ce risque infectieux, des recommandations concernant le dépistage et la prise en charge maternelle ont été proposées dans la plupart des pays industrialisés, notamment aux Etats-Unis dès 1996 (recommandations révisées en 2002) (Anonymous, 1996, Schrag *et al.*, 2002) et en France en 2001 (Anonymous, 2001). La stratégie préventive préconisée par l'ANAES repose sur un dépistage systématique du portage vaginal de *S. agalactiae* en fin de grossesse, entre la 34e et la 38e semaines d'aménorrhée. Le test de dépistage est la culture en boîte de Petri sur gélose au sang, sans enrichissement sélectif, d'un prélèvement vaginal (incluant le balayage des parois de la moitié inférieure du vagin jusqu'au vestibule) réalisé à l'écouvillon. La technique d'ensemencement est celle des quadrants qui permet l'obtention d'un résultat semi-quantitatif (Anonymous, 2001).

En cas de dépistage positif, de même qu'au cours des situations obstétricales à risque infectieux, une antibioprophylaxie *per-partum* doit être systématiquement administrée. Celle-ci consiste en une première injection par voie intra-veineuse (IV) de 5 millions d'UI de pénicilline G dès le déclenchement du travail, suivi de 2,5 millions d'UI (IV) toutes les 4 heures jusqu'à l'expulsion, ou d'amoxicilline IV (2g puis 1g toutes les 4 heures). En cas d'allergie à la pénicilline, l'administration d'érythromycine IV (500 mg toutes les 6 heures) ou d'une céphalosporine est recommandée en France. L'allergie aux pénicillines justifie la réalisation d'un

antibiogramme car près d'un quart des souches isolées actuellement en France sont résistantes aux macrolides en France (Poyart *et al.*, 2008). Aux Etats-Unis, les recommandations revues de 2002 préconisent l'utilisation de la cefazoline en raison du faible risque d'allergie croisée, de son intérêt écologique (antibiotique de première ligne) et surtout de l'absence de résistance (Anonymous, 2002).

2.2.4. Impact de la politique de prévention des infections néonatales

Des estimations récentes ont montré que les infections néonatales précoces avaient diminué selon les pays de 50 à 80% depuis la mise en place des recommandations des instances nationales. Aux Etats-Unis, l'incidence, qui était de 1,7 cas pour 1000 naissances avant 1996, est passée à 0,6 en 1998, puis s'est stabilisée à 0,4 cas pour 1000 naissances (Anonymous, 2009, Anonymous, 2002). En France, les taux sont plus faibles mais suivent la même tendance : l'incidence qui était de 0,69 cas pour 1000 naissances avant 2001, est passée à 0,47 en 2002, puis à 0,23 cas pour 1000 naissances en 2006, ce qui représente 180 cas par an (Jourdan-Da Silva *et al.*, 2008). Ces données ne suffisent pas à attribuer cette diminution des infections néonatales à *S. agalactiae* à cette seule stratégie. Toutefois, les enquêtes réalisées dans les maternités confortent cette hypothèse (Jourdan-Da Silva *et al.*, 2008).

2.2.5. Questions en suspens

Les chiffres cités ci-dessus masquent une grave disparité ethnique observée dans le « melting-pot » américain: si l'incidence des infections est restée stable les cinq dernières années dans les populations blanches (0,26 à 0,29 cas pour 1000 naissances), celle-ci a augmenté dans les populations afro-américaines (0,53 à 0,86 cas pour 1000 naissances) (Anonymous, 2009). C'est ainsi que les nouveaux-nés afro-américains ont trois fois plus de risque de faire une infection à *S. agalactiae* que les nouveaux-nés blancs (Jordan *et al.*, 2008). Cette différence ethnique serait probablement plus le reflet des conditions socio-économiques que de spécificités génétiques propres à une origine ethnique, mais aucune étude n'a été réalisée en ce sens à ce jour.

Les politiques de prévention semblent sans effet sur l'incidence des infections tardives dont les taux sont restés inchangés depuis dix ans (environ 0,30 cas pour 1000 naissances aux Etats-Unis et de 0,10 en France, soit 70 cas par an),

probablement en raison d'une transmission post-natale (Anonymous, 2009, Jourdan-Da Silva *et al.*, 2008).

Par ailleurs, une étude récente multicentrique réalisée aux Etats-Unis relève que plus de 60% des nouveaux-nés infectés par *S. agalactiae* sont nés de mères dépistées négatives (Van Dyke *et al.*, 2009). Ce chiffre souligne la proportion importante de faux négatifs, lesquels peuvent s'expliquer par le phénomène de colonisation transitoire, mais également par le défaut de sensibilité de la méthode de screening par culture (Van Dyke *et al.*, 2009, Yow *et al.*, 1980). Certains auteurs soulignent de ce fait l'intérêt des méthodes de screening rapide basées sur la PCR en temps réel et réalisées extemporanément entre le début du travail et la délivrance ; celles-ci n'ont néanmoins pas fait la preuve de leur rapport coût-efficacité (Bergeron *et al.*, 2000, Daniels *et al.*, 2009, El Helali *et al.*, 2009).

2.3. *S. agalactiae*, pathogène émergent chez les adultes (femmes enceintes exclues)

2.3.1. Evolution de l'épidémiologie

Plus récemment, *S. agalactiae* a pris une nouvelle place en tant qu'agent infectieux émergent chez les sujets adultes âgés et/ou immunodéprimés (Edwards & Baker, 2005, Farley, 2001, Phares *et al.*, 2008, Schuchat, 1998, Schwartz *et al.*, 1991, Skoff *et al.*, 2009). En effet, l'observation des quinze dernières années montre une recrudescence des infections invasives à *S. agalactiae* dans la population adulte (femmes enceintes exclues) de nombreux pays tels que la France, l'Espagne, le Danemark, les Etats-Unis, ou Taiwan, avec des taux allant de 4,4 à 23 cas pour 100 000 adultes selon les pays (Trivalle *et al.*, 1998, Farley, 2001, Blancas *et al.*, 2004, Lefebvre *et al.*, 2007, Lee *et al.*, 2005, Ekelund *et al.*, 2005). Aux Etats-Unis, un doublement de l'incidence a été observé entre 1990 et 2007 ; le taux est passé de 3,6 à 7,3 cas pour 100 000 adultes (Skoff *et al.*, 2009).

2.3.2. Formes cliniques

Les bactériémies primaires sont la forme la plus courante d'infections invasives à *S. agalactiae* et représentent près de 40% des manifestations cliniques ; elles sont suivies par les infections de la peau et/ou des tissus mous (près de 26% des manifestations cliniques), les pneumonies et les infections urinaires (Skoff *et al.*, 2009).

Les pneumonies et les infections urinaires sont plus courantes chez les personnes âgées (Edwards & Baker, 2005). Des cas de spondylodiscite, d'arthrite septique et, plus rarement, d'endocardite ou de méningite ont été rapportés (Farley, 2001, Nolla *et al.*, 2003, Sambola *et al.*, 2002, Varon, 2009). Les infections sont soit communautaires soit nosocomiales ; elles sont polymicrobiennes dans moins de 10% des cas (Easmon *et al.*, 1981, Farley, 2001, Jackson *et al.*, 1995, Skoff *et al.*, 2009).

2.3.3. Facteurs de risque

Les infections invasives à *S. agalactiae* du sujet adulte touchent préférentiellement les sujets âgés ou atteints de maladies chroniques immunodépressives (Farley, 2001). Ces patients présentent des taux de colonisation identiques à ceux des femmes en âge de procréer ; par conséquent, ce sont surtout les facteurs de comorbidité, présents dans près de 90% des cas (au moins un facteur), qui modifient l'épidémiologie de cette tranche de la population (Skoff *et al.*, 2009, Edwards & Baker, 2005). Le diabète de type 2 est la maladie sous-jacente la plus fréquemment retrouvée ; les maladies cardio-vasculaires, hépatiques et néoplasiques sont d'autres facteurs de risque possibles (Farley, 2001, Schwartz *et al.*, 1991, Skoff *et al.*, 2009). Actuellement, *S. agalactiae* serait responsable de 23% des infections invasives chez les immunodéprimés (Ekelund *et al.*, 2005).

2.4. Epidémiologie des sérotypes de *S. agalactiae*

Le sérotype capsulaire est l'un des premiers marqueurs utilisés dans les études épidémiologiques descriptives (Brimil *et al.*, 2006, Dogan *et al.*, 2005, Ekelund & Konradsen, 2004, Gherardi *et al.*, 2007, Harrison *et al.*, 1998, Phares *et al.*, 2008, Quentin *et al.*, 1995). Les méthodes de sérotypie moléculaire tendent actuellement à supplanter les méthodes classiques basées sur les antisérums (Cieslewicz *et al.*, 2005, Kong *et al.*, 2002, Poyart *et al.*, 2007). La sérotypie moléculaire a d'ailleurs contribué à découvrir de nouveaux sérotypes. Dix sérotypes sont actuellement dénombrés : le sérotype Ia, Ib, II, III, IV, V, VI, VII, VIII et le IX récemment décrit (Jelinkova & Motlova, 1985, Jensen, 1979, Motlova *et al.*, 1986, Ryc *et al.*, 1988, Slotved *et al.*, 2007).

2.4.1. Sérotypes et origine écologique des souches

On observe une distribution préférentielle des sérotypes en fonction de l'origine écologique des souches. En portage vaginal maternel, ce sont, les sérotypes III et Ia sont les plus fréquents, suivis par le sérotype II ou le V (Barcaite *et al.*, 2008, Phares *et al.*, 2008, van der Mee-Marquet *et al.*, 2009). Pour les autres sites de portage de *S. agalactiae* (peau, gorge et marge anale), ce sont également les sérotypes III et Ia qui prédominent (van der Mee-Marquet *et al.*, 2008). Chez les nouveaux-nés, il faut distinguer les infections précoces des infections tardives. La distribution des sérotypes dans les infections néonatales précoces est globalement similaire à celle des souches de portage vaginal des femmes enceintes : dans l'ordre, on retrouve les sérotypes Ia et III (30% des souches chacun), les sérotypes V (18% des souches) et les sérotypes II (13% des souches) (Phares *et al.*, 2008). Dans les infections néonatales tardives, le sérotype III représente entre la moitié et 83% des souches, suivi par les sérotypes Ia (24% des souches) et V (14% des souches) (Castor *et al.*, 2008, Phares *et al.*, 2008).

2.4.2. Sérotypes et origine géographique des souches

Les prévalences rapportées ci-dessus proviennent d'études réalisées dans les pays où la surveillance épidémiologique est la plus forte, à savoir l'Europe et les Etats-Unis. Des spécificités géographiques sont néanmoins observées de part le monde : par exemple, au Japon, ce sont les sérotypes VI et VIII qui prédominent en portage vaginal, aux Emirats Arabes Unis le sérotype IV, et le sérotype V dans plusieurs pays africains (Gambie, Zimbabwe, Sénégal et République Centrafricaine) (Amin *et al.*, 2002, Suara *et al.*, 1994, Moyo *et al.*, 2002, Lachenauer *et al.*, 1999, Brochet *et al.*, 2009a). Les disparités géographiques sont plus marquées chez les souches responsables d'infections invasives de l'adulte (hors femmes enceintes). Aux Etats-Unis, ces souches sont majoritairement de sérotype V (entre un tiers et la moitié des souches), puis viennent les sérotypes Ia (entre 24 et 39% des souches), II (entre 12 et 36%) et III (entre 12 et 27%) (Phares *et al.*, 2008, Skoff *et al.*, 2009). Au Japon, le sérotype Ib, classiquement peu fréquent, prédomine (31% des souches), suivi par le sérotype V (19%) (Murayama *et al.*, 2009). De même, les souches de sérotypes VI sont rarement isolées dans ce contexte hormis au Japon (Matsubara *et al.*, 2006).

2.4.3. Sérotypes et pouvoir pathogène

Certains sérotypes semblent significativement associés à certaines entités cliniques. C'est le cas du sérotype III associé aux méningites, ou du sérotype Ia associé à des formes mortelles d'infections néonatales (Phares *et al.*, 2008, Castor *et al.*, 2008). L'émergence des infections invasives de l'adulte est à relier à celle du sérotype V qui date des années 90 (Blumberg *et al.*, 1996, Elliott *et al.*, 1998). Celle-ci correspond à la diffusion d'un clone au niveau mondial (Le Thomas-Bories *et al.*, 2001). La distribution des sérotypes est donc également soumise à des variations temporelles en plus des variations géographiques.

2.5. De l'épidémiologie moléculaire à la phylogénie de *S. agalactiae*

2.5.1. Historique

Historiquement, les marqueurs utilisés pour le typage épidémiologique des souches en circulation étaient la sérotypie, la biotypie (avec l'étude de la fermentation du lactose notamment), et la lysotypie (Christensen & Christensen, 1978, Stringer & Maxted, 1979, Norcross & Oliver, 1976, Wilkinson, 1978). Ces marqueurs, peu discriminants, se sont rapidement révélés insuffisants pour tracer la diffusion des clones (Norcross & Oliver, 1976, Denning *et al.*, 1989, Band *et al.*, 1981, Haug *et al.*, 1981). Avec l'avènement des méthodes moléculaires et leur application en microbiologie (Tompkins, 1992), la diversité génétique de l'espèce a été plus largement explorée, et les approches moléculaires se sont diversifiées.

2.5.2. Structure des populations de *S. agalactiae* établie par Multilocus Enzyme Electrophoresis (MLEE)

2.5.2.1. Généralités

L'étude du polymorphisme allélique des isoenzymes ou MLEE (Multilocus Enzyme Electrophoresis) a été la première méthode phénotypique capable de refléter fidèlement un variant allélique spécifique de *loci* génomiques (Selander *et al.*, 1986). C'était la première méthode capable de fournir des informations sur la fréquence des allèles et sur les génotypes multilocus, éléments requis pour l'analyse de la structure génétique des populations (Selander *et al.*, 1986).

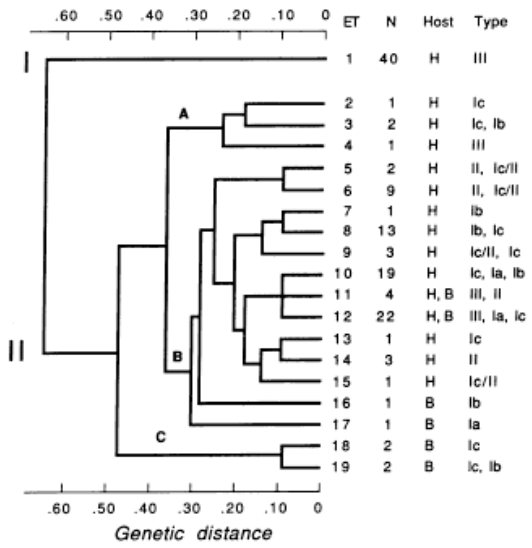


Figure 2. Structure de la population de *S. agalactiae* établie par Multilocus Enzyme Electrophoresis aux Etats-Unis (Musser *et al.*, 1989).

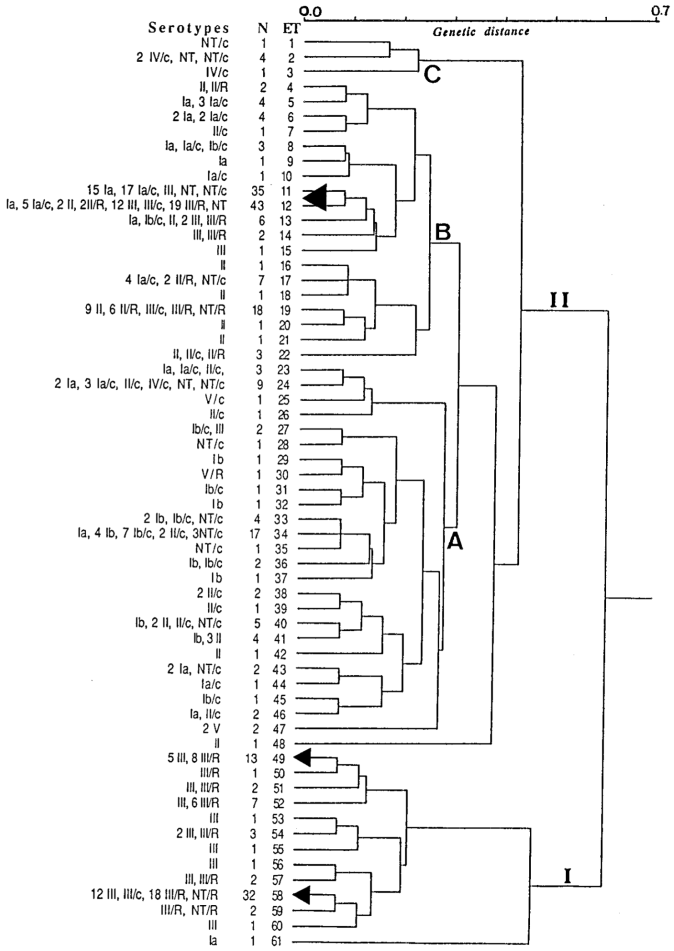


Figure 3. Structure de la population de *S. agalactiae* établie par Multilocus Enzyme Electrophoresis en France (Quentin *et al.*, 1995).

2.5.2.2. Principe

Le principe repose sur la variation de la mobilité électrophorétique des enzymes qui reflète les changements de leur séquence en acides aminés, modifiant leur migration. Les isolats qui possèdent des allèles identiques à tous les *loci* étudiés appartiennent au même électrophorétype (ET). Cette méthode fournit des informations sur les distances génétiques entre les isolats testés et permet l'étude la structure des populations bactériennes (Ochman & Selander, 1984). L'inconvénient majeur de la technique reste sa lourdeur de réalisation.

2.5.2.3. Mise en évidence de clone(s) hyperinvasif(s)

Musser *et al.* ont révélé la structure clonale de la population de *S. agalactiae* par MLEE (Musser *et al.*, 1989), en montrant deux grandes divisions phylogénétiques à l'intérieur de l'espèce (**Figure 2**) : la première division très homogène composée d'un seul ET (ET1) ne regroupant que le sérotype III ; la seconde division hétérogène composée de 18 ET et des sérotypes variés (dont le sérotype III) (Musser *et al.*, 1989). Le clone ET1 avait la particularité de regrouper 81% des souches invasives néonatales, lesquelles composaient 85% des souches de cette lignée (Musser *et al.*, 1989). En mettant en évidence l'existence de ce clone hautement virulent pour le nouveau-né, la corrélation entre un clone et le pouvoir pathogène était démontrée pour la première fois chez *S. agalactiae*.

Les études réalisées ultérieurement en France et au Danemark ont i) confirmé la nature clonale de la bactérie en retrouvant les deux grandes divisions phylogénétiques, et ii) mis en évidence un plus grand polymorphisme dans les populations de *S. agalactiae* circulant en Europe (Helmig *et al.*, 1993, Quentin *et al.*, 1995). En France, les souches invasives néonatales étaient distribuées dans les deux divisions phylogénétiques, entre tous les ET de la division I et les électrophorétypes ET11 et ET12 de la division II (**Figure 3**), laissant penser que plus d'évènements génétiques avaient modelé les populations de *S. agalactiae* en Europe (Quentin *et al.*, 1995).

2.5.3. Structure des populations de *S. agalactiae* établie par les méthodes génotypiques

2.5.3.1. Restriction enzymatique de l'ADN chromosomique : technique de Restriction Fragment Length Polymorphism (RFLP)

La première étude d'épidémiologie moléculaire utilisant la restriction du génome total par des endonucléases date de 1989 (Denning *et al.*, 1989). Cette étude a montré que certains profils étaient retrouvés dans des zones géographiques distinctes, ce qui suggérait la dissémination possible d'un nombre limité de clones (Denning *et al.*, 1989). L'exploitation phylogénétique du RFLP réalisée avec l'enzyme *Hind*III a identifié deux lignées à risque pour les infections précoces néonatales au Japon, appelées RDP Ia-3 et RDP III-3 (RDP pour Restriction Digest Pattern) (Nagano *et al.*, 1991). Ces deux RDP représentaient des lignées phylogénétiquement éloignées au sein des sérotypes Ia et III (Nagano *et al.*, 1991). Par la suite, la démarche a été de simplifier les profils RFLP par l'hybridation sélective avec des sondes spécifiques afin i) de déterminer des marqueurs génétiques de lignée, et ii) de déterminer des corrélations avec des gènes de virulence (Bohnsack *et al.*, 2002, Hauge *et al.*, 1996, Takahashi *et al.*, 1998, Takahashi *et al.*, 2002). Parmi les gènes cibles, les opérons codant les ARN ribosomaux (ARNr) ont été fréquemment choisis ; la ribotypie a identifié les mêmes groupes à risque que le MLEE et confirmé l'homogénéité génétique de la lignée à haut potentiel invasif (Chatellier *et al.*, 1996). L'ajout de six autres sondes ciblant notamment des facteurs de virulence a permis de subdiviser les grandes divisions en six lignées phylogénétiques lesquelles étaient congruentes avec la distribution des sérotypes ; ceci laissait supposer une origine génétique propre à chaque sérotype (Hauge *et al.*, 1996). Au sein du sérotype III, il est clairement apparu que la lignée hautement invasive RDP III-3 présentait des gènes ou des propriétés particulières lui conférant un caractère hautement virulent (Bohnsack *et al.*, 2002, Takahashi *et al.*, 1998).

2.5.3.2. Par électrophorèse en champ pulsé : technique de Pulsed Field Gel Electrophoresis (PFGE)

Le PFGE permet d'obtenir une diminution des fragments à analyser grâce à l'utilisation d'une endonucléase à faible fréquence de coupure (Gordillo *et al.*, 1993). Les fragments, en petit nombre mais de grande taille, sont séparés par une technique d'électrophorèse utilisant un champ électrique d'orientation variable qui permet la migration de fragments de grande taille. Cette méthode, qui présente un grand pouvoir discriminant, est particulièrement adaptée aux études

d'épidémiologie locale telles que l'investigation des infections nosocomiales. Actuellement toujours appliquée à *S. agalactiae*, le PFGE a permis i) d'expliquer des phénomènes localisés, comme l'émergence clonale du sérotype V aux Etats-Unis (Elliott *et al.*, 1998), et ii) de mettre en évidence la diversité génétique des isolats circulant dans un pays ou provenant de sites anatomiques particuliers (Bergseng *et al.*, 2009, Bidet *et al.*, 2003, Moyo *et al.*, 2002, Oliveira *et al.*, 2006, Thong *et al.*, 2004, van der Mee-Marquet *et al.*, 2009, von Both *et al.*, 2008). Le PFGE, qui n'est pas la méthode la plus adaptée à l'étude de la structure des populations, n'a pas permis de retrouver les deux grandes sous-populations de *S. agalactiae*. Il a néanmoins conforté la très grande homogénéité de la lignée néonatale hautement invasive correspondant au clone ET1 en MLEE et RDP III-3 en RFLP (Rolland *et al.*, 1999).

2.5.4. Structure des populations de *S. agalactiae* établie par les méthodes nucléotidiques

2.5.4.1. Amplification aléatoire de différentes parties du génome : technique de Randomly Amplified Polymorphic DNA (RAPD)

De même que le PFGE ou le RFLP, le RAPD explore le génome bactérien dans son ensemble. Cette méthode repose sur l'utilisation de courtes amorces de séquences aléatoires qui permettent l'amplification de plusieurs fragments de taille variable (Caetano-Anolles *et al.*, 1991). Appliqué à *S. agalactiae*, le RAPD a confirmé l'existence des trois lignées à plus haut risque infectieux pour le nouveau-né précédemment définies par MLEE en France (Chatellier *et al.*, 1997, Quentin *et al.*, 1995).

2.5.4.2. Amplification de fragments spécifiques de lignées

Les approches génomiques globales (RAPD, PFGE ou RFLP) ont mis en évidence des fragments spécifiques de lignées grâce à l'analyse différentielle des profils obtenus dans les populations de *S. agalactiae* (Chatellier *et al.*, 1996, Rolland *et al.*, 2002, van der Mee-Marquet *et al.*, 2006). Ainsi, l'amplification spécifique de clusters de gènes codant des ARN de transfert (ARNt) localisés en 3' des opérons des ARNr a permis d'identifier deux des trois lignées à haut risque infectieux pour le nouveau-né en France (Rolland *et al.*, 2002). De plus, la présence

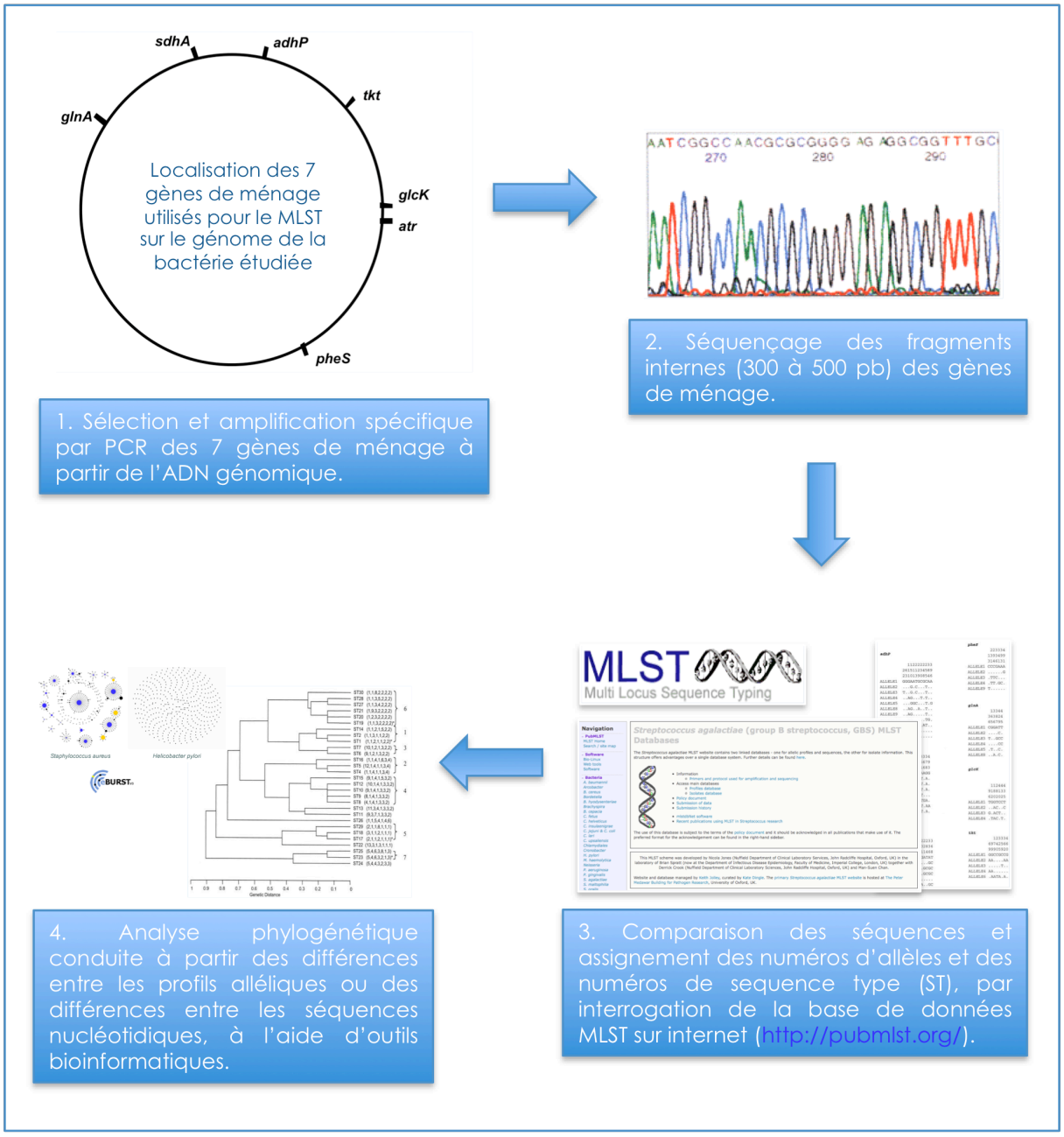


Figure 4. Principe du multilocus sequence typing (MLST).

de la séquence d'insertion IS*1548* dans le gène *hylB*, et de l'intron du groupe I GBSi1 dans le fragment intergénique *scpB-lmb*, identifie respectivement les souches ET12 (définies en MLEE en France) et les souches RDP III-3 définies en RFLP (correspondant en MLEE à l'ET1 aux Etats-Unis ou à la division I en France) (Bohnsack *et al.*, 2002, Rolland *et al.*, 1999, Takahashi *et al.*, 2002). Enfin, trois fragments d'ADN prophagiques sont significativement associés aux souches à haut risque pour le nouveau-né (van der Mee-Marquet *et al.*, 2006). Au total, il est apparu que des éléments génétiques particuliers, associés à la mobilité, étaient des marqueurs de lignées, le corollaire étant que ces éléments pouvaient jouer un rôle direct dans le potentiel infectieux de *S. agalactiae* (Bohnsack *et al.*, 2002, Rolland *et al.*, 1999, Takahashi *et al.*, 2002, van der Mee-Marquet *et al.*, 2006).

2.5.4.3. Amplification-séquençage pour l'étude du polymorphisme unilocus

L'avènement du séquençage automatisé a ouvert un nouveau champ d'étude : l'analyse du polymorphisme nucléotidique des gènes, dont l'exploitation a permis des avancées en phylogénie bactérienne, notamment avec l'analyse du gène de l'ARNr 16S. Chez *S. agalactiae*, l'analyse du gène *infB*, codant un facteur d'initiation de la traduction, a mis en évidence six allèles (pour la portion de gène analysée) correspondant aux six lignées phylogénétiques précédemment décrites (Hauge *et al.*, 1996, Hedegaard *et al.*, 2000).

2.5.4.4. Amplification-séquençage pour l'étude du polymorphisme multilocus : technique de Multilocus Sequence Typing (MLST)

2.5.4.4.1. Généralités

L'automatisation des méthodes de séquençage a permis d'envisager une approche multilocus (appelée MLST), utilisée pour la première fois chez *Neisseria meningitidis* (Maiden *et al.*, 1998). Cette méthode est proche conceptuellement du MLEE, mais l'identification des allèles de gènes de ménage se fait directement par séquençage nucléotidique d'un fragment interne (entre 400 et 500 pb) des gènes. Le principe repose sur le fait que chaque séquence unique est affectée d'un numéro d'allèle (**Figure 4**) ; la combinaison des numéros d'allèle de chaque *locus* donne un profil allélique appelé sequence type (ST) ; chaque profil allélique unique

est affecté d'un numéro de ST. La recherche des numéros d'allèle et de ST est réalisée en ligne sur le site internet dédié au système MLST d'une espèce bactérienne donnée. Si de nouveaux allèles et/ou ST sont trouvés, le modérateur du système MLST attribue les numéros après vérification des séquences.

Le MLST permet de détecter un plus grand nombre d'allèles par *locus* par rapport au MLEE. En effet, toutes les variations nucléotidiques, y compris celles qui n'entraînent pas de modification des séquences d'acides aminés, sont prises en compte. En revanche, le nombre de *loci* étudiés (généralement sept) est moindre qu'en MLEE. Les progrès majeurs apportés par le MLST sont une très bonne reproductibilité et la possibilité de consulter librement des banques de données, permettant une comparaison des séquences dans le monde entier.

2.5.4.4.2. L'épidémiologie de *S. agalactiae* à l'échelle mondiale

Cinq ans après la première publication sur le MLST de *N. meningitidis*, un système MLST adapté à *S. agalactiae* a été créé (Jones *et al.*, 2003). Depuis, de nombreux travaux ont été publiés, concernant tous les continents et de très nombreux pays comme l'Amérique du Nord (Etats-Unis et Canada) (Bohnsack *et al.*, 2008, Davies *et al.*, 2004, Lin *et al.*, 2006, Manning *et al.*, 2008, Manning *et al.*, 2009, Springman *et al.*, 2009), l'Europe (Royaume-Uni, France, Allemagne, Italie, Portugal, et Suède) (Eickel *et al.*, 2009, Gherardi *et al.*, 2007, Jones *et al.*, 2003, Jones *et al.*, 2006b, Luan *et al.*, 2005, Martins *et al.*, 2007, Poyart *et al.*, 2008, van der Mee-Marquet *et al.*, 2008), le Proche-Orient (Israël) (Marchaim *et al.*, 2006), l'Afrique (Sénégal et la République Centrafricaine) (Brochet *et al.*, 2009a), et l'Australasie (Sun *et al.*, 2005).

La base de données mondiale (http://pubmlst.org/sagalactiae/) recense à ce jour plus de 400 ST et constitue une ressource importante pour l'épidémiologie mondiale (Jones *et al.*, 2003). Cinq grands groupes clonaux, appelés complexes clonaux (CC) sont retrouvés à travers le monde : CC17, CC19, CC2, CC23 et CC12 (Bohnsack *et al.*, 2008, Manning *et al.*, 2009, Springman *et al.*, 2009, Jones *et al.*, 2003, Jones *et al.*, 2006b, Luan *et al.*, 2005, Marchaim *et al.*, 2006, Brochet *et al.*, 2009a). La diffusion mondiale de ces clones interroge sur la cause de leur expansion. L'analyse de la diversité génétique des souches de colonisation a mis en évidence un tropisme préférentiel de certains ST pour des sites anatomiques particuliers (van der Mee-Marquet *et al.*, 2008). Seul le tractus génital présente un panel de ST identiques à

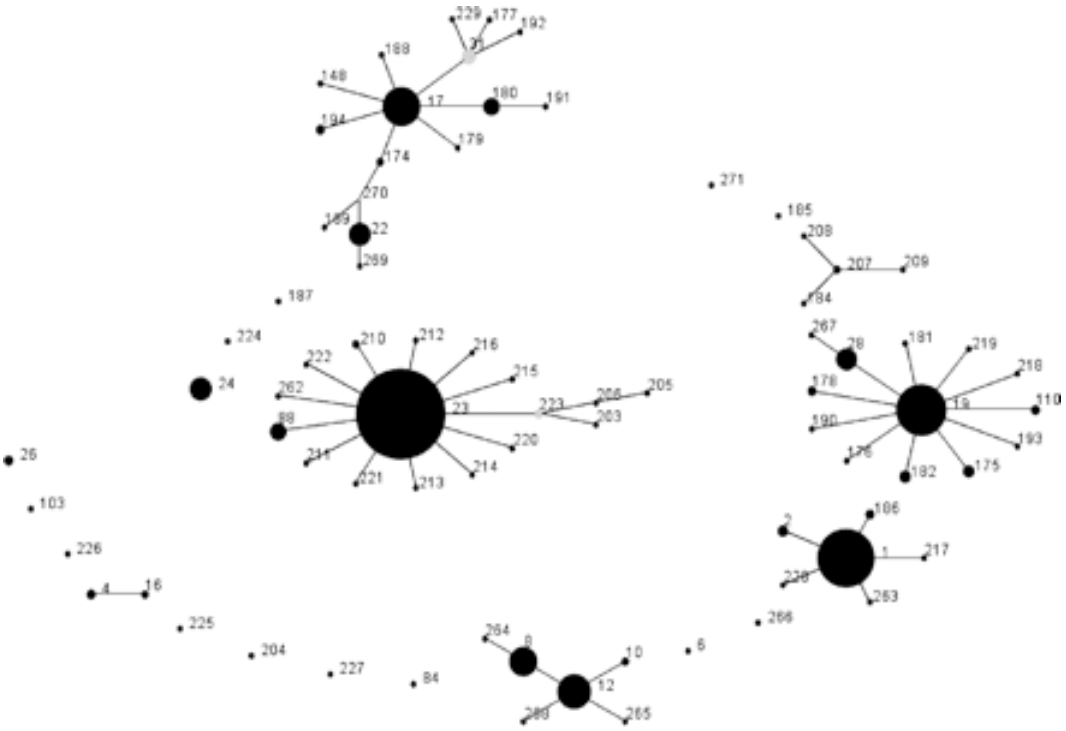


Figure 5. Représentation graphique en « snap-shot », obtenue avec le logiciel eBURST, représentant la population de *S. agalactiae* sous forme de complexes clonaux (CC). Chaque point central (17, 23, 19, 1 et 12) correspond au sequence type (ST) fondateur hypothétique du CC. La taille du point central est proportionnelle au nombre de souches de ce ST. Les autres ST rattachés au ST fondateur sont des variants unilocus ou bilocus du ST fondateur et appartiennent au CC. On dénombre cinq CC : CC17, CC19, CC12, CC1 et CC23 [image de Bohnsack *et al.* (Bohnsack *et al.*, 2008)].

ceux retrouvés dans l'épidémiologie mondiale, ce qui suggère que la transmission sexuelle pourrait jouer un rôle dans la dissémination de la bactérie (van der Mee-Marquet *et al.*, 2008). Quelle que soit l'incidence des infections néonatales à *S. agalactiae* dans les régions où les études ont été réalisées, toutes identifient le clone ST-17 comme un clone hautement invasif responsable d'infections néonatales (Bisharat *et al.*, 2005, Bohnsack *et al.*, 2008, Jones *et al.*, 2003, Luan *et al.*, 2005, Manning *et al.*, 2009, Poyart *et al.*, 2008). Le ST-17 serait préférentiellement associé aux infections néonatales tardives (Manning *et al.*, 2009, Poyart *et al.*, 2008), et aux méningites (Manning *et al.*, 2009).

2.5.4.4.3. Apport du MLST à l'étude de la structure des populations de *S. agalactiae*

La structure des populations de *S. agalactiae* établie par MLST retrouve les résultats des techniques historiques avec l'identification de deux grandes lignées de sérotype III, ST-17 et ST-19 (Jones *et al.*, 2003). Le ST-17 correspondrait au clone hyperinvasif ET1 identifié en MLEE, et correspondrait à l'homologue de RDP III-3 identifié en RFLP (Musser *et al.*, 1989, Nagano *et al.*, 1991, Jones *et al.*, 2003). Contrairement à ST-17, le potentiel infectieux de ST-19 ne fait pas consensus : certaines études le rattachent préférentiellement aux infections néonatales précoces (Manning *et al.*, 2009), alors que d'autres travaux le relient au portage asymptomatique (Jones *et al.*, 2003). Un troisième clone humain, ST-23, est également discuté en fonction des études en raison de sa position incertaine entre commensal préférentiel et clone néonatal invasif (Martins *et al.*, 2007, Jones *et al.*, 2003).

Des logiciels ont été développés spécifiquement pour exploiter les données du MLST. Le logiciel eBURST permet notamment d'exploiter les profils alléliques en retrouvant l'ancêtre hypothétique fondateur d'un CC (**Figure 5**) (Feil *et al.*, 2004). Le logiciel SplitsTree permet notamment de représenter la population sous forme de réseau et de mettre ainsi en évidence les recombinaisons existant entre les CC (Huson & Bryant, 2006). Ces analyses, impossibles à réaliser avec les résultats du MLEE par exemple, ouvrent des perspectives intéressantes pour la compréhension de l'évolution de la bactérie. C'est ainsi que l'hypothèse d'ancestralité bovine du clone hyperinvasif ST-17 a été pour la première fois proposée avec le MLST (Bisharat *et al.*, 2004).

B. LES ÉLÉMENTS GÉNÉTIQUES MOBILES

1. Généralités sur les éléments génétiques mobiles

1.1. Historique

Les génomes de tous les organismes ont la capacité de subir des réarrangements qui vont permettre de modifier ou diversifier (par acquisition ou délétion) le pool de leurs gènes et/ou le niveau de leur expression. Ces mécanismes de recombinaison génétique sont essentiels à l'évolution des espèces. La transposition est le mécanisme capable de générer le plus de diversité. Il n'y a en effet aucune similitude de séquence entre les sites échangés, contrairement à la recombinaison homologue. Les premiers éléments transposables ont été identifiés il y a près de 60 ans par Barbara McClintock (Prix Nobel de Médecine en 1983) dans le cadre d'études sur des mutations instables du maïs (Mc Clintock, 1950). Chez les bactéries, les éléments transposables ont été décrits pour la première fois dans les années 60. Les premiers travaux concernaient la découverte du bactériophage Mu responsable de multiples mutations par insertion chez *Escherichia coli* K12 (Taylor, 1963). Ultérieurement, les équipes de Shapiro et de Starlinger ont rapporté l'existence de mutants *E. coli* de l'opéron galactose et relié ces mutants spontanés à l'insertion de matériel génétique défini comme séquences d'insertion (Shapiro, 1969, Michaelis *et al.*, 1969). Finalement, la présence d'éléments transposables chez les bactéries a été reliée à la notion de transmission des résistances aux antibiotiques, notamment par rapprochement avec les plasmides R (comme résistance) décrits au Japon (Watanabe & Fukasawa, 1962). Avec l'évolution des connaissances, il est devenu de plus en plus évident que ces éléments génétiques mobiles faisaient partie intégrante du paysage génomique des procaryotes.

1.2. Définition et terminologie

Les éléments transposables sont des fragments d'ADN codant des enzymes ou d'autres protéines permettant la mobilité de l'ADN au sein des génomes (mobilité intracellulaire) ou entre les bactéries (mobilité intercellulaire) (Frost *et al.*, 2005). Appliquée aux procaryotes, cette définition concerne les trois grands mécanismes de mouvement génique intercellulaire (la transformation, la conjugaison et la transduction) et englobe par conséquent tous les éléments génétiques qui y

participent : les séquences d'insertion, les transposons, les plasmides, les phages et les introns du groupe II. Le terme éléments génétiques mobiles (EGM), souvent utilisé en bactériologie, regroupe l'ensemble de ces entités. Par analogie avec le transcriptome, le protéome etc …, le terme mobilome désigne la totalité des EGM présents dans un génome donné (Frost *et al.*, 2005, Koonin & Wolf, 2008, Siefert, 2009).

1.3. Nomenclature

Le premier système de nomenclature publié reconnaissait trois classes d'EGM : les séquences d'insertion, les transposons plus complexes et les épisomes doués d'auto-épissage (Campbell *et al.*, 1979). Deux ans plus tard, Kleckner *et al.* revoyait la nomenclature sur la base de la structure génétique, des propriétés mécaniques et des homologies de séquences ADN (Kleckner, 1981). Trois nouvelles classes d'EGM étaient identifiées : la classe I comprenant tous les éléments « IS-like », c'est-à-dire les séquences d'insertion (IS) *stricto sensu* et les transposons composites ; la classe II comprenant la famille de transposons Tn3 qui sont des transposons non composites appelés par la suite transposons unitaires ; et la classe III comprenant les bactériophages (Kleckner, 1981).

Les résultats des séquençages des génomes ont rapidement rendu obsolète cette classification. L'annotation des génomes a mis en évidence des éléments à l'organisation fonctionnelle très complexe, comme le transposon Tn7 (Orle & Craig, 1991), et des éléments avec des structures chimériques, comme les transposons conjugatifs qui partagent des propriétés de plasmide et de phage (Salyers *et al.*, 1995b). Ainsi, la liste des EGM s'est enrichie des îlots génomiques mobiles, des transposons mobilisables, des IStrons, des introns des groupes I et II, … (Roberts *et al.*, 2008). De fait, divers systèmes de dénomination ont été publiés, aboutissant à la complexification voire à la confusion des termes employés.

Au cours des dix dernières années, deux courants se sont opposés pour la classification des EGM. L'un, basé sur les caractéristiques fonctionnelles des EGM, prend en compte la nature mosaïque des EGM et les classe sur la présence de modules impliqués dans trois catégories de fonctions globales: la mobilité intracellulaire, la mobilité intercellulaire et la stabilité ; ces catégories sont ensuite subdivisées suivant leurs fonctions locales (transposase, recombinase, lyse cellulaire, contact cellulaire, toxine, antibiorésistance, etc …), elles-mêmes

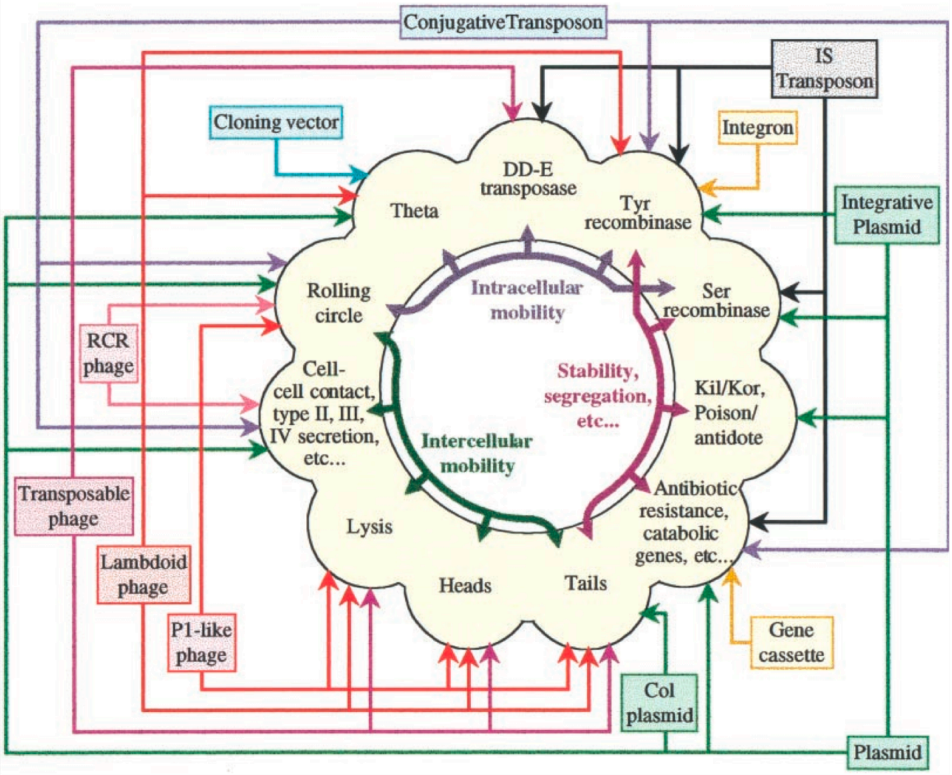


Figure 6. Représentation des EGM en modules fonctionnels (Toussaint & Merlin, 2002).

subdivisées en familles d'enzymes (Toussaint & Merlin, 2002, Osborn & Boltner, 2002) (**Figure 6**). L'autre système de classification repose strictement sur la nomenclature (en faisant abstraction de l'aspect fonctionnel) et vise à universaliser le système de dénomination des EGM déjà décrits et de ceux à venir (Roberts *et al.*, 2008). Le principe est le suivant : tous les EGM sont désignés par « Tn » suivi d'un numéro séquentiel écrit en italique ; un registre de tous les EGM décrits est accessible sur un site internet (http://www.ucl.ac.uk/eastman/tn/) et interrogeable en ligne pour demander l'affectation d'un nouveau numéro pour tout nouvel EGM découvert (Roberts *et al.*, 2008).

Aucun système de classification des EGM n'étant officiellement reconnu, nous avons fait le choix d'opter pour une présentation simplifiée des différents EGM identifiés chez *S. agalactiae* en fonction de leurs caractéristiques fonctionnelles et en respectant les dénominations parues dans les publications correspondantes.

1.4. Impact des éléments génétiques mobiles dans la virulence bactérienne

Par leur transposition, les EGM peuvent influer sur la virulence bactérienne, soit par leur action sur des gènes de virulence pré-existants, soit par leur participation à l'acquisition et à la dissémination de nouveaux gènes.

1.4.1. Eléments génétiques mobiles et gènes de virulence pré-existants

L'insertion ou l'excision d'un EGM directement dans un gène de virulence peut induire la modulation de l'expression du gène. Par exemple, l'expression du gène *stx* codant une shigatoxine chez les souches vérotoxiques d'*E. coli* est contrôlée par l'insertion/excision de l'IS*1203V* dans *stx* entraînant respectivement l'inactivation ou la récupération de la synthèse de la toxine (Kusumoto *et al.*, 2000). Ce mécanisme peut également provoquer des alternances antigéniques dénommées variations de phase. Chez *N. meningitidis*, le gène *siaA*, qui code une enzyme essentielle pour la synthèse de lipooligosaccharides de la capsule, est régulé par l'insertion/excision de l'IS*1301 (Hammerschmidt et al., 1996)*. La capsule favorise la survie et la dissémination de la bactérie, mais l'inactivation réversible de *siaA* permet à la bactérie de franchir les barrières mucosales et d'infester les méninges (Hammerschmidt *et al.*, 1996).

L'insertion d'EGM dans des séquences régulatrices des gènes influe également sur l'expression de la virulence. Ainsi, certaines souches de

Mycobacterium tuberculosis présentent une virulence accrue suite à l'insertion de l'IS*6110* dans la région promotrice du gène *phoP,* qui fait partie du système régulateur à deux composantes (Soto *et al.*, 2004). L'insertion de l'IS*6110* entraîne l'apparition d'une nouvelle région promotrice ayant un pouvoir transcriptionnel plus fort que le promoteur sauvage (Soto *et al.*, 2004).

A contrario, l'insertion d'un EGM peut conduire à une réduction de la virulence, comme pour *Ralstonia solanacearum,* chez lequel l'insertion de l'IS*Rso4* inactive le régulateur pléiotropique *phcA* et entraîne une hypovirulence de ce phytopathogène (Jeong & Timmis, 2000).

1.4.2. Eléments génétiques mobiles et acquisition de gènes de virulence

Les EGM, notamment les transposons composites et les îlots de pathogénicité, peuvent véhiculer des gènes impliqués dans la pathogénicité et disséminer entre les bactéries par transfert horizontal. Les îlots de pathogénicité ont été identifiés la première fois comme de grands fragments d'ADN perdus spontanément par le chromosome de souches d'*E. coli* uropathogènes (Hacker *et al.*, 1990). Ils sont présents dans le génome des souches pathogènes et le plus souvent absents des variants non pathogènes de la même espèce (Hacker *et al.*, 1997). Par exemple, chez des souches d'*E. coli* entéropathogènes, un îlot de pathogénicité spécifique, porteur notamment du gène codant l'intimine, un facteur d'adhérence, a été mis en cause dans les lésions des entérocytes de type « attachement-effacement » (Hacker *et al.*, 1997). Chez *Shigella flexneri*, l'opéron aérobactine est localisé dans l'îlot de pathogénicité SHI-2 et code un système de sidérophore permettant aux bactéries de capter le fer présent en quantité limitée dans l'environnement et nécessaire à la colonisation et à l'invasion de l'hôte (Vokes *et al.*, 1999). Le génome de *Yersinia pestis*, connu pour sa plasticité, est très riche en EGM et de multiples facteurs de virulence ont été apportés par les îlots de pathogénicité, qualifiés d' « îlots de haute pathogénicité » (Hacker *et al.*, 1997, Parkhill *et al.*, 2001). La faculté de perdre ou de récupérer ces îlots de pathogénicité serait la stratégie adoptée par la bactérie pour s'adapter à des hôtes très variés, comme les arthropodes ou les animaux à sang chaud (Hacker *et al.*, 1997). Au final, les îlots de pathogénicité contribueraient, non seulement à la virulence, mais également au pouvoir d'adaptation, aussi bien chez les bactéries à Gram négatif que les bactéries à Gram positif (Hacker *et al.*, 1997).

2. Les éléments génétiques mobiles dans le génome de *S. agalactiae*

La revue des EGM au sein d'une espèce comme *S. agalactiae* permet d'aborder les différentes caractéristiques de ces entités génétiques qui forment un groupe très hétérogène. Le génome de *S. agalactiae* comporte en effet un panel varié d'EGM, principalement localisés dans les zones de haute variabilité génétique mises en évidence par les cartes génomiques comparatives (Glaser *et al.*, 2002, Tettelin *et al.*, 2005, Tettelin *et al.*, 2002). Six séquences d'insertion (IS*861,* IS*1381,* IS*1548*, IS*Sag2*, IS*Sag1* et ISSa4), un intron du groupe II (GBSi1), des transposons composites (comme Tn*917*/Tn*916*-like et Tn*3706*) et plus d'une trentaine de transposons conjugatifs et dérivés ont été identifiés chez *S. agalactiae* (Achard & Leclercq, 2007, Achard *et al.*, 2005, Brochet *et al.*, 2008a, Brochet *et al.*, 2009b, Franken *et al.*, 2004, Franken *et al.*, 2001, Granlund *et al.*, 2001, Granlund *et al.*, 1998, Horaud *et al.*, 1996, Luan *et al.*, 2003, Puopolo *et al.*, 2007, Rubens *et al.*, 1989, Spellerberg *et al.*, 2000, Tamura *et al.*, 2000). Le séquençage systématique des génomes a par ailleurs révélé de nombreux gènes typiques de phages ou de plasmides souvent intégrés dans des blocs génomiques caractéristiques appelés îlots génomiques (Glaser *et al.*, 2002, Tettelin *et al.*, 2005, Tettelin *et al.*, 2002).

2.1. Les séquences d'insertion

2.1.1. Généralités

Le nombre d'IS bactériennes identifiées a augmenté de manière exponentielle au cours des dernières années du fait de la multiplicité des séquençages de génomes. La base de données internationale référençant les IS, l'ISFinder (www-is.biotoul.fr), contient plus de 2 500 séquences annotées ce qui ne correspond qu'à une très faible part des IS existant dans la masse de données que représentent les génomes séquencés (Siguier *et al.*, 2006).

2.1.1.1. Définition

Les séquences d'insertion (IS), connues aussi sous le terme d'éléments IS, sont les formes d'EGM les plus frustres. Elles se définissent comme des segments ADN de petite taille (de 700 pb à 2,5 Kb), phénotypiquement cryptiques, avec une organisation génétique simple, capables de s'insérer à de multiples sites dans une molécule cible (Mahillon & Chandler, 1998).



Figure 7. Structure d'une séquence d'insertion (IS). L'IS est encadrée par deux répétitions inversées (IR), l'IRL située en amont du gène de la transposase, et l'IRR située en aval. Le domaine I des IR est impliqué dans le clivage et le transfert du brin ADN; le domaine II dans l'attachement de la transposase. Le promoteur p de la transposase est généralement situé dans l'IRL.

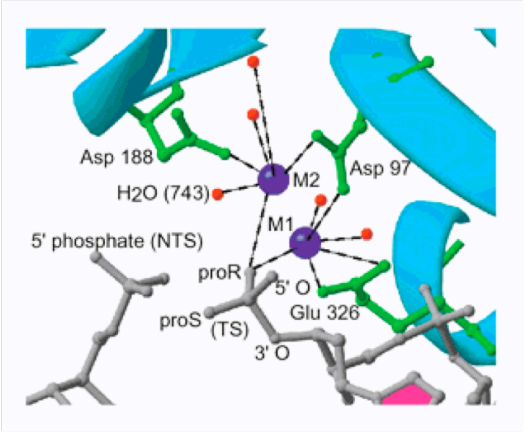


Figure 8. Site catalytique de la transposase. Représentation de l'interaction entre le motif DDE (Asp-Asp-Glu) (en vert), des deux ions Mn^{2+} (en violet foncé) et des molécules d'H_2O (en rouge) au niveau du site catalytique de la transposase Tn*5* (*image de Lovell et al., 2002*).

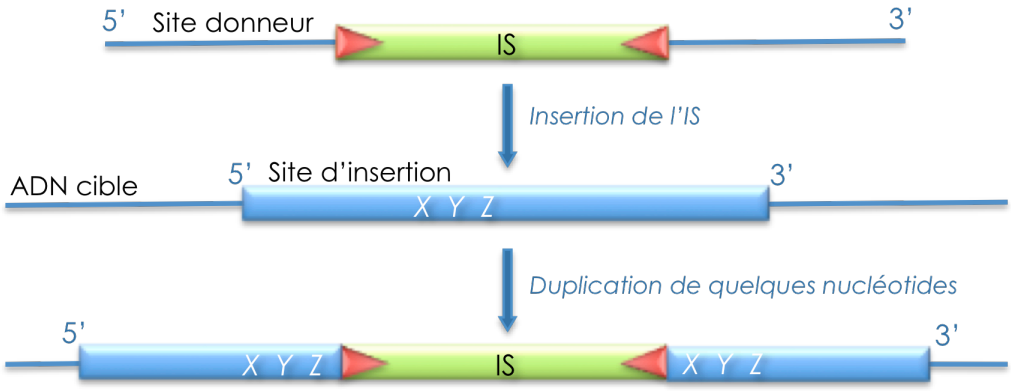


Figure 9. Duplication provoquée par l'insertion d'une IS. La transposase, encadrée par les séquences inversées répétées, reconnaît sa séquence cible sur la molécule d'ADN et vient s'y insérer. L'insertion entraîne la duplication des quelques nucléotides (X, Y, Z) adjacents au site d'insertion (*d'après Mahillon et Chandler, 1998*).

2.1.1.2. Nomenclature

Plusieurs systèmes de nomenclature coexistent. Le premier système adopté consistait à accoler à l'abréviation IS un nombre, attribué en fonction de son ordre d'apparition (IS*1*, IS*2*, IS*3*, etc …) (Lederberg, 1981). Pour exemple, l'IS*861* et l'IS*1548* de *S. agalactiae* correspondent à cette première nomenclature et leur numéro correspond à l'ordre chronologique de leur description (Granlund *et al.*, 1998, Rubens *et al.*, 1989). Ce système a été dépassé par l'ampleur du nombre d'IS découvertes. Un second système a consisté à accoler les initiales de la bactérie (première lettre du genre et les deux premières lettres de l'espèce) chez laquelle a été isolée l'IS (Mahillon & Chandler, 1998). L'IS*Sag1* et l'IS*Sag2* sont des exemples de cette nomenclature appliquée à *S. agalactiae* (Franken *et al.*, 2004, Franken *et al.*, 2001). Des dénominations « sauvages » sont par ailleurs retrouvées. Au final, le système incluant les initiales de l'espèce bactérienne semble avoir pris le pas sur les autres, notamment sous l'impulsion de la base de données internationale, l'ISFinder (http://www-is.biotoul.fr/is.html) (Siguier *et al.*, 2006).

2.1.1.3. Organisation typique d'une séquence d'insertion

En plus d'être de petite taille, les IS sont génétiquement compactes (**Figure 7**). Dans leur grande majorité, elles ne codent que des fonctions impliquées dans leur mobilité.

2.1.1.3.1. Les répétitions inversées terminales (IR)

La plupart des IS ont de courtes répétitions inversées à leurs extrémités (**Figure 7**). Ces IR ont une longueur comprise entre 10 et 40 pb. Par convention, l'IR située en amont du gène transposase est appelée IRL (Inverted Repeat Left) alors que celle située en aval est l'IRR (Inverted Repeat Right). Les IR comportent deux domaines. Le domaine I, qui comprend les 2 ou 3 dernières paires de bases de la séquence d'insertion, est impliqué dans le clivage et le transfert du brin permettant la transposition. Le domaine II, positionné dans l'IR, permet l'attachement de la transposase. Les promoteurs des IS sont fréquemment positionnés, au moins partiellement, dans l'IRL. Cet agencement permet une autorégulation de la synthèse de la transposase (Mahillon & Chandler, 1998).

Tableau I. Caractéristiques majeures des principales familles de séquences d'insertion.

Famille	Sous-groupes	Longueur de l'IS (pb)	Longueur des répétitions de l'ADN cible (pb)	IR	Commentaires
IS*1*	–	770	9 (8–11)	O	intégrase de phage λ ?
IS*1595*	IS*Pna2*	1000-1150	8	O	
	IS*H4*	1000	8	O	
	IS*1016*	700-745	7-9	O	
	IS*1595*	900-1100	8	O	
	IS*Sod11*	1080-1200	8	O	
	IS*Nw1*	3450-7900	8	O	
	IS*Nha5*	3450-7900	8	O	
IS*3*	IS*2*	1300–1350	5	O	DDE
	IS*3*	1200–1300	3 (4)		
	IS*51*	1300–1400	3 (4)		
	IS*150*	1400–1550	3–5		
	IS*407*	1200–1250	4		
IS*481*	–	950-1300	4-15		
IS*4*	–	1300–1950	9–12	O	DDE
IS*701*	–	1400-1550	4	O	DDE
IS*H3*	–	1225-1500	4-5	O	DDE
IS*1634*	–	1500-2000	5-6	O	DDE
IS*5*	IS*5*	1100–1350	4	O	DDE
	IS*427*	800–1000	2–3		
	IS*903*	1000–1100	9		
	IS*1031*	850–950	3		
	ISH*1*	900–1150	8		
	ISL*2*	800–1100	2–3		
IS6	–	750–900	8	O	DDE
IS*21*	–	1950–2500	4 (5, 8)	O	DDE
IS*30*	–	1000–1250	2-3	O	DDE
IS66	–	2500–2700	8	O	
IS*91*	–	1500–1850	0	N	Protéine Rep associée
IS*110*	–	1200–1550	0	N	Similitudes avec les recombinases
IS*200*/IS*605*	–	700–2000	0	N	Organisation complexe
IS*607*	–	1700-2500	0	N	
IS*256*	–	1300–1500	8–9	O	Similitudes avec le DDE eucaryote
IS*630*	–	1100–1200	2	O	Similitudes avec le DDE eucaryote
IS*982*	–	1000	ND	O	DDE
IS*1380*	–	1380 – 1650	4	O	DDE
ISA*s1*	–	1200–1350	8	O	–
ISL*3*	–	1300–1550	8	O	–
Tn*3*			5	O	DDE

ND : non déterminée. Entre parenthèses, les longueurs des répétitions cibles plus rarement observées**. O** : présence ou **N** : absence des répétitions inversées terminales (IR). **DDE** : motif conservé, retrouvé dans la transposase (d'après Mahillon et Chandler, 1998).

2.1.1.3.2. Les transposases

La transposase est une enzyme codée généralement par un, deux, voire trois cadre ouvert de lecture (ORF pour open Reading frame) qui occupe presque la totalité de la séquence d'insertion. La transposase est l'élément clé dans la mobilité des EGM. Le site catalytique de cette activité est généralement composé par une triade d'acides aminés caractéristique appelée motif DDE (Asp Asp Glu) dont le rôle essentiel est de coordonner les cations divalents, indispensables à la catalyse (**Figure 8**) (Fayet *et al.*, 1990). Les séquences impliquées dans l'attachement à l'ADN sont localisées dans la région N-terminale et le domaine catalytique est situé en C-terminal de la protéine. Cette organisation permettrait l'interaction de la protéine naissante avec ses sites d'attachement sur l'IS, couplant ainsi l'expression et l'activité de l'enzyme. L'expression de la protéine sous la forme fonctionnelle nécessite parfois un décalage du cadre de lecture (Mahillon & Chandler, 1998). Certaines transposases ont la capacité de générer des formes multimériques nécessaires à leur activité.

2.1.1.3.3. Les répétitions cibles

Une des caractéristiques des IS est de provoquer la duplication directe de quelques nucléotides au niveau de la zone d'insertion (**Figure 9**). La séquence obtenue, qui varie de 2 à 14 pb, est caractéristique de l'IS. Une IS donne généralement une duplication de longueur fixe (Reimmann & Haas, 1990).

2.1.1.4. Classification : les familles de séquences d'insertion

Les IS ont été regroupées en 18 familles établies sur la base des homologies de séquence et sur leur organisation génétique (**Tableau I**) (Mahillon & Chandler, 1998). Celle-ci inclut la disposition des ORF codant pour la transposase, la longueur et la similitude des séquences inversées terminales (IR), et le nombre de pb de la séquence ADN cible. Si la plupart des familles majeures d'IS sont relativement homogènes, d'autres sont éclatées en différents sous-groupes (Mahillon & Chandler, 1998, Mahillon *et al.*, 1999).

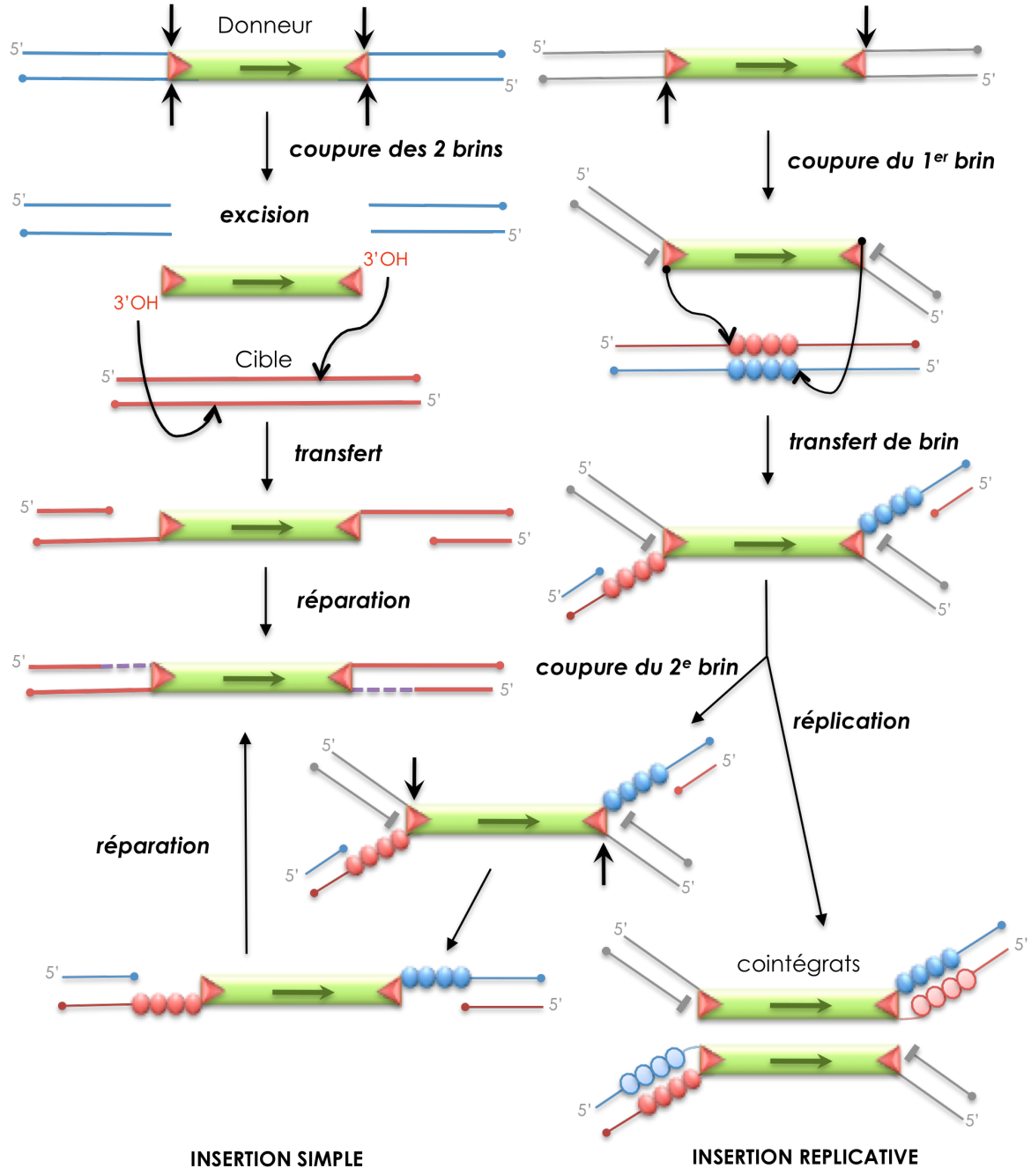


Figure 10. Modes de transposition des séquences d'insertion (IS). L'insertion simple (ou mode « couper-coller ») provient soit de la coupure des deux brins suivie du transfert de l'IS, soit de la coupure du premier brin suivie du transfert de l'IS puis de la coupure du second brin. Le mode réplicatif (ou mode « copier-colller ») suppose la coupure du premier brin, suivie du transfert et de la réplication de l'IS.

2.1.1.5. Le mécanisme réactionnel

La transposition se définit comme une réaction de recombinaison de l'ADN résultant de la translocation d'un segment d'ADN (l'IS ou le transposon) à partir d'un site donneur sur un site cible non homologue. Le mécanisme se déroule en 4 étapes. La réaction commence avec l'attachement de la transposase aux extrémités de l'IS. Il s'élabore alors un complexe actif ou transposome, qui comprend l'enzyme, les deux extrémités de l'IS, et éventuellement des protéines accessoires. Le recrutement de l'ADN cible est soit concomitant, soit consécutif à cette étape. Le transposome permet le clivage et le transfert des deux extrémités de l'IS dans l'ADN cible. Une fois que les extrémités de l'IS sont insérées dans l'ADN cible, l'IS dans son intégralité est prise en charge et intégrée dans l'ADN cible (Mizuuchi, 1992).

Il existe deux modes de transposition des IS (**Figure 10**). Le premier implique le clivage direct des deux brins d'ADN puis le transfert de l'IS dans l'ADN cible. Le deuxième mode implique un découplage dans le clivage des deux brins. Après clivage du premier brin, les extrémités 3'OH sont transférées sur l'ADN cible. Deux mécanismes sont alors possibles. Soit un second clivage intervient et l'IS se trouve libérée totalement du site donneur pour être insérée au site cible et fusionnée après réparation du site cible; soit l'IS est répliquée et les structures obtenues, appelées cointégrats, sont constituées de la molécule donneuse et de la molécule cible reliées par l'IS. Au total, la transposition est soit réplicative (mode dit « copier-coller »), soit simple c'est-à-dire non réplicative (mode dit « couper-coller »). Dans ce dernier cas, le site donneur perd la séquence d'insertion et il ne reste alors que la duplication comme signature génétique du passage de l'IS (Mahillon & Chandler, 1998).

2.1.1.6. La spécificité de cible

Le degré de spécificité des IS pour leur cible est variable. Certaines IS répondent à une séquence nucléotidique précise, d'autres à un simple dinucléotide (TA, par exemple), d'autres encore à la conformation de l'ADN, liée par exemple, à son degré de super enroulement (Mahillon *et al.*, 1999). Toutes bactéries confondues, le pouvoir de dissémination d'une IS est inversement proportionnel à sa spécificité.

2.1.2. Les séquences d'insertion chez *S. agalactiae*

Le génome de *S. agalactiae* recèle de nombreux gènes de transposases (Glaser *et al.*, 2002, Tettelin *et al.*, 2005, Tettelin *et al.*, 2002), dont certains correspondent à des IS, parfois fonctionnelles. Plusieurs IS ont été bien caractérisées : l'IS*861*, l'IS*1381*, l'IS*1548*, l'IS*Sag2*, l'IS*Sag1* et l'IS*Sa4*.

2.1.2.1. L'IS*861*

2.1.2.1.1. Caractéristiques générales

L'IS*861* est la première IS décrite chez *S. agalactiae* (Rubens *et al.*, 1989). Il s'agit d'une IS de 1442 pb, comportant des séquences inversées répétées de 22 et 26 pb (IR), deux ORF codant des protéines de 141 et 277 acides aminés. La séquence cible, ACA, est dupliquée lors de l'insertion. L'IS*861* appartient à la famille IS*3* et l'ORF2 montre 50% et 29% d'homologie respectivement avec l'IS*3* et l'IS*150* d'*E. coli* (**Tableau I**) (données de l'ISFinder : http://www-is.biotoul.fr/is.html).

2.1.2.1.2. Localisation génomique et impact sur le phénotype

Rubens *et al.* ont prouvé que l'IS*861* était fonctionnelle (Rubens *et al.*, 1989). Neuf copies de cette IS ont été identifiées dans le chromosome de la souche COH-1, dont une proche du locus *cps* codant la synthèse de la capsule (Rubens *et al.*, 1989). L'hypothèse alors envisagée était que l'insertion de l'IS*861* pouvait provoquer une surexpression de la capsule chez COH-1, souche hautement virulente car productrice d'une grande quantité d'antigènes capsulaires (Rubens *et al.*, 1989, Rubens *et al.*, 1987). Le positionnement exact de l'IS*861* n'ayant pas été défini précisément, nous avons réalisé une recherche par BLAST (http://www.ncbi.nlm.nih.gov/sites/entrez) dans le génome de la souche récemment séquencé (Tettelin *et al.*, 2005) ; 12 copies de l'IS*861* ont été retrouvées, dont une positionnée en amont de l'opéron *cps* au niveau du gène *LysR* (SAN1302) qui pourrait correspondre à un régulateur transcriptionnel de l'opéron *cps* (Koskiniemi *et al.*, 1998). L'impact de l'IS*861* à ce locus reste à préciser, même s'il paraît probable que son insertion dans *LysR* pourrait lever l'inhibition de la synthèse de la capsule comme l'avaient suggéré Rubens *et al.* (Rubens *et al.*, 1989).

Après avoir montré, au contraire de Rubens *et al.*, qu'il n'y avait pas de corrélation entre le phénotype capsulaire et la localisation et/ou le nombre de

copies de l'IS*861* (Sellin *et al.*, 1995), Sellins *et al.* ont mis en évidence une insertion de l'IS*861* dans le gène *cpsD* codant une galactosyltransférase essentielle pour l'assemblage de la capsule type III (Rubens *et al.*, 1993, Sellin *et al.*, 2000). La souche étudiée avait pour particularité d'être non typable de manière irréversible ce qui a été relié à l'insertion de l'IS*861* dans *cpsD* (Sellin *et al.*, 2000). Au final, ces résultats confortent l'hypothèse du rôle de l'IS*861* sur la virulence de *S. agalactiae*.

2.1.2.2. L'IS*1381*

2.1.2.2.1. Caractéristiques générales

L'IS*1381* a d'abord été décrite chez *Streptococcus pneumoniae* (Sanchez-Beato *et al.*, 1997). L'IS*1381* identifiée chez *S. agalactiae* partage 94% d'homologie avec l'IS de *S. pneumoniae* (Tamura *et al.*, 2000). Elle appartient à la famille IS*5,* mesure 900 pb de longueur, a des IR de 23 et 24 pb, et comporte deux ORF codant des protéines de 128 et 144 acides aminés (**Tableau I**) (données de l'ISFinder : http://www-is.biotoul.fr/is.html).

2.1.2.2.2. Localisation génomique et impact sur le phénotype

L'IS*1381* est présente dans le génome de *S. agalactiae* avec une à huit copies (Tamura *et al.*, 2000). Sa présence a été reliée à la fois à la virulence de *S. agalactiae* et à sa résistance aux antibiotiques. L'insertion de l'IS*1381* à proximité du gène *bac* entraîne l'expression de plus faibles taux de Bac, phénomène observé chez les souches de colonisation vaginale ; Bac est une protéine de surface C β-antigénique dont l'expression est augmentée chez les souches virulentes (Nagano *et al.*, 2006). L'insertion de l'IS*1381* dans le gène *cylA,* qui code un transporteur de type ATP-binding cassette (ABC) permettant l'exportation de la β-hémolysine, a été retrouvée chez des mutants naturels non-hémolytiques (Spellerberg *et al.*, 1999a). L'insertion de l'IS*1381* dans le gène *cpsE,* entraîne l'interruption de l'expression de la glycosyltransférase ; elle se retrouve chez les souches non typables (NT) (Ramaswamy *et al.*, 2006). Enfin, l'insertion de l'IS*1381* dans le gène atténuateur de l'*erm*(A) entraîne l'expression constitutive de la résistance croisée aux macrolides (Culebras *et al.*, 2005).

2.1.2.3. L'IS*1548*

2.1.2.3.1. Caractéristiques générales

L'IS*1548* a été décrite pour la première fois chez *S. agalactiae* par Granlund *et al.* qui ont retrouvé également cette IS chez *Streptococcus pyogenes* (Granlund *et al.,* 1998). L'IS*1548* appartient à la famille IS*As1*, comporte 1316 pb, deux ORF codant pour des protéines de 377 et 62 acides aminés, des IR de 20 et 24 pb et s'insère en provoquant une duplication de 14 pb (**Tableau I**) (données de l'ISFinder : http://www-is.biotoul.fr/is.html).

2.1.2.3.2. Localisation génotypique et impact sur le phénotype

L'IS*1548* a été détectée en premier lieu dans le gène *hylB* qui code la hyaluronidase (Granlund *et al.*, 1998, Sukhnanand *et al.*, 2005). L'insertion de l'IS*1548* dans *hylB* a été retrouvée préférentiellement chez des souches d'endocardite sans que le mécanisme physiopathologique n'ait pu être expliqué (Granlund *et al.*, 1998). Le qualificatif de facteur de virulence est d'ailleurs discuté pour la hyaluronidase car il a été montré que cette enzyme n'était pas requise pour les infections chez l'homme ; en effet, toutes les souches humaines qui portent l'IS*1548* ont perdu leur activité hyaluronidase et cela quelque soit leur potentiel invasif (Sukhnanand *et al.*, 2005). En plus de l'insertion dans *hylB*, au moins deux à sept autres sites d'insertion existent dans le génome de *S. agalactiae ;* en effet, entre trois et huit copies de l'IS ont été mises en évidence par Southern blot (Granlund *et al.*, 1998). Un des sites préférentiels d'insertion est situé entre le gène *scpB* codant la C5a-peptidase et le gène *lmb* codant une protéine d'attachement cellulaire, exactement à -9 pb du promoteur -35 du gène *lmb* (voir § 2.1.3.2) (Granlund *et al.*, 2001, Granlund *et al.*, 1998). L'impact de cette insertion n'a pas été déterminé mais on peut spéculer une augmentation de l'activité transcriptionnelle du gène *lmb* qui code un facteur de virulence avéré (Tenenbaum *et al.*, 2007). Par ailleurs, il est intéressant de souligner que les autres IS de la famille de l'IS*1548*, l'IS*As1*, sont aussi associés à des gènes codant des composants de la surface cellulaire de leurs hôtes respectifs (Mahillon & Chandler, 1998). Enfin, l'IS*1548* a été retrouvée insérée dans le gène *cpsD* de l'opéron capsulaire chez des souches NT (Sellin *et al.*, 2000).

2.1.2.4. L'IS*Sa4*

2.1.2.4.1. Caractéristiques générales

L'IS*Sa4* a été décrite par Spellerberg *et al.* qui retrouvaient entre 3 et 15 copies de l'IS chez *S. agalactiae* (Spellerberg *et al.*, 2000). Il s'agit d'une IS de 963 pb, comportant des IR de 25 pb, un ORF codant une transposase de 287 acides aminés qui comporte un motif DDE typique et dont l'insertion entraîne une duplication de 9 pb. L'IS*Sa4* appartient à la famille IS*982* décrite au départ chez *Lactococcus lactis* (**Tableau I**) (données de l'ISFinder : http://www-is.biotoul.fr/is.html).

2.1.2.4.2. Localisation génomique et impact sur le phénotype

L'IS*Sa4* a été découverte par son impact sur l'activité hémolytique de *S. agalactiae*. Son insertion dans le gène *cylB,* qui code le domaine membranaire du transporteur ABC de l'hémolysine, entraîne l'inactivation de l'activité β-hémolytique (Spellerberg *et al.*, 2000). Neuf autres sites d'insertion ont été mis en évidence (dans des gènes de ménage, des régions intergéniques ou des gènes codant des protéines hypothétiques) ; ces sites d'insertion sont pour la plupart spécifiques de souche et sans lien avec le phénotype (Dmitriev *et al.*, 2004).

2.1.2.5. L'IS*Sag2*

2.1.2.5.1. Caractéristiques générales

Il s'agit d'une IS de 1220 pb, comportant des IR de 33 et 45 pb, deux ORF codant des protéines de 96 et 259 acides aminés qui comportent un motif DDE typique. L'IS*Sag2* appartient à la famille IS*3* et à la sous-famille IS*150* et présente 97% d'homologie avec l'IS*Sdy2* identifiée chez *Streptococcus dysgalactiae* (**Tableau I**) (données de l'ISFinder : http://www-is.biotoul.fr/is.html) (Franken *et al.*, 2001).

2.1.2.5.2. Localisation génomique et impact sur le phénotype

L'IS*Sag2* a été découverte dans le génome de *S. agalactiae* par Franken *et al.* qui ont révélé la présence de deux copies de l'IS encadrant le fragment *scpB-lmb* déjà connu pour accueillir l'IS*1548* ou l'intron GBSi1 (voir § 2.1.2.3 et § 2.1.6) (Franken *et al.*, 2001). L'ISSag2, qui encadre les gènes *scpB* et *lmb,* serait en fait

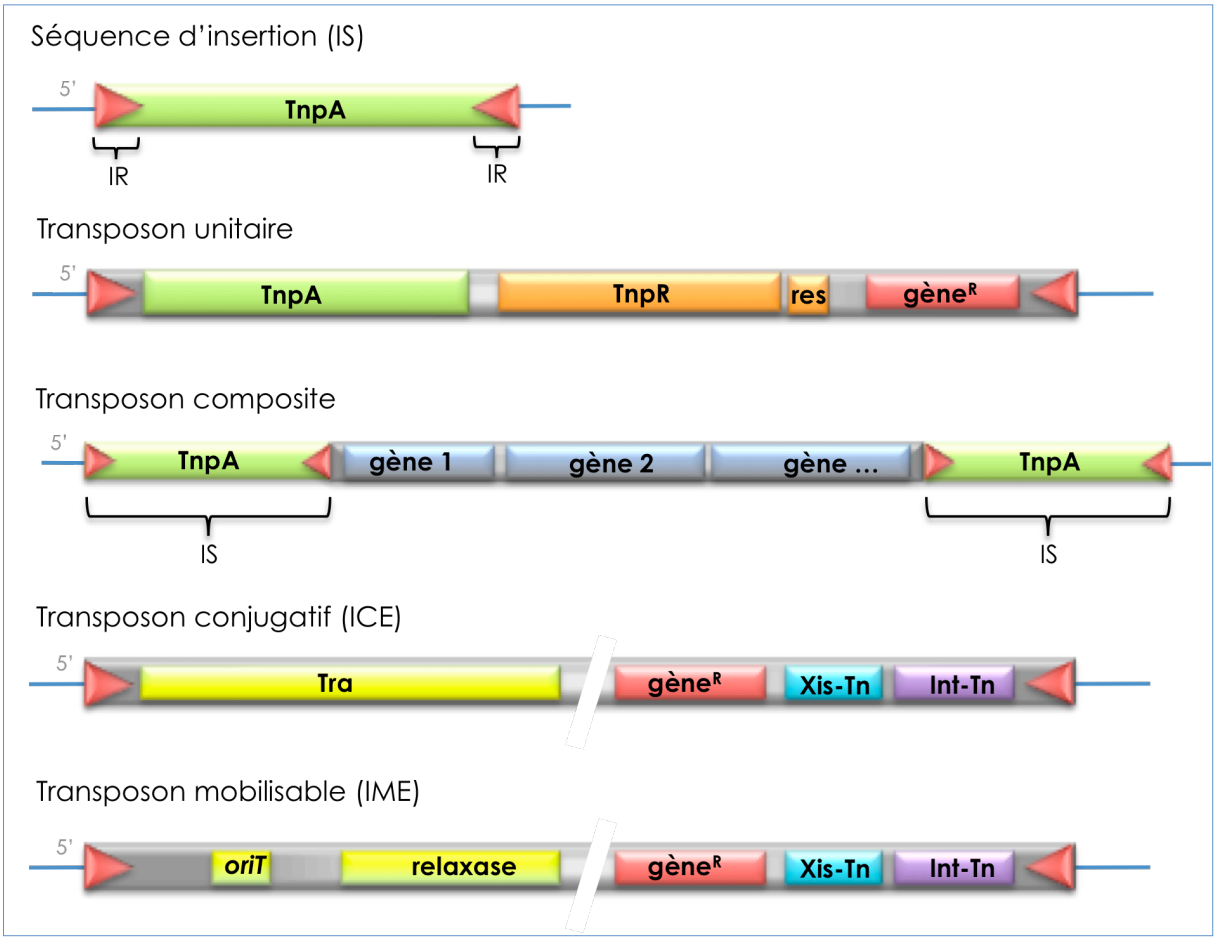


Figure 11. Les différentes catégories de transposons rencontrées chez *S. agalactiae*.
IR, inverted repeat ; TnpA, transposase ; TnpR, résolvase ; res, site d'action de la résolvase ; ICE, intégrative and conjugative élément ; Tra, fonction catalysant le transfert par conjugaison ; Xis-Tn, excisionase ; Int-Tn, intégrasse ; *oriT,* origine de transfert.

partie intégrante d'un transposon composite dont l'assemblage aurait eu lieu chez *S. dysgalactiae* et qui aurait été acquis ensuite par transfert horizontal par *S. agalactiae* (Broker & Spellerberg, 2004, Franken *et al.*, 2001). L'IS*Sag2* aurait ainsi contribué à l'acquisition de facteurs indispensables à la colonisation des écosystèmes humains car 100% des souches humaines portent ce transposon mais pas pour les écosystèmes bovins car la grande majorité des souches bovines ne portent pas ce transposon ou alors sous forme tronquée (Franken *et al.*, 2001).

2.1.2.6. L'IS*Sag1*

2.1.2.6.1. Caractéristiques générales

Il s'agit d'une IS de 1251 pb, comportant des IR de 37 pb et deux ORF. Le premier ORF code une transposase putative qui comporte un motif de type « leucine zipper ». Le second ORF présente un codon stop interrompant le motif DDE. L'IS*Sag1* appartient à la famille IS*3* et présente des similitudes avec l'IS*1627*w1 identifiée chez *Bacillus anthracis* (**Tableau I**) (Franken *et al.*, 2001).

2.1.2.6.2. Localisation génomique et impact sur le phénotype

Chez les souches humaines, toutes les souches IS*Sag1*+ ne comportent qu'une seule copie située dans le transposon composite *scpB-lmb* en aval du gène *lmb* (Franken *et al.*, 2004). Chez les souches bovines qui n'ont pas le transposon, l'IS*Sag1* n'est pas présent dans le génome (Franken *et al.,* 2004). Du fait du codon stop présent dans un des ORF, la transposase serait non fonctionnelle et l'IS*Sag1* incapable de transposer (Franken *et al.*, 2004). Aucun impact potentiel sur le phénotype de *S. agalactiae* n'a été rapporté à ce jour.

2.1.2.7. *tISSag10*, à la frontière entre IS et transposon

2.1.2.7.1. Caractéristiques générales des transposons unitaires

Dans la définition classique des IS, elles ne codent que pour des fonctions liées à leur mobilité. Le gène de la transposase couvre généralement la totalité de l'IS avec très peu ou pas d'ADN entre les ORF de la transposase et les séquences inversées répétées (**Figure 11**). Des éléments « IS-like » portant des ORF codant des protéines autres que la transposase ont été identifiés (Siguier *et al.*, 2009). Ces

éléments sont dénommés transposons unitaires ou transposons non composites. Parmi les ORF additionnels, il y a fréquemment un ORF codant une résolvase, enzyme qui permettra de terminer la réaction de transposition initiée par la transposase en catalysant une étape de recombinaison à un site spécifique (res) porté par le transposon.

Avec l'augmentation du nombre d'IS identifiées, il est donc apparu que le système de classification était de moins en moins adapté à certaines familles d'IS et que la délimitation entre les IS et les transposons unitaires devenait incertaine. Ainsi, il a été proposé d'introduire le préfixe « *t* » pour distinguer les IS *stricto sensu* des transposons unitaires (Siguier *et al.*, 2009). Chez *S. agalactiae* par exemple, l'IS*Sag10* vient d'être ré-annotée en « *t*IS*Sag10* » pour signifier la présence d'un gène passager (Siguier *et al.*, 2009).

2.1.2.7.2. Caractéristiques particulières de *t*IS*Sag10*

*t*IS*Sag10* appartient à la famille IS*1595* famille nouvellement définie et au groupe IS*Pna2* (**Tableau I**) (Siguier *et al.*, 2009). Elle comporte 1724 pb, des IR de 22 et 25 pb et entraîne une duplication de 8 pb par son insertion (données de l'ISFinder : http://www-is.biotoul.fr/is.html). Elle a été découverte insérée dans le locus *cps* d'une souche de *S. agalactiae* résistante aux lincosamides mais sensible aux macrolides et streptogramines (souche UCN36) (Achard *et al.,* 2005). La présence, en aval de l'ORF de la transposase, d'un ORF codant une 0-lincosamide nucléotidyl transférase, apportait l'explication de cette résistance peu commune chez *S. agalactiae* (Achard *et al.*, 2005). Cette *t*IS présente des similitudes génétiques et fonctionnelles avec la *t*IS*Cpe8* décrite chez *Clostridium perfringens* ce qui peut faire soupçonner un transfert horizontal entre *S. agalactiae* et *C. perfringens* (Siguier *et al.*, 2009, Lyras *et al.*, 2009).

L'autre particularité de cet EGM est qu'il contient un site *oriT* fonctionnel au sein de son gène de résistance *lnu*(C) (Achard & Leclercq, 2007). *t*IS*Sag10* (également appelé MTnLNU ou MTnSag1) serait donc mobilisable par un transposon conjugatif de type Tn*916* qui permettrait l'intégration dans un site cible de la forme préalablement excisée et circularisée du *t*IS (Achard & Leclercq, 2007). Cet EGM mobilisable pourrait être à l'origine d'un transfert entre souches de la résistance aux lincosamides (Achard & Leclercq, 2007).

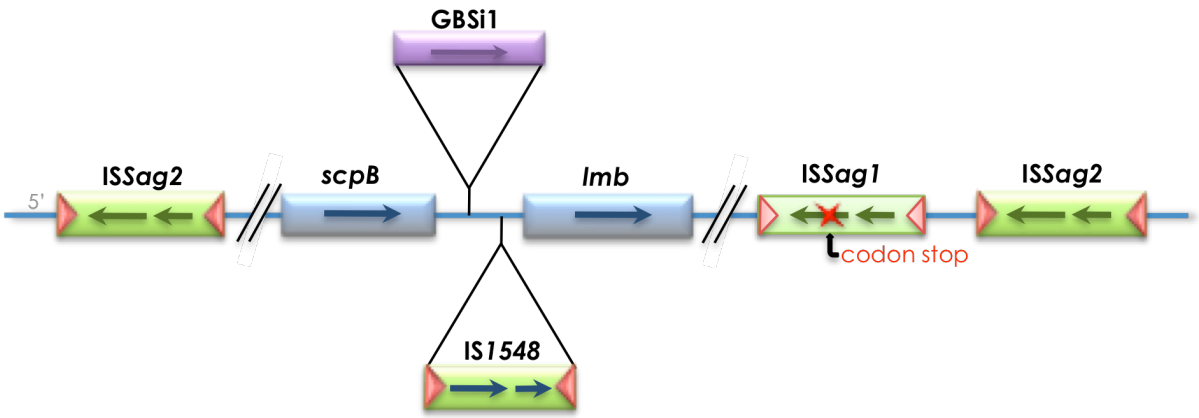


Figure 12. Structure du transposon composite *scpB-lmb* de *S. agalactiae*.

2.2. Les transposons composites

2.2.1. Caractéristiques générales

Une particularité importante des IS est qu'elles peuvent collaborer entre elles lors de leur transposition. Cette propriété permet de mobiliser toute séquence d'ADN comprise entre deux copies d'une même IS, en orientation directe ou inverse. La structure déplacée est appelée transposon composite (**Figure 11**). La catalyse nécessaire à l'excision et à l'incision provient des transposases des IS (Roberts *et al.*, 2008). Les transposons composites sont un important moyen de dissémination des gènes de résistance. La collaboration entre deux copies d'une même IS peut provoquer des remaniements multiples et complexes des génomes en fonction des extrémités impliquées et de la localisation du site cible.

2.2.2. Les transposons composites de *S. agalactiae*

Comme cela a été souvent le cas chez de nombreuses espèces, un des premiers transposons composites identifiés chez *S. agalactiae* a été signalé chez une souche présentant une résistance particulière aux antibiotiques, en l'occurrence de haut niveau à la gentamicine (Horaud *et al.*, 1996). Ce transposon, Tn*3706*, a été détecté en double exemplaire sur le plasmide pIP501 de *S. agalactiae* (Horaud *et al.*, 1996). Tn*3706* est composé d'un fragment central contenant le gène *aac6'-aph2* ; ce fragment est flanqué par deux copies de l'IS*256* en 5' du gène de résistance et une copie inversée à l'extrêmité 3' (Horaud *et al.*, 1996). L'IS*256* a été décrite originellement chez *Staphylococcus aureus* (Lyon *et al.*, 1984).

Plus récemment, un deuxième transposon composite a été associé à une résistance aux antibiotiques. Il s'agit du Tn*917*/Tn*916*-like qui porte le gène *erm*B codant la résistance croisée aux macrolides et à la clindamycine et flanqué par une seule IS en 3', l'IS*1381*, ce qui explique sa faible mobilité observée *in vitro*, laquelle nécessiterait probablement l'intervention d'un autre EGM (Puopolo *et al.*, 2007). Ce transposon de structure complexe a été décrit chez *S. pneumoniae* ce qui suggère que cette résistance ait pu être acquise suite au transfert horizontal du transposon (Puopolo *et al.*, 2007).

Enfin, un troisième transposon composite a été largement décrit chez *S. agalactiae* car il constitue un locus cible pour de nombreuses IS (voir § 2.1.2.3, § 2.1.2.6 et § 2.1.6) ; il s'agit du transposon composite *scpB-lmb* (**Figure 12**). L'existence

d'un transposon composite à ce locus a été pour la première fois évoquée par Franken *et al.* en 2001 qui ont découvert la présence de deux copies de l'IS*Sag2* flanquant *scpB-lmb* (Franken *et al.*, 2001). Ce transposon a été associé à la capacité de *S. agalactiae* à coloniser voire infecter l'homme. En effet, le gène *lmb* code une protéine de surface liant la laminine humaine qui favorise l'adhésion aux cellules épithéliales et microendothéliales ainsi que le franchissement de la barrière hématoencéphalique (Elsner *et al.*, 2002, Tenenbaum *et al.*, 2007). Le gène *scpB* code la C5a-peptidase qui a deux fonctions : elle interfère sur le recrutement des leucocytes en clivant la fraction C5a du complément et elle favorise l'invasion dans la cellule eucaryote en se liant à la fibronectine (Chen & Cleary, 1990, Cheng *et al.*, 2002). L'interaction hôte-parasite qui en résulte est spécifique car seul le sérum humain induit l'expression de la C5a peptidase des streptocoques (Gleich-Theurer *et al.*, 2009).

2.3. Les transposons conjugatifs et autres transposons dérivés

2.3.1. Caractéristiques générales

2.3.1.1. Définition et structure

Les transposons conjugatifs forment une famille très homogène. Leurs extrémités sont formées de séquences inversées répétées de 20 à 30 pb. Selon la nomenclature la plus récente, les transposons conjugatifs (CTn), également appelés éléments intégratifs conjugatifs (ICE), sont définis par la présence de gènes impliqués dans la mobilité, incluant des gènes pour l'excision, pour le transfert par conjugaison et pour l'intégration, ainsi que par la présence d'une grande variété de gènes, dont des gènes de résistance aux antibiotiques (**Figure 11**) (Roberts *et al.*, 2008). La plupart des transposons conjugatifs code une recombinase spécifique de site appartenant à la famille des intégrases à tyrosine (Salyers *et al.,* 1995b). Les fonctions catalysant la conjugaison (Tra) sont très similaires à celles portées par les plasmides conjugatifs *d'E. coli* (Salyers *et al.*, 1995b). L'originalité de ces transposons réside dans le fait qu'ils sont capables de se propager verticalement mais également horizontalement par conjugaison (Salyers *et al.*, 1995b). Ils sont donc très souvent associés aux phénomènes de propagation des résistances aux antibiotiques (Salyers & Shoemaker, 1995).

2.3.1.2. Mécanisme de transposition

La première étape de la transposition (l'excision) conduit à la circularisation précise de l'élément et est contrôlée par l'intégrase (Int-Tn) et éventuellement l'excisionase (Xis-Tn) qui catalysent la recombinaison entre des séquences identiques présentes au niveau des sites de recombinaison *attL* et *attR* bordant l'EGM (Burrus *et al.*, 2002a, Burrus *et al.*, 2002b). L'intermédiaire circulaire est non réplicatif. L'intégration du transposon dans une nouvelle cible est dépendante de l'intégrase de l'élément. Elle est sensée être très similaire à ce qui est observé pour l'excision. Elle peut intervenir dans le même génome ou après conjugaison dans un nouvel hôte. L'intégration a lieu principalement au niveau des régions 3' des gènes codant des protéines ou des ARNt (Burrus *et al.*, 2002a, Burrus *et al., 2002b)*.

2.3.1.3. Les dérivés des transposons conjugatifs

Les transposons mobilisables (MTn), ou éléments intégratifs et mobilisables (IME), possèdent des gènes pour l'excision ou pour l'intégration, mais à la différence des ICE font appel à d'autres éléments codant les fonctions nécessaires à la conjugaison pour pouvoir être transmis d'une bactérie à l'autre (**Figure 11**) (Roberts *et al.*, 2008). Ainsi, la plupart porte une origine de transfert *oriT* et des gènes codant des protéines mobilisatrice comme la relaxase, mais ils utilisent la machinerie conjugative d'autres éléments (mobilisation en *trans*) (Salyers *et al.*, 1995a). Dans le continuum des transposons, d'autres types d'EGM apparaissent après délétion de certains modules, intégratif ou conjugatif, et/ou, au contraire, l'accrétion d'autres modules. C'est par exemple le cas des éléments mobilisables en *cis* (CIME) (Pavlovic *et al.*, 2004).

2.3.2. Les transposons conjugatifs et leurs dérivés chez *S. agalactiae*

La comparaison des génomes séquencés de *S. agalactiae* a montré que la plasticité de certaines régions génomiques provenait avant tout des EGM (Glaser *et al.*, 2002, Tettelin *et al.*, 2005, Tettelin *et al.*, 2002). Les ICE et leurs dérivés représentent une part importante de cette plasticité ; en effet, 12 ICE, 6 IME, 13 CIME ont été définis après analyse des huit génomes séquencés (Brochet *et al.*, 2008a). Ces éléments joueraient un rôle important dans le transfert de matériel génétique entre les espèces et auraient influés sur l'évolution de *S. agalactiae* (Brochet *et al.*, 2008a).

Récemment, une famille atypique de transposons conjugatifs a été décrite chez *S. agalactiae* (Brochet *et al.*, 2009b). Cette famille nommée TN*GBSs* associe au module de conjugaison une transposase de type DDE (Brochet *et al.*, 2009b). Ces transposons s'insèrent dans des fragments inter géniques 15 ou 16 pb en amont des séquences en -35 des promoteurs, minimisant ainsi l'impact sur le génome et son expression (Brochet *et al.*, 2009b). L'originalité de cette famille de transposons tient au fait que l'intégrase est remplacée par une transposase. Les trois copies du plasmide intégratif pNEM316-1 présentes chez la souche NEM316 ainsi que l'îlot génomique X appartiendraient à cette famille atypique de transposons conjugatifs ; ils ont donc été renommés respectivement *Tn*GBS1 et *Tn*GBS2 (Brochet *et al.*, 2009b, Glaser *et al.*, 2002). Ces transposons sont fréquents chez les souches de *S. agalactiae* en particulier d'origine bovine, ce qui suggère qu'ils apportent un avantage sélectif pour la colonisation de ce réservoir (Brochet *et al.*, 2009b).

2.4. Les îlots génomiques et les îlots de pathogénicité

2.4.1. Caractéristiques générales

Le concept des îlots génomiques (GI) est dérivé de celui d'îlots de pathogénicité (PI) développé par Hacker *et al.* pour décrire les régions génomiques des *E. coli* uropathogènes qui portent les facteurs de virulence et qui sont spontanément excisables (Hacker *et al.*, 1990). Ces régions portent des séquences caractéristiques qui permettent de les distinguer du reste du génome (Hacker & Kaper, 2000). Elles sont le plus souvent insérées dans des gènes codant des ARNt (régions génomiques fréquemment réceptrices d'ADN étranger), et sont mobiles. Cette mobilité est médiée par des répétitions directes (DR) qui sont homologues à des sites d'attachement de phages et qui favorisent l'intégration dans le génome bactérien ou l'excision du génome bactérien (Hacker *et al.*, 1997). Elles portent fréquemment des IS et des séquences dérivées de phages et/ou de plasmides, comme les gènes de recombinases à tyrosine ou à sérine qui catalysent leur excision propre et leur intégration. Elles présentent un GC% différent du reste du génome, témoin de leur acquisition par transfert horizontal. Les PI portent en plus des GI des gènes de virulence. Les GI portent par contre des gènes associés à une très large variété de phénotypes (Roberts *et al.*, 2008).

2.4.2. Les îlots génomiques et les îlots de pathogénicité chez *S. agalactiae*

Deux îlots génomiques, appelés PI-1 et PI-2, codant des structures antigéniques pilus-like ont été identifiés ; ces deux GI de 16 et 11 kb présentent des répétitions directes aux extrémités ainsi qu'un régulateur transcriptionnel de type *AraC,* des gènes homologues de transposase et des fragments d'ADN prophagique (Rosini *et al.*, 2006).

Parmi les 11 à 14 îlots génomiques décrits dans le génome de *S. agalactiae*, certains ont la potentialité d'être mobiles ce qui le classe dans la catégorie des EGM d'après la nomenclature récemment établie (Roberts *et al.*, 2008). Certains même portent des gènes de virulence ce qui les définit alors comme des îlots de pathogénicité. Herbert *et al.* ont trouvé que sur tous les GI définis seuls quatre étaient de véritables PI : les îlots I, VI, X et XII (Herbert *et al.*, 2005). Deux particularités ont été rapportées mettant en lien ces PI et l'origine des souches : i) la présence de l'îlot VI chez toutes les souches bovines. ; ii) la présence l'îlot X chez quelques souches responsables d'infection néonatale précoce et son absence chez les souches de colonisation ou responsables d'infection néonatale tardive. Cependant, aucune corrélation n'a pu être formellement établie entre la présence d'un de ces quatre îlots et un quelconque pouvoir pathogène spécifique (Herbert *et al.*, 2005).

2.5. Les introns du groupe II

2.5.1. Généralités

Initialement découverts dans les organites des protistes, des végétaux et des champignons, les EGM ont été décrits en 1993 chez les bactéries (Ferat & Michel, 1993). L'hypothèse est qu'ils étaient en fait initialement présents chez les bactéries dont certaines espèces seraient les ancêtres des mitochondries et des chloroplastes (Ferat & Michel, 1993). Grâce aux programmes de séquençage à grande échelle des génomes, on a découvert que les introns du groupe II étaient très largement répandus parmi les procaryotes. D'une vingtaine d'espèces bactériennes recensées en 2003, le nombre est monté à près d'une centaine en 2008 avec plus de 200 introns du groupe II complets identifiés (Simon *et al.*, 2008, Dai & Zimmerly, 2003).

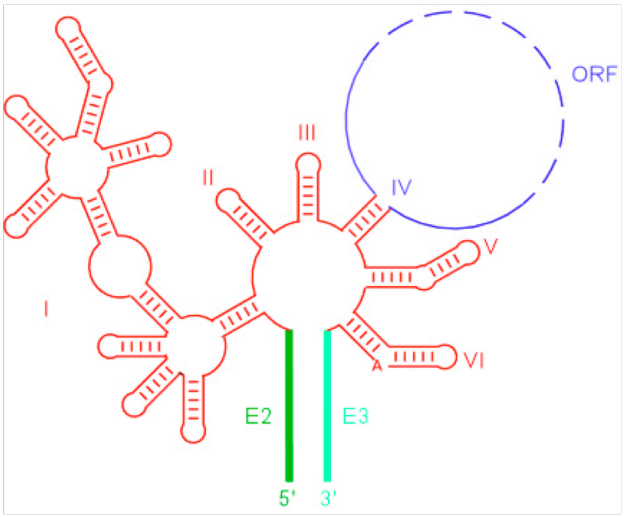


Figure 13. Structure d'un intron du groupe II. Les introns du groupe II se replient en une structure secondaire conservée constituée de six domaines arrangés autour d'une boucle centrale (Michel & Ferat, 1995). Le domaine V est le plus conservé et correspond au site actif du ribozyme. L'ORF est localisé dans la boucle du domaine IV et code une reverse transcriptase ou d'autres protéines liées (*image du site http://www.fp.ucalgary.ca)*.

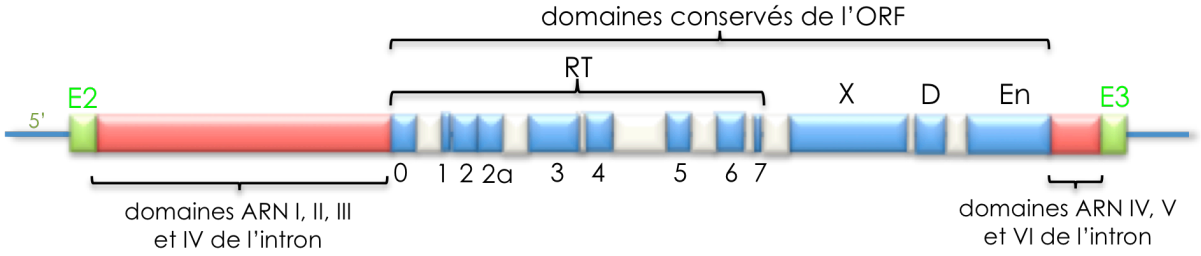


Figure 14. Structure de l'ORF d'un intron du groupe II. L'ORF contient les sept domaines communs à toutes les reverse transcriptases (RT) ainsi qu'un domaine X et un domaine En. Le domaine X est associé à l'activité de maturase de la protéine. Le domaine D est la région d'attachement à l'ADN, et le domaine En a une activité de nucléase utilisée lors de la réaction de mobilité (Michel & Ferat, 1995)

2.5.1.1. Définition et nomenclature

Les introns de groupe II sont une classe très particulière d'EGM ; ceux-ci transposent via un intermédiaire ARN, appelé ribozyme, qui agit comme une enzyme. Le ribozyme catalyse *in vitro* sa propre excision à partir du transcrit primaire par un procédé d'auto-épissage. La plupart d'entre eux porte en plus la séquence codante d'une reverse transcriptase qui agit comme une maturase en coopérant avec le ribozyme pour faciliter la mobilité de l'élément (Michel & Ferat, 1995, Ferat *et al.*, 1994). Les introns du groupe II bactériens sont localisés en majorité non pas dans des gènes conservés mais en dehors des ORF, et le plus souvent sur des ADN mobiles, tels que les plasmides ou les îlots de pathogénicité.

Une classification a été établie sur la base des phylogénies construites à partir des séquences des ORF. Huit classes ont été définies : CL, ML, A, B, C, D, E et F. Toutes ces classes sont retrouvées chez les bactéries. Chaque classe d'ORF est ensuite associée à un type de structure secondaire de l'ARN (Toor *et al.*, 2001, Zimmerly *et al.*, 2001).

2.5.1.2. Structure des introns du groupe II

Les introns du groupe II se replient en une structure secondaire conservée constituée de six domaines arrangés autour d'une boucle centrale (**Figure 13**) (Michel & Ferat, 1995). Le domaine V est le plus conservé et correspond au site actif du ribozyme. L'ORF est localisé dans la boucle du domaine IV et code une reverse transcriptase (RT) ou d'autres protéines liées. L'ORF contient les sept domaines communs à toutes les RT ainsi qu'un domaine X et un domaine En. Le domaine X est associé à l'activité de maturase de la protéine. Le domaine En a une activité de nucléase utilisée lors de la réaction de mobilité. Il existe aussi un domaine D qui est la région d'attachement à l'ADN (**Figures 14 et 16**) (Michel & Ferat, 1995).

2.5.1.3. Mécanisme réactionnel

2.5.1.3.1. Auto-épissage et assemblage du complexe ribonucléoprotéique

Le chemin réactionnel de l'auto-épissage est identique à celui des introns des ARN pré-messagers nucléaires et aboutit, après clivage et ligation des deux

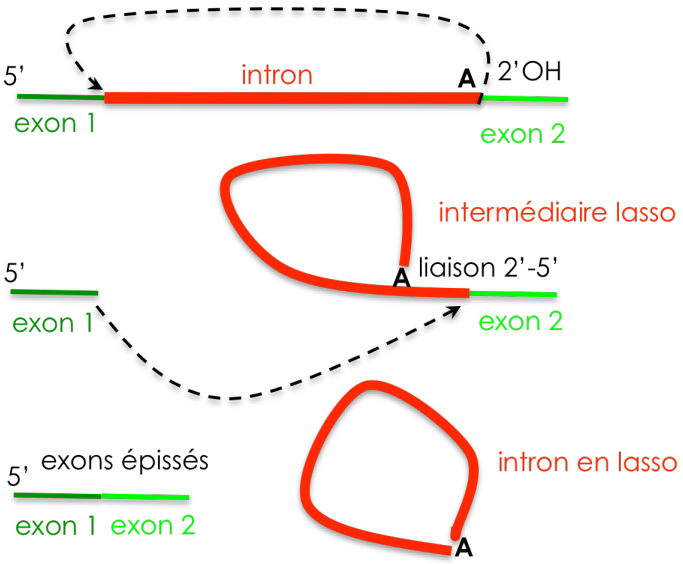


Figure 15. Mécanisme réactionnel de l'autoépissage des introns du groupe II.

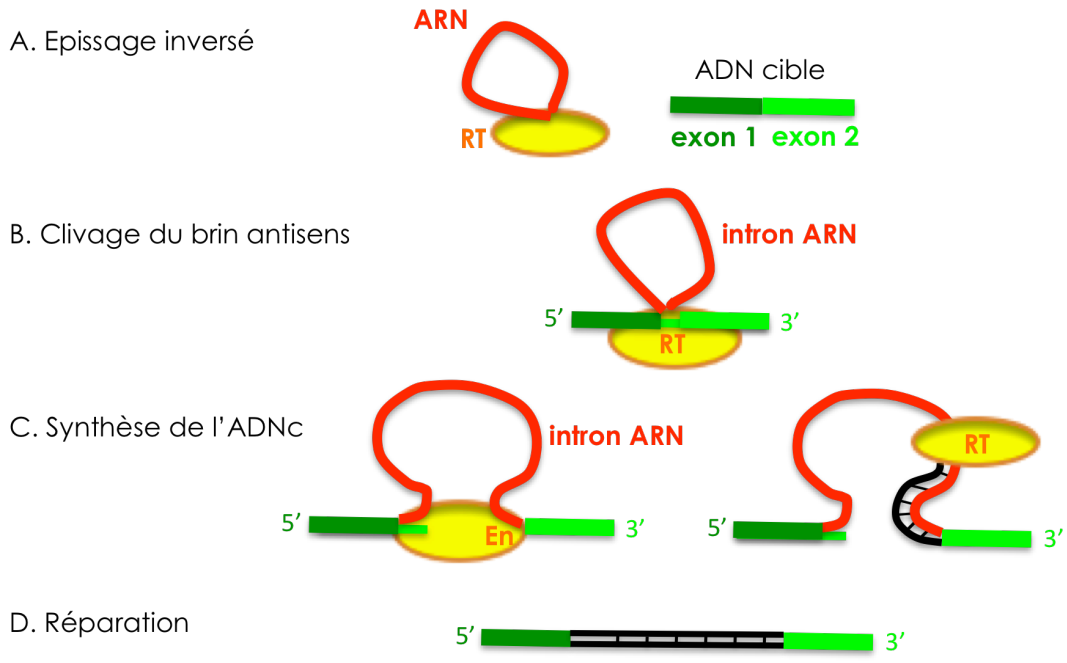


Figure 16. Mécanisme réactionnel du transfert des introns du groupe II.

extrémités 5' et 3', à la formation d'un lasso (**Figure 15**) (Lambowitz & Zimmerly, 2004). *In vivo*, l'auto-épissage requiert l'assistance d'une protéine, la RT, qui joue un rôle de maturase. La RT se lie au domaine 4 de l'intron. Les interactions protéine-ARN, qui se créent, entraînent un changement de conformation de l'intron aboutissant à son auto-épissage (**Figure 15**). Après épissage, la RT reste liée fortement à l'intron et constitue la fraction active des réactions de mobilité (Lambowitz & Zimmerly, 2004).

2.5.1.3.2. Mobilité des introns du groupe II

Le mécanisme de mobilité est caractérisé de « homing » ou « rétrohoming » (Cousineau *et al.*, 1998). Le complexe ribonucléoprotéique (RNP) reconnaît un site ADN spécifique dépourvu de l'intron et l'ARN s'insère dans le brin sens de l'ADN au niveau de la jonction des exons (**Figure 16**) (Michel & Ferat, 1995). La réaction est catalysée par le ribozyme mais facilitée par la RT de la même manière que l'épissage. Le domaine En de la transposase clive le brin antisens et le domaine RT reverse transcrit l'intron en ADNc en prenant l'ADN clivé comme amorce (Lambowitz & Zimmerly, 2004). En général, la mobilité des introns du groupe II est hautement spécifique d'une cible d'environ 35 pb identique au site de départ mais dépourvue de tout intron du groupe II. Les introns du groupe II sont aussi capables de s'insérer des sites ectopiques à un plus faible taux ; il s'agit alors du mécanisme de rétrotransposition (Cousineau *et al.*, 1998, Martinez-Abarca & Toro, 2000). Une troisième possibilité de mobilité existe chez les introns de la classe C où l'intron s'insère dans des motifs précis (Dai & Zimmerly, 2002, Granlund *et al.*, 2001, Roberts *et al.*, 2008).

2.5.2. Les introns du groupe II chez *S. agalactiae*

D'après la liste proposée sur la base de données internationale des introns du groupe II (http://www.fp.ucalgary.ca/group2introns/index.htm, Zimmerly Lab Web Site), l'espèce *S. agalactiae* compte deux introns du groupe II : GBSi1 (code d'accès Genbank AJ292930) et un ADN du prophage LambdaSa1 ayant une activité de maturase et de RT (code d'accès Genbank AE014217) (Dai *et al.*, 2003). Sur ces deux introns, seul GBSi1 a été bien caractérisé (Granlund *et al.*, 2001).

2.5.2.1. Caractéristiques générales de GBSi1

GBSi1 est un intron du groupe II de classe C de 1857 pb dont la transposase de 436 acides aminés possède un domaine RT, un domaine X mais pas de domaine endonucléase En, bien qu'une activité d'auto-épissage ait été démontrée *in vitro* (Granlund *et al.*, 2001). Le site d'insertion de GBSi1 est soit un site cible typique d'intron de groupe II (IBS1), soit un motif spécifique correspondant à une séquence répétée inversée avec une queue poly(A) comme c'est le cas pour le premier site d'insertion décrit (Granlund *et al.*, 2001, Luan *et al.*, 2003). La plupart des introns du groupe II ont eux-mêmes été associés à des EGM (Dai & Zimmerly, 2002, Klein & Dunny, 2002). L'hypothèse concernant GBSi1 est qu'il aurait été acquis par *S. agalactiae* par l'intermédiaire d'un plasmide intégratif dont quelques fragments sont retrouvés systématiquement chez les souches possédant GBSi1 (Luan *et al.*, 2003).

2.5.2.2. Impact sur le génome et le phénotype

Trois sites d'insertion de GBSi1 ont été caractérisés dans le génome de *S. agalactiae* : le premier correspond au fragment intergénique *scpB-lmb* (appelé également site X) (Granlund *et al.*, 2001), le deuxième correspond au fragment intergénique *ftsY-sag0728* (appelé également site Y) (Bohnsack *et al.*, 2002, Luan *et al.*, 2003), et le troisième correspond au fragment intergénique *lplA-1-sag0883* (appelé également site Z) (Luan *et al.*, 2003). Ces trois sites d'insertion sont donc à proximité de gènes codant des facteurs de virulence avérés (comme la C5a-peptidase ou la protéine LMB qui se fixe à la laminine humaine) ou putatifs (comme la lipoate ligase); cependant, aucune corrélation n'a pu être établie avec la pathogénicité des souches et/ou des manifestations cliniques particulières. Il faut souligner que GBSi1 partage le site d'insertion X avec l'IS*1548* mais jamais de manière concomitante, ce qui a donné lieu au concept de distribution exclusive mutuelle de ces EGM (Granlund *et al.*, 2001). Ce site apparaît comme un « hot-spot » d'insertion des EGM car il porte aussi parfois une autre IS, l'IS*Sag1* (voir § 2.1.2.6 et **Figure 13**) (Franken *et al.*, 2004). Il a aussi la particularité de présenter la structure d'un transposon composite étant encadré par deux IS*Sag2* (voir § 2.1.2.5 et **Figure 13**) (Franken *et al.*, 2001).

III-PARTIE EXPERIMENTALE

Partie 1.

Structure génétique de *Streptococcus agalactiae*

Partie 1. Structure génétique de *Streptococcus agalactiae*

A-Introduction

Streptococcus agalactiae est une bactérie au pouvoir pathogène complexe : à la fois commensale des tractus génito-urinaire et gastro-intestinaux de l'adulte sain (Bliss *et al.*, 2002), c'est un pathogène majeur responsable de sepsis chez la femme enceinte, le fœtus et le nouveau-né, et un pathogène émergent chez les personnes âgées (Henning *et al.*, 2001). Cette double facette de l'espèce ne peut être analysée qu'au moyen d'outils épidémiologiques permettant l'étude des traits génétiques de la bactérie et capables de différencier les souches commensales des souches pathogènes.

L'exploration de la génétique des populations bactériennes a été étudiée pendant longtemps par MLEE pour un grand nombre d'espèces pathogènes, telles *Neisseria meningitidis, Escherichia coli* ou *Legionella* spp. (Selander *et al.*, 1986). Il en est de même pour *S. agalactiae* dont la structure des populations a été décrite pour la première fois par MLEE (Musser *et al.*, 1989). Le MLEE a permis la compréhension de l'épidémiologie de la bactérie à travers le monde (Hauge *et al.*, 1996, Helmig *et al.*, 1993, Musser *et al.*, 1989, Quentin *et al.*, 1995), et a ainsi mis en évidence une lignée hyperinvasive regroupant des souches ayant un tropisme pour le système nerveux central des nouveaux-nés (Musser *et al.*, 1989). Les techniques de typage utilisées ultérieurement, comme le PFGE, le RAPD, ou le RFLP ont au final surtout confirmé les résultats du MLEE (Bingen *et al.*, 1992, Bohnsack *et al.*, 2002, Musser *et al.*, 1989, Quentin *et al.*, 1995, Rolland *et al.*, 1999, Sun *et al.*, 2005).

Plus récemment est apparu le MLST dont l'utilisation s'est largement répandue car cette nouvelle approche moléculaire, bien que coûteuse, s'est rapidement révélée très intéressante. Le MLST est une approche multilocus qui se rapproche en cela du MLEE, mais qui, à la différence du MLEE, offre des résultats non ambigus favorisant les comparaisons inter laboratoires via l'interrogation de base centrale de données en libre accès sur Internet (Maiden *et al.*, 1998). Ainsi, depuis l'année de la première publication du système MLST de *S. agalactiae*, la plupart des études portant sur la structure des populations s'appuie sur des données MLST.

Cependant, toute nouvelle méthode de typage doit faire l'objet d'une validation avec la technique de référence avant son utilisation à grande échelle (Maiden *et al.*, 1998). C'est en effet le seul moyen d'apporter la preuve, par exemple pour le MLST, que l'étude du polymorphisme nucléotidique de gènes de ménage peut être pertinente (Turner & Feil, 2007). Cela n'est pas le cas avec le schéma MLST établi pour *S. agalactiae*, qui n'a jamais été formellement comparé au

MLEE en dehors de quelques considérations spéculatives (Jones *et al.*, 2003, Jones *et al.*, 2006a).

L'objectif de ce travail expérimental est de déterminer dans un premier temps si le MLST est superposable au MLEE, puis dans un second temps, d'analyser si le MLST peut apporter de nouvelles données sur la structure des populations de *S. agalactiae*. Dans cette perspective, une collection de souches préalablement caractérisées par MLEE a été analysée en MLST et les données du MLEE réinterprétées à la lumière des outils bioinformatiques développés pour le MLST (Quentin *et al.*, 1995).

Partie 1. Structure génétique de *Streptococcus agalactiae*

B-Matériel et Méthodes

Tableau II. Les 113 souches de *S. agalactiae* analysées pour l'étude comparée multilocus sequence typing (MLST) *vs* multilocus enzyme electrophoresis (MLEE).

Souche	Séro-type	MLST								MLEE												
		ST	*adhP*	*pheS*	*atr*	*glnA*	*sdhA*	*glcK*	*tkt*	ET	PG1	PG2	G3P	PGI*	NSP	CAK	ADK	GLU	HEX	PLP	LGG	EST
G12	Ib	400	12	1	4	1	3	3	2	37	7	3	6	5	5	5	5	5	5	5	5	4
G17	Ib	12	10	1	4	1	3	3	2	35	5	2	6	5	5	5	5	5	5	5	5	4
G25	NT	2	1	1	3	1	1	2	2	1	6	4	6	5	7	5	5	5	5	6	4	3
G30	Ia	23	5	4	6	3	2	1	3	8	5	4	6	5	5	6	5	5	5	5	5	7
G32	II	28	1	1	3	5	2	2	2	18	5	1	6	5	5	6	5	5	5	6	5	7
G35	NT	197	1	1	3	1	2	2	2	14	5	5	6	5	5	6	5	5	5	5	5	7
G37	II	26	1	1	5	4	1	4	6	24	5	5	6	5	5	5	5	5	5	5	4	6
G40	Ib	10	9	1	4	1	3	3	2	27	5	3	6	5	5	6	5	5	5	5	6	4
G41	II	78	9	1	4	1	3	11	2	42	5	3	6	5	5	5	5	5	5	5	6	7
G42	II	10	9	1	4	1	3	3	2	22	5	3	6	5	5	5	5	5	5	6	5	7
G47	II	407	13	3	1	1	1	1	1	7	5	2	7	5	5	6	5	5	5	5	5	6
G49	III	17	2	1	1	2	1	1	1	53	6	4	4	5	5	5	7	6	5	5	6	7
G5	Ib	8	4	1	4	1	3	3	2	32	5	3	6	5	5	5	5	5	5	5	4	4
G51	V	398	1	1	4	1	1	4	2	47	5	2	6	5	5	5	5	5	7	5	6	1
G61	V	202	40	1	3	1	1	2	2	25	5	5	6	5	5	5	5	5	5	6	4	4
G64	II	2	1	1	3	1	1	2	2	26	5	5	7	5	5	5	5	5	5	6	4	6
G69	III	17	2	1	1	2	1	1	1	57	6	4	6	5	5	5	7	6	5	6	6	7
G71	III	17	2	1	1	2	1	1	2	56	6	4	6	5	5	5	7	6	5	6	6	6
G77	III	188	2	1	1	2	1	1	2	55	6	4	5	5	5	6	7	6	5	5	6	7
G8	Ia	7	10	1	2	1	3	2	2	45	5	4	6	5	5	5	5	5	5	5	4	6
G81	Ib	10	9	1	4	1	3	3	2	31	5	5	6	5	5	5	5	5	5	5	4	4
G82	III	46	1	1	4	1	3	3	2	15	5	5	6	5	5	6	5	6	5	5	5	6
G87	NT	10	9	1	4	1	3	3	2	2	5	5	6	5	7	5	5	5	5	6	4	3
L1	III	201	2	1	25	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L10	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L11	Ib	6	9	1	2	1	3	2	2	13	5	5	6	5	5	5	5	5	5	5	5	6
L12	II	28	1	1	3	5	2	2	2	13	5	5	6	5	5	5	5	5	5	5	5	6
L13	III	17	2	1	1	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L14	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L15	IV	196	1	1	3	1	1	12	2	2	5	5	6	5	7	5	5	5	5	6	4	3
L16	III	23	5	4	6	3	2	1	3	11	5	4	6	5	5	6	5	5	5	5	5	6
L17	III	195	9	1	1	1	3	2	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L18	Ib	10	9	1	4	1	3	3	2	34	5	3	6	5	5	5	5	5	5	5	5	4
L19	Ia	23	5	4	6	3	2	1	3	11	5	4	6	5	5	6	5	5	5	5	5	6
L2	III	17	2	1	1	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L20	III	17	2	1	1	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L21	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L22	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L23	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L24	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L25	IV	196	1	1	3	1	1	12	2	3	5	6	6	5	7	6	5	5	5	6	4	1
L26	Ia	23	5	4	6	3	2	1	3	11	5	4	6	5	5	6	5	5	5	5	5	6
L27	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L28	Ia	223	5	4	6	3	2	1	1	11	5	4	6	5	5	6	5	5	5	5	5	6
L29	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L3	III	17	2	1	1	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L30	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L31	III	193	1	1	3	2	2	2	1	12	5	5	6	5	5	6	5	5	5	5	5	6
L32	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L33	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L34	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L35	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L36	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L37	III	17	2	1	1	2	1	1	1	12	5	5	6	5	5	6	5	5	5	5	5	6
L38	III	164	1	1	2	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L39	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L40	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L41	III	17	2	1	1	2	1	1	1	52	6	5	5	5	5	6	7	6	5	6	6	7
L42	Ia	406	5	4	6	6	2	1	3	11	5	4	6	5	5	6	5	5	5	5	5	6
L43	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
L44	Ia	23	5	4	6	3	2	1	3	12	5	5	6	5	5	6	5	5	5	5	5	6
L45	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L46	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L47	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L48	II	19	1	1	3	2	2	2	2	19	5	5	6	5	5	6	5	5	5	6	5	6
L49	II	86	1	1	3	2	2	1	2	17	5	4	6	5	5	6	5	5	5	6	5	6
L5	III	17	2	1	1	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L50	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L51	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L52	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L53	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L54	III	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
L55	III	44	1	1	3	2	1	2	1	12	5	5	6	5	5	6	5	5	5	5	5	6
L6	II	19	1	1	3	2	2	2	2	17	5	4	6	5	5	6	5	5	5	6	5	6
L7	III	17	2	1	1	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L8	III	17	2	1	1	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
L9	III	19	1	1	3	2	2	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
V101	V	41	10	1	12	1	3	2	2	30	5	5	6	5	5	5	5	5	5	5	6	4
V104	II	28	1	1	3	5	2	2	2	16	5	1	6	5	5	6	5	5	5	6	5	4
V105	III	404	5	1	3	2	2	1	3	13	5	5	6	5	5	5	5	5	5	5	5	6
V11	Ib	84	10	1	2	1	3	1	2	33	5	5	6	5	5	5	5	5	5	5	5	4
V113	Ia	221	5	4	6	3	2	2	3	43	5	4	6	5	5	5	5	5	5	5	5	6
V114	Ib	10	9	1	4	1	3	3	2	29	5	3	6	5	5	5	5	5	5	5	6	4
V12	III	405	2	1	4	2	1	1	2	50	6	4	5	5	5	6	7	6	5	6	6	6
V17	Ia	199	5	4	6	3	2	1	2	11	5	4	6	5	5	6	5	5	5	5	5	6
V18	III	201	2	1	25	2	1	1	1	49	6	4	5	5	5	6	7	6	5	6	6	7
V19	Ia	23	5	4	6	3	2	1	3	10	5	4	6	5	4	6	5	5	5	5	5	1
V20	III	315	2	1	1	2	31	1	1	51	6	4	5	5	5	6	7	6	5	6	7	7
V21	II	19	1	1	3	2	2	2	2	20	5	5	6	5	5	6	5	5	5	6	6	6
V23	NT	17	2	1	1	2	1	1	1	59	6	4	5	5	5	5	7	6	5	6	6	6
V24	IV	7	10	1	2	1	3	2	2	24	5	5	6	5	5	5	5	5	5	5	4	6
V25	NT	19	1	1	3	2	2	2	2	19	5	5	6	5	5	6	5	5	5	6	5	6
V27	Ia	23	5	4	6	3	2	1	3	5	5	5	7	5	5	6	5	5	5	5	5	6
V29	II	10	9	1	4	1	3	3	2	4	5	5	7	5	5	6	5	5	5	6	5	6
V30	II	226	16	1	2	2	9	1	3	39	5	3	7	5	5	5	5	5	5	5	4	6
V31	Ia	23	5	4	6	3	2	1	3	44	5	4	7	5	5	5	5	5	5	5	5	6
V37	Ia	221	5	4	6	3	2	2	3	9	5	4	6	5	5	6	5	5	5	5	5	9
V38	Ia	403	5	4	6	3	2	2	2	17	5	4	6	5	5	6	5	5	5	6	5	6
V4	Ia	314	16	1	2	2	9	2	2	61	5	5	6	5	5	5	3	6	5	6	6	3
V44	Ia	402	2	1	3	1	3	2	2	23	5	5	6	5	5	5	5	5	5	6	4	6
V5	Ib	6	9	1	2	1	3	2	2	12	5	5	6	5	5	6	5	5	5	5	5	6
V57	III	17	2	1	1	2	1	1	1	60	6	4	5	5	5	5	7	6	7	6	6	7
V64	III	17	2	1	1	2	1	1	1	52	6	5	5	5	5	6	7	6	5	6	6	7
V7	Ib	12	10	1	4	1	3	3	2	34	5	3	6	5	5	5	5	5	5	5	5	4
V71	Ib	8	4	1	4	1	3	3	2	36	5	3	7	5	5	5	5	5	5	5	5	4
V80	II	12	10	1	4	1	3	3	2	41	5	3	6	5	5	5	5	5	5	5	5	6
V82	III	17	2	1	1	2	1	1	1	54	6	4	5	5	5	5	7	6	5	5	6	7
V87	NT	401	1	8	4	1	5	3	2	28	5	3	6	5	5	6	5	5	5	5	5	4
V88	II	10	9	1	4	1	3	3	2	40	5	3	6	5	5	5	5	5	5	5	6	6
V89	NT	2	1	1	3	1	1	2	2	2	5	5	6	5	7	5	5	5	5	6	4	3
V91	II	49	1	1	2	1	2	2	2	46	5	4	6	5	5	6	5	5	5	5	4	6
V93	NT	17	2	1	1	2	1	1	1	58	6	4	5	5	5	5	7	6	5	6	6	7
V99	II	399	10	1	4	1	3	1	2	38	5	3	6	5	5	5	5	5	5	5	4	6

*G, souches gastriques, L, souches isolées de LCR, V, souches vaginales

1. Souches bactériennes

La collection utilisée pour notre travail comprenait 113 souches humaines de *S. agalactiae*, épidémiologiquement non reliées, isolées en France entre 1986 et 1990, et provenant de 25 hôpitaux. Ces souches ont été utilisées précédemment pour établir la structure de la population de *S. agalactiae* et sont représentatives de la diversité génétique de l'espèce (Quentin *et al.*, 1995, Chatellier *et al.*, 1997, Chatellier *et al.*, 1996, Rolland *et al.*, 1999, Rolland *et al.*, 2002).

Les souches provenaient de prélèvements réalisés à visée diagnostique ou de dépistage, selon la répartition suivante (**Tableau II**): 54 souches isolées de méningites néonatales, 23 souches isolées de liquide gastrique de nouveau-nés asymptomatiques, et 36 souches de colonisation vaginale. Cette collection comprenait donc 54 souches invasives et 59 souches de colonisation.

Les souches ont été conservées à -80°C en bouillon K3-Schaedler additionné de 15% de glycérol jusqu'à utilisation.

2. Extraction de l'ADN génomique bactérien

2.1. Extraction de l'ADN total

Les cultures bactériennes ont été mises en suspension dans 1 mL de tampon de lyse (NaCl 1 M, tris HCl 10 mM pH 7,5, EDTA 5 mM pH 8, triton 0.5X). Vingt microlitres de lysozyme à 100 mg/mL et 20 µl de protéinase K à 20 mg/mL ont été ajoutés à 700 µL de cette suspension bactérienne, et incubés successivement une heure à 37°C puis une heure à 60°C. L'extraction de l'ADN a été réalisée par addition d'un volume égal du mélange phénol/chloroforme/alcool isoamylique (25/24/1 en volumes) (BioSolve BV, Netherland, Waalreweg). Après agitation, la suspension obtenue a été centrifugée pendant 15 min à 15000 x g. L'ADN contenu dans la phase aqueuse supérieure a été récupéré et une nouvelle extraction réalisée dans des conditions identiques. Après une centrifugation de 3 minutes à 15000 x g, l'ADN contenu dans la phase supérieure a été précipité par addition de 2 volumes d'éthanol absolu –20°C, en présence d'acétate de sodium 3 M pH 5,6 (10 % du volume final), agitation vigoureuse et incubation à –80°C pendant 10 minutes. Après une centrifugation de 10 min à 15000 x g, le surnageant a été éliminé et le

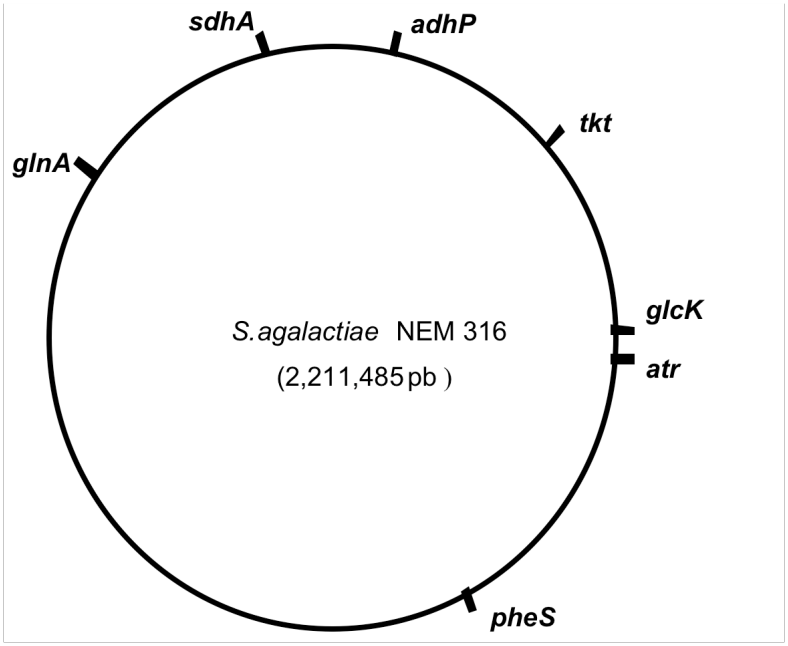


Figure 17. Représentation schématique de la répartition des sept loci (*adhP, pheS, atr, glnA, sdhA, glcK,* et *tkt*) étudiés en multilocus sequence typing sur le génome de la souche de *S. agalactiae* NEM 316.

culot séché. Enfin, l'ADN a été remis en solution dans 100 µl d'eau ultrapure et conservé à 4°C jusqu'à utilisation.

2.2. Contrôle de la pureté et de la concentration des ADN

La concentration et la pureté des ADN extraits ont été déterminées par mesure de l'absorbance au spectrophotomètre (Beckmann) à quatre longueurs d'ondes : 260 nm (longueur d'onde maximale d'absorbance des acides nucléiques), 280 nm (longueur d'onde maximale d'absorbance des protéines), 230 nm (longueur d'onde d'absorbance des polysaccharides) et 320 nm (l'absorbance à cette longueur d'onde mesure le bruit de fond). Par application de la loi de Beer-Lambert, une unité d'absorbance (Abs) à 260 nm correspond à une concentration d'ADN bicaténaire de 50 µg/mL. Pour le contrôle de pureté, les rapports R1 [(Abs $_{260}$ – Abs $_{320}$) / (Abs $_{280}$ – Abs $_{320}$)] et R2 [(Abs $_{260}$ – Abs $_{320}$) / (Abs $_{230}$ – Abs $_{320}$)] ont été calculés. Un rapport R1 inférieur à 1,8 indique une contamination par des protéines, un rapport R2 inférieur à 2 indique une contamination par des polysaccharides.

3. Sérotypage des souches

Les sérotypes capsulaires des souches ont été préalablement déterminés par la technique de double diffusion et/ou l'agglutination de particules de latex selon le protocole décrit par Quentin *et al.* (Quentin *et al.*, 1995). La répartition des sérotypes était la suivante (**Tableau II**): Ia (*n* = 16), Ib (*n* = 12), II (*n* = 18), III (*n* = 53), IV (*n* = 3), V (*n* = 3), non typables (*n* = 8) (Quentin *et al.*, 1995).

4. Caractérisation génétique des souches par multilocus sequence typing

4.1. Le multilocus sequence typing de *S. agalactiae*

La technique de typage par séquençage multilocus consiste en l'analyse du polymorphisme allélique de gènes de ménage. Pour *S. agalactiae,* le système a été développé par Jones *et al.* et déposé sur un site internet accessible en libre consultation (http://pubmlst.org/sagalactiae/) (Jones *et al.*, 2003). Les gènes de ménage (ou « house keeping genes ») analysés sont au nombre de sept, répartis sur l'ensemble du génome de *S. agalactiae* (**Figure 17**) : *adhP* (gène codant l'alcool déshydrogénase), *pheS* (gène codant la phénylalanine ARNt synthétase), *atr* (gène codant un transporteur des acides aminés), *glnA* (gène codant la glutamine

synthétase), *sdhA* (gène codant la sérine déshydratase), *glcK* (gène codant une glucose kinase) et *tkt* (gène codant une transcétolase) (**Tableau II**).

4.2. Amplification des gènes de ménage

L'amplification a été obtenue en utilisant les amorces dessinées par Jones *et al.* (Jones *et al.*, 2003) (**Tableau III**). Les conditions optimales d'amplification ont été obtenues, dans un volume de 50 µL, avec le mélange réactionnel suivant : 10 mM de tampon PCR (Tris-HCl, KCl 50 mM, pH 8,3) (Applied Biosystems, Foster City, USA), 1,5 mM de $MgCl_2$ (Applied Biosystems), 200 µM de chacun des dNTPs (Amersham Biosciences, Buckinghamshire, England), 0,5 µM de chaque amorce (Eurogentec, Seraing, Belgique), et 2,5 unités d'*AmpliTaq* DNA Polymerase (Applied Biosystems), auxquels ont été ajoutés 50 ng d'ADN. La réaction de PCR a été réalisée dans un thermocycleur GeneAmp PCR System 9600 (Perkin-Elmer, Norwalk, USA). Après un premier cycle de dénaturation de l'ADN à 94°C de 1 minute et 30 secondes, 30 cycles (comprenant une dénaturation de 30 secondes à 94°C, une hybridation des amorces de 30 secondes à 50°C et une élongation de 1 minute à 72°C) ont été réalisés, suivis d'un cycle d'élongation finale de 7 minutes à 72°C. Les produits de PCR ont été contrôlés par électrophorèse sur gel d'agarose (Eurobio) à 1,5%.

Tableau III. Amorces utilisées pour l'analyse en multilocus sequence typing.

Locus	Amorce sens (5' à 3')	Amorce anti-sens (5' à 3')	Taille de l'amplicon (pb)	Taille du fragment étudié (pb)
adhP	GGTGTGTGCCATACTGATTT	ACAGCAGTCACAACCACTCC	611	498
pheS	ATATCAACTCAAGAAAAGCT	TGATGGAATTGATGGCTATG	597	501
atr	ATGGTTGAGCCAATTATTTC	CCTTGCTCAACAATAATGCC	665	501
glnA	AATAAAGCAATGTTTGATGG	GCATTGTTCCCTTCATTATC	594	498
sdhA	AACATAGCAGAGCTCATGAT	GGGACTTCAACTAAACCTGC	602	519
glcK	GGTATCTTGACGCTTGAGGG	ATCGCTGCTTTAATGGCAGA	593	459
tkt	ACACTTCATGGTGATGGTTG	TGAACCTAGGTCATGAGCTTT	588	480

(D'après Jones *et al.*, (Jones *et al.*, 2003): revu et corrigé)

4.3. Purification des produits de PCR

Les produits de PCR ont été purifiés à l'aide du kit Montage™ PCR Centrifugal

Filter Devices (Millipore Corporation, Bedford, USA). Les produits amplifiés ont été déposés dans la colonne, et centrifugés 15 minutes à 1700 × g, puis, après ajout de 20 µL d'eau ultrapure (Millipore), de nouveau centrifugés 2 minutes à 1700 × g. Les produits purifiés ont ensuite été dosés par spectrophotométrie U.V..

4.4. Séquençage des produits de PCR

La réaction de séquençage a été réalisée avec le kit BigDye™ Terminator Mix v.3.0 (Applied Biosystems) dans un thermocycleur GeneAmp PCR System 9600 (Perkin-Elmer). Deux microlitres du BigDye™ Terminator Mix (contenant une *AmpliTaq*® DNA Polymerase), et 5 µM de l'amorce sens ou antisens ont été ajoutés à 300 ng d'ADN. Le programme appliqué a été le suivant : une étape de dénaturation de dix secondes à 94°C, une hybridation des amorces de dix secondes à 50°C, et une étape d'élongation de quatre minutes à 60°C, ce cycle de trois étapes étant répété 30 fois. Le séquençage a été réalisé sur les deux brins de chaque ADN (sens et antisens).

Les produits séquencés ont été purifiés par ajout de 20 µL d'eau déionisée et 60 µL d'isopropanol absolu 100% (Carlo Erba Reagents, Val de Reuil, France), précipitation 15 minutes à température ambiante et centrifugation de 30 minutes à 3000 × g. La purification a été suivie d'un lavage à l'isopropanol 70%, d'une deuxième centrifugation de dix minutes à 2000 × g, puis d'un séchage et d'une mise en solution dans 10 µL de formamide (Applied Biosystems). Les séquences nucléotidiques ont été analysées par le séquenceur capillaire automatique ABI Prism® 3100-*Avant* Genetic Analyser (Applied Biosystems).

4.5. Analyse *in silico*

4.5.1. Analyse des séquences nucléotidiques

L'analyse des séquences nucléotidiques a été réalisée à l'aide du logiciel d'exploitation ABI Prism® DNA Sequencing Analysis (Applied Biosystems). Le logiciel BioEdit (Biological sequence alignment editor), disponible en ligne (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), a été utilisé pour l'ajustement des séquences ; leur alignement a été réalisé avec le logiciel ClustalW Multiple alignement. Les séquences obtenues ont été comparées aux séquences de *S. agalactiae* déposées sur le site internet : http://pubmlst.org/sagalactiae/.

4.5.2. Assignement des allèles et des profils alléliques

Pour chaque locus, toute mutation dans la séquence nucléotidique, silencieuse ou non, définit un nouvel allèle. A chaque variant allélique correspond un numéro d'identification attribué par Nicola Jones, « webmaster » du site MLST de *S. agalactiae* (Jones *et al.*, 2003).

La combinaison de numéros correspondant aux allèles des sept loci étudiés définit un profil allélique ou « Sequence Type » (ST) (Maiden *et al.*, 1998). L'ordre d'affectation des numéros aux sept loci a été fixé arbitrairement (Jones *et al.*, 2003) : *adhP*, *pheS*, *atr*, *glnA*, *sdhA*, *glcK* et *tkt*. A chaque profil allélique correspond un numéro d'identification (ST-1, ST-2, … etc.), attribué par le webmaster (Jones *et al.*, 2003).

L'identification des allèles et du profil allélique de chacune des 113 souches de la collection a été réalisée en ligne par interrogation directe de la base de données accessible à l'adresse : http://pubmlst.org/sagalactiae/. Pour les souches dont les séquences nucléotidiques et/ou les profils alléliques n'étaient pas encore référencés dans la base de données, les séquences ont été soumises au webmaster (à l'issue d'une vérification par une nouvelle PCR et une nouvelle réaction de séquençage).

4.5.3. Analyse phylogénétique

L'analyse MLST fournit deux types de données, les profils alléliques des souches et les séquences nucléotidiques des gènes de ménage. Les relations évolutives entre les souches ont été analysées à ces deux niveaux.

4.5.3.1. Analyses à partir des profils alléliques

4.5.3.1.1. Analyse basée sur les complexes clonaux

Le programme eBURST v3 (http://eburst.mlst.net/) a permis d'identifier les complexes clonaux (CC) au sein du jeu de données correspondant aux 113 souches (Feil *et al.*, 2004). L'identification des complexes clonaux permet de définir un ancêtre hypothétique commun au sein du complexe qui est appelé le ST fondateur. Les ST assignés à un même CC possèdent au moins cinq allèles en commun sur les sept loci étudiés avec au moins un autre membre du complexe (Feil *et al.*, 2004).

Ainsi, au sein d'un CC, les ST sont soit des « variants simple locus » (SLV) (6 allèles identiques sur 7), soit des « variants double locus » (DLV) (5 allèles identiques sur 7) d'au moins un autre ST du complexe. Les ST qui ne sont affiliés à aucun CC sont appelés singletons.

4.5.3.1.2. Analyse basée sur la méthode des distances

Le logiciel Sequence Type Analysis and Recombinational Test (START2), accessible en ligne (http://www.startsoftware.net) (Jolley *et al.*, 2001), a été utilisé pour la construction d'un arbre phylogénétique réalisé à partir de la matrice de distances issue des profils alléliques en choisissant la méthode d'agrégation Unweighted Pair Group Method with Arithmetic Averages (UPGMA).

4.5.3.2. Analyses à partir des séquences nucléotidiques

4.5.3.2.1. Phylogénie basée sur la méthode des distances

Le calcul des matrices de distances a été réalisé grâce au logiciel Molecular Evolutionary Genetics Analysis (MEGA version 4.1 Beta, (http://www.megasoftware.net) (Kumar *et al.*, 2004, Kumar *et al.*, 2008), en utilisant l'indice de Kimura comme modèle évolutif. Deux méthodes de reconstruction phylogénétique ont été utilisées : la méthode Neighbor-Joining (NJ) et la méthode UPGMA. La robustesse des arbres phylogénétiques obtenus a été évaluée par la procédure de ré-échantillonnage ou « bootstrap », (1000 combinaisons). Les séquences des sept gènes ont été concaténées artificiellement en un « supergène » pour une analyse globale des séquences nucléotidiques des gènes.

4.5.3.2.2. Phylogénie basée sur le maximum de vraisemblance

Les analyses basées sur le maximum de vraisemblance ou « maximum-likelihood » (ML) ont été conduites à partir des séquences nucléotidiques du supergène en prenant un représentant par ST et non la totalité de la collection de 113 souches. Ces analyses ont été réalisées à l'aide du logiciel PAUP* 4.0 beta 10 (Swofford, 2003). Le logiciel Modeltest 3.8 accessible librement sur le serveur web Modeltest (http://darwin.uvigo.es/software/modeltest.html) a permis de sélectionner par des tests de "hierarchical likelihood ratio" (hLRT) le modèle de substitution le plus

approprié à l'analyse ML des séquences nucléotidiques étudiées (Posada, 2006, Posada & Crandall, 1998).

4.5.3.2.3. Choix de l'outgroup

Pour toutes les analyses phylogénétiques basées sur les séquences nucléotidiques, l'espèce *Streptococcus pyogenes* a été choisie comme séquence externe ou « outgroup ». Les sept gènes du système MLST de *S. agalactiae*, orthologues chez *S. pyogenes*, ont été choisis sur le génome de *S. pyogenes* MGAS10394 (numéro d'accès Genbank NC_006086) (http://www.ncbi.nlm.nih.gov).

4.5.4. Evaluation du phénomène de recombinaison homologue dans la population de *Streptococcus agalactiae* à partir des données MLST

4.5.4.1. Exploitation des résultats issus de l'analyse eBURST

Turner *et al.* ont développé trois tests permettant d'évaluer la proportion recombinaison/mutation à partir des données eBURST (Turner *et al.*, 2007). Le premier test consiste à évaluer l'existence d'un groupe clonal très volumineux dans la population représentée sous-forme de « snapshot » (représentation graphique obtenue par l'eBURST) ; le deuxième test consiste à révéler les liens de type SLV entre les ST d'un groupe clonal donné qui ne seraient pas apparus dans le représentation graphique et à regarder si ces liens parcourent de grands trajets à travers le groupe clonal ; le troisième test consiste à évaluer si la proportion des ST inclus dans le grand groupe clonal excède 25% de la totalité des ST. Une réponse positive aux trois tests fait suspecter une population dans laquelle il y aurait un rapport recombinaison/mutation important.

4.5.4.2. Analyse de la structure de la population sous forme de réseau

Les relations évolutives entre les souches ont été analysées par une représentation sous forme de réseau. Le logiciel SplisTree4 (http://www.splitstree.org/) a été utilisé en appliquant l'algorithme neighbor-net sur les sites informatifs (retrouvés parmi les 3456 nucléotides du supergène) pour l'analyse par parcimonie (Huson & Bryant, 2006). La recombinaison entre les souches a été évaluée par l'index appelé « pairwise homoplasy index » (PHI) (Bruen *et al.*, 2006).

4.5.4.3. Evaluation du ratio recombinaison/mutation

La méthode développée par Feil *et al.* permet d'évaluer directement le taux de recombinaison et le taux de mutation à partir des données MLST (Feil *et al.*, 2000). Cette méthode a été appliquée sur les CC définis par eBURST dans la population. Pour chaque CC, les SLV du ST fondateur ont été recherchés et pour chaque SLV il a été regardé si la variation allélique provenait d'un nucléotide ou de plus de un nucléotide. L'hypothèse est que les variations alléliques multi-sites (c'est-à-dire sur plusieurs nucléotides) proviennent toujours d'une recombinaison homologue alors que les variations uni-sites (sur un nucléotide) peuvent provenir aussi bien d'une mutation ponctuelle que d'une recombinaison faisant suite à une mutation. Après avoir défini la variation allélique (uni- ou multi-site), il a été recherché si le variant allélique existait (SLV partagé) ou non (SLV nouveau) dans les autres CC, à partir de l'hypothèse selon laquelle les variants uni-sites nouveaux sont issus d'une mutation ponctuelle alors que les variants uni-sites partagés sont issus d'une mutation suivie d'une recombinaison. Au total, le nombre de recombinaisons a été calculé en faisant la somme des variants multi-sites et des variants uni-sites partagés. Le nombre de mutations correspondant au nombre de variants uni-sites non partagés. Ce calcul a été réalisé pour chaque CC et pour l'ensemble des STs regroupés en CC.

5. Caractérisation génétique des souches par multilocus enzyme electrophoresis

L'analyse des 113 souches par multilocus enzyme electrophoresis (MLEE) avait été réalisée précédemment, montrant l'existence de 58 electrophorétypes (ET) sur la base d'un système d'analyse de 12 isoenzymes (**Tableau II**) (Quentin *et al.*, 1995). Nous avons ré-analysé les données obtenues par MLEE en utilisant le logiciel d'analyse START (voir § 4.5.3.1.2) (Jolley *et al.*, 2001). Cet outil bioinformatique, qui a été conçu pour l'analyse des données MLST, est en effet parfaitement applicable à l'analyse des données MLEE puisqu'il exploite les profils alléliques. L'application BURST contenue dans ce logiciel a permis de définir des lignées clonales, appelées groupes clonaux (CG) en référence à la terminologie du MLST (« complexes clonaux »). Les souches partageant au moins 10 allèles identiques sur les 12 étudiés en MLEE ont été regroupés au sein d'un même CG.

6. Analyses statistiques

6.1. Clonalité de la population

La clonalité de la population étudiée a été testée avec l'application Linkage Disequilibrium utilisable en ligne (http://linux.mlst.net/link_dis/index.htm). Cette application calcule l'indice d'association I_A à partir des variances attendue (V_e) et observée (V_o) de la distribution des mésappariements d'allèles [$I_A=(V_o / V_e) -1$]. Une population est clonale si I_A est supérieur à 1, c'est-à-dire s'il existe un déséquilibre de liaison significatif (Smith *et al.*, 1993).

6.2. Pouvoir discriminant

Le pouvoir discriminant de chacune des deux méthodes de typage, MLST et MLEE, a été mesuré par l'indice de discrimination (D) de Simpson, décrit par Hunter et Gaston, qui calcule la probabilité que deux souches non reliées aient des génotypes différents, les génotypes étant les ET en MLEE et les ST en MLST (Hunter & Gaston, 1988). La formule est la suivante :

$$D = 1 - [1/N(N-1)] \sum_{j=1}^{S} nj(nj-1)$$

où **N** est le nombre de souches dans la population totale, **S** le nombre total de types décrits et **nj** le nombre de souches appartenant au type **j**.

6.3. Congruence entre les techniques de typage

La concordance entre MLST et MLEE a été analysée par deux tests statistiques.

6.3.1. Analyse en classification croisée

Une analyse en classification croisée a été réalisée d'après la méthodologie décrite par Robinson *et al.* (Robinson *et al.*, 1998). Toutes les souches ont été appariées deux à deux dans une matrice. Pour chacune des paires de souches identifiées, il a été déterminé si elles appartenaient à la même lignée clonale ou non, selon que l'analyse soit menée en MLST ou en MLEE. Ensuite, pour une même paire de souches, il a été établi si les résultats du MLST et le MLEE étaient concordants ou discordants. Un tableau de correspondances a été établi pour l'ensemble des paires identifiées dans la population de 113 souches et le pourcentage de correspondance calculé.

6.3.2. Analyse des correspondances

L'analyse des correspondances a été réalisée selon la méthode développée par Robinson *et al.* (Robinson *et al.,* 1998). Un tableau a été établi en comparant la correspondance entre les principales lignées clonales identifiées en MLST et en MLEE. Le tableau des correspondances a été analysé par un test du Chi-deux.

Partie 1. Structure génétique de *Streptococcus agalactiae*

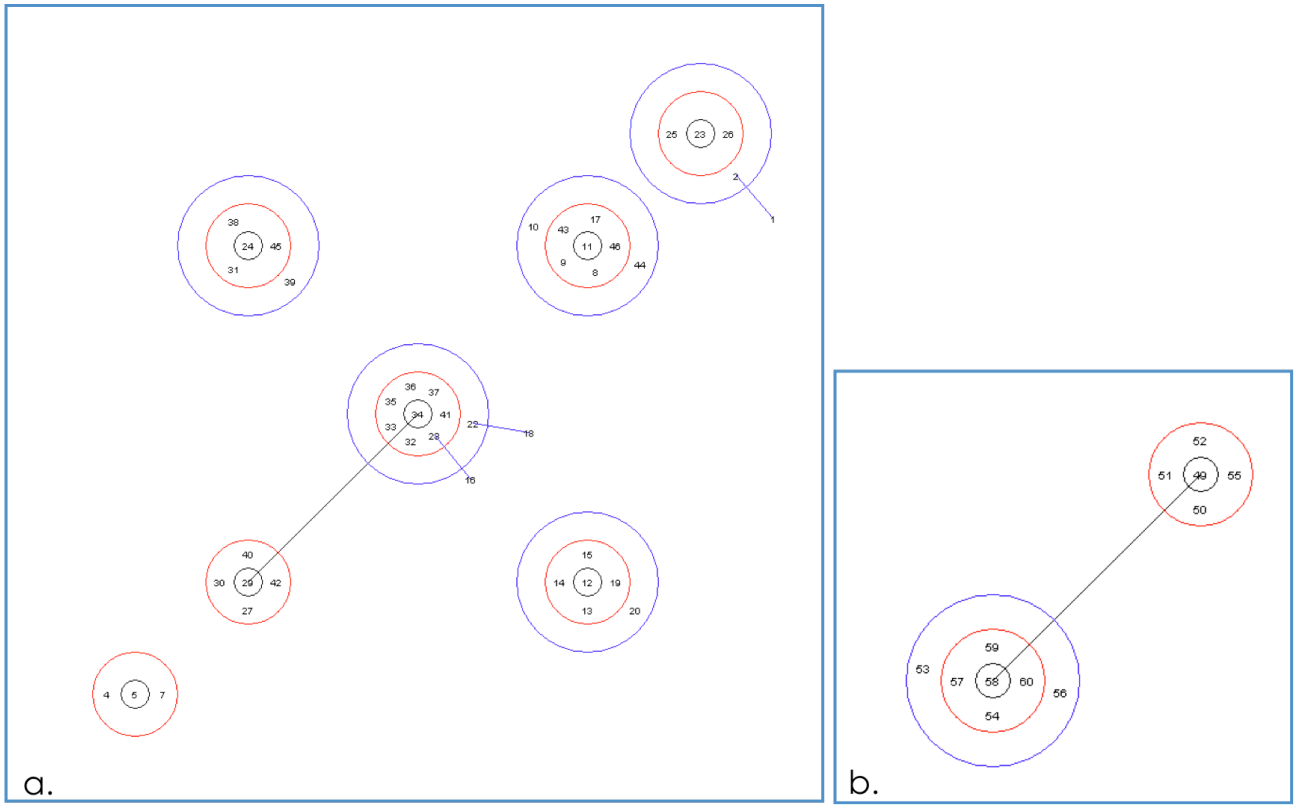


Figure 18. Structure de la population de *S. agalactiae* déterminée par multilocus enzyme electrophoresis (MLEE). Les deux lignées principales identifiées par l'algorithme BURST sont représentées (a,b). Les électrophorétypes (ET) qui diffèrent par un ou deux allèles dans leurs profils électrophorétiques (les variants unilocus sont dans les cercles rouges et les variants bilocus sont dans les cercles bleus) sont regroupés en cercle autour de l'ET fondateur (ET dans les cercles noirs). Ces graphiques ont été créés à l'aide du logiciel START v2 (http://www.mlst.net) à partir des 58 profils MLEE trouvés dans la population de 113 souches de *S. agalactiae* (Jolley *et al.*, 2001).

1. Structure génétique de la population définie par multilocus enzyme electrophoresis (MLEE) après ré-analyse des données

La collection de souches utilisée pour notre travail avait préalablement fait l'objet d'une analyse phylogénétique par multilocus enzyme electrophoresis (MLEE), sur la base de l'étude du polymorphisme électrophorétique de 12 enzymes métaboliques (Quentin *et al.*, 1995). Avec le MLEE, chaque souche a été caractérisée par la combinaison des allèles des loci de 12 enzymes : la nucléoside phosphorylase, la phosphoglucose isomérase, la carbamylate kinase, la glycéraldéhyde-3-phosphate déshydrogénase, la glutamate déshydrogénase, l'hexokinase, les phosphoglucomutases 1 et 2, la phénylalanyl leucine peptidase, la leucyl glycylglycine peptidase, l'adénylate kinase et l'estérase (**Tableau II**). Cinquante-huit électrophorétypes (ET) différents ont été identifiés parmi les 113 souches composant la population étudiée. Cette population de *S. agalactiae* se divisait en deux groupes phylogénétiques principaux relativement distants : la division I homogène comprenant 13 ET, et la division II plus hétérogène comprenant 45 ET (Quentin *et al.*, 1995).

Nous avons ré-analysé les données du MLEE en utilisant l'algorithme BURST, initialement conçu pour exploiter les profils alléliques issus du MLST (**Figure 18**). Les 58 ET ont été regroupés en sept groupes clonaux (CG), incluant les variants alléliques qui ne diffèrent que par deux loci au maximum. Cinq ET n'ont été affiliés à aucun CG et identifiés comme singletons. Le CG le plus important en taille (*n* = 39 souches) est le CG49, appelé du nom de l'ancêtre hypothétique commun défini par BURST (**Figure 18a**) ; il est lui-même divisé en deux sous-groupes, appelés subCG49 et subCG58, composés respectivement de quatre SLV de ET49 et de quatre SLV et deux DLV de ET58. Par BURST, CG49, qui regroupe 12 ET, est défini comme une lignée phylogénétique en tant que telle. Ce groupe clonal correspond à la division I précédemment décrite par Quentin *et al.* (Quentin *et al.*, 1995). La seconde lignée phylogénétique compte six CG regroupés autour de l'ancêtre hypothétique ET34 : CG12 (*n* = 23 souches), composé de quatre SLV et un DLV de ET12, soit six ET; CG11 (*n* = 15 souches), composé de cinq SLV et deux DLV de ET11, soit huit ET ; CG34 (*n* = 15 souches), composé de deux sub-CG, sub-CG34 (comptant sept SLV et un DLV de ET34) et sub-CG29 (comptant quatre SLV de ET29), soit 14 ET ; CG 24 (6 souches ; 5 ET), CG23 (6 souches ; 5 ET), and CG5 (3 souches ; 3 ET) (**Figure 18b**).

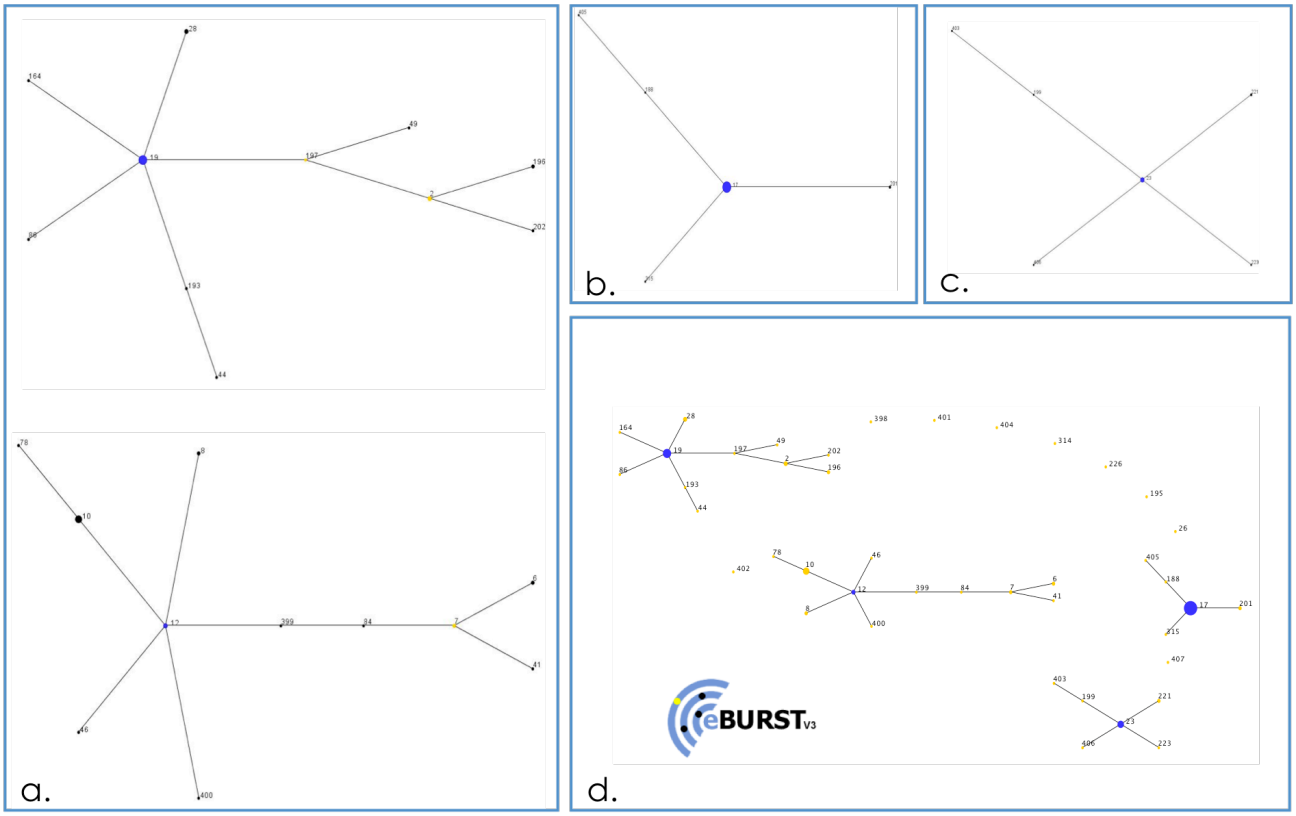


Figure 19. Structure de la population de *S. agalactiae* déterminée par multilocus sequence typing (MLST). Les trois lignées principales identifiées par l'algorithme eBURST v3 sont représentées (a,b,c) ainsi que l'ensemble de la population qui contient neuf singletons (d). Les sequence types (ST) qui varient par un ou deux allèles (variant uni- ou bi-locus) sont reliés au ST fondateur (point bleu). Le graphique a été créé en se basant sur les profils alléliques des 113 souches de *S. agalactiae* à l'aide du logiciel eBURST v3 (http://eburst.mlst.net) (Feil *et al.*, 2004).

2. Structure génétique de la population définie en multilocus sequence typing (MLST)

Les séquences nucléotidiques des sept loci du système MLST défini par Jones *et al.* pour *S. agalactiae* ont été déterminées pour les 113 souches et les profils alléliques ont été assignés (Jones *et al.*, 2003). La taille de la séquence nucléotidique analysée varie de 459 pb (*glcK*) à 519 pb (*sdhA*) et le nombre d'allèles pour chaque locus est compris entre quatre (*pheS* et *tkt*) et dix (*adhP*) (**Tableau IV**). Quarante-deux profils alléliques différents ont été identifiés dont 29 avec un représentant unique. Quatre profils alléliques, ST-17 (*n* = 33), ST-19 (*n* = 13), ST-10 (*n* = 8) et ST-23 (*n* = 8), regroupent près de 55% des souches. Les profils alléliques des souches sont présentés dans le **Tableau II.**

Tableau IV. Caractéristiques des loci du MLST appliqué à la collection de 113 souches de *S. agalactiae*.

Locus	Taille du fragment séquencé (pb)	Nombre d'allèles identifiés	Nombre de sites nucléotidiques polymorphiques
adhP	498	10	11
pheS	501	4	6
atr	501	8	12
glnA	498	6	7
sdhA	519	6	12
glcK	459	6	9
tkt	480	4	8

L'indice d'association I_A calculé à 2,185 (V_o=4,659, V_e=1,4628) montre un déséquilibre de liaison statistiquement significatif (P<0,001), indiquant ainsi la nature clonale de la population. L'algorithme eBURST a mis en évidence trois lignées phylogénétiques divisées en quatre complexes clonaux (**Figure 19**). La première lignée regroupe 51 souches phylogénétiquement reliées par l'ancêtre fondateur hypothétique, ST19. Cette lignée se divise en deux complexes clonaux : CC12 qui compte 11 ST et 23 souches, et CC19 qui compte 11 ST et 28 souches (**Figure 19a**). La deuxième lignée phylogénétique (*n* = 39 souches) est regroupée en un seul CC autour du fondateur hypothétique, ST17. Ce CC17 est composé de

Tableau V. Tableau croisé des concordances et discordances entre les paires de souches (*n* = 12656) classées en groupes clonaux (GC) par MLEE et en complexes clonaux (CC) par MLST.

	CG*		
	Concordances	Discordances	
CC			Total
Concordances	2123 (677)	905 (2351)	3028
Discordances	707 (2123)	8921 (7475)	9628
Total	2830	9826	12656

*observés (calculés)

Tableau VI. Tableau de contingence des correspondances entre le typage MLEE et le typage MLST.

	Lignées MLEE*					
Lignées MLST*	CG11	CG49	CG12	CG34	Autres CG	Total
CC23	12 (1,85)	0 (4,83)	1 (2,85)	0 (1,85)	1 (2,6)	14
CC17	0 (5,17)	38 (13,46)	1 (7,93)	0 (5,17)	0 (7,24)	39
CC19	3 (3,72)	0 (9,66)	17 (5,69)	0 (3,71)	8 (5,20)	28
CC12	0 (3,18)	0 (8,28)	3 (4,88)	14 (3,18)	7 (4,46)	24
Autres CC	0 (1,06)	1 (2,76)	1 (1,63)	1 (1,06)	5 (1,48)	8
Total	15	39	23	15	21	113

*observées (calculées)

trois SLV de ST17 (ST198, ST201 et ST315), et d'un DLV (ST405) (**Figure 19b**). Enfin, la troisième lignée phylogénétique (*n* = 14) correspond elle aussi à un seul complexe clonal, CC23, avec ST23 en fondateur hypothétique, quatre SLV (ST199, ST221, ST223 and ST406), et un DLV (ST403) (**Figure 19c**). Sur l'ensemble de la population, neuf souches correspondent à des singletons (**Figure 19d**).

3. Pouvoir résolutif des deux méthodes de typage

Le MLEE a fourni une plus grande diversité de données que le MLST. En effet, 58 ET ont été définis contre 42 ST. Le génotype le plus fréquent en MLEE, ET49, représente 15,9% de la population, alors que le génotype le plus fréquent en MLST, ST17, représente 29,2% de la population. L'indice de discrimination des génotypes a été calculé pour les deux méthodes. Le MLEE montre un meilleur pouvoir discriminant égal à 0,939 contre 0,897 pour le MLST, sachant que le pouvoir discriminant exigé pour les études épidémiologiques est idéalement supérieur ou égal à 0,90 (Hunter & Gaston, 1988).

4. Congruence entre les lignées phylogénétiques définies par les deux méthodes de typage

La classification croisée des souches basée sur la correspondance et la discordance entre complexes clonaux du MLST et groupes clonaux du MLEE montre que les deux systèmes de typage sont concordants à 87,3% sur l'ensemble des quatre lignées phylogénétiques principales, confirmant la bonne congruence entre les deux méthodes (**Tableau V**). En d'autres termes, deux souches distinguées par l'une des méthodes de typage le sont aussi, dans la grande majorité des cas, par l'autre méthode de typage. A noter toutefois que le MLEE apparaît un peu plus prédictif du MLST que l'inverse. En effet, près de 75% (2123/2830) des paires de souches ayant le même CG ont aussi le même CC, alors que 70% (2123/3028) des paires de souches ayant le même CC ont le même CG (**Tableau V**).

Par ailleurs, un tableau de contingence classant les souches en fonction des deux systèmes de typage a permis de confirmer la concordance entre MLEE et MLST (**Tableau VI**). Une association statistiquement significative ($P < 2.2e-16$) a été mise en évidence entre les lignées définies en MLEE et celles définies en MLST. Près de 94% (14/15) des souches appartenant au groupe clonal CG34 défini en MLEE appartiennent au complexe clonal CC12 défini en MLST, 74% (17/24) des souches

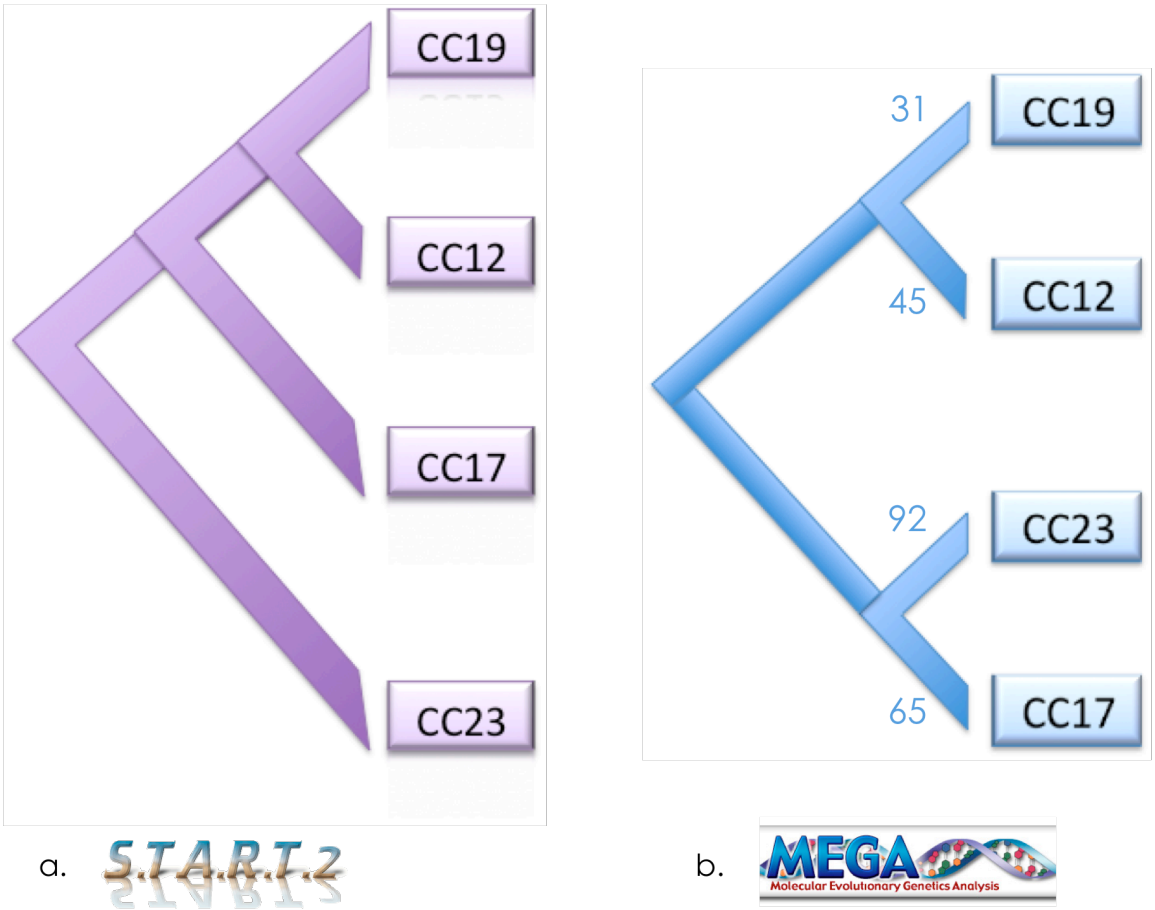


Figure 20. Deux modèles d'évolution de la population de *S. agalactiae* obtenus à l'issue de l'analyse par by multilocus sequence typing (MLST). Le premier modèle (a) provient de l'analyse des profils alléliques avec le logiciel START v2 (http://www.mlst.net), avec l'algorithme de clusterisation UPGMA et la correction de Kimura (Jolley *et al.*, 2001). Le second modèle (b) provient de l'analyse des séquences nucléotidiques du supergène obtenu selon la méthode de Jones *et al.* (Jones *et al.,* 2003), en utilisant le logiciel MEGA v4.1 Beta (http://www.megasoftware.net), avec les algorithmes de clusterisation UPGMA et NJ et la correction de Kimura (Kumar *et al.,* 2008). Les dendrogrammes reconstruits avec la méthode Njet UPGMA ont la même architecture. Les valeurs des bootstraps de chaque complexe clonal (CC) sont indiquées à côté des branches.

appartenant au groupe clonal CG12 défini en MLEE appartiennent au complexe clonal CC19 défini en MLST, 80% (12/15) des souches appartenant au groupe clonal CG11 défini en MLEE appartiennent au complexe clonal CC23 défini en MLST, et 97% (38/39) des souches appartenant au groupe clonal CG49 défini en MLEE appartiennent au complexe clonal CC17 défini en MLST (**Tableau VI**). On observe une bonne concordance entre les résultats du MLST et ceux du MLEE et ce, pour une population de *S. agalactiae* de sérotypes et d'origines géographiques et temporelles multiples.

5. Relations évolutives entre les lignées clonales

5.1. Lignées clonales

Un dendrogramme a été construit à partir des profils alléliques du MLST à l'aide du logiciel START en utilisant le même algorithme (UPGMA) que celui utilisé par Quentin *et al.*, l'UPGMA (**Figure 20a**) (Quentin *et al.*, 1995). On observe l'existence de trois lignées principales : une première regroupant CC19 et CC12, une seconde correspondant à CC17, et une troisième correspondant à CC23. Il est à noter que, comme en MLEE, CC17 (équivalent de CG49 en MLEE, **Tableau VI**) est éloigné de CC19 et CC12 (équivalents de CG12 et CG34 en MLEE, **Tableau VI**). En revanche, à l'inverse du MLEE, CC23 (équivalent de CG11 en MLEE, **Tableau VI**) n'est pas inclus dans une grande division phylogénique comprenant tous les CC excepté CC17, mais constitue un cluster bien individualisé.

Pour compléter l'analyse de la structure génétique de la population, les relations phylogénétiques entre les souches ont été déterminées à partir des distances génétiques existant entre les séquences nucléotidiques du supergène obtenu par la concaténation des sept *loci*. Les séquences nucléotidiques disponibles pour les 113 souches ont été analysées selon les différents algorithmes du logiciel MEGA version 4.1, permettant de confirmer l'existence de la troisième lignée clonale CC23. Cependant sa position est différente de celle observée par analyse des profils alléliques, puisque CC23 se branche dans la même division phylogénétique que le CC17, que ce soit par la méthode UPGMA ou par la méthode NJ (**Figure 20b**). Ainsi, la position phylogénétique de CC23 reste non résolue avec les outils bioinformatiques usuels dédiés à l'analyse MLST (START et MEGA).

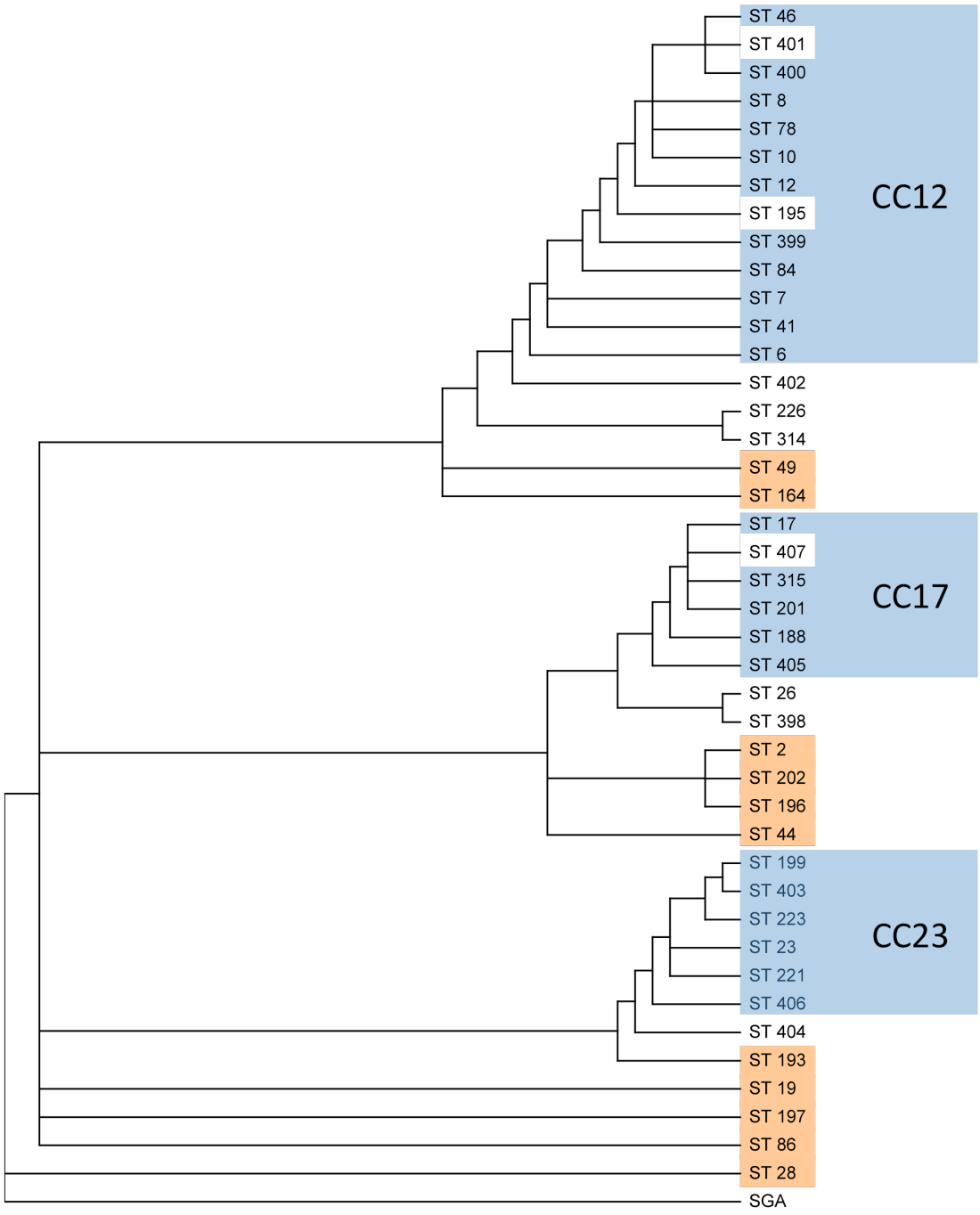


Figure 21. Arbre phylogénétique reconstruit à l'aide du logiciel PAUP version 4 beta 10 avec la méthode de maximum de vraisemblance à partir des séquences concaténées générées après la combinaison des sept loci étudiés dans le système MLST (Jones *et al.*, 2003). Il montre les relations phylogénétiques des 42 profils alléliques (ST) définis par MLST pour les 113 souches de *S. agalactiae*. Les séquences orthologues de *Streptococcus pyogenes* (SGA) ont été prises comme outgroup. Les ST membres du complexe clonal CC19 sont surlignés en orange, ceux de CC12, CC17 et CC23 en bleu.

5.2. Positionnement de CC23

Afin de préciser la position phylogénétique de CC23, nous avons déterminé le modèle d'évolution le plus approprié grâce au logiciel ModelTest version 3.08, en utilisant hLRT comme critère. Les paramètres du modèle d'évolution le mieux adapté au supergène ont ensuite été saisis dans le logiciel PAUP* version 4.0 beta 10 pour déterminer le meilleur arbre en analyse ML (**Figure 21**). Le mode de regroupement des souches confirme la nature hautement clonale de CC17 (CG39 ou division I en MLEE) et de CC12. L'existence du troisième groupe monophylétique composé des souches du CC23 est confirmée par cette dernière analyse qui positionne CC23 dans une division phylogénétique spécifique (**Figure 21**). La différence majeure par rapport au MLST est la disparition de CC19 car les souches qui sont censées le composer sont dispersées dans l'ensemble des branches de l'arbre phylogénétique. Cette observation concorde avec les faibles bootstraps observés pour CC19 avec le logiciel MEGA (31% en UPGMA et 24% en NJ, **Figure 20**), mais elle interroge sur l'existence réelle du complexe clonal CC19. Au total, l'analyse en maximum de vraisemblance conforte le MLST en identifiant la lignée clonale CC23, lignée qui n'avait pas été mise en évidence par MLEE (Quentin *et al.*, 1995).

5.3. Niveau de recombinaison dans la population

5.3.1. Evaluation de la pertinence de l'analyse eBURST

L'identification des complexes clonaux et de leur ST fondateur a été réalisée en MLST à l'aide du logiciel eBURST (voir § 2.). La fiabilité de l'analyse par eBURST a été testée en appliquant la série des trois tests simples préconisés par Turner *et al.* sur l'ensemble des données MLST disponibles au niveau mondial (**Figure 22**) (Turner *et al.*, 2007):

1) la représentation graphique (« snap-shot ») des populations de *S. agalactiae* met en évidence la présence d'un très grand groupe, le CC19 (**Figure 22a**),

2) la mise en évidence des liens de type SLV non représentés dans la première représentation graphique (en rose sur la **Figure 22b**) montre que certains relient des points (correspondant à des ST) éloignés au sein du CC19,

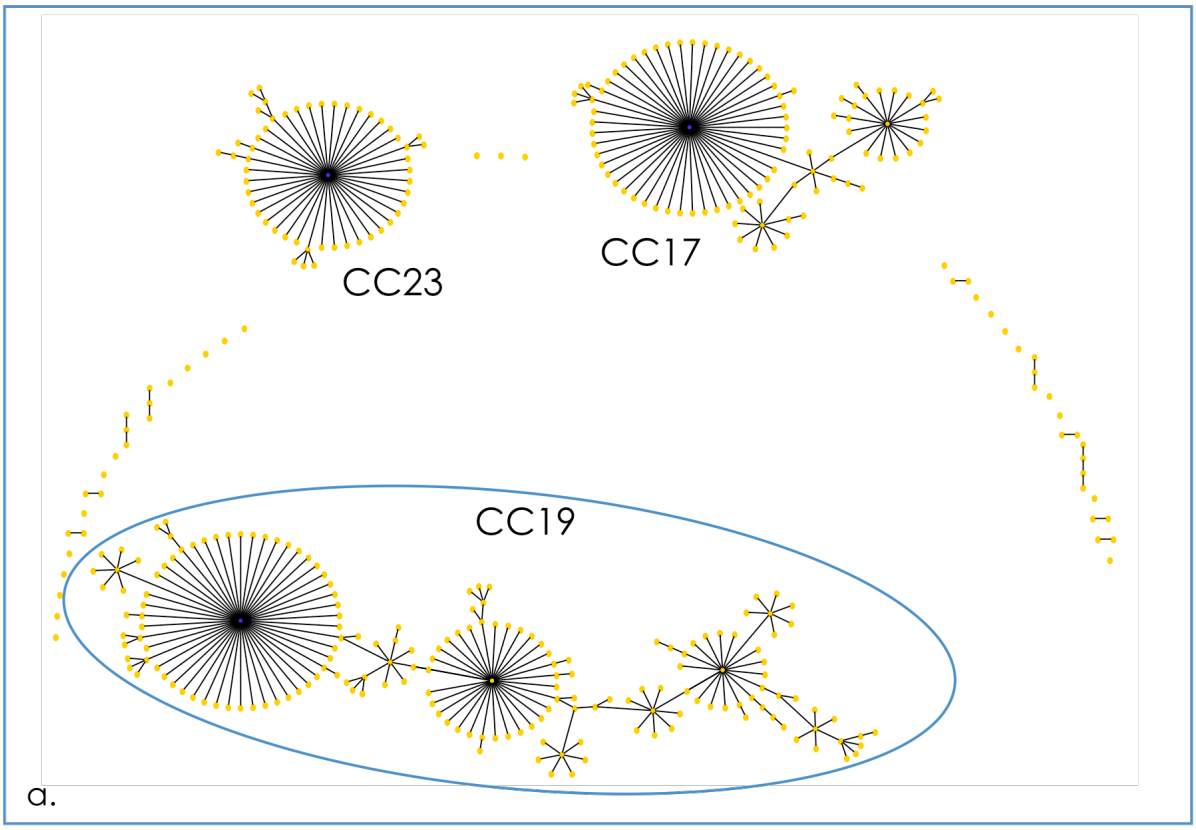


Figure 22. Analyse de la structure des populations de *S. agalactiae* déterminée par MLST après analyse avec le logiciel eBURST v3 (http://eburst.mlst.net) (Feil *et al.*, 2004).

a) La représentation graphique « snap-shot » de la population mondiale (277 ST en août 2009) montre la présence d'un complexe clonal prédominant, CC19. CC19 regroupe à lui seul plus de 25% des ST de la population.
b) La révélation des liens masqués (en rose) de type « SLV » (variant simple locus) fait apparaître des relations entre des ST éloignés au sein du CC19.

3) le calcul de la proportion des ST inclus dans le CC19 démontre qu'il représente à lui seul plus de 25% des ST de la population (181 ST sur les 377 de la population, soit 48%).

Ces tests sont en faveur d'un niveau important de recombinaison en comparaison du niveau de mutation ce qui pourrait minorer les performances de l'analyse eBURST dans le décryptage des populations de *S. agalactiae*.

5.3.2. Analyse de la structure de la population

Le logiciel SplitsTree4 a été utilisé pour préciser les relations évolutives liant les 113 souches de la collection. La représentation sous forme de réseau, obtenue d'après l'analyse par neighbor-net des 46 sites informatifs en parcimonie (retenus sur les 3456 nucléotides du supergène), montre l'existence de nombreuses recombinaisons, notamment au niveau de CC19 (**Figure 23**). Le test PHI, réalisé grâce à SplitsTree4, montre l'existence statistiquement significative de recombinaisons homologues au sein de la population (*P*=3,48E-13).

5.3.3. Evaluation du ratio recombinaison/mutation

La méthode de Feil *et al.* permet d'évaluer directement le taux de recombinaison et le taux de mutation à partir des données MLST (Feil *et al.*, 2000). Cette méthode a été appliquée sur les quatre CC définis par eBURST dans la population, CC19, CC12, CC17 et CC23 (**Tableau VII**).

Dix-sept SLV ont été retrouvés sur l'ensemble des CC. Sur les 17, dix étaient différents du ST fondateur de plus de deux nucléotides et sept d'un seul nucléotide. Sur les sept variants uni-sites, trois avaient un polymorphisme allélique partagé par des ST provenant d'un autre CC. Au total, en appliquant la théorie de Feil *et al.* à notre population de 113 souches (Feil *et al.*, 2000), on évalue le nombre de recombinaisons homologues à 13, contre quatre mutations ponctuelles. Ceci correspond à un ratio recombinaison-mutation de l'ordre de 3 : 1 sur l'ensemble de la population de *S. agalactiae*, ce qui confirme à la fois le test PHI et les tests de Turner *et al.* (Turner *et al.*, 2007) en faveur d'un haut niveau de recombinaison homologue dans la population de *S. agalactiae*. Il existe de grandes différences en fonction des CC analysés : pour le CC19 et CC12 le ratio est à 4 : 1, tandis que pour le CC17 le ratio est à 1 : 2.

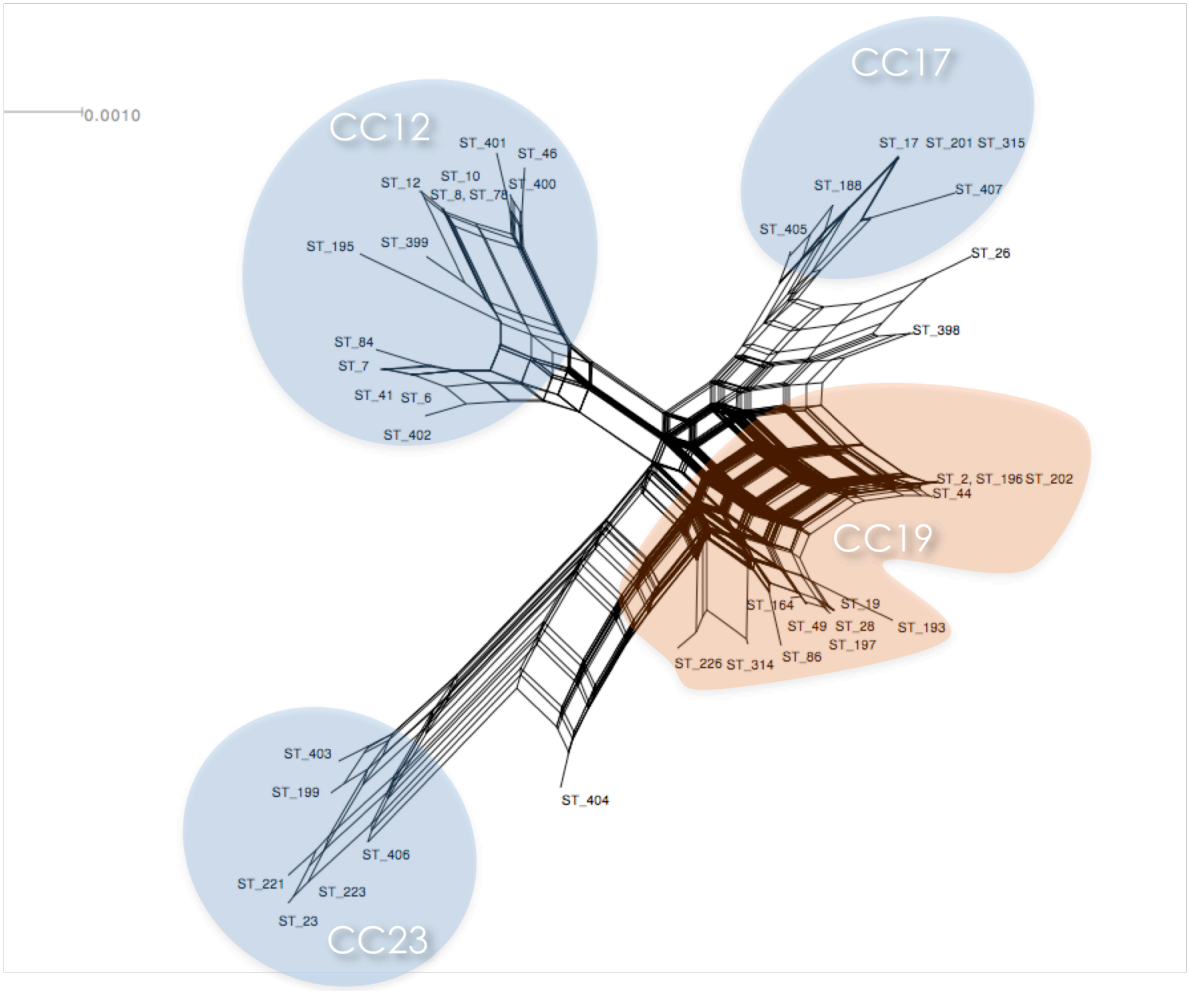


Figure 23. Représentation de la structure de la population de *S. agalactiae* sous forme de réseau phylogénétique interconnecté. Le réseau a été obtenu à l'aide du logiciel Splitstree4 (http://www.splitstree.org/) (Huson et Bryant, 2006), en appliquant la méthode neighbor-net sur les 46 sites informatifs retrouvés dans les séquences nucléotidiques du supergène du système MLST développé par Jones *et al.* (Jones *et al.*, 2003). La représentation sous forme de réseau réticulé permet de mettre en évidence les signaux qui perturbent la phylogénie représentée sous forme de dendrogramme, tels que ceux occasionnés par les recombinaisons homologues. La figure met en évidence le très fort maillage (représenté par les nombreux parallélogrammes) existant entre les profils alléliques (ST) du CC19 (surligné en orange).

Tableau VII. Répartition des variants simple-locus (SLV) des quatre complexe clonaux (CC) en fonction de la différence du nombre de nucléotides (un ou plus) avec la séquence du sequence type (ST) fondateur du CC.

Complexe clonal	Profil allélique MLST								Nombre de SLV ayant un seul nucléotide (nt) différent		Nombre de SLV ayant plusieurs nt différents	
	ST	*adhP*	*pheS*	*atr*	*glnA*	*sdhA*	*glcK*	*tkt*	nouveau	partagé	nouveau	partagé
CC19												
ST fondateur	19	1	1	3	2	2	2	2				
SLV	28	1	1	3	5	2	2	2	1	0	0	0
	86	1	1	3	2	2	1	2	0	0	0	1
	164	1	1	2	2	2	2	2	0	1	0	0
	193	1	1	3	2	2	2	1	0	0	0	1
	197	1	1	3	1	2	2	2	0	1	0	0
								sous-totaux	*1*	*2*	*0*	*2*
CC12												
ST fondateur	12	10	1	4	1	3	3	2				
SLV	8	4	1	4	1	3	3	2	0	0	1	0
	10	9	1	4	1	3	3	2	1	0	0	0
	46	1	1	4	1	3	3	2	0	0	0	1
	399	10	1	4	1	3	1	2	0	0	0	1
	400	12	1	4	1	3	3	2	0	0	1	0
								sous-totaux	*1*	*0*	*2*	*2*
CC17												
ST fondateur	17	2	1	1	2	1	1	1				
SLV	188	2	1	1	2	1	1	2	0	0	0	1
	201	2	1	25	2	1	1	1	1	0	0	0
	315	2	1	1	2	31	1	1	1	0	0	0
								sous-totaux	*2*	*0*	*0*	*1*
CC23												
ST fondateur	23	5	4	6	3	2	1	3				
SLV	199	5	4	6	3	2	1	2	0	0	0	1
	221	5	4	6	3	2	2	3	0	0	0	1
	223	5	4	6	3	2	1	1	0	1	0	0
	406	5	4	6	6	2	1	3	0	0	1	0
								sous-totaux	*0*	*1*	*1*	*2*
								totaux	*4*	*3*	*3*	*7*

Partie 1. Structure génétique de *Streptococcus agalactiae*

D-Discussion

Dès la publication de leur premier article décrivant le MLST, Enright et Spratt s'interrogeaient sur les outils bioinformatiques permettant une exploitation optimale des données générées par cette technique (Enright & Spratt, 1999, Maiden *et al.*, 1998). Nos résultats illustrent cette difficulté chez *S. agalactiae* : l'analyse réalisée à partir des profils alléliques par START présente CC23 comme une lignée bien individualisée (**Figure 20a**), alors que l'analyse des séquences nucléotidiques du supergène par MEGA le branche dans le même cluster que CC17 (**Figure 20b**). Certes, cette discordance a déjà été observée dans la littérature, puisque, selon les travaux publiés, CC23 se retrouve soit dans la même division que CC17 (Bohnsack *et al.*, 2008, Jones *et al.*, 2006b, Manning *et al.*, 2009), soit distant des autres clusters (Luan *et al.*, 2005, Jones *et al.*, 2003, Sun *et al.*, 2005). Pour tenter de clarifier cette situation, nous avons conduit une reconstruction phylogénétique en maximum de vraisemblance grâce au logiciel PAUP. L'utilisation de l'application Modeltest a permis de sélectionner le modèle évolutif le mieux adapté aux séquences nucléotidiques parmi 56 modèles, sans *a priori*. Cette nouvelle analyse confirme l'existence des complexes clonaux CC12, CC17 et CC23, et démontre que CC23 est distant de CC17 et CC12. Il est intéressant de noter que ce résultat est en accord avec l'analyse menée en utilisant la technique de RAPD (technique qui explore des évènements génétiques plus nombreux que le MLST), qui montrait une clusterisation particulière des souches ET11, équivalent du CC23 en MLST (Chatellier *et al.*, 1997).

L'analyse avec PAUP a également montré que la clonalité de CC19 n'était pas soutenue en maximum de vraisemblance : les souches qui composent CC19 sont dispersées au sein du dendrogramme (**Figure 21**), suggérant que le fond génétique de CC19 est partagé par la majorité de la population. Les faibles valeurs de bootstraps que nous avons observées pour CC19 (**Figure 21**), et qui sont comparables à celles retrouvées par Manning *et al.* (Manning *et al.*, 2009), corroborent ce résultat. De même, dans le travail de Brochet *et al.*, le dendrogramme issu de l'analyse comparative des génomes par hybridation (CGH) montre une dispersion des souches appartenant au CC19 qui se répartissent dans différents clusters (Brochet *et al.*, 2006). *A contrario*, il existe une bonne congruence des deux systèmes de typage CGH et MLST pour les souches appartenant à CC23, CC12 et CC17. Dans le modèle d'évolution proposé ultérieurement par les auteurs, un seul clone, génétiquement proche de CC19, aurait disséminé dans la population

et les différents complexes clonaux en seraient dérivés après remplacement de grandes régions chromosomiques (Brochet *et al.*, 2008b).

Une hypothèse permettant d'expliquer la dispersion des souches du CC19 est fournie par les tests de Turner, qui montrent un niveau de recombinaison élevé comparé au niveau de mutation. Ceci pourrait minorer les performances de l'analyse eBURST dans le décryptage des populations de *S. agalactiae* (Turner *et al.*, 2007). Le test PHI et le calcul du ratio recombinaison/mutation confirment le haut niveau de recombinaison dans la population analysée, ce qui recoupe les résultats des analyses par approche génomique (Tettelin *et al.*, 2005, Brochet *et al.*, 2008b). Le ratio recombinaison/mutation calculé pour chaque complexe clonal démontre que le niveau de recombinaison est inégal entre les lignées. Il est supérieur pour CC19 (4:1) en comparaison de CC17 (1:2), pour lequel les mutations ponctuelles sont à l'origine de l'émergence de la lignée, ce que confirme la haute clonalité attribuée à CC17 (Bisharat *et al.*, 2004, Jones *et al.*, 2006b). Le haut niveau de recombinaison observé pour CC19 serait à l'origine de l'extinction de certains signaux phylogénétiques et à l'instabilité de ce clone dans le temps. Les différences de niveau de recombinaison observées entre les génotypes pourraient provenir des différences de régulation ou du niveau d'échanges de grands fragments d'ADN (Brochet *et al.*, 2006).

Le niveau important des recombinaisons dans la population de *S. agalactiae*, ainsi que la faible diversité génétique des séquences analysées en MLST (Manning *et al.*, 2009), expliquent la résolution imparfaite de la reconstruction phylogénétique. La complexité des relations phylogénétiques entre les différents clusters, masquée par l'apparente simplicité des dendrogrammes obtenus le plus souvent en UPGMA (méthode de reconstruction simplificatrice de la phylogénie), a été soulignée par Bisharat *et al.* (Bisharat *et al.*, 2004). L'exemple de *S. agalactiae* illustre en ce sens la difficulté d'obtenir un dendrogramme dans lequel toutes les relations aient un sens phylogénétique, d'autant que le dendrogramme est déjà une représentation simplifiée des relations entre des souches de génotypes proches (Feil *et al.*, 2001, Feil & Spratt, 2001, Spratt *et al.*, 2004). Lefébure et Stanhope, qui ont étudié l'évolution de 26 génomes de streptocoques, ont indiqué que la plupart des branches du dendrogramme représentant les liens entre les huit génomes séquencés de

S. agalactiae étaient non résolues (Lefebure & Stanhope, 2007). La représentation de la structure de la population sous forme de réseau interconnecté se révèle alors particulièrement contributive (Huson & Bryant, 2006). Ainsi, l'utilisation de SplitsTree4 permet de faire apparaître les signaux phylogénétiques contradictoires, tels que ceux occasionnés par les recombinaisons, les réseaux se révélant particulièrement denses au niveau de CC19 (**Figure 23**).

Certains auteurs mettent en cause l'exploitation des données MLST issues des études épidémiologiques pour l'interprétation de la biologie des populations en arguant du fait qu'elles compliquent l'utilisation du MLST en microbiologie appliquée (Maiden, 2006). Au contraire d'autres auteurs défendent l'idée selon laquelle l'augmentation des données permettra à terme de corréler lignées phylogénétiques et virulence (Turner & Feil, 2007). Ainsi, notre analyse par MLST a mis en évidence l'existence de la lignée clonale CC23, alors qu'elle n'était pas différenciée par le MLEE. Ceci est d'autant plus intéressant que les souches qui composent la lignée CC23 présentent diverses particularités biologiques. En effet, les souches de sérotype Ia, qui donnent significativement plus d'infections néonatales mortelles (Phares *et al.*, 2008), se regroupent préférentiellement dans le CC23 (sur l'ensemble de notre collection 13 des 16 souches de sérotype Ia appartiennent à CC23 qui compte 14 souches au total) (**Tableau II**). Cette corrélation CC23-Ia a d'ailleurs été établie dans d'autres études (Martins *et al.*, 2007, Luan *et al.*, 2005, Jones *et al.*, 2006b, Bohnsack *et al.*, 2008, van der Mee-Marquet *et al.*, 2008). De plus, le potentiel infectieux de la lignée CC23 a été associé aux infections néonatales précoces (Martins *et al.*, 2007). Mais d'autres spécificités biologiques ont été récemment rapportées chez les souches appartenant à CC23 par différents travaux réalisés dans notre équipe. Premièrement, seules les souches *S. agalactiae* de CC23 résistent à l'attaque des phages isolés des souches de CC17 (Domelier *et al.*, 2009). Deuxièmement, elles présentent en spectrométrie de masse une plus grande diversité de profils des masses moléculaires des protéines exprimées par rapport aux souches des autres complexes clonaux (Lartigue *et al.*, 2009). Troisièmement, en lien plus direct avec le potentiel de virulence, les souches *S. agalactiae* de CC23 sont les seules pour lesquelles il existe une discordance entre leur faible capacité d'adhésion au fibrinogène humain, et la forte prévalence du gène *fbsB* (Rosenau *et al.*, 2007).

Partie 2.

Corrélation entre la phylogénie de *Streptococcus agalactiae* et la distribution des EGM

Partie 2. Corrélation entre la phylogénie de *Streptococcus agalactiae* et la distribution des EGM

A-Introduction

L'évolution singulière de l'épidémiologie des infections à *S. agalactiae* durant le siècle dernier suggère un lien possible entre les réservoirs humain et bovin. *S. agalactiae* a d'abord été reconnu vers la fin du XIXe siècle par Nocard et Mollereau comme le «streptocoque des mastites bovines» (Nocard & Mollereau, 1887). Par la suite, avec l'apparition des antibiotiques et leur utilisation en médecine vétérinaire, les mastites à *S. agalactiae* ont considérablement régressé dans la plupart des pays industrialisés (Keefe *et al.*, 1997). De manière presque concomitante, *S. agalactiae* est apparu comme une cause majeure d'infection chez le nouveau-né (Eickhoff *et al.*, 1964), et demeure à l'heure actuelle la première cause d'infection néonatale (Phares *et al.*, 2008). Plus récemment, cette bactérie a également émergé chez les sujets immunodéprimés (Edwards & Baker, 2005, Skoff *et al.*, 2009, Ekelund *et al.*, 2005). Par ailleurs, l'ancestralité bovine d'une lignée clonale humaine hautement invasive a récemment été soulevée (Bisharat *et al.*, 2004), hypothèse qui reste toutefois discutée (Brochet *et al.*, 2006).

Or, l'origine des souches à haut potentiel invasif reste incertaine. La description phylogénétique des souches à potentiel invasif ne suffit pas à caractériser de manière exhaustive les lignées à haut risque infectieux. De même, les principaux facteurs de virulence de *S. agalactiae* connus à ce jour sont présents chez l'ensemble des souches, à la différence d'autres espèces bactériennes pathogènes pour l'homme telles que *Staphylococus aureus* ou *Escherichia coli*. L'exploration de la structure génétique de l'espèce doit donc être poursuivie afin de déterminer de nouveaux marqueurs susceptibles d'apporter un éclairage sur l'évolution et l'émergence des lignées invasives de *S. agalactiae*.

Compte tenu de l'influence des éléments génétiques sur la plasticité et l'évolution des génomes bactériens (Kazazian, 2004), il était licite d'explorer l'hypothèse selon laquelle les EGM pouvaient constituer des marqueurs voire des moteurs de l'évolution de *S. agalactiae*. Pour confirmer cette hypothèse, nous avons déterminé la prévalence de divers EGM identifiés chez *S. agalactiae* au sein d'une collection de souches d'origine humaine et bovine. Ces résultats ont été interprétés à la lumière de la structure génétique de la population déterminée par MLST, la première partie du travail expérimental ayant validé cette approche par comparaison avec le MLEE. En établissant un lien entre EGM et phylogénie, nos résultats ont permis de formuler une hypothèse pour expliquer la chronologie d'acquisition des EGM au cours de l'évolution de l'espèce *S. agalactiae*.

Partie 2. Corrélation entre la phylogénie de *Streptococcus agalactiae* et la distribution des EGM

B-Matériel et Méthodes

Tableau VIII. Origines écologiques des 98 souches de *S. agalactiae* de la collection issues des réservoirs humain et bovin. Cette collection comprend 21 souches isolées de LCR, 17 souches isolées dans un contexte d'endocardite, 20 souches de colonisation vaginale, 5 souches isolées de liquide gastrique et 35 souches isolées dans un contexte de mammite. Pour chaque souche sont indiqués le sérotype et le complexe clonal (CC) d'appartenance.

Souches humaines					Souches bovines			
Invasives	**LCR :**				**Mammites**			
		L1	III	CC17		N1	NT	
		L9	III	CC19		N2	II	CC61
		L10	III	CC19		N3	NT	
		L13	III	CC17		N6	NT	
		L14	III	CC19		N7	III	CC19
		L15	IV	CC2		N8	II	CC61
		L16	III	CC23		N9	V	
		L17	III	CC7		N10	II	
		L19	Ia	CC23		N11	Ib	CC7
		L21	III	CC17		N12	III	CC23
		L22	III	CC19		N13	NT	CC61
		L23	III	CC17		N14	NT	CC2
		L24	III	CC17		N15	NT	CC2
		L29	III	CC19		N16	NT	CC61
		L34	III	CC19		N17	II	CC61
		L36	III	CC19		N18	Ia	CC7
		L44	Ia	CC23		N19	NT	CC61
		LJ1	III	CC19		N23	II	
		LJ4	III	CC19		N24	NT	CC2
		LJ5	Ib	CC10		N25	III	CC61
		LJ6	III	CC19		N26	III	CC23
						N27	II	CC2
	Endocardites :					N28	NT	CC2
		H1	III	CC17		N29	III	CC23
		H2	NT	CC2		N30	IV	CC2
		H3	NT	CC2		N31	III	
		H4	Ib	CC10		N32	NT	CC61
		H5	III	CC17		N33	III	
		H6	III	CC23		N34	III	CC23
		H7	II	CC10		N35	NT	CC61
		H8	III	CC17		N36	II	CC61
		H9	V	CC19		N37	III	CC23
		H10	Ib	CC10		N38	Ib	CC10
		H11	NT	CC2		N39	NT	CC61
		H12	III	CC19		N40	NT	CC61
		H13	Ib	CC10				
		H14	II					
		H15	III	CC23				
		H16	III	CC23				
		H17	Ib	CC10				
Colonisation	**Vaginales :**							
		V4	Ia					
		V5	Ib	CC7				
		V7	Ib	CC10				
		V18	III	CC17				
		V20	III	CC17				
		V21	II	CC19				
		V23	NT	CC17				
		V24	IV	CC7				
		V25	NT	CC19				
		V64	III	CC17				
		V71	Ib	CC10				
		V80	II	CC10				
		V88	II	CC10				
		V101	V	CC7				
		V104	II	CC19				
		V114	Ib	CC10				
		VV2	III	CC19				
		VV4	III	CC19				
		VV105	III	CC19				
		VV188	III	CC19				
	Gastriques :							
		G8	Ia	CC7				
		G35	NT	CC19				
		G49	III	CC17				
		G61	V	CC2				
		G64	II	CC2				

1. Souches bactériennes

La collection comprenait 98 souches de *S. agalactiae*, épidémiologiquement non reliées, isolées pour la plupart en France entre 1966 et 2003, et issues de deux réservoirs écologiques distincts, l'Homme et le bovin (**Tableau VIII**).

1.1. Souches humaines

A partir du souchier initial composé de 113 souches humaines utilisées dans la précédente étape de notre travail (voir Partie 1, §1.), nous avons sélectionné 38 souches les plus représentatives de la diversité de la population. Nous avons secondairement complété ce souchier avec 25 nouvelles souches (dont 17 souches responsables d'endocardites).

Ainsi, nous disposions de 63 souches humaines provenant de prélèvements réalisés à visée diagnostique ou de dépistage, avec la répartition suivante : 21 souches responsables de méningites néonatales précoces, 17 souches responsables d'endocardites de l'adulte, 5 souches isolées de liquide gastrique de nouveaux-nés asymptomatiques, et 20 souches de colonisation vaginale. Au total, le souchier humain comprenait 38 souches invasives et 25 souches de colonisation.

1.2. Souches bovines

Les souches bovines nous ont été amicalement transmises par Pascal Rainard de l'INRA de Nouzilly (Laboratoire d'Infection et d'Immunité Mammaire, Unité de Pathologie Infectieuse et Immunologie). Ces 35 souches provenaient de prélèvements de mamelle réalisés dans un contexte de mammite chez l'animal en France, en Allemagne, aux Etats-Unis, en Grande-Bretagne et en Malaisie.

Les souches d'origine bovine, conservées sous forme lyophilisée, ont été réhydratées puis cultivées sur gélose trypticase-soja à 5% de sang de cheval (bioMérieux), pendant 24 à 48 heures, à 37°C, en aérobiose, sous atmosphère enrichie à 5,5 % de CO_2. L'identification de l'espèce *S. agalactiae* a été contrôlée par l'étude morphologique des colonies, la présence d'une hémolyse de type β et la réaction de catalase, puis confirmée par le sérogroupage en B réalisé avec le kit Slidex Strepto Kit® (bioMérieux) et la positivité du CAMP-test réalisé à l'aide de la souche de référence de *Staphylococcus aureus* CIP 65.8 (Darling, 1975). Les souches ont ensuite été conservées par congélation à –80°C en milieu Schaedler vitaminé K3

(bioMérieux, Marcy l'Etoile, France) à 10% de glycérol jusqu'à utilisation.

2. Extraction de l'ADN génomique bactérien

Les ADN des souches bactériennes ont été extraits, dosés et conservés comme précédemment décrit (voir Partie 1, §2.2).

3. Sérotypage des souches

Le sérotypage capsulaire a été réalisé par méthode génotypique selon la technique décrite par Kong *et al.* permettant la détermination des sérotypes Ia, Ib, III, IV, V et VI (Kong *et al.*, 2002). Le sérotype II a été testé par agglutination de particules de latex au moyen du kit Pastorex selon les instructions du fabricant (Biorad, Hercules, USA).

3.1. Amplification par PCR des gènes responsables des sérotypes

Les amorces utilisées pour le sérotypage moléculaire sont indiquées dans le **Tableau IX**. Les conditions optimales d'amplification ont été obtenues, dans un volume de 50 µL, avec le mélange réactionnel suivant : 10 mM de tampon PCR (Tris-HCl, KCl 50 mM, pH 8,3) (Applied Biosystems, Foster City, USA), 1,5 mM de $MgCl_2$ (Applied Biosystems), 200 µM de chacun des dNTPs (Amersham Biosciences, Buckinghamshire, England), 0,5 µM de chaque amorce (Eurogentec, Seraing, Belgique), et 2,5 unités d'*AmpliTaq* DNA Polymerase (Applied Biosystems), auxquels ont été ajoutés 50 ng d'ADN. La réaction de PCR a été réalisée dans un thermocycleur GeneAmp PCR System 9600 (Perkin-Elmer, Norwalk, USA). Après une étape initiale de dénaturation de l'ADN à 94°C pendant une minute et trente secondes, 30 cycles ont été réalisés, chacun comprenant une dénaturation à 94°C pendant 30 secondes, une hybridation des amorces à 60°C pendant 30 secondes et une élongation de 4 minutes à 72°C. Une élongation de 10 minutes à 72°C a été ajoutée en étape finale.

Tableau IX. Amorces utilisées pour le sérotypage par PCR des souches de *S. agalactiae*.

Séro-type	Amorces (5' à 3')*	Ampli-con (pb)	Hybrida-tion (T°C)
Ia	GGCCTGCTGGGATTAATGAATATAGTTCCAGGTTTGC (s) GTATAACTTCTATCAATGGATGAGTCTGTTGTAGTACGG (as)	354	55
Ib	GATAATAGTGGAGAAATTTGTGATAATTTATCTCAAAAAGACG (s) CCTGATTCATTGCAGAAGTCTTTACGATGCGATAGGTG (as)	523	55
III	GAATACTATTGGTCTGTATGTTGGTTTTATTAGCATCGC (s) GTATAACTTCTATCAATGGATGAGTCTGTTGTAGTACGG (as)	641	55
IV	CCCAAGTATAGTTATGAATATTAGTTGGATGGTTTTTGG (s) GGGTCAATTGTATCGTCGCTGTCAACAAAACCAATCAAATC (as)	379	55
V	CCCAGTGTGGTAATGAATATTAGTTGGCTAGTTTTTGG (s) CCCCCCATAAGTATAAATAATATCCAATCTTGCATAGTCAG (as)	374	55
VI	CCTTATTGGGCAAGGTATAAGAGTTCCCTCCAGTGTG (s) GAAGCAAAGATTCTACACAGTTCTCAATCACTAACTCCG (as)	360	55

* (s), amorce sens ; (as), amorce anti-sens (d'après Kong *et al.*, 2002)

3.2. Migration et révélation des produits amplifiés

Les amplicons ont été séparés par électrophorèse en gel d'agarose (Eurobio, Les Ulis, France) à 1% (m/v), en tampon TBE 1X (Tris-base 90 mM, acide borique 89 mM, EDTA 2 mM, pH 8), pendant 45 minutes à voltage constant de 110 Volts. Au moment du dépôt, cinq microlitres de tampon de charge (glycérol 30%, bleu de bromophénol 0,25%, tampon Tris-base 40 mM, pH 8, EDTA 2 mM) ont été ajoutés à chacun des échantillons. En parallèle des échantillons, le marqueur de taille moléculaire Smart Ladder (Eurogentec) a été déposé afin de contrôler la taille des fragments obtenus. Les amplicons ont été révélés par transillumination U.V. après immersion du gel dans une solution de bromure d'éthidium à 0,1 µg/mL.

4. Prévalence des éléments génétiques mobiles (EGM)

Nous avons déterminé la prévalence de 11 éléments génétiques, apparentés aux EGM, au sein de notre souchier. Parmi ces éléments, sept sont des séquences d'insertion déjà décrites chez *S. agalactiae* (IS*1548*, IS*861*, IS*1381*, IS*Sa4*, IS*Sag1*, IS*Sag2* et IS*Sag*9), trois sont des gènes (IS*Sag9*, SAG00448, SAG1241 et SAG0610) codant une transposase et correspondant à des restes d'IS appelés « IS remnants », et un est un intron du groupe II, GBSi1. Pour des raisons de clarté, nous avons choisi de regrouper ces 11 éléments sous l'appellation « EGM », même si tous ne répondent pas à la définition *stricto sensu* (voir Revue de la littérature, Partie B. § 1.2).

4.1. Identification des EGM par PCR

Les amorces utilisées pour les gènes étudiés sont présentées au **Tableau X**. L'amplification a été obtenue dans un volume de 50 µL grâce au mélange réactionnel détaillé au paragraphe 2.3 et le programme le suivant : une étape initiale de dénaturation de l'ADN à 94°C pendant une minute et trente secondes; 30 cycles comprenant chacun une dénaturation à 94°C pendant 30 secondes, une hybridation des amorces à la température d'hybridation (Tableau 2-III) pendant 30 secondes et une élongation de 1 minutes à 72°C ; une étape finale d'élongation à 72°C de 10 minutes. La taille des amplicons obtenus a été estimée, après électrophorèse en gel d'agarose (Eurobio) à 2% et tampon TBE 1X (Tris-base 90 mM, acide borique 89 mM, EDTA 2 mM, pH 8) pendant 1 heure à 90V, marquage par le bromure d'éthidium et visualisation en transillumination sous UV et comparaison avec le marqueur de taille 100 pb (Amersham Biosciences).

Tableau X. Amorces utilisées pour l'amplification par PCR des éléments génétiques mobiles (EGM) de *S. agalactiae*.

EGM	Code d'accès Genbank*	Amorces (5' à 3')**	Ampli-con (pb)	Hybri-dation (T°C)	Référence
IS*1381*	AF064785	GTTAAGCTTAGAAGATCTCCTCATGG (s) CCTTGATAACCACTGTCTGCCA (as)	425	55	(Dmitriev *et al.*, 2004)
IS*861*	M22449	ATGACAGAGCTCGAGCGACTT (s) TATCAGCCTTCTTACCAACCTCA (as)	447	55	(Dmitriev *et al.*, 2004)
IS*1548***	Y14270	TTGCGCAGTTGAATTGGATAG (s) TTCTCTAACTTCAATCTGTCCCCTA (as)	218	55	(Dmitriev *et al.*, 2004)
IS*1548***	Y14270	CCAAACGCATGAAGAAACTTTTGTTTTCAG (s) CTTCATCCTTTTGTGCCCGCACATC (as)	1287	52	(Granlund *et al.*, 1998)
GBSi1	AJ292930	CGAGTGGATTGTGGACATTG (s) GTGATAGATTCCCGCCTTGA (as)	197	55	(Bidet *et al.*, 2003)
IS*Sa4*	AF165983	CTGAGCTAGGTATTAAGTCGCAAC (s) GCGTGTCTCAATGGTTGGT (as)	563	55	(Dmitriev *et al.*, 2004)
IS*Sag2*	AF329275	CACACTTTCACGTCGTCGG (s) TGTTTATCAAGAAACCTTTCA (as)	504	50	(Franken *et al.*, 2001)
IS*Sag1*	AF329276	GTACTGTACAGGCGACTGGTC (s) CTGGCTTTGTCGGACACGTAC (as)	563	55	(Franken *et al.*, 2004)
IS*Sag9*	NC_004116	GACCGTAATGACATCCAATCTGG (s) GGCTTGAAATGGACACGGTT (as)	600	50	Cette étude
SAG0448	NC_004116	CCACAGAATTACTTAACTTCCTAGC (s) CGACCAACCTTATTGAATCACT (as)	994	50	Cette étude
SAG1241 (*orfA*)	NC_004116	CTCTTGTTAACCTTTATCAAACTGG (s) CCTAAAAAAAGCAAGTGCCA (as)	218	50	Cette étude
SAG0610	NC_004116	TACTCCACATTAGCTAAAGAACAGG (s) GGCCTGATTTAACCGAGATT (as)	394	50	Cette étude

* NC_004116 fait référence à la souche 2603 V/R (REF), ** (s), amorce sens ; (as), amorce anti-sens, *** la présence de l'IS*1548* a été recherchée par les deux couples d'amorces décrits dans la littérature

4.2. Séquençage des fragments amplifiés

Nous avons confirmé la spécificité des produits amplifiés par séquençage. Ainsi, pour chacun des 11 EGM, les produits de PCR ont été séquencés avec le kit BigDye™ Terminator Mix v.3.0 et analysés avec le séquenceur capillaire automatique ABI Prism® 3100-*Avant* Genetic Analyser (Applied Biosystems). L'interrogation de la banque de données du National Center for Biotechnology Investigation avec le programme BLAST (http://www.ncbi.nih.gov/blast) a permis de contrôler les séquences nucléotidiques des EGM amplifiés.

5. Analyse de la structure génétique de la population par multilocus sequence typing

L'analyse par MLST a été conduite selon la méthodologie précédemment décrite dans la première partie (voir Partie 1, § 5.)

6. Analyses statistiques

6.1. Analyse factorielle des correspondances

Une analyse factorielle des correspondances (AFC) a été réalisée en appliquant la méthodologie décrite par Picard *et al.* (Picard *et al.*, 1999). La méthode AFC est une méthode statistique d'analyse des données dont le principe est de n'avoir aucun *a priori* sur les données et de les décrire en analysant la hiérarchisation de l'information présente dans celles-ci. En cela, elle se rapproche de l'analyse en composante principale, mais elle en diffère par l'utilisation d'une matrice de covariance basée sur les distances du chi-deux et non sur les distances euclidiennes. L'AFC étudie l'inertie du nuage de points ayant pour coordonnées les valeurs présentes sur les lignes d'un tableau de données et décrit la dispersion et la forme du nuage de points représentant n objets (ici, les souches de la collection) dans un espace multidimensionnel, après avoir centré et réduit le tableau de données originales que l'on affecte d'un système de masse. Les variances attendues se hiérarchisent par ordre décroissant, axe d'inertie par axe d'inertie, en fonction de ces nouvelles coordonnées. Un plan est déterminé par les deux principaux axes de l'analyse ; le premier axe, F1, regroupe la plus grande part de la variance et le second axe, F2, orthogonal à F1, regroupe la plus grande part de la variance non représentée par F1.

Appliquée à notre problématique, l'AFC a été réalisée à partir des données résumées dans un tableau comprenant 98 lignes (une pour chaque souche de *S. agalactiae*) avec le nombre de colonnes correspondant à tous les caractères (appelés variables) suivants : 11 colonnes correspondant aux EGM étudiés (présence/absence), trois colonnes correspondant à l'appartenance à CC17, CC19 et CC61 (oui/non), et deux colonnes correspondant à l'origine humaine ou bovine de la souche (oui/non), soit 16 colonnes au total. Une AFC a été réalisée sur ce tableau. Les 11 EGM ont été pris comme variables actives ; les CC et les origines écologiques comme variables illustratives. Les projections des variables et des souches sur le plan F1/F2 ont été analysées et les liens possibles entre les caractères recherchés à travers leur proximité sur le graphique. Le logiciel SPAD.N (Centre International de Statistiques et Informatique Appliquées, St. Mandé, France) a permis de conduire cette analyse.

6.2. Autres analyses statistiques

Les analyses statistiques restant a été conduit avec le logiciel Epi Info version 6.04 disponible en ligne (http://www.epiconcept.fr/html/epiinfo.html).

Partie 2. Corrélation entre la phylogénie de *Streptococcus agalactiae* et la distribution des EGM

1. Distribution des sérotypes

La répartition des sérotypes pour les 98 souches est détaillée dans le **Tableau XI**. Au sein des souches humaines on observe une nette prédominance du sérotype III dont la prévalence était de 50.8% (n=32), suivi des sérotypes Ib (n=9 soit 14,3% des souches humaines) et II (n=7 soit 11,1% des souches humaines). Seuls 9,5% (n=6) des souches humaines étaient non typables.

Par contraste, 40% des souches bovines (n=14) n'étaient pas typables (NT). Parmi les souches bovines typables, les six sérotypes ont été retrouvés, avec en premier lieu les sérotypes III (n=9 soit 25,7% des souches bovines) et II (n=7 soit 20% des souches bovines).

Tableau XI. Prévalence des sérotypes au sein des souches humaines et bovines de *S. agalactiae*.

Sérotype*	Nombre (%) de souches positives		
	Souches humaines (n=63)	Souches bovines (n=35)	Total (n=98)
Ia	4 (6,3)	1 (2,9)	5 (5,1)
Ib	9 (14,3)	2 (5,7)	11 (11,2)
II	7 (11,1)	7 (20)	14 (14,3)
III	32 (50,8)	9 (25,7)	41 (41,8)
IV	2 (3,2)	1 (2,9)	3 (3,1)
V	3 (4,8)	1 (2,9)	4 (4,1)
NT	6 (9,5)	14 (40)	20 (20,4)

*NT, non typable

2. Etude des éléments génétiques mobiles

2.1. Prévalence des éléments génétiques mobiles

Les 11 EGM ont été recherchés par PCR. Dans tous les cas, la taille obtenue correspondait à la taille attendue. De plus, le séquençage a confirmé la spécificité de la réaction de PCR.

L'IS*861* a été trouvée chez 80,6% des souches (n=79), GBSi1 chez 56,1% des souches (n=55), l'IS*1381* chez 51% des souches (n=50), l'IS*1548* chez 25,5% des souches (n=25), l'IS*Sa4* chez 13,3% des souches (n=13%), l'IS*Sag2* chez 90,8% des

souches (n=89), l'IS*Sag1* chez 75,5% des souches (n=74), l'IS*Sag9* chez 67,3% des souches (n=66), SAG0448 chez 63,3% des souches (n=62), SAG1241 chez 91,8% des souches (n=90) et SAG0610 chez 89,8% des souches (n=88) (Tableau 2-V).

2.2. Eléments génétiques mobiles et sérotypes

La prévalence des EGM en fonction du sérotype des souches est détaillée dans le **Tableau XII**. Il n'a pas été observé de distribution des EGM exclusive ou prédominante d'un sérotype et tous les EGM sont associés à au moins trois sérotypes et (Tableau 2-V).

Tableau XII. Prévalence des éléments génétiques mobiles en fonction des sérotypes étudiés chez les 98 souches de *S. agalactiae*.

EGM	Nombre (%) de souches							Nombre total de souches
	Ia (n=5)	Ib (n=11)	II (n=14)	III (n=41)	IV (n=3)	V (n=4)	NT* (n=20)	
IS*1381*)	4 (80)	10 (90.9)	7 (50)	17 (41.5)	2 (66.7)	3 (75)	7 (35)	50
IS*861*	3 (60)	10 (90.9)	13 (92.9)	35 (85.4)	1 (33.3)	2 (50)	15 (75)	79
IS*1548*	1 (20)	1 (9.1)	4 (28.6)	16 (39)	0 (0)	1 (25)	2(10)	25
GBSi1	2 (40)	2 (18)	11 (78.6)	22 (53.7)	1 (33.3)	2 (50)	15 (75)	55
IS*Sa4*	0 (0)	0 (0)	3 (21.4)	5 (12.2)	1 (33,3)	0 (0)	4 (20)	13
IS*Sag2*	5 (100)	10 (90,9)	12 (85,7)	38 (92,7)	3 (100)	3 (75)	18 (90)	89
IS*Sag1*	3 (60)	10 (90,9)	12 (85,7)	36 (87,8)	3 (100)	3 (75)	7 (35)	74
IS*Sag9*	4 (80)	9 (81,8)	9 (64,3)	33 (80,5)	2 (66,7)	3 (75)	6 (30)	66
SAG0448	4 (80)	4 (36,4)	9 (64,3)	24 (58,5)	3 (100)	3 (75)	15 (75)	62
SAG1241	3 (60)	10 (90,9)	12 (85,7)	39 (95,1)	3 (100)	4 (100)	19 (95)	90
SAG0610	5 (100)	11 (100)	11 (78,6)	37 (90,2)	3 (100)	4 (100)	17 (85)	88

*NT,non typable

2.3. Eléments génétiques mobiles et potentiel invasif

Les souches humaines ont été divisées en trois groupes en fonction de leur potentiel invasif : les souches responsables de méningites néonatales (n=21), les souches responsables d'endocardites (n=17) et les souches de colonisation (souches gastriques et vaginales ; n=25) (**Tableau XIII**). La distribution des EGM au sein de ces trois groupes n'a pas montré d'association préférentielle entre un EGM donné et un potentiel invasif particulier.

Tableau XIII. Prévalence des éléments génétiques mobiles (EGM) en fonction du potentiel de virulence et de l'origine écologique des souches de *S. agalactiae*.

EGM	Nombre (%) de souches positives					p^b
	Souches humaines (n=63)				Souches bovines (n=35)	
	LCR (n=21)	Endocardites (n=17)	Colonisation (n=25)	Total (n=63)	Mammites	
IS*1381* (n=50)	15 (71,4)	10 (58,8)	20 (80)	45 (71,4)	5 (14,3)	< 0,00001
IS*861* (n=79)	18 (85,7)	11 (64,7)	21 (84)	50 (79 ,4)	29 (82,9)	NS
IS*1548* (n=25)	10 (47,6)	4 (23,5)	10 (40)	24 (38,1)	1 (2,9)	< 0,001
GBSi1 (n=55)	7 (33,3)	5 (29,4)	8 (32)	20 (31,7)	35 (100)	< 0,00001
IS*Sa4* (n=13)	0 (0)	3 (17,7)	2 (8)	5 (7,9)	8 (22,9)	< 0,05
IS*Sag2* (n=89)	21 (100)	12 (70,6)	24 (96)	57 (90,5)	32 (91,4)	NS
IS*Sag1* (n=74)	20 (95,2)	16 (94,1)	21 (84)	57 (90,5)	17 (48,6)	< 0,00001
IS*Sag9* (n=66)	20 (95,2)	17 (100)	13 (52)	61 (96,8)	5 (14,3)	< 0,00001
SAG0448 (n=62)	16 (76,2)	4 (23,5)	21 (84)	33 (52,4)	29 (82,9)	< 0,01
SAG1241 (n=90)	18 (85,7)	16 (94,1)	23 (92)	55 (87,3)	35 (100)	< 0,05
SAG0610 (n=88)	20 (95,2)	17 (100)	24 (96)	60 (95,2)	28 (80)	< 0,05

[a], souches vaginales et isolées de liquides gastriques

[b], test du χ^2, comparant la totalité des souches humaines aux souches bovines, $p<0,05$, NS, non significatif

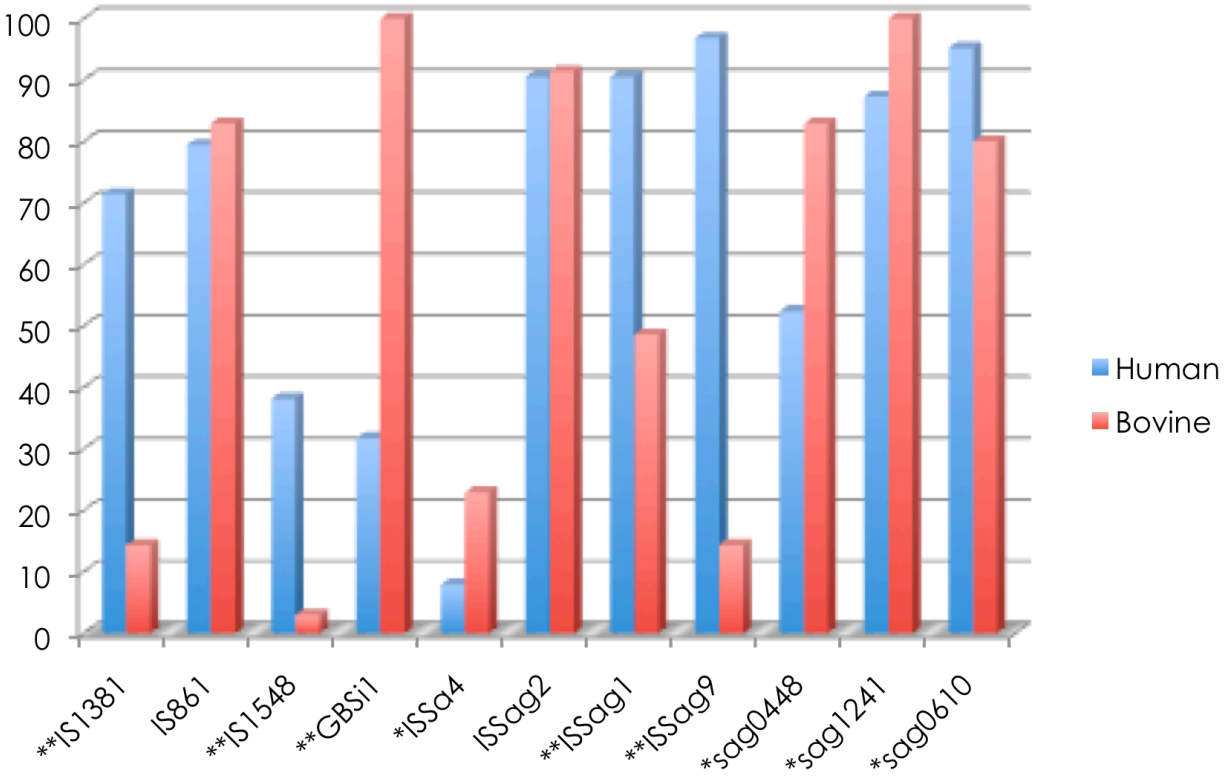


Figure 24. Histogramme représentant la prévalence de 11 éléments génétiques mobiles au sein de la collection de souches humaines et bovines de *S. agalactiae*. Les différences significatives (sans correction du biais due aux comparaisons multiples) sont indiquées par un astérisque pour les valeurs de *P* inférieures à 0,05 et par deux astérisques pour les valeurs de *P* inférieures à 0,001.

2.4. Eléments génétiques mobiles et réservoir écologique des souches

La distribution des 11 EGM a été analysée en fonction de l'origine humaine ou bovine des souches et est détaillée dans le **Tableau XIII**. Il existe une différence statistiquement significative entre la distribution de neuf des onze EGM (IS*1381*, IS*1548*, GBSi1, IS*Sa4*, IS*Sag1*, IS*Sag9*, SAG0448, SAG1241 et SAG0610) au sein du réservoir humain et du réservoir animal ($p < 0,05$; test du χ^2) (**Tableau XIII** et **Figure 24**). Ainsi, l'IS*1381*, l'IS*1548*, l'IS*Sag1*, l'IS*Sag9* et SAG0610 sont majoritairement présentes au sein des souches d'origine humaine, alors que GBSi1, l'IS*Sa4*, SAG0448 et SAG1241 sont principalement présents au sein des souches d'origine bovine.

3. Structure génétique de la population

3.1. Polymorphisme allélique des sept *loci*

Les séquences nucléotidiques des sept *loci* ont été déterminées pour les 98 souches et les profils alléliques ont été assignés. La taille de la séquence nucléotidique analysée varie de 459 pb (*glcK*) à 519 pb (*sdhA*) et le nombre d'allèles pour chaque locus est compris entre cinq (*pheS*) et douze (*glnA*).

Quarante-sept profils alléliques différents ont été identifiés dont 35 ont un représentant unique. Quatre profils alléliques, ST-19 (n=18), ST-17 (n=10), ST-10 (n=6) et ST-23 (n=6), regroupent 40% des souches. Les caractéristiques des profils alléliques sont présentées dans le **Tableau XIV**.

3.2. Profils alléliques et sérotypes

Il n'existe pas de corrélation étroite entre profil allélique et sérotype. Sur les 41 souches de sérotype III, 29 (soit 70,8% des souches) sont regroupées dans les trois ST majoritaires, ST-17, ST-19 et ST-23 (**Tableau XIV**). Parmi les 34 souches appartenant à ces trois profils alléliques, 29 (soit 85,3% des souches) sont de sérotype III.

Tableau XIV. Distribution des profils alléliques (ST) des souches de *S. agalactiae* en fonction de l'origine écologique et du sérotype.

ST	Profils alléliques							Nombre (%) de souches positives					Sérotypes (nombre de souches)
								Souches bovines	Souches humaines			Total	
	adhP	*pheS*	*atr*	*glnA*	*sdhA*	*glcK*	*tkt*	Mammites (n=35)	LCR (n=21)	Endocardites (n=17)	Colonisation (n=25)	Total (n=63)	
19	1	1	3	2	2	2	2	1 (2.9)	10 (47.6)	1 (5.9)	6 (24)	17 (26.9)	18 (18.4)	III (16), II (1), NT (1)
17	2	1	1	2	1	1	1	-	4 (19)	2 (11.8)	3 (12)	10 (15.9)	10 (10.2)	III (9), NT (1)
23	5	4	6	3	2	1	3	2 (5.7)	3 (14.3)	1 (5.9)	-	4 (6.3)	6 (6.1)	III (4), Ia (2)
10	9	1	4	1	3	3	2	1 (2.9)	1 (4.8)	2 (11.8)	2 (8)	5 (7.9)	6 (6.1)	Ib (4), II (2)
12	10	1	4	1	3	3	2	-	-	2 (11.8)	2 (8)	4 (6.3)	4 (4.1)	Ib (3), II (1)
2	1	1	3	1	1	2	2	1 (2.9)	-	1 (5.9)	1 (4)	2 (3.2)	3 (3.1)	II (2), NT (1)
7	10	1	2	1	3	2	2	1 (2.9)	-	-	2 (8)	2 (3.2)	3 (3.1)	Ia (2), IV (1)
61	13	1	1	13	1	1	1	3 (8.6)	-	-	-	-	3 (3.1)	NT (2), III (1)
250	1	1	3	1	12	2	2	3 (8.6)	-	-	-	-	3 (3.1)	NT (3)
301	13	1	1	13	1	28	5	3 (8.6)	-	-	-	-	3 (3.1)	II (2), NT (1)
67	13	1	1	13	1	1	5	2 (5.7)	-	-	-	-	2 (2)	NT (2)
201	2	1	25	2	1	1	1	-	1 (4.8)	-	1 (4)	2 (3.2)	2 (2)	III (2)
1	1	1	2	1	1	2	1	-	-	1 (5.9)	-	1 (1.6)	1 (1)	NT (1)
6	9	1	2	1	3	2	2	-	-	-	1 (4)	1 (1.6)	1 (1)	Ib (1)
8	4	1	4	1	3	3	2	-	-	-	1 (4)	1 (1.6)	1 (1)	Ib (1)
22	13	3	1	3	1	1	1	-	-	1 (5.9)	-	1 (1.6)	1 (1)	II (1)
26	1	1	5	4	1	4	6	1 (2.9)	-	-	-	-	1 (1)	V (1)
28	1	1	3	5	2	2	2	-	-	-	1 (4)	1 (1.6)	1 (1)	II (1)
41	10	1	12	1	3	2	2	-	-	-	1 (4)	1 (1.6)	1 (1)	V (1)
63	13	1	1	1	1	10	1	1 (2.9)	-	-	-	-	1 (1)	II (1)
64	13	1	1	2	1	1	4	1 (2.9)	-	-	-	-	1 (1)	II (1)
85	13	1	17	15	1	1	5	1 (2.9)	-	-	-	-	1 (1)	NT (1)
131	1	1	3	2	2	3	2	-	-	1 (5.9)	-	1 (1.6)	1 (1)	V (1)
186	1	1	2	1	1	2	5	-	-	1 (5.9)	-	1 (1.6)	1 (1)	NT (1)
195	9	1	1	1	3	2	1	-	1 (4.8)	-	-	1 (1.6)	1 (1)	III (1)
196	1	1	3	1	1	12	2	-	1 (4.8)	-	-	1 (1.6)	1 (1)	IV (1)
197	1	1	3	1	2	2	2	-	-	-	1 (4)	1 (1.6)	1 (1)	NT (1)
198	13	4	6	3	2	1	3	-	-	1 (5.9)	-	1 (1.6)	1 (1)	III (1)
199	5	4	6	3	2	1	2	-	-	1 (5.9)	-	1 (1.6)	1 (1)	III (1)
200	9	1	4	1	3	3	5	-	-	1 (5.9)	-	1 (1.6)	1 (1)	Ib (1)
202	40	1	3	1	1	2	2	-	-	-	1 (4)	1 (1.6)	1 (1)	V (1)
226	16	1	2	2	9	1	2	1 (2.9)	-	-	-	-	1 (1)	II (1)
300	56	23	34	37	3	9	4	1 (2.9)	-	-	-	-	1 (1)	NT (1)
302	56	23	6	33	3	9	4	1 (2.9)	-	-	-	-	1 (1)	NT (1)
303	56	23	35	37	3	9	4	1 (2.9)	-	-	-	-	1 (1)	NT (1)
304	9	1	2	4	3	2	2	1 (2.9)	-	-	-	-	1 (1)	Ib (1)
305	5	4	6	32	2	1	3	1 (2.9)	-	-	-	-	1 (1)	III (1)
306	1	1	3	1	12	2	1	1 (2.9)	-	-	-	-	1 (1)	NT (1)
307	5	7	6	3	2	1	1	1 (2.9)	-	-	-	-	1 (1)	III (1)
308	1	1	2	34	1	29	3	1 (2.9)	-	-	-	-	1 (1)	IV (1)
309	13	1	1	2	1	9	1	1 (2.9)	-	-	-	-	1 (1)	III (1)
310	13	1	1	13	1	9	1	1 (2.9)	-	-	-	-	1 (1)	III (1)
311	5	4	6	35	2	1	3	1 (2.9)	-	-	-	-	1 (1)	III (1)
312	13	1	1	13	1	9	5	1 (2.9)	-	-	-	-	1 (1)	NT (1)
313	13	1	1	2	1	28	5	1 (2.9)	-	-	-	-	1 (1)	II (1)
314	16	1	2	2	9	2	2	-	-	-	1 (4)	1 (1.6)	1 (1)	Ia (1)
315	2	1	1	2	31	1	1	-	-	-	1 (4)	1 (1.6)	1 (1)	III (1)

3.3. Profils alléliques et potentiel infectieux

Dix des 21 souches isolées de LCR (soit 47.6%) appartiennent au ST-19 (**Tableau XIV**). Ce profil allélique ne rassemble par contre que six des 25 souches de colonisation (soit 24% des souches) et une des 17 souches d'endocardites (soit 5.9% des souches). Ainsi, il existe une association statistiquement significative entre les souches responsables de méningites et le profil allélique ST-19 ($p<0.01$, test du χ^2), contrairement aux souches provenant de colonisation et aux souches isolées d'endocardite. Par ailleurs, les autres ST ne présentent pas d'association préférentielle avec un groupe particulier de souches.

Le groupe composé de souches isolées de LCR est plus homogène que les deux autres groupes écologiques. En effet, les 21 souches qui le composent sont groupées dans sept profils alléliques différents (ST-19, ST-17, ST-23, ST-10, ST-195, ST-196, et ST-201), alors que les 25 souches de colonisation sont groupées dans 16 profils alléliques différents et les 17 souches d'endocardites dans 13 profils alléliques différents (**Tableau XIV**).

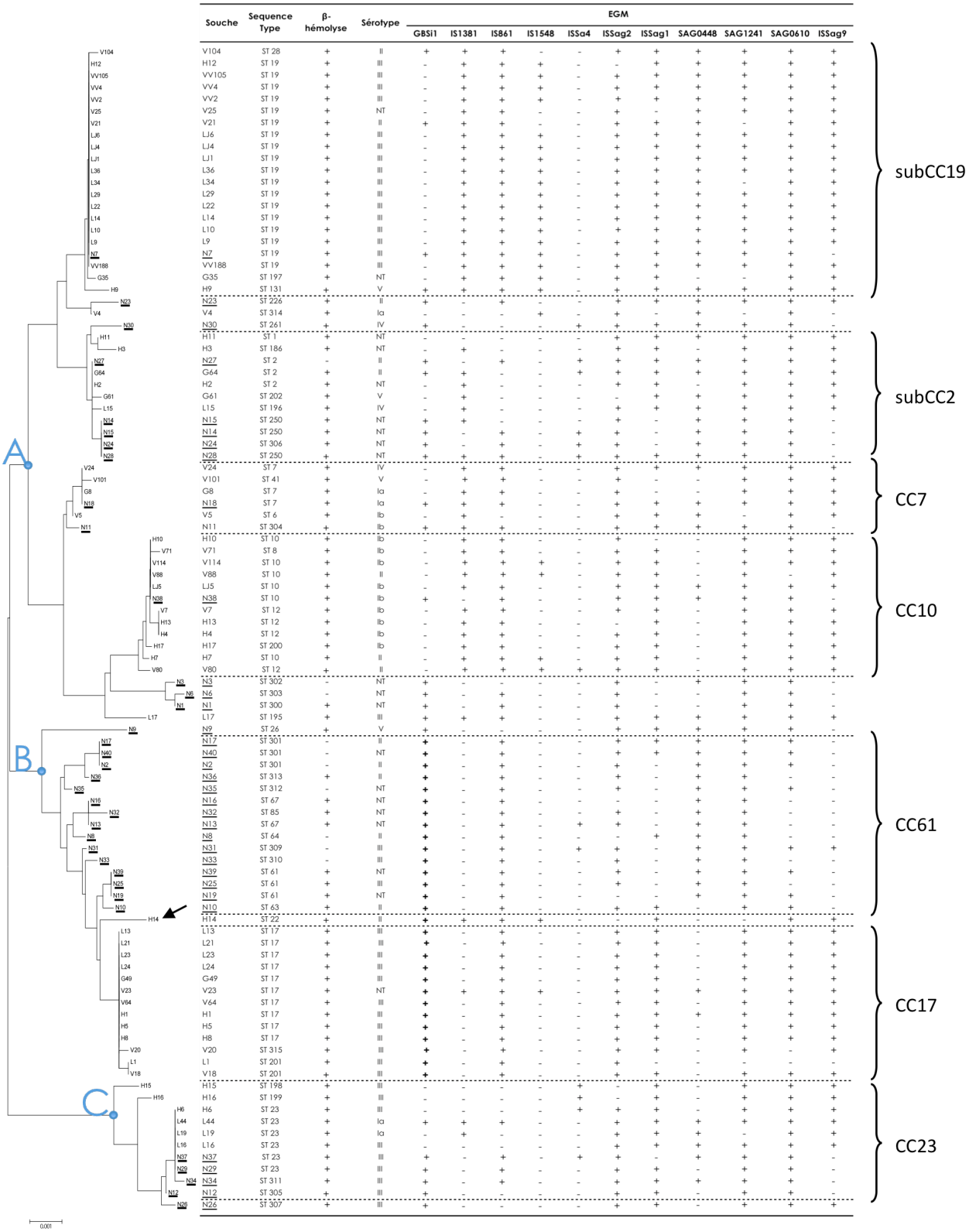
3.4. Profils alléliques et réservoir écologique des souches

Sur les 47 ST représentés dans la collection, 22 sont spécifiques des souches bovines, 20 sont spécifiques des souches humaines et cinq sont communs aux deux réservoirs écologiques. Alors que seuls trois ST (ST-19, ST-17 et ST-10) concentrent plus de la moitié des souches humaines, les souches bovines sont réparties de façon homogène sur l'ensemble des ST bovins. Ainsi, le réservoir bovin semble comporter plus de diversité que le réservoir humain.

3.5. Liens phylogénétiques entre les souches

3.5.1. Au sein de la collection

Les liens évolutifs entre les 98 souches de *S. agalactiae* sont représentés sous forme de dendrogramme obtenu avec le logiciel MEGA v4.1 beta en utilisant l'algorithme NJ et l'indice de Kimura sur les séquences nucléotidiques du supergène obtenu par concaténation des allèles des sept loci (**Figure 25**). La phylogénie



Souche	Sequence Type	β-hémolyse	Sérotype	EGM										
				GBSi1	IS1381	IS861	IS1548	ISSa4	ISSag2	ISSag1	SAG0448	SAG1241	SAG0610	ISSag9
V104	ST 28	+	II	+	+	+	-	-	+	+	+	+	+	+
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	+	-	+	-	-	+	+	+	+	+	+
V4	ST 314	+	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	-	+	-	-	-	+	+	+	+	+	+
N15	ST 250	+	NT	+	+	-	-	-	+	-	+	+	+	-
N14	ST 250	+	NT	+	-	+	-	+	+	-	+	+	+	-
N24	ST 306	+	NT	+	-	+	-	+	+	-	+	+	+	-
N28	ST 250	+	NT	+	+	+	-	+	+	+	+	+	+	-
V24	ST 7	+	IV	-	+	+	-	-	+	+	+	+	+	+
V101	ST 41	+	V	-	+	+	-	-	+	-	-	+	+	+
G8	ST 7	+	Ia	-	+	+	-	-	+	-	-	+	+	+
N18	ST 7	+	Ia	+	+	+	-	-	+	+	+	+	+	+
V5	ST 6	+	Ib	-	+	-	-	-	+	+	+	-	+	+
N11	ST 304	+	Ib	+	+	+	-	-	+	+	+	+	+	-
H10	ST 10	+	Ib	-	+	+	-	-	+	-	-	+	+	+
V71	ST 8	+	Ib	-	+	+	-	-	+	+	-	+	+	+
V114	ST 10	+	Ib	-	+	+	+	-	+	+	-	+	+	+
V88	ST 10	+	II	-	+	+	+	-	+	+	-	+	-	+
LJ5	ST 10	+	Ib	-	+	+	-	-	+	+	+	+	+	+
N38	ST 10	+	Ib	+	-	+	-	-	+	+	+	+	+	-
V7	ST 12	+	Ib	-	+	+	-	-	+	+	-	+	+	+
H13	ST 12	+	Ib	-	+	+	-	-	-	+	-	+	+	+
H4	ST 12	+	Ib	-	+	+	-	-	+	+	-	+	+	+
H17	ST 200	+	Ib	-	+	+	-	-	+	+	-	+	+	+
H7	ST 10	+	II	-	+	+	+	-	+	+	-	+	+	+
V80	ST 12	+	II	-	+	+	+	+	+	+	-	+	+	+
N3	ST 302	-	NT	+	-	-	-	-	+	-	+	+	+	-
N6	ST 303	-	NT	+	-	+	-	-	+	-	-	+	+	-
N1	ST 300	+	NT	+	-	+	-	-	+	-	-	+	+	-
L17	ST 195	+	III	+	+	+	-	-	+	+	+	+	+	+
N9	ST 26	+	V	+	-	-	-	-	+	+	+	+	+	-
N17	ST 301	-	II	+	-	+	-	-	+	+	+	+	+	-
N40	ST 301	-	NT	+	-	+	-	-	+	+	+	+	+	-
N2	ST 301	-	II	+	-	+	-	-	+	-	+	+	+	-
N36	ST 313	+	II	+	-	+	-	-	+	-	+	+	-	-
N35	ST 312	-	NT	+	-	+	-	-	+	-	+	+	+	-
N16	ST 67	+	NT	+	-	+	-	-	-	-	+	+	-	-
N32	ST 85	+	NT	+	-	+	-	-	+	-	+	+	-	-
N13	ST 67	+	NT	+	-	+	-	+	+	-	+	+	-	-
N8	ST 64	-	II	+	-	+	-	-	-	+	+	+	-	-
N31	ST 309	-	III	+	-	+	-	+	+	-	+	+	+	+
N33	ST 310	-	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	+	-	+	-	-	+	+	-	+	+	+
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 315	+	III	+	-	+	-	-	+	+	-	+	-	+
L1	ST 201	+	III	+	-	+	-	-	+	-	-	-	-	-
V18	ST 201	+	III	+	-	+	-	-	+	+	-	+	+	+
H15	ST 198	+	III	-	-	-	-	+	-	+	-	+	+	+
H16	ST 199	+	III	-	-	-	-	+	-	+	-	+	+	+
H6	ST 23	+	III	-	-	-	-	+	+	+	-	+	+	+
L44	ST 23	+	Ia	+	+	+	-	-	+	+	+	+	+	+
L19	ST 23	+	Ia	-	+	-	-	-	+	+	+	-	+	+
L16	ST 23	+	III	-	-	-	-	-	+	+	+	+	+	+
N37	ST 23	+	III	+	-	+	-	+	+	-	+	+	+	-
N29	ST 23	+	III	+	-	+	-	-	+	+	-	+	+	-
N34	ST 311	+	III	+	-	+	-	-	+	+	+	+	+	-
N12	ST 305	+	III	+	-	-	-	-	+	+	-	+	+	-
N26	ST 307	+	III	+	-	-	-	-	+	+	-	+	+	+

Figure 25. Dendrogramme représentant les liens évolutifs entre les 98 souches de *S. agalactiae* obtenu avec la méthode de neighbor-joining avec le logiciel MEGA v4.0 et l'indice de Kimura à partir des séquences nucléotidiques du supergène obtenues la méthode d'analyse multilocus sequence typing décrite par Jones *et al.* (Jones *et al.*, 2003). Les complexes clonaux (CC) ont été définés par l'algorithme eBURST. Pour chaque souche, le sequence type (ST), l'hémolyse β (+ ou -), le sérotype, et la présence (+) ou l'absence (-) de 11 éléments génétiques mobiles (EGM) sont indiqués. Les CC sont aussi indiqués sur la droite de la figure (certaines souches sont des singletons et n'appartiennent à aucun CC). La position particulière de H14 est indiquée par une flèche noire. Les trois grandes divisions phylogénétiques (A, B et C) sont notées en bleu. Les souches humaines sont nommées L ou LJ pour celles isolées du LCR, H pour celles isolées dans un contexte d'endocardite, V ou VV pour le portage vaginal ou G pour le portage gastrique; les souches bovines (N) sont soulignées . NT, non typable en sérotypage.

obtenue à partir des profils alléliques des 98 souches avec le logiciel START en UPGMA donne une représentation identique.

Sur le dendrogramme présenté à la **Figure 25**, trois divisions phylogénétiques principales sont identifiées, A, B et C. Le programme eBURST v3 a identifié sept lignées clonales sur la base des profils alléliques de la population : subCC19 (n=21) et subCC2 (n=11) qui sont des subdivisions du grand complexe clonal CC19, CC10 (n=12), CC7 (n=7), CC17 (n=13), CC67 (n=16) et CC23 (n=11) (**Figure 25**).

3.5.2. Au sein de la banque de données mondiale

Un des avantages du MLST est de permettre une comparaison des résultats obtenus au laboratoire à ceux de la database mondiale. Nous avons donc réalisé une première recherche des complexes clonaux dans la population mondiale à partir des données disponibles en février 2006 : les 277 ST disponibles dans la base de données MLST de *S. agalactiae* (http://pubmlst.org/sagalactiae/) ont été regroupés par eBURST en neuf complexes clonaux (comprenant chacun au moins six ST) ont été définis : CC19 (43 ST différents), CC17 (36 ST différents), CC23 (32 ST différents), CC1 (21 ST différents) (équivalent du CC2 dans notre population d'étude), CC10 (15 ST différents), CC67 (12 ST différents), CC22 (9 ST différents), CC4 (8 ST différents) et CC7 (6 ST différents) (**Annexe A**). Ainsi, sur les neuf lignées clonales principales constituant la population mondiale de *S. agalactiae*, sept sont retrouvées dans la population analysée par notre travail. Une deuxième recherche effectuée en août 2009 (à partir des 449 ST déposés dans la database mondiale) n'a pas identifié de modification dans la structure de la population de *S. agalactiae*. Au contraire, l'enrichissement de la base de données montre une consolidation des trois grands groupes clonaux : i) le plus important en nombre de ST englobe CC19, CC10, CC1, CC7 et CC4, ii) le deuxième englobe CC17 et CC67, iii) le troisième englobe CC23 et CC22 (**Annexe A**).

3.6. Distribution des souches humaines et bovines au sein des lignées clonales

Sur les sept lignées clonales décrites, trois regroupent à la fois des souches humaines et bovines : CC2, CC7 et CC23 (**Tableau XV**). Trois lignées clonales sont quasi exclusives des souches humaines : CC19 (dont 20 des 21 souches sont humaines), CC10 (dont 11 des 12 souches sont humaines) et CC17 (dont toutes les souches sont humaines). CC67 est la seule lignée clonale exclusivement composée

Tableau XV. Répartition des origines écologiques, des éléments génétiques mobiles (EGM) et des sérotypes au sein des sept complexes clonaux (CC) définis en MLST dans la population de *S. agalactiae*.

	Nombre (%) de souches par complexe clonal						
	subCC19 (n=21)	subCC2 (n=11)	CC10 (n=12)	CC7 (n=6)	CC17 (n=13)	CC61 (n=15)	CC23 (n=10)
Réservoir écologique :							
Human	20 (95,2)	6 (54,6)	11 (91,7)	4 (66,7)	13 (100)	-	6 (60)
Bovin	1 (4,8)	5 (45,4)	1 (8,3)	2 (33,3)	-	15 (100)	4 (40)
EGM:							
IS*1381*	21 (100)	7(63,6)	11 (91,7)	6 (100)	1 (7,7)	-	2 (20)
IS*861*	21 (100)	4 (36,4)	12 (100)	5 (83,3)	13 (100)	15 (100)	4 (40)
IS*1548*	18 (85,7)	-	4 (33,3)	-	1 (7,7)	1 (6,7)	-
GBSi1	4 (19)	6(54,6)	1 (8,3)	2 (33,3)	13 (100)	15 (100)	5 (50)
IS*Sa4*	-	5 (45,4)	1 (8,3)	-	-	2 (13,3)	4 (40)
IS*Sag2*	20 (95,2)	10 (90,9)	11 (91,7)	6 (100)	13 (100)	12 (80)	8 (80)
IS*Sag1*	20 (95,2)	8 (72,7)	11 (91,7)	4 (66,7)	12 (92,3)	4 (26,7)	9 (90)
IS*Sag9*	20 (95,2)	7(63,6)	11 (91,7)	5 (83,3)	12 (92,3)	1 (6,7)	6 (60)
SAG0448	20 (95,2)	9 (81,8)	2 (16,6)	5 (83,3)	2 (15,4)	14 (93,3)	5 (50)
SAG1241	18 (85,7)	11 (100)	12 (100)	5 (83,3)	12 (92,3)	15 (100)	9 (90)
SAG0610	21 (100)	11 (100)	11 (91,7)	6 (100)	11 (84,6)	8 (53,3)	10 (100)
Sérotype:							
Ia	-	-	-	2 (33,3)	-	-	2 (20)
Ib	-	-	9 (75)	2 (33,3)	-	-	-
II	2 (9,5)	2 (18,2)	3 (25)	-	-	5 (33,3)	-
III	17 (80,9)	-	-	-	12 (92,3)	3 (20)	8 (80)
IV	-	1 (9,1)	-	1 (16,7)	-	-	-
V	1 (4,8)	1 (9,1)	-	1 (16,7)	-	-	-
NT	1 (4,8)	7 (63,6)	-	-	1 (7,7)	7 (46,7)	-

de souches bovines. CC19 et CC10 sont intégrées dans la division A, alors que CC17 est intégré avec CC67 dans la division B (**Figure 25**). L'analyse phylogénétique met ainsi en évidence une plus forte proximité phylogénétique de la lignée humaine CC17 avec la lignée bovine CC67 par rapport aux deux autres lignées humaines, CC19 et CC10.

3.7. Distribution des éléments génétiques mobiles au sein des lignées clonales

Nous avons observé une forte corrélation entre certains EGM et les lignées obtenues par MLST. Ainsi, l'IS*1381* est significativement associée aux complexes clonaux subCC19, subCC2, CC10 et CC7 (45 des 51 souches [88.2%] IS*1381* + appartiennent à ces complexes clonaux), alors que cette IS est quasi absente de CC17, CC67 et CC23 (p<0,00001) (**Tableau XV**). De même, l'IS*1548* est significativement associée à CC19 (18 des 25 souches [72%] IS*1548* + appartiennent à CC19) (p<0,00001) (Tableau 2-VIII). Enfin, l'intron GBSi1 est significativement associé à CC17 et à CC67 (100% des souches GBSi1+ appartiennent à ces complexes clonaux contre 25,5% des souches des complexes subCC19, subCC2, CC10 et CC7 et 50% des souches de CC23) (p<0.00001) (**Tableau XV** ; **Figure 25**).

A l'inverse, SAG0448 est présent dans tous les complexes clonaux sauf CC10 et CC17. De même, l'IS*Sag9* et de l'IS*Sag1* sont présentes dans toutes les lignées à l'exclusion de la lignée bovine CC67. En revanche, les autres EGM présentent une distribution homogène au sein des différents complexes clonaux (**Tableau XIV** ; **Figure 25**).

3.8. Confrontation entre complexes clonaux, éléments génétiques mobiles, sérotypes et origines écologiques

La confrontation de l'ensemble des données disponibles pour les 98 souches (sérotype, EGM, origine écologique) avec l'analyse phylogénétique (**Tableau XV** ; **Figure 25**) singularisent trois lignées clonales majeures. Deux lignées sont à composante humaine, CC19 et CC17 ; la troisième lignée CC67 à composante bovine. La lignée humaine CC19 est caractérisée par le sérotype III (80% des souches), la présence de l'IS*1381* et SAG0610 (100% des souches), de l'IS*Sag2*, de l'IS*Sag1*, de l'IS*Sag9* et SAG0448 (95,2% des souches), de l'IS*1548* et SAG1241 (85,7% des souches). La deuxième lignée humaine CC17 est également caractérisée par le sérotype III (92.3% des souches), mais diffère de CC19 par l'absence de l'IS*1381*, de l'IS*1548* et SAG0448, mais la présence de GBSi1 (100% des souches).

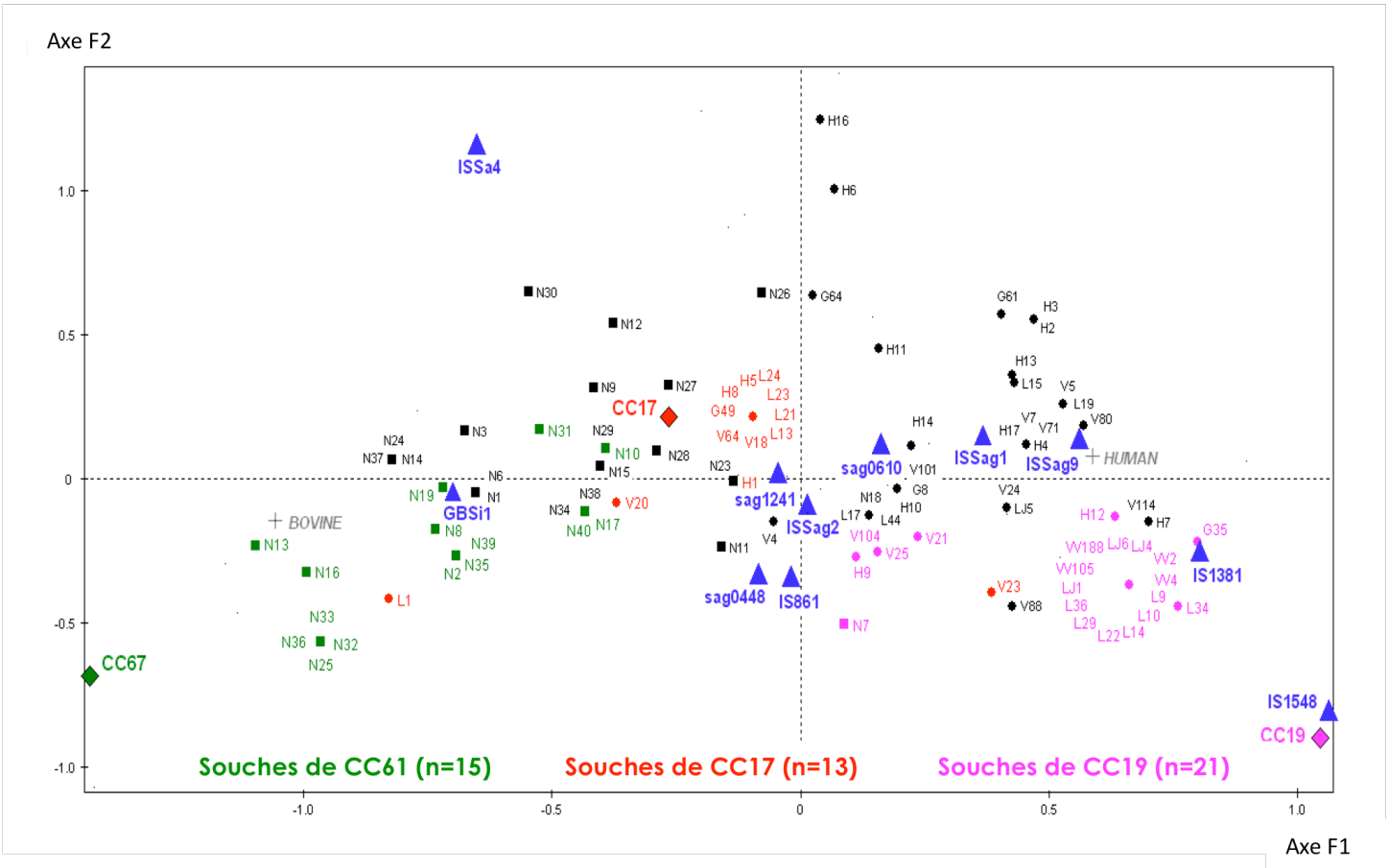


Figure 26. Analyse factorielle des correspondances multiples entre les 98 souches de *S. agalactiae* basée sur la répartition des éléments génétiques mobiles (EGM). Onze variables actives correspondant aux EGM (IS*1381*, IS*861*, IS*1548*, GBSi1, IS*Sa4*, IS*Sag2*, IS*Sag1*, SAG0448, SAG1241, SAG0610, et IS*Sag9*), cinq variables illustratives correspondant aux trois complexes clonaux principaux (CC19 [diamant rouge], CC17 [diamant rose] et CC61 [diamant vert] et aux deux origines écologiques des souches (humaine [cercles] ou bovine [carré]), ainsi que les 98 souches sont projetés sur le plan F1/F2. Ce plan obtenu grâce au logiciel SPAD.N est défini par deux axes principaux : l'axe F1 qui représente la majorité de la variance, et F2, le second axe, qui représente le reste de la variance non prise en compte par F1. F1 et F2 représentent ensemble 40,9% de la variance totale. Les souches sont représentées en rouge pour celles appartenant à CC17 (n=13), en vert pour celles appartenant à CC61 (n=15) et en rose pour celles appartenant à CC19 (n=21).

Enfin, CC61, seule lignée bovine pure, est caractérisée par une proportion de souches NT non négligeable au regard de la collection entière (43.8% contre 20%), la présence de GBSi1 (100% des souches) et l'absence de l'IS*1381*, de l'IS*1548,* de l'IS*Sag1* et de l'IS*Sag9*.

Afin de préciser ces corrélations, une analyse factorielle des correspondances a été réalisée sur l'ensemble des données en prenant les 11 EGM comme variables actives, les trois complexes clonaux principaux (CC19, CC61 et CC17) et l'origine écologique (humaine ou bovine) comme données illustratives (**Figure 26**). La projection des variables illustratives et actives et des souches sur l'axe F1/F2 représentant 40,97% de la variance totale montre une ségrégation des variables et des souches en deux groupes. GBSi1 et l'IS*Sa4* (et à un moindre degré SAG0448 et SAG1241), l'origine bovine, CC17 et CC61 et 94,3% des souches bovines ont été projetées du côté des valeurs négatives de F1. L'IS*1548*, l'IS*1381*, l'IS*Sag9* et l'IS*Sag1* (et à un moindre degré SAG0610), l'origine humaine, CC19 et 79,4% des souches humaines ont été projetées du côté des valeurs positives de F1. L'AFC réalisée avec les EGM comme seules variables actives permet donc de retrouver la clusterisation des souches mise en évidence par l'analyse phylogénétique ce qui prouve que les EGM sont des marqueurs des lignées. Plus précisément, l'analyse met en évidence le marquage différentiel des deux lignées humaines principales. CC19 et très fortement associée à l'IS*1548* (**Figure 26**) alors que CC17 est fortement associé à GBSi1. GBSi1 est d'ailleurs quasiment équidistant de CC17 et CC61. Il est remarquable de constater que toutes les souches humaines projetées du côté négatif de l'axe F1, c'est-à-dire avec la majorité des souches bovines, appartiennent à CC17 (à l'exception d'une, V23) (**Figure 26**). Le prisme de l'analyse de la répartition des EGM montre que les souches de CC17 sont plus proches de la lignée bovine CC61 que de la lignée humaine CC19, qui est située du côté diamétralement opposé à CC67.

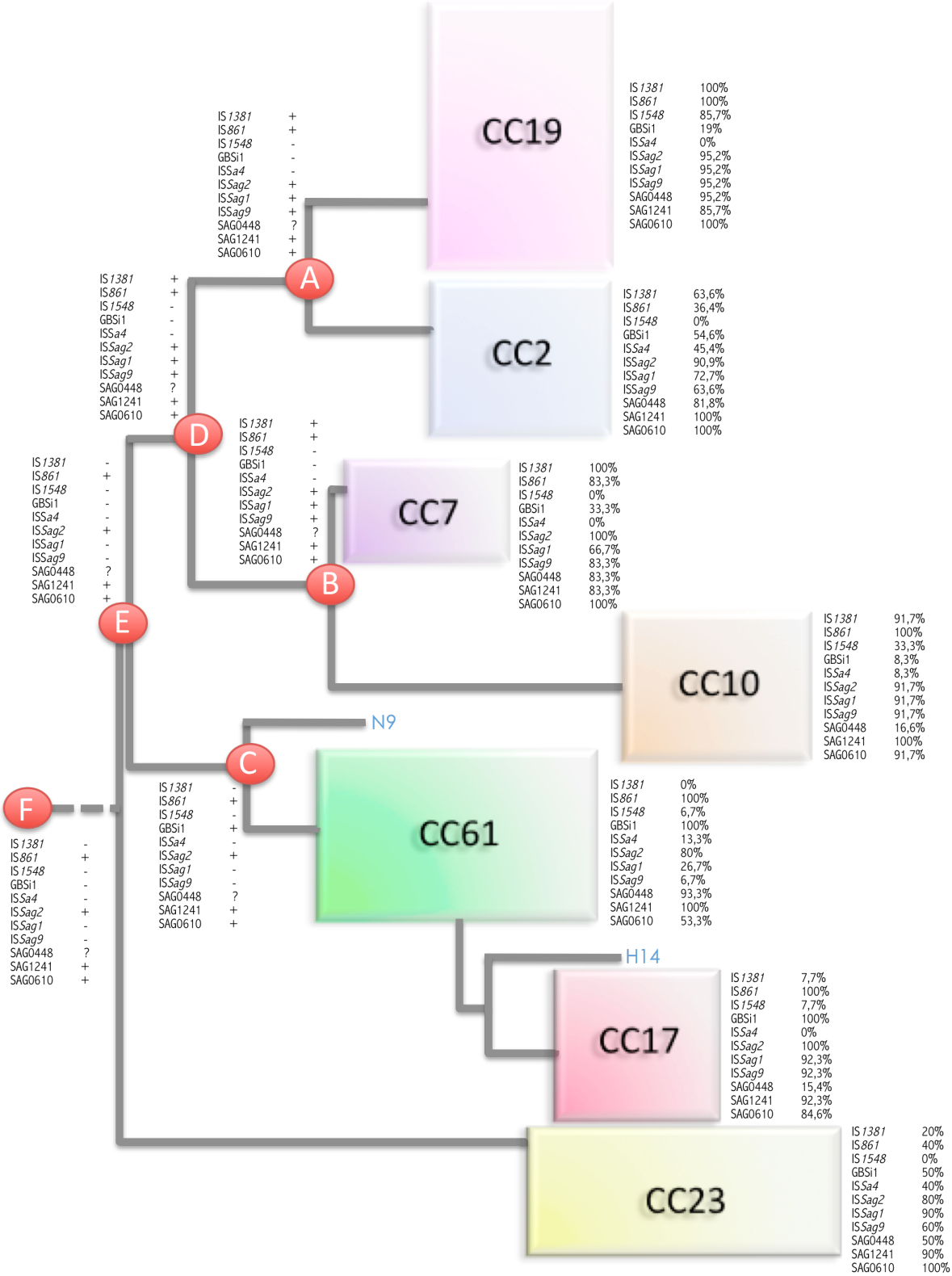


Figure 27. Reconstitution du statut EGM (présence ou absence des 11 Eléments Génétiques Mobiles) aux principaux nœuds du dendrogramme obtenu en NJ (voir Figure 25).

3.9. Hypothèse sur la chronologie d'acquisition des EGM

La prévalence des 11 EGM a été calculée au sein de chaque complexe clonal défini par MLST (**Tableau XV**). Puis, le statut EGM (présence ou absence) a été déduit pour les ancêtres intermédiaires et l'ancêtre commun sur la base de l'hypothèse suivante : un EGM est supposé présent chez un ancêtre dont les deux lignées descendantes portent l'EGM (au moins un tiers des souches de la lignée ; les EGM présents dans moins d'un tiers des souches de la lignée sont supposés avoir été transmis par transfert horizontal). *A contrario*, si une seule des deux lignées présente l'EGM donné, l'ancêtre commun aux deux lignées est supposé non porteur de celui-ci. Dans un deuxième temps, sur la base du statut de chaque ancêtre, nous avons reconstitué les étapes de l'insertion de chaque EGM (**Figure 27**). Cette reconstitution a permis de formuler une hypothèse sur la chronologie d'acquisition des EGM au cours de l'évolution de l'espèce *S. agalactiae* (**Figure 27**).

Selon ce scénario d'évolution, l'ancêtre A, à l'origine de CC19 et CC2, est porteur de l'IS*1381*, l'IS*861*, l'IS*Sag2*, l'IS*Sag1*, l'IS*Sag9*, SAG0448, SAG1241 et SAG0610. L'ancêtre B, à l'origine de CC10 et CC7, est porteur de l'IS*1381*, l'IS*861*, l'IS*Sag2*, l'IS*Sag1*, l'IS*Sag9*, SAG1241 et SAG0610. En regroupant les données des ancêtres A et B, l'ancêtre D semble porteur de l'IS*1381*, l'IS*861*, l'IS*Sag2*, l'IS*Sag1*, l'IS*Sag9*, SAG1241 et SAG0610. De même, l'ancêtre C, à l'origine de CC67 et CC17, serait porteur de l'IS*861*, GBSi1, l'IS*Sag2*, SAG1241 et SAG0610. L'ancêtre intermédiaire E, à l'origine des ancêtres D et C, serait porteur de l'IS*861*, l'IS*Sag2*, SAG1241 et SAG0610. Enfin, l'ancêtre commun F, à l'origine de CC23 et de l'ancêtre E, serait porteur de l'IS*861*, l'IS*Sag2*, SAG1241 et SAG0610. En repartant de l'ancêtre commun F et en suivant les branches des différentes lignées clonales, on peut reconstituer l'apparition des EGM (Figure 2-F). L'IS*1548* serait apparue avec l'émergence de CC19, l'IS*Sa4* avec l'émergence de CC23 et CC2, GBSi1 avec l'apparition de l'ancêtre de CC67 et CC17. L'IS*Sag1* et l'IS*Sag9* seraient apparues, de façon concomitante, au niveau des branches donnant des CC « humains » et toujours concomitamment.

Deux hypothèses sont envisageables pour SAG0448. Dans la première, SAG0448 apparaitrait à quatre reprises et est perdue lors de l'émergence de CC17 (**Figure 28a**). La seconde hypothèse envisage la présence de SAG0448 dès l'ancêtre commun F, puis elle aurait été perdue lors de l'émergence de deux lignées, CC17 et CC12 (**Figure 28b**). Ce scénario est plus parcimonieux en termes d'évènements génétiques : il correspond à 13 évènements (11 acquisitions et deux pertes) contre

16 (15 acquisitions et une perte) pour le premier scenario (**Figure 28a,b**). L'ancêtre commun F serait donc porteur de l'IS*861*, de l'IS*Sag2* et des trois transposases (SAG0448, SAG0610 et SAG1241). Le scénario le plus parcimonieux suggère également que, excepté pour CC7, l'émergence d'un CC est toujours corrélée à au moins un « évènement EGM », qu'il s'agisse d'une acquisition ou d'une perte.

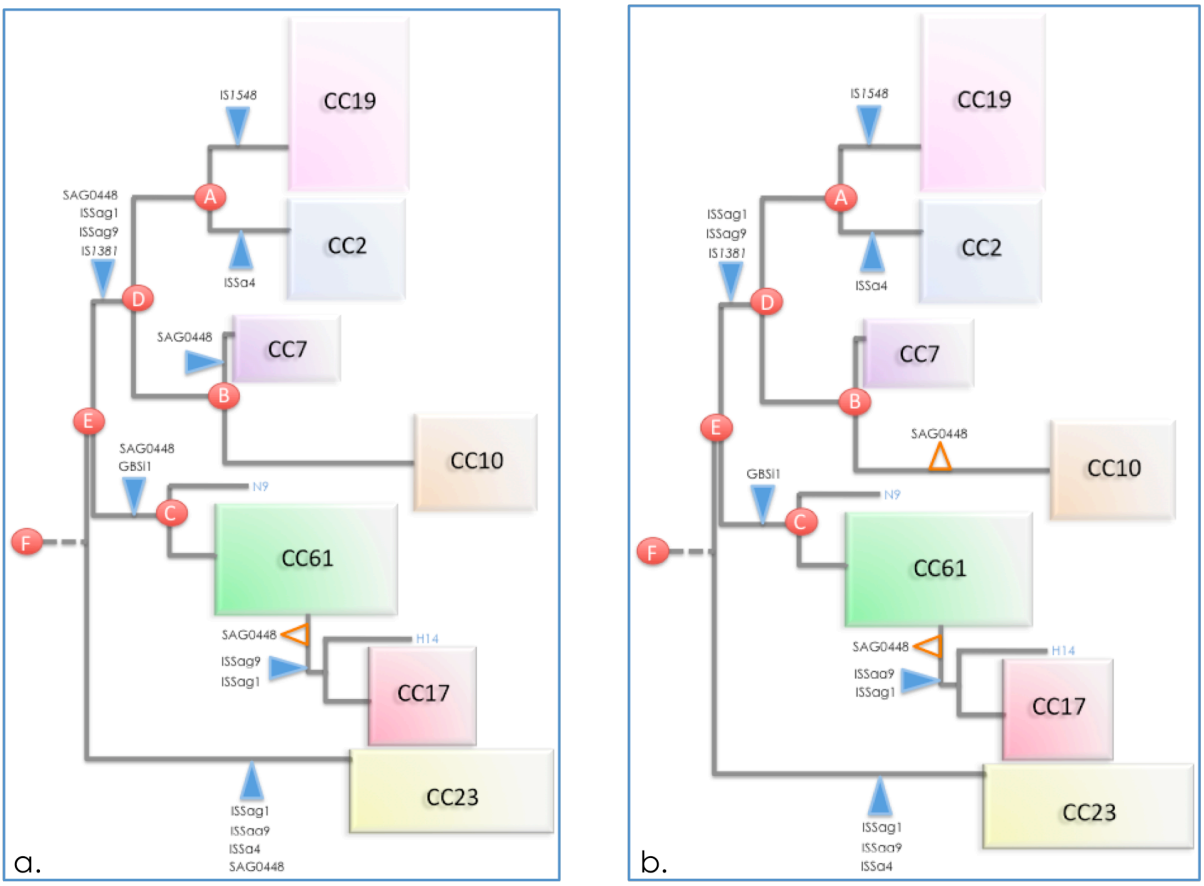


Figure 28. Chronologie d'acquisition des EGM. Deux hypothèses ont été envisagées en fonction de la présence (b.) ou non (a.) de SAG0448 chez l'ancêtre commun F. L'hypothèse présentée en b est la plus parcimonieuse en termes d'évènements génétiques (nombre d'acquisitions et de pertes d'éléments génétiques mobiles).

Partie 2. Corrélation entre la phylogénie de *Streptococcus agalactiae* et la distribution des EGM

D-Discussion

Notre analyse par MLST a mis en évidence une plus grande homogénéité génétique des souches responsables de méningites néonatales (« groupe LCR »). En effet, les 21 souches isolées de méningites sont réparties dans sept profils alléliques différents, alors que les 17 souches isolées d'endocardites sont regroupées dans 13 profils alléliques et les 25 souches de colonisation dans 15 profils alléliques (**Tableau XIV**). Cette observation, déjà faite en MLEE (Musser *et al.*, 1989, Quentin *et al.*, 1995), établit le lien entre la phylogénie de *S. agalactiae* et la physiopathologie des infections dues à cette bactérie. Ainsi, après contamination du fœtus *in utero* ou du nouveau-né lors de son passage par la filière génitale de la mère, les bactéries colonisent le tractus aéro-digestif, mais seule une population restreinte de souches semble capable de franchir les barrières anatomiques naturelles pour gagner le sang puis le système nerveux central. Ces souches dites à « plus haut risque infectieux néonatal » présentent vraisemblablement une communauté de propriétés (adhésion, destruction des matrices extracellulaires, échappement aux systèmes de défense,…), dont l'homogénéité génétique observée en MLST est le reflet. Il a été montré que les souches ST-17 portaient un ensemble de gènes de virulence qui leur est propre et qui code essentiellement des protéines de surface (gbs2018, gènes *sip* et *cpsA*) (Springman *et al.*, 2009, Brochet *et al.*, 2006). Il n'est pas surprenant que les déterminants génétiques de la lignée ST-17 soient principalement des gènes codant des protéines de surface car les souches invasives sont soumises à une forte pression de sélection du système immunitaire de l'hôte ; cette pression immunologique est révélée par le mosaïcisme des gènes de protéines de surface de *S. agalactiae* (Lachenauer *et al.*, 2000). Les souches responsables d'endocardites, qui sont également des affections invasives associées à un fort taux de mortalité (ce qui a motivé notre intérêt de les étudier et de les comparer aux souches responsables de méningite), présentent au contraire une hétérogénéité génétique plus proche de celle des souches de colonisation. Or les endocardites à *S. agalactiae* surviennent plus volontiers chez des patients présentant des facteurs de risque (diabète, SIDA, alcoolisme, cirrhose, cancer digestif) (Sambola *et al.*, 2002). Il est donc logique que le passage de souches responsables d'endocardites soit plus lié à la diminution des systèmes de défense de l'hôte qu'à des facteurs d'adhésion ou des facteurs de virulence spécifiques de ces souches.

A la différence des souches d'endocardites, le clone ST-17 responsable des méningites néonatales semble posséder des propriétés uniques favorisant son pouvoir invasif. L'origine de cette lignée hautement invasive reste cependant incertaine. Bisharat *et al.* ont proposé une ancestralité bovine pour ce clone sur la base du MLST (Bisharat *et al.*, 2004). A l'inverse, par analyse comparée des génomes de souches de ST-17 et ST-61, Brochet *et al.* ont conclut que l'ancêtre commun de ces deux lignées était génétiquement plus proche des souches humaines ST-17 que des souches bovines ST-61 (Brochet *et al.*, 2006). Notre analyse phylogénétique conduite à partir de deux sous-populations (humaine et bovine), associées dans le temps et dans l'espace, va dans le sens de l'hypothèse de Bisharat *et al.* (Bisharat *et al.*, 2004). En effet, notre analyse MLST montre que la lignée clonale CC17 est phylogénétiquement plus proche de la lignée bovine CC61 que des autres lignées humaines (**Figure 25**), ce résultat ayant été confirmé très récemment par une autre équipe (Springman *et al.*, 2009).

L'autre observation majeure de notre travail est la forte congruence entre la répartition des EGM et les lignées MLST, au point que certains EGM semblent correspondre à des marqueurs génétiques des principales lignées de l'espèce *S. agalactiae*. De fait, l'IS*1381* est presque exclusivement retrouvée associée aux complexes CC19, CC2, CC10 et CC7 (**Figure 25** : division A), l'IS*1548* au complexe CC19, et GBSi1 aux complexes CC17 et CC61 (**Figure 26**) ; à l'inverse, trois EGM se distinguent par leur absence dans certaines lignées : l'ISSag1 et l'ISSag9 dans la lignée CC61, et SAG0448 dans CC12 et CC17. Les rares exceptions à ces observations sont probablement liées à des transferts horizontaux (**Tableau XV**). Cette distribution très particulière des EGM est également observée en fonction du réservoir d'origine (**Tableau XIII**, **Figure 24**). Cette différence est particulièrement évidente pour l'IS*1381*, l'IS*1548*, l'IS*Sag1*, l'IS*Sag9* et GBSi1. L'IS*1548*, l'IS*1381*, l'IS*Sag1*, l'IS*Sag9* sont associés préférentiellement au réservoir humain alors que GBSi1 est associé préférentiellement au réservoir bovin. En effet, 100% des souches bovines portent GBSi1, alors que seulement 31,7% des souches humaines. L'AFC réalisée a permis de corréler EGM, lignées clonales et origine écologique en une seule analyse. Au final, les souches humaines se répartissent dans les deux lignées clonales principales : la lignée CC19 marquée par l'IS*1548* et la lignée CC17 marquée par GBSi1. De manière remarquable, il apparaît que la quasi totalité des souches

humaines possédant GBSi1 est regroupée dans CC17. Or si GBSi1 est présent chez l'ensemble des souches d'origine bovine, il peut être considéré comme un marqueur de l'écosystème bovin. De fait, les souches humaines du CC17 auraient conservées le marqueur bovin GBSi1 au cours de l'évolution. Ces résultats, obtenus sur la base de l'étude des EGM, renforcent l'hypothèse établie de l'émergence d'un clone invasif chez l'Homme à partir d'un ancêtre bovin (Bisharat *et al.*, 2004).

La distribution particulière des EGM entre les souches d'origines bovine et animale est probablement révélatrice de la barrière physique qui a réduit les échanges de matériel génétique entre les deux réservoirs. La plupart des études s'accordent sur le caractère indépendant des deux populations arguant de l'identification de sous-types spécifiques de chaque hôte (Sukhnanand *et al.*, 2005, Dogan *et al.*, 2005, Martinez *et al.*, 2000). La souche humaine H14, qui occupe une position particulière dans le dendrogramme car elle est située entre le CC61 « bovin » et le CC17 « humain » (**Figure 25**), présente deux particularités. Premièrement, le nombre de copies de GBSi1 y est bien supérieur à celui observé dans les autres souches humaines (12 copies contre au maximum quatre, voir Partie 3. Résultats. § 1.3). Parkhill *et al.* ont suggéré que les expansions d'IS se produisaient lors d'un rétrécissement (« bottleneck ») évolutif (Parkhill *et al.*, 2003, Parkhill *et al.*, 2001). Chez *Yersinia pestis,* l'expansion des IS coïncide avec un changement de niche écologique de l'espèce qui, de commensal intestinal, est devenue un pathogène systémique (Parkhill *et al.*, 2001). Chez *S. agalactiae*, CC17 est la seule lignée humaine pour laquelle un phénomène de pression sélective positive a été détecté, ce qui pourrait être lié au changement de niche écologique (Lefebure & Stanhope, 2007). Deuxièmement, la souche H14 porte simultanément GBSi1, marqueur de l'écosystème bovin, et l'IS*1548*, marqueur de l'écosystème humain. Or les travaux publiés, y compris le notre, soulignent la très faible prévalence de souches portant simultanément ces deux EGM (Granlund *et al.*, 2001, Luan *et al.*, 2003). Granlund *et al.* expliquent cette distribution exclusive mutuelle par la proximité des sites d'insertion de deux EGM sur le chromosome qui résulterait en une compétition entre les deux EGM (Granlund *et al.*, 2001). Cependant les sites d'insertion de GBSi1 et de l'IS*1548* sont séparées par 88 pb sur le chromosome bactérien (Luan *et al.*, 2003), ce qui ne permet pas d'expliquer l'incapacité de deux EGM à se retrouver côte à côte. L'explication plus plausible démontrée par notre

travail serait que les deux EGM ne sont quasiment jamais associés car ils sont issus de deux écosystèmes distincts, l'humain et le bovin.

Notre travail montre que les EGM sont des marqueurs de l'évolution de *S. agalactiae*. Certains EGM semblent présents chez l'ancêtre commun, tel l'IS*861*, qui est également retrouvé chez de nombreuses autres espèces de streptocoques telles que *S. equisimilis, S. pyogènes, S. pneumoniae, S. sanguinis, S. thermophilus,* ou *S. mutans* (recherche par BLAST - données non présentées). D'autres EGM ont été acquis au cours du temps, parfois par acquisition simultanée, comme l'IS*Sag1* et l'IS*Sag9*. La corrélation positive entre deux IS est le signe d'une transmission par l'intermédiaire d'un plasmide (Hartl & Sawyer, 1988). Ces deux IS sont par ailleurs liées à l'émergence de toutes les lignées humaines (Figure 2D). La survenue d'un événement génétique à de multiples reprises est le signe d'une mutation bénéfique offrant un avantage sélectif (Zhong *et al.*, 2004). Ces deux EGM auraient donc pu favoriser l'adaptation de différents clones de *S. agalactiae* à l'Homme, notamment de CC17, pour lequel l'histoire évolutive montre un changement de niche écologique. L'IS*Sag1* ne comporte qu'une seule copie chez les souches humaines et est toujours retrouvé dans le locus *scpB-lmb* (Franken *et al.*, 2004). Ce locus, qui a la structure d'un transposon composite comportant des gènes de virulence, est absent des souches bovines (Franken *et al.*, 2004). L'IS*Sag9* n'a pas été étudiée chez *S. agalactiae*, mais l'analyse des trois génomes séquencés et assemblés (NEM316, A909 et 2603 V/R) montre que les copies sont pour la plupart adjacentes à des gènes codant des transporteurs (notamment un « transporteur facilitateur majeur » correspondant à SAG1522 chez 2603 V/R et SAK_1546 chez A909) ; l'expression de ces gènes pourrait être modulée par l'IS*Sag9*. Or, l'importance des transporteurs dans l'adaptation écologique a été soulignée chez *S. uberis* (Ward *et al.*, 2009). D'autres acquisitions d'EGM sont encore plus récentes dans notre schéma évolutif, telle celle de l'IS*Sa4*. Cette IS, qui est portée par un faible nombre de souches dans notre collection, a été retrouvée uniquement chez des souches isolées après 1996 par Spellerberg *et al.* (Spellerberg *et al.*, 2000).

L'excision d'un EGM est également possible, sans que cela ne remettre en question la pertinence de l'utilisation des EGM comme marqueurs de l'évolution de *S. agalactiae*. L'excision d'IS a déjà été décrite chez diverses espèces bactériennes.

Ce mécanisme est un phénomène moins fréquent que l'insertion (cinq fois moins fréquent chez *Staphylococcus epidermidis*), mais qui ne requiert pas que la transposase soit fonctionnelle (Hennig & Ziebuhr, 2008). Suite à l'excision d'une IS, une bactérie peut retrouver la pleine expression d'un gène inactivé par l'insertion de cette IS (Hennig & Ziebuhr, 2008). Ainsi, l'émergence de CC10 et en partie de CC17 semble liée à l'excision de SAG0448. Il est intéressant de noter que SAG0448 est juxtaposée à un transporteur du magnésium de type CorA (SAG0447) chez la souche 2603 V/R. Les transporteurs CorA Mg^{2+} jouent un rôle important dans les interactions bactérie-environnement. Chez *Salmonella spp*. ce transporteur est associé à la virulence de la bactérie (Papp-Wallace *et al.*, 2008). Chez *S. agalactiae*, l'excision de SAG0448 pourrait avoir engendré la récupération de la fonctionnalité du transporteur CorA Mg^{2+} et favorisé ainsi l'émergence de nouveaux clones.

Partie 3.

Sites d'insertion des EGM dans le génome de *Streptococcus agalactiae*

Partie 3. Sites d'insertion des EGM dans le génome de *Streptococcus agalactiae*

A-Introduction

La pathogénèse des infections à *S. agalactiae* reflète sa capacité à coloniser différents épithéliums (intestinal, vaginal, respiratoire), à pénétrer à travers les barrières épithéliales ou le placenta, à résister au système immunitaire, à envahir le système sanguin, et plus spécifiquement pour les souches responsables de méningite, à traverser la barrière hémato-encéphalique. Pour s'adapter à ces différents environnements, la bactérie exprime un large panel de facteurs de virulence qui interagissent avec les cellules de l'hôte et interférent avec les mécanismes de défense immunitaire (Maisey *et al.*, 2008).

La génomique a montré que l'espèce possède un « pan-génome » composé d'un squelette génomique conservé ou « core-génome », représentant près de 80% du génome, et des zones de plasticité localisées dans des îlots, représentant parfois des îlots de pathogénicité (Herbert *et al.*, 2005, Glaser *et al.*, 2002, Tettelin *et al.*, 2005). Les gènes liés à la virulence pourraient résider dans cette partie variable du génome ce qui expliquerait l'échec des méthodes explorant le core-génome dans la caractérisation du potentiel invasif de toutes les lignées phylogénétiques (Tettelin *et al.*, 2005).

Les zones de variabilité génomique sont liées en partie à la présence des EGM qui représentent une source majeure de la plasticité génomique bactérienne en participant aux évènements de recombinaison homologue et aux transferts horizontaux. En raison de leur activité de transposition, les EGM agissent aussi directement sur les caractéristiques phénotypiques des bactéries. C'est ainsi que les EGM provoquent des modulations dans l'expression de gènes de virulence, soit en augmentant leur expression soit en la diminuant (Mahillon & Chandler, 1998). Chez *S. agalactiae*, plusieurs EGM ont été associés à des facteurs de virulence : l'IS*1548* a été retrouvée dans le gène de la hyaluronate lyase (Granlund *et al.*, 1998), l'ISSa4 dans l'opéron du transporteur de l'hémolysine (Spellerberg *et al.*, 2000), l'IS*861* dans l'opéron de la capsule (Sellin *et al.*, 2000), et GBSi1 entre les gènes *scpB* et *lmb* qui sont des facteurs d'adhésion à la matrice extracellulaire (Granlund *et al.*, 2001).

Compte tenu de la variabilité observée dans la distribution des EGM entre les souches de *S. agalactiae* (voir Partie 2.) et du lien rapporté chez d'autres espèces entre EGM et virulence, l'objectif de cette troisième partie est de préciser l'implication potentielle des EGM dans la physiopathologie des infections à *S. agalactiae* en identifiant les sites d'insertion des différentes copies des EGM sur le génome de la bactérie.

Partie 3. Sites d'insertion des EGM dans le génome de *Streptococcus agalactiae*

B-Matériel et Méthodes

Tableau XVI. Origine écologique, sérotype et prévalence de cinq éléments génétiques mobiles (EGM) pour les 52 souches de *S. agalactiae*.

			EGM				
Souche	Origine écologique	Sérotype	IS *1548*	GBSi1	IS *861*	IS *1381*	ISSa4
L1	LCR	III	-	+	+	-	-
L9	LCR	III	+	-	+	+	-
L10	LCR	III	+	-	+	+	-
L13	LCR	III	-	+	+	-	-
L14	LCR	III	+	-	+	+	-
L15	LCR	IV	-	-	-	+	-
L16	LCR	III	-	-	-	-	-
L17	LCR	III	-	+	+	+	-
L19	LCR	Ia	-	-	-	+	-
L22	LCR	III	+	-	+	+	-
L23	LCR	III	-	+	+	-	-
L24	LCR	III	-	+	+	-	-
L29	LCR	III	+	-	+	+	-
L34	LCR	III	+	-	+	+	-
L36	LCR	III	+	-	+	+	-
L44	LCR	Ia	-	+	+	+	-
LJ1	LCR	III	+	-	+	+	-
LJ4	LCR	III	+	-	+	+	-
LJ5	LCR	Ib	-	-	+	+	-
LJ6	LCR	III	+	-	+	+	-
G8	Gastrique	Ia	-	-	+	+	-
G35	Gastrique	II	+	-	+	+	-
G49	Gastrique	III	-	+	+	-	-
G61	Gastrique	V	-	-	-	+	-
G64	Gastrique	II	-	+	-	+	+
V5	Vagin	Ib	-	-	-	+	-
V18	Vagin	III	-	+	+	-	-
V21	Vagin	II	-	+	+	+	-
V25	Vagin	NT	-	-	+	+	-
V64	Vagin	III	-	+	+	-	-
V71	Vagin	Ib	-	-	+	+	-
VV2	Vagin	III	+	-	+	+	-
VV4	Vagin	III	+	-	+	+	-
VV105	Vagin	III	+	-	+	+	-
VV188	Vagin	III	+	-	+	+	-
H1	Endocardite	III	-	+	+	-	-
H2	Endocardite	NT	-	-	-	+	-
H3	Endocardite	NT	-	-	-	+	-
H4	Endocardite	Ib	-	-	+	+	-
H5	Endocardite	III	-	+	+	-	-
H6	Endocardite	III	-	-	-	-	+
H7	Endocardite	II	+	-	+	+	-
H8	Endocardite	III	-	+	+	-	-
H9	Endocardite	V	+	+	+	+	-
H10	Endocardite	Ib	-	-	+	+	-
H11	Endocardite	NT	-	-	-	-	-
H12	Endocardite	III	+	-	+	+	-
H13	Endocardite	Ib	-	-	+	+	-
H14	Endocardite	II	+	+	+	+	-
H15	Endocardite	III	-	-	-	-	+
H16	Endocardite	III	-	-	-	-	+
H17	Endocardite	Ib	-	-	+	+	-

1. Souches bactériennes

L'étude a été réalisée sur une collection composées de 52 souches de *S. agalactiae*, sélectionnées au sein du souchier utilisé dans l'étape précédente (voir Parties 2, § 1. et 2, § 2.) et choisies pour leur représentativité de la diversité de l'espèce.

2. Nombre de copies des EGM

Cinq EGM ont été étudiés : l'IS*1548*, l'IS*861*, l'IS*1381*, l'IS*Sa4* et GBSi1. Leur prévalence au sein de la collection a été déterminée au cours de l'étape précédente de notre travail (Partie 2, § 2.1) (**Tableau XVI**). Le nombre de copies des cinq EGM a été déterminé par hybridation selon Southern pour chaque souche (Southern, 1974).

2.1. Hybridation selon Southern

2.1.1. Microrestriction du génome bactérien

Pour chaque souche, trois microgrammes d'ADN ont été clivés par 30 UI d'enzyme de restriction *Eco*RI (BioLabs, Beverly, USA) dans un volume réactionnel de 100 µL à 10% de tampon NEBuffer 10X (Tris-HCl 100 mM, NaCl 50 mM, $MgCl_2$ 10 mM, Triton X-100 0.025%, pH 7,5) pendant 16 heures à 37°C.

2.1.2. Migration des fragments de restriction

Les produits de digestion ont été soumis à une migration en gel d'agarose à 0.8% à 40V pendant 20 heures. Les acides nucléiques ont été dépurinés dans un bain d'HCl 0,25 M pendant 10 minutes, puis dénaturés dans une solution de dénaturation (NaOH 0,5 M, NaCl 1,5 M) pendant 20 minutes, et enfin neutralisés dans une solution de neutralisation (NaCl 1,5 M, Tris-HCl 1 M, pH 8) pendant 20 minutes.

2.1.3. Transfert sur membrane

Le transfert a été réalisé sur membrane de nylon Hybon N chargée positivement (Roche Diagnostics GmbH, Manheim, Germany) suivant la méthode par capillarité décrite par Southern, en tampon SSC 20X (NaCl 3 M, citrate trisodique dihydraté 0,3 M, pH 7,4) pendant 24 heures à 4°C.

2.1.4. Hybridation des sondes marquées

Le marquage des sondes d'hybridation a été réalisé par sonde froide avec le kit AlkPhos Direct Labelling (Amersham). Les sondes ont été synthétisées en utilisant les produits de PCR générés par les cinq couples d'amorces précédemment utilisés pour la détection des EGM (Partie 2, § 4.1, **Tableau III**). Vingt microlitres de produit de PCR contenant 10 mg/mL d'ADN ont été portés à ébullition pendant 5 minutes, puis placés dans la glace. Vingt microlitres de tampon de réaction, 4 mL de réactif de marquage et 20 mL de réactif de liaison ont été successivement ajoutés à l'ADN, et incubés 30 minutes à 37°C. Après transfert, la membrane a été préhybridée pendant 30 minutes à 55°C dans un tampon d'hybridation AlkPhos direct contenant 4 % de réactif bloquant et 0,5 M NaCl, puis hybridée pendant une nuit à 55°C avec le même tampon contenant la sonde.

2.1.5. Révélation

Après hybridation, la membrane a subi successivement 2 lavages de 10 minutes à 55°C avec un tampon de lavage (NaH_2PO_4 50 mM pH 7, NaCl 150 mM, urée 2 M, SDS 0,1%, $MgCl_2$ 1 mM, agent bloquant 0,2%), puis 2 lavages de 5 minutes avec un tampon de conservation (TrisBase 1 M, NaCl 2 M). La membrane a ensuite été recouverte de 10 ml de CDP Star (Amersham) pendant 5 minutes, puis séchée et révélée par contact de la membrane avec un film X-OMAT (Amersham Biosciences) en chambre noire pendant 20 minutes.

2.2. Analyse des données *in silico*

L'absence du site de restriction de *Eco*RI dans les séquences nucléotidiques des gènes codant pour les cinq EGM étudiés a été vérifiée *in silico* à l'aide du logiciel Vector NTI 10 (Invitrogene). La corrélation entre la distance de migration et la taille des fragments amplifiés a été réalisée par calibration avec le marqueur de taille moléculaire SmartLadder (Eurogentec) ; l'analyse *in silico* a permis ensuite la localisation des copies des EGM sur les fragments de restriction.

3. Identification des sites d'insertion des EGM

Après avoir déterminé le nombre de copies des EGM au sein de chaque souche de la collection, l'étape suivante a été d'identifier les sites d'insertion des copies des EGM au sein du génome de *S. agalactiae*. Dans cette perspective, trois

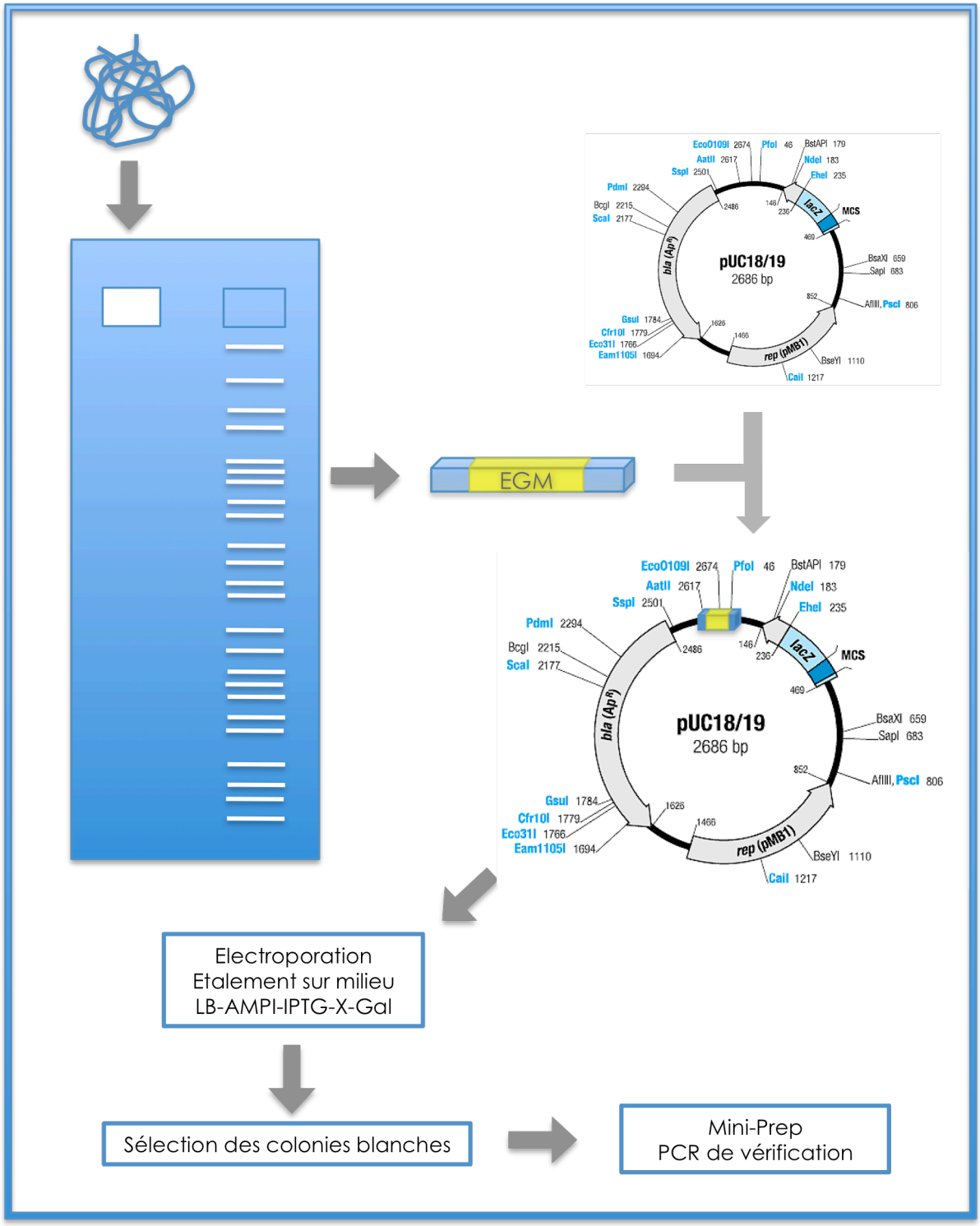


Figure 29. Principe du protocole A. Le génome de *S. agalactiae*, préalablement restreint par *Eco*RI, a été soumis à une migration en gel d'agarose pendant 16 heures à 40V. Des cubes de 0,5 cm^3 ont été découpés et les fragments de restriction ont été extraits. Ces derniers ont été ensuite ligués dans le vecteur pUC18 et électroporés dans *Escherichia coli* Top10 électrocompétentes. Les clones bactériens contenant un plasmide avec un insert ont été sélectionnés et une PCR avec des amorces spécifiques de l'élément génétique mobile (EGM) étudié a été réalisé.

stratégies expérimentales ont été développées et appliquées sur un nombre limité de souches afin i) de déterminer le protocole le plus adapté à notre problématique, ii) d'appliquer le protocole le plus contributif à l'ensemble des souches de la collection.

3.1. Extraction des fragments de restriction à partir du gel de migration (Protocole A)

3.1.1. Principe

Le principe de l'extraction des fragments de restriction à partir du gel de migration est de cloner des fragments issus de la restriction par l'endonucléase *Eco*RI, après électrophorèse et comparaison avec les fragments révélés par la sonde de l'EGM étudié en Southern blot (**Annexes B-F**). Après vérification par PCR de la présence de l'EGM dans le fragment cloné, le fragment est séquencé (**Figure 29**).

3.1.2. Restriction du génome et PCR de vérification

Trois microgrammes d'ADN correspondant à $9x10^8$ copies du génome de *S. agalactiae* ont été restreints par 30 UI d'enzyme de restriction *Eco*RI (Biolabs) dans un volume réactionnel de 30 % μL pendant 16 heures à 37°C, générant ainsi 500 fragments par génome. Les produits de restriction ont été séparés par électrophorèse en gel d'agarose (Eurobio) à 0,8% en tampon TAE 1X pendant 20 heures à 40V. Les fragments d'ADN ont été extraits du gel avec le kit DNA Gel Extraction (Millipore Corporation, Bedford, USA). La présence de l'EGM dans les produits de restriction a été confirmée par PCR réalisée avec le mélange réactionnel recommandé par le fabricant (Applied Biosystems, Foster City, USA), avec 200 μM de chacun des dNTPs (Amersham Biosciences, Buckinghamshire, England), 0,5 μM de chaque amorce (Eurogentec, Seraing, Belgique) (**Tableau XVII**), auxquels ont été ajoutés 25 ng d'ADN.

Tableau XVII. Amorces utilisées pour l'amplification par PCR des éléments génétiques mobiles (EGM) de *S. agalactiae*.

EGM	Code d'accès Genbank	Amorces (5' à 3')*	Amplicon (pb)	Hybri-dation (T°C)	Référence
IS*1381*	AF064785	GTTAAGCTTAGAAGATCTCCTCATGG (s) CCTTGATAACCACTGTCTGCCA (as)	425	55	(Dmitriev *et al.*, 2004)
IS*861*	M22449	ATGACAGAGCTCGAGCGACTT (s) TATCAGCCTTCTTACCAACCTCA (as)	447	55	(Dmitriev *et al.*, 2004)
IS*1548*	Y14270	TTGCGCAGTTGAATTGGATAG (s) TTCTCTAACTTCAATCTGTCCCCTA (as)	218	55	(Dmitriev *et al.*, 2004)
GBSi1	AJ292930	CGAGTGGATTGTGGACATTG (s) GTGATAGATTCCCGCCTTGA (as)	197	55	(Bidet *et al.*, 2003)
IS*Sa4*	AF165983	CTGAGCTAGGTATTAAGTCGCAAC (s) GCGTGTCTCAATGGTTGGT (as)	563	55	(Dmitriev *et al.*, 2004)

* (s), amorce sens ; (as), amorce anti-sens

3.1.3. Ligation et clonage des produits de digestion dans le plasmide PUC18

Les produits d'extraction positifs en PCR EGM ont été ligués dans le plasmide PUC18 préalablement ouvert par restriction *Eco*RI selon un rapport de trois inserts pour un vecteur. Cette réaction a été réalisée avec 200 UI de T4 ligase (Biolabs) dans un volume réactionnel de 10 µL et incubation à 4°C pendant 16 heures. Les produits de ligation ont été électroporés dans la souche *E. coli* Top 10 électrocompétente. Les bactéries électroporées ont été ensemencées sur milieu LB agar contenant 100 µg/mL d'ampicilline (GlaxoSmithKline, Marly-Le-Roi, France), 0,5 mM d'IPTG (Sigma-Aldrich, Lyon, France) et 50 mg/mL de X-Gal (Eurogentec), et incubées pendant 16 heures à 37°C. Les colonies blanches ont été repiquées en bouillon LB contenant 100 µg/mL d'ampicilline et incubées sous agitation à 37°C pendant 18 heures.

3.1.4. Lyse des clones bactériens et extraction plasmidique

Un millilitre et demi de culture a été centrifugé à 12 000 g pendant 30 minutes. Le cultot a été mis en suspension dans 350 µL de STET (0,1 M Nacl, 10 mM Tris-HCl pH8, 1 mM EDTA pH8, 5% Triton X-100) et 25 µL de lysozyme à 10 mg/mL. Après une incubation de 40 secondes dans de l'eau bouillante, le mélange a été centrifugé 10 minutes à 12 000 x g. Le surnageant a été ajouté à 420 µL d'isopropanol à 75% et 30 µL d'acétate de sodium à 3 M (pH 5,2). Après une incubation de 5 minutes à température ambiante, une centrifugation de 15 minutes a été effectuée à 12 000 x g et le culot séché puis repris par 30 µL d'eau milliQ. Une PCR de vérification

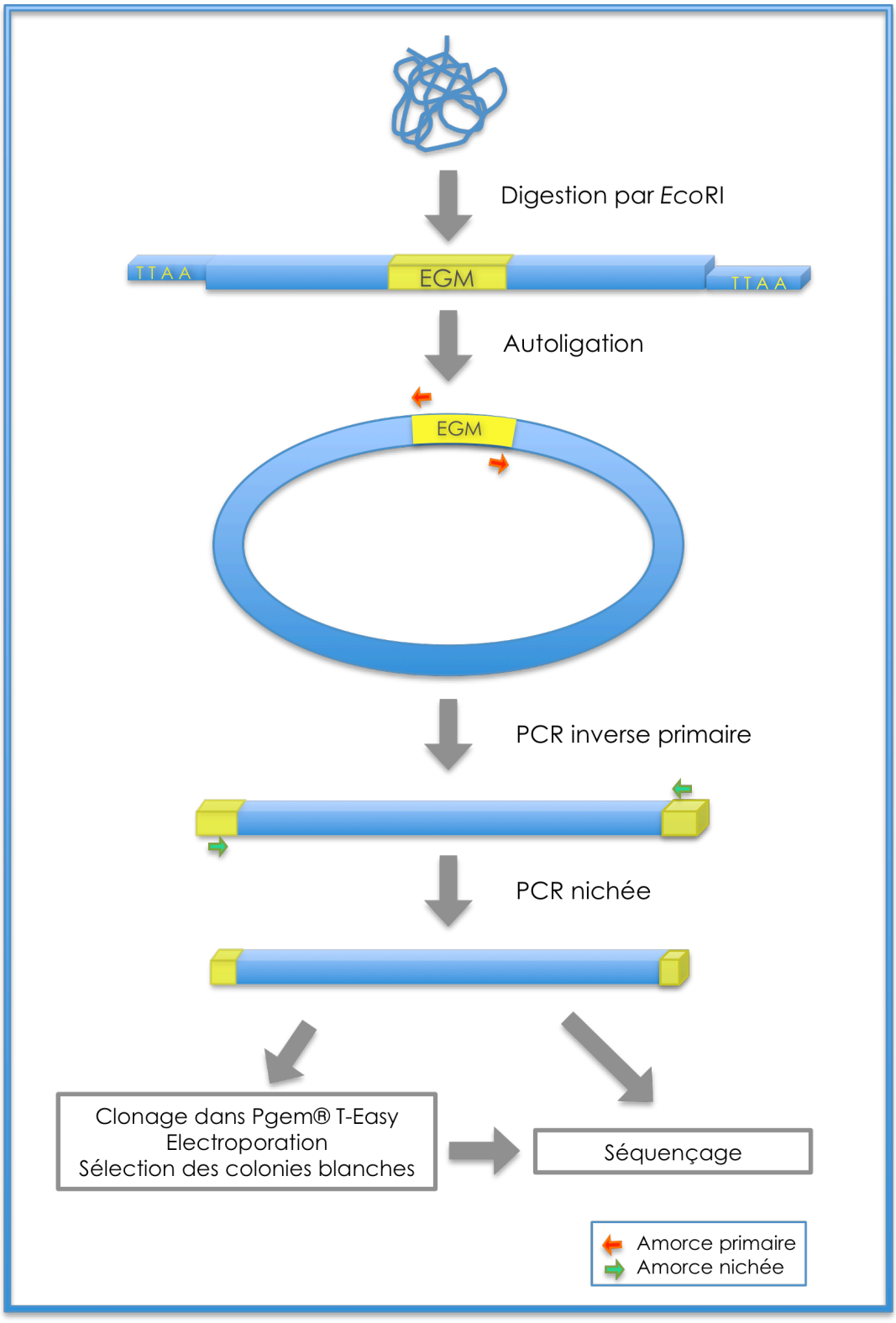


Figure 30. Principe du protocole B. Les fragments issus de la restriction par *Eco*RI du génome de *S. agalactiae* ont été autoligués pour permettre l'amplification des fragments contenant l'élément génétique mobile (EGM) étudié par des amorces spécifiques avec une PCR nichée. Les amplicons ont ensuite été clonés dans le vecteur Pgem® T-Easy puis séquencés, ou directement séquencés en cas de concentration en ADN suffisante.

a été réalisée sur les extraits d'ADN plasmidique dans les conditions décrites au paragraphe 2.4.1 suivie d'une restriction de 500 ng d'ADN plasmidique par 10 UI d'*Eco*RI (Biolabs) dans un volume réactionnel de 20 µL et 100 ng/µL de Ribonucléase A (Fermentas Life Science, St Leon Rot, Allemagne) pendant 3 heures à 37°C, afin de contrôler la présence de l'insert par électrophorèse. En cas de positivité de la PCR, l'extrait plasmidique a été séquencé (voir § 3.5.).

3.2. PCR inverse (Protocole B)

3.2.1. Principe

Le principe de la méthode de PCR inverse est l'autoligation des fragments issus de la restriction par *Eco*RI et l'amplification des fragments contenant les EGM d'intérêt par des amorces spécifiques (**Figure 30**).

3.2.2. Restriction du génome

Un microgramme d'ADN, c'est-à-dire $3x10^8$ copies du génome de *S. agalactiae*, ont été restreintes par *Eco*RI (Biolabs) dans les conditions décrites au paragraphe 3.1.2. Les produits de restriction ont été purifiés par une précipitation avec deux volumes d'éthanol absolu et un dixième de volume d'acétate de sodium 3 M pH 5,2 à -20°C pendant 30 minutes. Une centrifugation de 15 minutes à 15 000 × g a été effectuée et le culot séché au bain-marie à sec puis repris par 100 µL d'eau milliQ.

3.2.3. Réaction d'autoligation

La réaction d'autoligation a été réalisée sur $1,5x10^8$ copies du génome restreint avec 200 UI de T4 DNA Ligase (Biolabs) dans un volume réactionnel de 400 µL (Biolabs) pendant 16 heures à 4°C. Les produits autoligués ont été purifiés comme décrit au paragraphe 3.2.2 et repris dans 100 µL d'eau milliQ.

3.2.4. PCR primaire

L'amplification a été réalisée avec la Taq polymérase Takara TaqTM Hot start version (Takara, Cambrex Verviers, Belgique) dans un mélange réactionnel de 50 µL réalisé selon les instructions du fabricant avec un µM de chaque amorce (Eurogentec) (**Tableau XVIII**) et 100 ng d'ADN autoligué. Après une étape initiale de

dénaturation de l'ADN à 94°C pendant 4 minutes, 35 cycles d'amplification ont été réalisés, comprenant une dénaturation à 94°C pendant 30 secondes, une hybridation des amorces à la température optimale du couple d'amorces pendant 30 secondes, et une élongation de 1 minute à 68°C. Une élongation de 7 minutes à 68°C a été ajoutée en étape finale.

3.2.5. PCR nichée

Une PCR nichée a été réalisée avec un µL de produit d'amplification obtenu lors de la PCR primaire, en utilisant les mêmes conditions que pour la PCR primaire avec des amorces internes (**Tableau XVIII**). Les produits d'amplification ont été séparés par électrophorèse. L'extraction des fragments d'ADN a été réalisée avec le kit DNA Gel Extraction (Millipore Corporation). Les produits purifiés ont été dosés par spectrophotométrie à 260 nm, puis, soit séquencés directement si la concentration en acides nucléiques était supérieure à 40 ng/µL, soit clonés dans le vecteur Pgem®-T Easy (Promega) (voir § 3.5).

Tableau XVIII. **Amorces utilisées pour l'amplification par PCR inverse des cinq éléments génétiques mobiles (EGM) de *S. agalactiae*.**

EGM	Amorces primaires (5' à 3')*	Hybridation (T°C)	Amorces nichées(5' à 3')*	Hybridation (T°C)
IS*1381*	GTTAAGCTTAGAAGATCTCCTCATGG (s) CCTTGATAACCACTGTCTGCCA (as)	55	TGATCAACCGTGAACTAGGA (s) CGCAGGAAGTCTATTGTACCATA (as)	59
IS*861*	ATGACAGAGCTCGAGCGACTT (s) TATCAGCCTTCTTACCAACCTCA (as)	55	GCTTTGTCGTGTCGGATATAG (s) TGCAGTTCATGTCTCAGAACTG (as)	49
IS*1548*	TTGCGCAGTTGAATTGGATAG (s) TTCTCTAACTTCAATCTGTCCCCTA (as)	55	GCCCTTGGTGGTAATCTAGT (s) AAAGTTTCTTCATGCGTTTG (as)	52
GBSi1	CGAGTGGATTGTGGACATTG (s) GTGATAGATTCCCGCCTTGA (as)	55	CTCATGGCTATGGTTGTGAGTTTAGTAATG (s) ACGTTTCTGTTCGTAGAGCCACGAT (as)	52
IS*Sa4*	CTGAGCTAGGTATTAAGTCGCAAC (s) GCGTGTCTCAATGGTTGGT (as)	55	CTAGCACCACGGGTTAAGATA (s) AGATTAACTAGCACCACGGGTTA (as)	52

*(s), amorce sens ; (as), amorce antisens

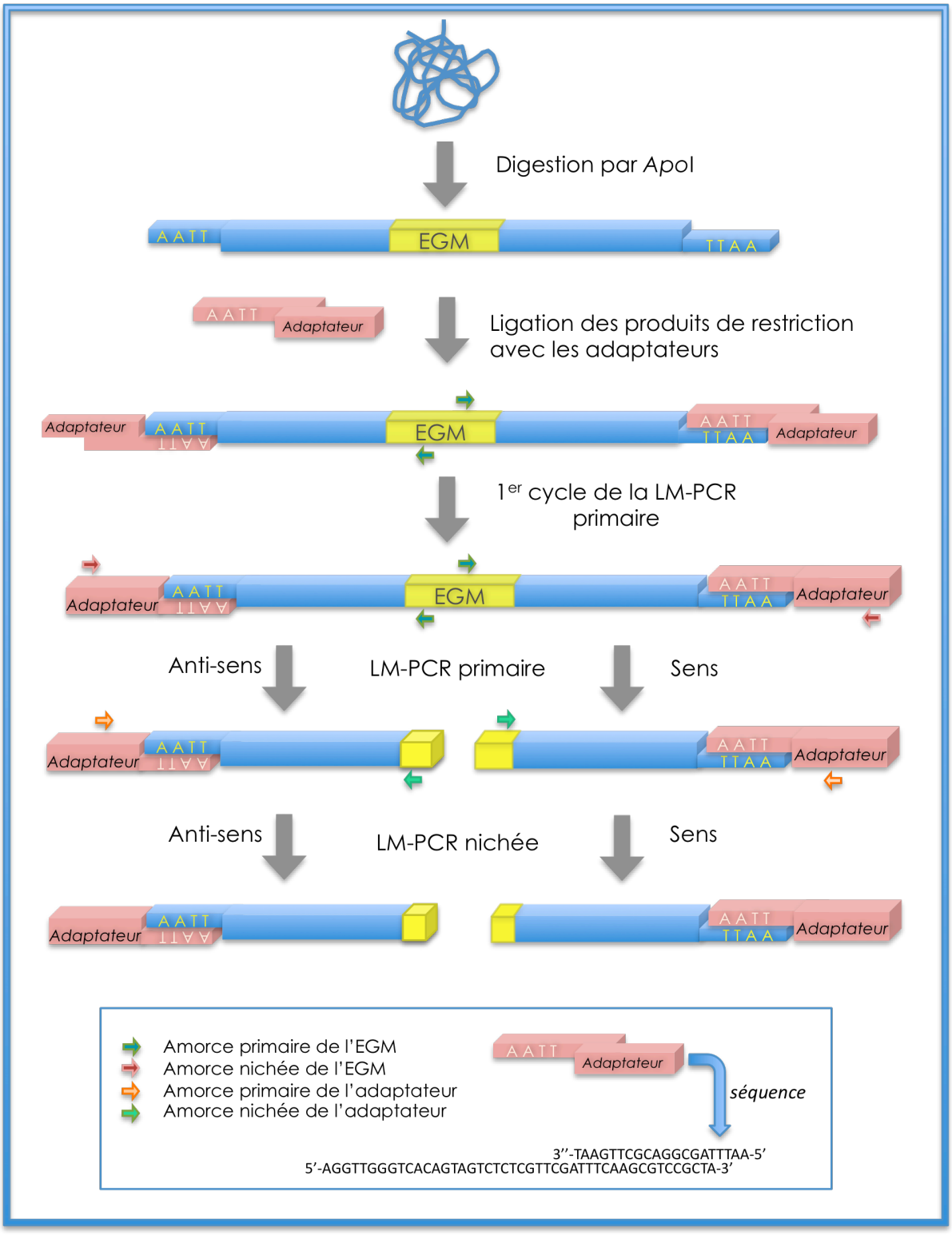


Figure 31. Principe du protocole C. Les fragments de restriction issus de la digestion par ApoI du génome de *S. agalactiae* ont été ligués à des adaptateurs spécifiques à bouts collants. Une PCR avec des amorces spécifiques à l'adaptateur et à l'élément génétique mobile (EGM) testé a été effectué ainsi qu'une PCR nichée. Les amplicons ont ensuite été clonés dans le vecteur Pgem® T-Easy puis séquencés, ou directement séquencés en cas de dosage suffisant.

3.3. Ligation-Mediated PCR (Protocole C)

3.3.1. Principe

Le principe de la ligation-mediated PCR (LM-PCR) repose sur la ligation d'un adaptateur de séquence connue, aux fragments issus de la restriction de l'ADN génomique de *S. agalactiae* par l'endonucléase *Apo*I, ce qui permet dans un second temps l'amplification spécifique des fragments compris entre l'adaptateur et l'EGM (**Figure 31**).

3.3.2. Restriction du génome

Un microgramme d'ADN, c'est-à-dire $3x10^8$ copies du génome de *S. agalactiae*, ont été restreintes par *Apo*I (Biolabs) dans un volume réactionnel de 30 µL pendant 2 heures à 50°C générant 5 000 fragments par copie de génome. Une purification a été réalisée par le kit Montage™ PCR Centrifugal Filter Devices (Millipore Corporation) de manière à éliminer les composants réactionnels en excès.

3.3.3. Assemblage de l'adaptateur

L'assemblage de l'adaptateur a été réalisé dans un volume réactionnel de 100 µL de TE 1X contenant 4 nM d'oligonucléotides simple brin. La présence en quantité équimolaire de deux oligonucléotides complémentaires de tailles différentes a permis de générer des adaptateurs à bouts cohésifs de type *Apo*I. L'hybridation a été obtenue par cinq minutes de chauffage à 95°C et descente progressive en température.

3.3.4. Ligation de l'adaptateur avec les produits de digestion

Les conditions optimales de ligation correspondent à un rapport 1/10 entre la quantité des produits de restriction et celle de l'adaptateur. Cent picogrammes soit $1,5x10^8$ fragments issus de la restriction équivalents à 30 000 copies du génome de *S. agalactiae* ont été mélangés à $1,5x10^9$ copies d'adapteurs. Après une pré-incubation à 55°C pendant trois minutes du mélange réactionnel, la ligation a été obtenue en présence de T4 DNA Ligase (Biolabs) pendant 16 heures à 4°C. La purification réalisée avec le kit Montage™ PCR Centrifugal Filter Devices (Millipore Corporation) a permis d'éluer l'excès d'adaptateurs.

3.3.5. LM-PCR primaire

L'originalité de cette technique réside dans le premier cycle de la réaction de PCR (**Figure 31**). L'utilisation d'une amorce spécifique de l'EGM permet à la polymérase de combler la partie simple brin de l'adaptateur, rendant possible l'hybridation de l'amorce spécifique de l'adaptateur aux cycles suivants. Ainsi, seuls les fragments contenant l'EGM sont amplifiés. La LM-PCR primaire a été réalisée sur la totalité des produits de ligation purifiés avec la Taq polymérase Takara TaqTM Hot start version (Takara) et un μM d'amorces purifiées par HPLC-RT dont une spécifique à l'adaptateur (5'-AGGTTGGGTCACAGTAGTCTCTCGTTCG-3') et l'autre spécifique à l'EGM étudié (**Tableau XIX**). Après une étape de dénaturation de l'ADN de cinq minutes à 94°C, un programme « Touchdown » a été réalisé dont la caractéristique est la diminution progressive de la température de 67°C à 62°C sur dix cycles, pour terminer par 30 cycles comportant une dénaturation à 94°C pendant 30 secondes, une hybridation des amorces à 60°C pendant 30 secondes et une élongation à 68°C durant cinq minutes. L'utilisation de ce programme et d'amorces ultra-purifiées permet d'éviter les réactions aspécifiques et de réduire les artefacts qui surviendraient dans des conditions normales de PCR. Les amplicons ont ensuite été purifiés pour éluer les amorces en excès et pour éviter la formation d'hétéroduplexes avec les amorces de la PCR nichée.

3.3.6. LM-PCR nichée

Une PCR nichée a été réalisée avec un microlitre d'amplicons purifiés issus de la LM-PCR primaire, dans des conditions réactionnelles identiques, mais avec des amorces nichées spécifiques de l'EGM étudié (**Tableau XIX**) et de l'adaptateur (5'-CTCGTTCGATTTCAAGCGTCCG-3'). Après une dénaturation de cinq minutes à 94°C, le programme d'amplification suivant a été appliqué : 30 cycles comportant une dénaturation à 94°C pendant 30 secondes, une hybridation des amorces à 60°C pendant 30 secondes et une élongation à 68°C durant cinq minutes. Les amplicons ont été séparés en gel d'agarose à 2%. L'extraction des bandes d'ADN a été réalisée avec le kit DNA Gel Extraction (Millipore Corporation). Les produits purifiés ont ensuite été dosés par spectrophotométrie et séquencés directement pour une concentration supérieure à 40 ng/μL, sinon clonés dans le vecteur Pgem®-T Easy (Promega) (voir § 3.4).

Tableau XIX. Amorces utilisées pour l'amplification par Ligation-Mediated PCR des éléments génétiques mobiles (EGM) de *S. agalactiae*.

EGM	Amorces primaires (5' à 3')*	Amorces nichées(5' à 3')*
GBSi1	GCGAGGACTGGTCAGTTGCTTAGATTACT (s)	ACTTCATCACTACTATGGGTTCGGCTGAC (s)
	AGTTGAACCGCCGTATGCCGA (as)	ACCTCACGGGATAAGTCGGACATCT (as)
IS*1548*	GAAATCGGTGTCCCTCGCTCAATTA (s)	TTCCGTAAAACGCTTGTCAATCCAT (s)
	AGATTGACGACTATCCAATTCAACTGCG (as)	GCGTTTGGCGTGTGGATTGG (as)
IS*1381*	CCGTGAACTAGGATTTTAGTTTCGCAGTAACTC (s)	CCGTGAACTAGGATTTTAGTTTCGCAGTAACTC (s)**
	GATCCCTTGCATCCACAATCACCG (as)	GGCCACCTTTTGCGTGTTTACG (as)

*(s), amorce sens ; (as), amorce antisens

**amorce identique à celle utilisée en PCR primaire (PCR semi-nichée)

3.4. Ligation des amplicons dans le vecteur Pgem®-T Easy (Promega)

La ligation des amplicons a été réalisée sur 16 heures à 4°C dans le vecteur Pgem®-T Easy (Promega) selon les conditions recommandées par le fabricant. Les produits de ligation ont été électroporés dans *E. coli* Top 10 compétente. Les clones ont été sélectionnés suivant la méthode décrite au paragraphe 3.1.3. et une restriction par *Eco*RI a été réalisée pour vérifier la présence et la taille de l'insert.

3.5. Séquençage

La réaction de séquençage a été réalisée selon le protocole décrit précédemment (voir Partie 1, § 4.4.). L'interrogation par le programme BLAST du National Center for Biotechnology Investigation (http ://www.ncbi.nih.gov/blast) a permis l'identification des séquences nucléotidiques que nous avons obtenues, sur la base des données disponibles pour les génomes complets des souches de *S. agalactiae* NEM316 (code d'accès Genbank NC_004368) et 2603V/R (code d'accès Genbank NC_004116) (Glaser *et al.*, 2002, Tettelin *et al.*, 2002).

3.6. PCR de jonction entre l'EGM et son site d'insertion

Afin de confirmer la localisation des sites d'insertion supposés des copies des EGM, des amorces spécifiques de chacun des sites d'insertion identifiés ont été définies sur la base des résultats du séquençage des régions flanquantes de l'EGM (**Tableau XX**). Ce travail a été réalisé *in silico* avec le logiciel Vector NTI 10

(Invitrogene). Chacune de ces amorces, couplée à une amorce spécifique de l'EGM, a permis de tester par PCR la présence ou non de l'EGM au site étudié.

Tableau XX. Amorces utilisées pour l'amplification des sites d'insertion des éléments génétiques mobiles (EGM) chez *S. agalactiae*.

EGM	Site d'insertion	Amorces s'hybridant à l'EGM (5' à 3')	Amorces s'hybridant sur le gène (5' à 3')	Hybridation (T°C)	Amplicon (pb)
GBSi1	*lmb*	TGAACCCAATAGCGACTCC	CTATCATTACTAAACTCACAAC	55	1997
	ftsY	ACTTCATCACTACTATGGGTTCGGCTGAC	CTGTTAAATTCATTGGTTTCGCGA	58	452
	lplA-1	ACTTCATCACTACTATGGGTTCGGCTGAC	GTTGGAAGGTGTCCGCTACGATTACA	58	765
IS*1548*	*hylB*	AAAGTTTCTTCATGCGTTTG	GCATCTTATTCATGTTGGCAATATCGG	58	509
	sag0155	AAAGTTTCTTCATGCGTTTG	CTCTAATACAACTGCCGCCAAAGTAACAA	58	906
	sag1963	GCCCTTGGTGGTAATCTAGT	TCTGTACGCTCATCTTGTGGCAATG	58	1108
	sag0911	GCCCTTGGTGGTAATCTAGT	AGTAGCTGACGGTGCAATGGGAAC	58	601
	gbs0667	GCCCTTGGTGGTAATCTAGT	CGGTCCAATCAAACCACTGATACCA	58	456
	lmb	CCAAACGCATGAAGAAACTTTTGTTTTCAG	CTATCATTACTAAACTCACAAC	55	1352
	cpsD	ACTACTTTGATTGATTTTGTCATTATGAAT	AATCCTACCATTACGACCTAC	52	800

Partie 3. Sites d'insertion des EGM dans le génome de *Streptococcus agalactiae*

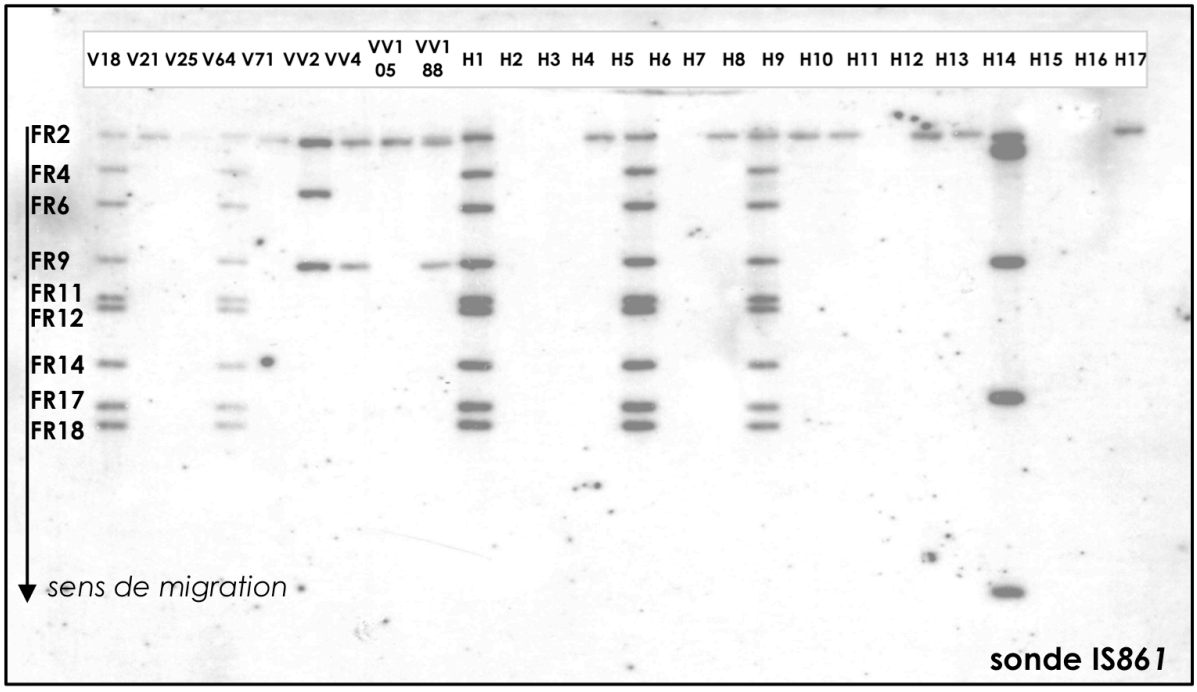


Figure 32. Image de film autoradiographique révélant l'hybridation de la sonde IS*861* après réalisation d'un Southern blot sur les souches de *S. agalactiae*. La restriction des génomes a été réalisée avec l'enzyme *Eco*RI, et la migration a été réalisée en gel d'agarose à 0,8% pendant 18h. L'image montre les profils de restriction obtenus pour 26 souches (indiquées en haut) et précise pour la première souche (V18) les fragments de restriction (FR) révélés par la sonde IS*861*.

1. Nombre de copies des EGM et profils d'hybridation

L'ADN chromosomique des 52 souches, digéré par *EcoR*I, a été hybridé avec les sondes IS*1381,* IS*861,* GBSi1, IS*1548* et ISSa4. *In silico,* il a été vérifié au préalable qu'aucune des séquences nucléotidiques des EGM ne porte le site de restriction de l'enzyme *EcoR*I. Ainsi, le nombre de fragments hybridés est une évaluation minimale du nombre de copies des EGM présentes sur chaque souche, car on ne peut exclure une hybridation de la sonde sur deux (ou plusieurs) fragments de même taille.

1.1. L'IS*1381*

Trente-sept souches possèdent l'IS*1381*. Le nombre de copies varie de un à dix selon les souches (**Annexe B**). Dix-huit fragments de taille différente ont permis de définir 31 profils d'hybridation, ce qui montre la grande hétérogénéité des profils IS*1381* répertoriés. Chaque profil est quasiment spécifique d'une souche. Dans ce contexte, aucun fragment de restriction (FR) ne semble associé à un potentiel de virulence particulier.

1.2. L'IS*861*

Quarante souches possèdent l'IS*861*. Le nombre de copies varie de un à dix (**Figure 32**, Annexe **C**). Vingt fragments de taille différente ont permis de définir 16 profils d'hybridation. La localisation génomique précise de l'IS*861* a été identifiée uniquement pour une copie présente au niveau du locus *cpsD* (Sellin *et al.*, 2000). Ce locus pourrait correspondre, dans notre étude, au fragment de restriction FR 7. Quatorze des 17 souches isolées de LCR possédant IS*861* présentent le FR1 alors qu'aucune des 12 souches de colonisation possédant IS*861* ne présente ce FR. Le FR1 est donc potentiellement associé à un facteur de virulence.

1.3. GBSi1

Seize souches possèdent GBSi1. Le nombre de copies de GBSi1 varie de un à 12 (**Annexe D**). Treize fragments de taille différente ont permis de définir 5 profils d'hybridation. Trois fragments apparaissent plus fréquemment : FR2, FR4 et FR6. Ces FR pourraient correspondre aux trois sites d'insertion de GBSi1 déjà rapportés dans la littérature (Granlund *et al.*, 2001, Luan *et al.*, 2003) : le site X (entre *scpB* et *lmb*), le site Y (entre *ftsY* et sag0278) et le site Z (entre *lplA*-1 et sag0883).

1.4. L'IS*1548*

Dix-neuf souches possèdent l'IS*1548*. Le nombre de copies de l'IS*1548* varie de un à neuf (**Annexe E**). Quinze fragments de restriction de taille différente ont permis de définir 15 profils d'hybridation. Trois sites d'insertion sont actuellement identifiés pour l'IS*1548* : au niveau du fragment intergénique situé entre le gène *scpB* (gène codant pour la C5a-peptidase) et le gène *lmb* (gène codant pour la protéine liant la laminine), appelé site X; dans le gène *hylB* qui code pour la hyaluronate lyase ; enfin, dans le gène *cpsD* qui appartient à l'opéron codant pour la capsule de la bactérie (Granlund *et al.*, 2001, Luan *et al.*, 2003, Sellin *et al.*, 2000). L'insertion de l'IS*1548* dans *hylB* et *cpsD* provoque l'inactivation de ces gènes. *In silico*, ces trois loci ont été respectivement retrouvés sur les fragments FR1, FR14 et FR7. Ceci est en faveur d'une insertion de l'IS*1548* à leur niveau, mais reste à vérifier par clonage et séquençage. L'IS*1548* a été retrouvée en copie unique chez une seule souche (LJ1). Chez cette souche, l'IS*1548* est retrouvée au niveau du FR1 son site d'insertion étant potentiellement le site X. Pour la majorité des souches IS*1548*+, FR1 et FR14 sont retrouvés. Cependant, deux souches (H9 et H14) n'ont ni FR1 ni FR14, et trois souches n'ont pas hybridé en FR14 (LJ1, G35 et H7).

1.5. L'IS*Sa4*

Quatre souches possèdent l'IS*Sa4*. Le nombre de copies est de trois (pour deux souches), quatre (pour une souche) et 20 (pour une souche) (**Annexe F**). Vingt fragments de taille différente ont permis de définir quatre profils. Le site d'insertion identifié au niveau du locus *cylB* (gène impliqué dans le transport de l'hémolysine) est potentiellement présent sur le FR2, lequel est retrouvé uniquement chez la souche G64 (Spellerberg *et al.*, 2000).

1.6. Bilan du Southern blot

L'étude par Southern blot de la distribution des EGM dans le génome de la bactérie met en évidence une variabilité importante dans la distribution des EGM entre les souches. Par ailleurs, si l'approche globale en terme de prévalence des EGM n'a pas montré de corrélation évidente entre un EGM donné et un potentiel infectieux particulier, le Southern blot a mis en évidence des fragments de restriction porteurs de l'EGM potentiellement corrélé au pouvoir pathogène de la souche.

Tableau XXI. Sites d'insertion des éléments génétiques mobiles (EGM) identifiés par PCR inverse chez *S. agalactiae*.

EGM	Souche	Longueur du fragment (Kb)	Gène	Longueur de l'alignement (pb)/% d'identité	Site d'insertion	Référence
GBSi1	L23	10	*scpB/lmb*	500/98	Site X*	Luan *et al.*, 2003
IS*861*	G49	3	*cpsX/cpsY*	445/99	En aval du gène *cps*X	Koskiniemi *et al.*, 1998

*site défini **Figure 33**

Tableau XXII. Sites d'insertion de GBSi1 identifiés par Ligation-Mediated PCR pour différentes souches de *S. agalactiae*.

Souche	Longueur du fragment (Kb)*	Gène	Longueur de l'alignement (pb)/% d'identité	Site d'insertion**	Référence
L17	0,45 (as)	*lplA-1/sag0883*	450/100	Site Z	Luan *et al.*, 2003
	0,3 (as)	*ftsY/sag0728*	290/97	Site Y	Luan *et al.*, 2003
	0,6 (as)	*scpB/lmb*	450/99	Site X	Luan *et al.*, 2003
G64	0,5 (as)	*scpB/lmb*	460/99	Site X	Luan *et al.*, 2003
	0,3 (as)	*ftsY/sag0728*	250/96	Site Y	Luan *et al.*, 2003
H9	0,2 (s)	*scpB/lmb*	180/97	Site X	Luan *et al.*, 2003
H14	0,45 (as)	*lplA-1/sag0883*	479/99	Site Z	Luan *et al.*, 2003

*as, utilisation de l'amorce anti-sens, s, utilisation de l'amorce sens

sites définis **Figure 33

2. Comparaison des trois protocoles testés

Nous avons successivement testé trois approches expérimentales et comparé leur rendement respectif. Pour l'étape de mise au point, GBSi1 a été choisi en raison du nombre limité de copies présentes dans le génome de *S. agalactiae*. Au préalable, les protocoles ont été testés sur les souches L23, G64, H9 et H14 car elles sont représentatives de la diversité des fragments de restriction obtenus en Southern blot (**Annexe D**). En cas de succès, le protocole testé a été appliqué à l'ensemble du souchier et autres EGM.

2.1. Extraction des fragments de restriction (Protocole A)

Au terme des six essais réalisés, 52 clones ont été retenus car ils contenaient un plasmide présentant un insert de taille comprise entre 6 et 10 Kb, c'est-à-dire compatible avec la taille des fragments de restriction porteurs des copies de GBSi1 selon les évaluations du Southern blot. Après réalisation des PCR de vérification, aucun clone n'a montré la présence de l'EGM. Ce résultat a conduit à l'abandon de cette approche.

2.2. PCR inverse (Protocole B)

Un fragment portant GBSi1 a été amplifié et séquencé après la PCR nichée (**Tableau XXI**). Cette copie de GBSi1 était insérée entre les gènes *scpB* et *lmb* codant respectivement une C5a-peptidase et un récepteur à la laminine. Ce fragment d'insertion correspondait au site X défini par Luan *et al.* (Luan *et al.*, 2003).

Le protocole a ensuite été appliqué pour les quatre autres EGM mais, en dépit d'une optimisation des étapes de restriction et d'autoligation, un seul site d'insertion a pu être identifié. Il s'agit de l'insertion de l'IS*861* entre les gènes *cps*Y et *cps*X qui codent des régulateurs transcriptionnels impliqués dans la synthèse de la capsule de *S. agalactiae* (Koskiniemi *et al.*, 1998).

Au final, compte tenu de la faible rentabilité de la technique, nous avons choisi de ne pas poursuivre cette approche.

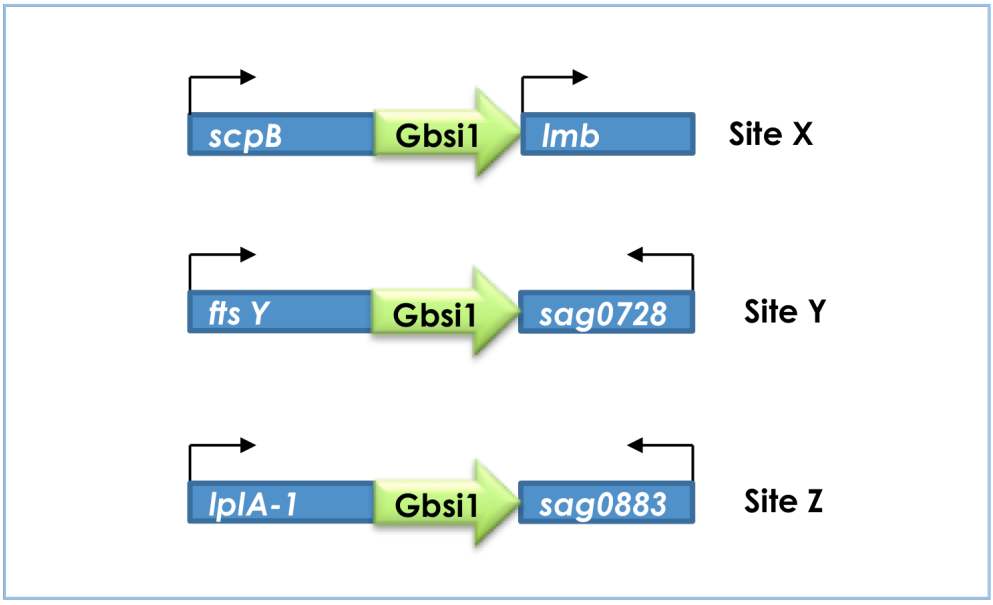


Figure 33. Représentation schématique des régions flanquant GBSi1 chez *S. agalactiae* (d'après Luan *et al.*, 2003). Les gènes concernés codent les fonctions suivantes : *ftsY* code une particule de signalisation; *sag0728* code une protéine chaperonne pour le transport ABC; *sag0883* code une hypothétique enzyme acide cohybrique synthase; *lplA-1* code une hypothétique lipoate ligase; *scpB* code la C5-a peptidase; *lmb* code un récepteur à la laminine.

Tableau XXIII. Sites d'insertion de l'IS*1548* identifiés par Ligation-Mediated PCR pour différentes souches de *S. agalactiae*.

Souche	Longueur du fragment (Kb)*	Gène	Longueur de l'alignement (pb)/% d'identité	Rôle	Référence
G35	0,5 (as)	*hylB*	418 / 97	inactive la synthèse de la hyaluronate lyase	(Granlund *et al.*, 1998)
	0,3 (as)	*gbs0667 (sag0696)*	290 / 97	inactive la synthèse de l'enzyme IIBC du système PTS	(Glaser *et al.*, 2002)
H9	1,4 (as)	*sag0911 (gbs0922)*	400 / 99	inactive la synthèse d'un transporteur ABC cationique ATPase	(Tettelin *et al.*, 2002)
	0,6 (as)	*gbs0667 (sag0696)*	374 / 97	inactive la synthèse de l'enzyme IIBC du système PTS	(Glaser *et al.*, 2002)
	1,6 (as)	*mfd /sag0009*	346 / 88	insertion en aval du gène *mfd* : rôle ?	(Tettelin *et al.*, 2002)
	1,2 (as)	*sag0156 /adcR*	419 / 100	insertion entre ces deux gènes : rôle ?	(Tettelin *et al.*, 2002)
L9	1,3 (as)	*sag1963 (gbs1950)*	398 / 94	inactive la synthèse d'un transporteur ABC phosphate	(Tettelin *et al.*, 2002)
	1,7(as)	*sag2048 (gbs2005)*	397 / 99	inactive la synthèse d'une hypothétique 5-méthyltétrahydrofolate-homocystéine méthyltransférase	(Tettelin *et al.*, 2002)

*as, utilisation de l'amorce antisens

2.3. Ligation Mediated-PCR (Protocole C)

Appliqué initialement à l'identification des sites d'insertion de GBSi1, ce protocole a permis d'identifier les trois principaux sites d'insertion (cf. § 3.1.1). Compte tenu des bons résultats obtenus pour GBSi1, la LM-PCR a été appliquée aux autres EGM et a permis d'identifier, à ce jour, les sites d'insertion de l'IS*1548* et de l'IS*1381,* en plus de ceux de GBSi1.

3. Identification des sites d'insertion des EGM par LM-PCR

3.1.1. Identification des sites d'insertion de GBSi1

Treize fragments de tailles différentes portant au moins une copie de GBSi1 ont été observés en Southern Blot (**Annexe D**). La LM-PCR a permis, après clonage et séquençage des amplicons et analyse *in silico* par BLAST, d'identifier trois sites d'insertion pour GBSi1 (**Tableau XXII**) :

- un site correspondant au site X précédemment cité situé à la jonction des gènes *scpB* et *lmb,*
- un site correspondant à l'insertion de GBSi1 entre le gène *fts*Y codant une protéine de reconnaissance et *sag0728* codant une protéine chaperonne pour les substrats des transporteurs de type ABC (« ATP-Binding Cassette ») (Tettelin *et al.*, 2002),
- un site correspondant à l'insertion de GBSi1 entre le gène *lpl*A-1 codant une lipoate ligase et *sag0883* codant une enzyme, l'acide cohybrique synthase (Tettelin *et al.*, 2002).

Ces trois sites correspondent respectivement au site X, au site Y et au site Z décrits précédemment par Luan *et al.* (**Figure 33**) (Luan *et al.*, 2003). L'identification du site X pour la souche H9 et la souche L23, qui correspondrait respectivement au FR4 et au FR1 observés par Southern blot (**Annexe D**), permet d'émettre l'hypothèse de deux exemplaires du site X pour les souches L13, L17 et G64 qui présentent ces deux fragments de restriction. Une analyse par Southern blot avec des sondes recouvrant GBSi1 et *lmb* ou/qet *scp*B permettrait de confirmer cette hypothèse.

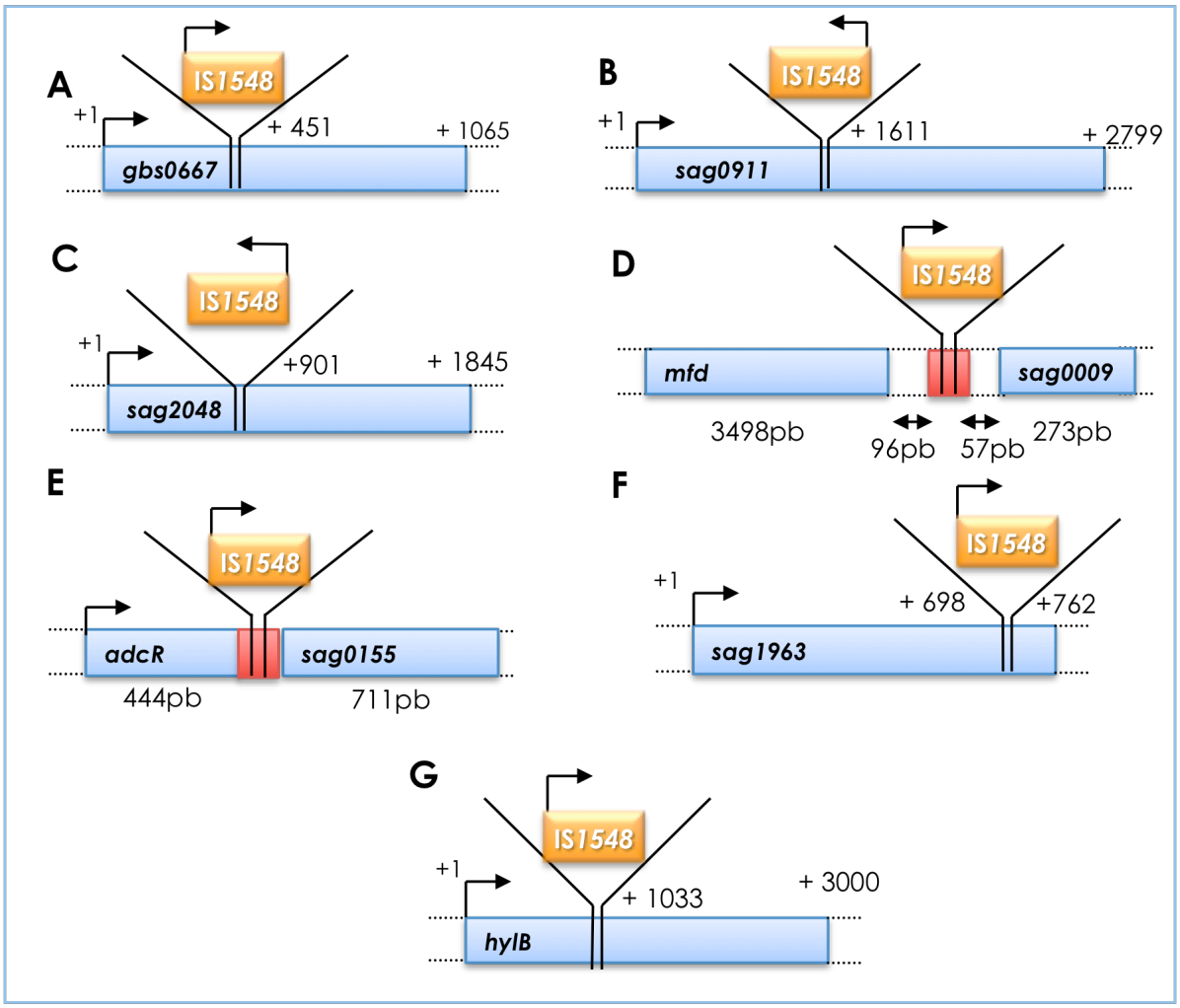


Figure 34. Représentation schématique des régions flanquant l'IS*1548* chez *S. agalactiae*.
L'IS*1548* s'insère dans le gène *gbs0667* (A) codant une hypothétique enzyme IIBC d'un système PTS, dans le gène *sag0911* (B) codant un transporteur cationique ATPase, dans le gène *sag2048* (C) codant une hypothétique 5-méthyltétrahydrofolate-homocystéine méthyltransférase, en aval du gène *mfd* (D) un facteur de translocation, entre le gène *adcR et sag0155* (E) codant respectivement un répresseur et un transporteur ABC du zinc, dans le gène *sag1963* (F) codant un transporteur ABC phosphate et dans le gène *hylB* (G) codant la hyaluronate lyase.

Concernant plus spécifiquement H14, unique par son profil GBSi1 composé de 12 fragments (**Annexe D**), elle possède outre les trois sites d'insertion décrits ci-dessus, neuf autres sites qui n'ont pu être identifiés par l'analyse *in silico*, car malgré l'obtention de trois autres produits d'amplification le séquençage n'a pas permis l'identification du site d'insertion.

3.1.2. Identification des sites d'insertion de l'IS*1548*

Sur les 15 sites d'insertion potentiels reconnus par Southern blot (**Annexe E**), la LM-PCR suivie du clonage, du séquençage et BLAST sur le site de la NCBI, a permis d'identifier sept sites d'insertion de l'IS*1548*, dont un seul déjà décrit, le gène *hyl*B codant la hyaluronate lyase (**Tableau XXIII**).

Parmi les six nouveaux sites identifiés (**Figure 34**), quatre correspondent à l'insertion de l'IS dans l'ORF d'un gène. Il s'agit :

- du gène *gbs0667* qui code une enzyme IIBC d'un système PTS dévolu au transport actif d'un substrat (Glaser *et al.*, 2002),
- du gène *sag1963* codant un transporteur de type ABC permettant le transport actif du phosphate inorganique (Tettelin *et al.*, 2002),
- du gène *sag0911* codant un transporteur cationique ATPase (Tettelin *et al.*, 2002),
- du gène *sag2048* impliqué dans la biosynthèse de la méthionine codant une 5-méthyltétrahydrofolate-homocystéine méthyltransférase (Tettelin *et al.*, 2002).

Les deux derniers sites d'insertion identifiés sont localisés dans des fragments intergéniques :

- un site situé entre le gène *sag0155* codant un transporteur ABC du zinc et son opéron répresseur *adc*R (Tettelin *et al.*, 2002),
- un site situé à 96 pb en aval du gène *mfd* qui code un facteur de transcription permettant l'arrêt de la transcription lorsque l'ADN est endommagé (Tettelin *et al.*, 2002).

3.1.3. Identification des sites d'insertion de l'IS*1381*

Sur les 18 fragments de restriction IS*1381*+ identifiés par Southern blot (**Annexe B**), seuls deux sites d'insertion de l'IS ont été identifiés :

Tableau XXIV. Distribution des copies de l'IS*1548* et de GBSi1 au sein de la collection (n=52) de *S. agalactiae*

		Site d'insertion de GBSi1			Site d'insertion de l'IS*1548*						
Souche	Sérotype	*ftsY /sag0728*	*ipiA-1 /sag0883*	*scpB /lmb*	*scpB /lmb*	*hylB*	*cpsD*	*sag0155 / adcR*	*gbs0667*	*sag0911*	*sag2048*
L1	III	+	+	+	/	/	/	/	/	/	/
L9	III	/	/	/	+	+	-	+	+	+	+
L10	III	/	/	/	+	+	-	+	+	-	+
L13	III	+	+	+	/	/	/	/	/	/	/
L14	III	/	/	/	+	+	-	+	+	+	-
L15	IV	/	/	/	/	/	/	/	/	/	/
L16	III	/	/	/	/	/	/	/	/	/	/
L17	III	+	+	-*	/	/	/	/	/	/	/
L19	Ia	/	/	/	/	/	/	/	/	/	/
L22	III	/	/	/	+	+	-	-	-	+	-
L23	III	-	-	+	/	/	/	/	/	/	/
L24	III	+	+	+	/	/	/	/	/	/	/
L29	III	/	/	/	+	+	-	+	+	+	+
L34	III	/	/	/	+	+	-	+	+	+	+
L36	III	/	/	/	+	+	+	-	-	-	+
L44	Ia	+	+	+	/	/	/	/	/	/	/
LJ1	III	/	/	/	+	-	-	-	-	-	-
LJ4	III	/	/	/	+	+	-	-		-	+
LJ5	Ib	/	/	/	/	/	/	/	/	/	/
LJ6	III	/	/	/	+	+	-	-	+	+	+
G8	Ia	/	/	/	/	/	/	/	/	/	/
G35	III	/	/	/	+	+	+	-	+	-	-
G49	III	+	+	+	/	/	/	/	/	/	/
G61	V	/	/	/	/	/	/	/	/	/	/
G64	II	+	+	+	/	/	/	/	/	/	/
V5	Ib	/	/	/	/	/	/	/	/	/	/
V18	III	+	+	+	/	/	/	/	/	/	/
V21	II	-	-	+	/	/	/	/	/	/	/
V25	NT	/	/	/	/	/	/	/	/	/	/
V64	III	+	+	+	/	/	/	/	/	/	/
V71	Ib	/	/	/	/	/	/	/	/	/	/
VV2	III	/	/	/	+	+	+	-	-	+	+
VV4	III	/	/	/	+	+	+	-	-	+	+
VV105	III	/	/	/	+	+	+	-	+	+	+
VV188	III	/	/	/	+	+	+	-	-	+	+
H1	III	+	+	+	/	/	/	/	/	/	/
H2	NT	/	/	/	/	/	/	/	/	/	/
H3	NT	/	/	/	/	/	/	/	/	/	/
H4	Ib	/	/	/	/	/	/	/	/	/	/
H5	III	+	+	+	/	/	/	/	/	/	/
H6	III	/	/	/	/	/	/	/	/	/	/
H7	II	/	/	/	+	+	-	-	-	+	-
H8	III	+	+	+	/	/	/	/	/	/	/
H9	V	-	-	+	-	-	-	-	+	+	+
H10	Ib	/	/	/	/	/	/	/	/	/	/
H11	NT	/	/	/	/	/	/	/	/	/	/
H12	III	/	/	/	+	+	+	-	-	-	+
H13	Ib	/	/	/	/	/	/	/	/	/	/
H14	II	+	+	+**	-	+	-	-	-	-	-
H15	III	/	/	/	/	/	/	/	/	/	/
H16	III	/	/	/	/	/	/	/	/	/	/
H17	Ib	/	/	/	/	/	/	/	/	/	/

/, souches ne contenant pas l'élément génétique mobile (EGM) étudié
+, souches possédant le site d'insertion de l'EGM
-, souches possédant l'EGM dans leur génome mais pas dans le site d'insertion concerné
*la souche est négative en PCR de jonction l'identification de ce site a été effectuée pour cette souche (Tableau 3-VI)
**la PCR de jonction montre deux fragments à des tailles différentes.

- un site correspondant à l'insertion de l'IS*1381* dans le gène *sag1433* codant une protéine de fonction inconnue (Tettelin *et al.*, 2002)
- un site correspondant à l'insertion de l' IS*1381* dans le gène *cps*Y, régulateur transcriptionnel pour la synthèse de la capsule (Koskiniemi *et al.*, 1998).

4. Distribution des copies des EGM sur le génome

Au cours d'une étape ultérieure, la démarche a été de déterminer la prévalence de ces sites d'insertion de GBSi1 et de l'IS*1548* pour la collection d'étude comptant 52 souches humaines. C'est dans cet objectif que des PCR spécifiques à chaque site d'insertion ont été réalisées.

4.1. Sites d'insertion de GBSi1

Le site d'insertion X (insertion entre *scp*B et *lmb*) est présent chez toutes les souches GBSi1+, quel que soit le contexte clinique d'isolement (infection invasive ou portage asymptomatique) (**Tableau XXIV**). Par conséquent, ce site d'insertion ne semble pas avoir de corrélation directe avec le potentiel infectieux des souches.

Le site Y (insertion entre *fts*Y et *sag0728*) et le site Z (insertion entre *lpl*A-1 et *sag0883*) sont présents chez 13 des 16 souches GBSi1+ et sont toujours retrouvés associés (**Tableau XXIV**). De la même manière que le site X, il ne semble que les sites Y et Z aient une quelconque corrélation avec le potentiel infectieux des souches.

4.2. Sites d'insertion de l'IS*1548*

Pour l'IS*1548*, en plus des sites nouvellement décrits dans ce travail expérimental, deux sites d'insertion précédemment rapportés dans la littérature ont été également recherchés : le fragment intergénique *scp*B-*lmb* (Granlund *et al.*, 1998) aussi dénommé site X (Luan *et al.*, 2003) et le gène *cps*D (Sellin *et al.*, 2000).

Le site X est retrouvé dans 17 des 19 souches possédant l'IS*1548* (**Tableau XXIV**). Il faut noter que les deux souches [IS*1548*+ ; site X -] (H9 et H14) sont les seules à porter simultanément l'IS*1548* et GBSi1, avec GBSi1 présent au site X. L'insertion de l'IS*1548* dans *hyl*B, *gbs0667*, *sag0911* et *sag2048* est retrouvée respectivement chez 16 souches, 9 souches, 12 souches et 13 souches. Ces souches possèdent l'IS*1548*, mais il n'existe pas de corrélation évidente avec le potentiel infectieux.

En revanche, l'analyse de la prévalence du site d'insertion de l'IS*1548* dans le fragment intergénique *adc*R-*sag0155* montre qu'il n'est retrouvé que chez des souches responsables de méningites néonatales sur cinq des dix souches possédant l'IS*1548* isolées de LCR (**Tableau XXIV**). De même, le gène *cps*D codant une galactosyl transférase, impliquée dans la synthèse de la capsule, est un site d'insertion pour sept des 19 souches possédant IS*1548* dont cinq sont des souches de colonisation vaginale ou gastrique (**Tableau XXIV**).

Partie 3. Sites d'insertion des EGM dans le génome de *Streptococcus agalactiae*

D-Discussion

Dans notre travail, la recherche des sites d'insertion des EGM chez *S. agalactiae* a mis en évidence des sites d'insertion non publiés à ce jour, dont cinq situés dans la séquence codante d'un gène, avec pour conséquence probable l'interruption de la transcription du gène. Ces gènes sont impliqués, pour la plupart, dans des fonctions métaboliques de la bactérie (transport de substrats cationiques ou carbonés, biosynthèse d'acide aminé). L'IS*1548* par exemple, s'insère dans deux transporteurs ABC, l'un spécifique au transport actif des cations (*sag0911*), l'autre spécifique au transport du phosphate inorganique (*sag1963*). Ces deux substrats jouent un rôle majeur dans la physiologie bactérienne en intervenant dans la transduction intracellulaire du signal ou en tant que cofacteurs de nombreuses enzymes (Novak *et al.*, 1999, Janulczyk *et al.*, 1999). Un autre gène cible, *sag2048*, intervient notoirement dans le métabolisme bactérien en codant une enzyme impliquée dans la biosynthèse de la méthionine. Cependant, l'inactivation de ces gènes ne semble pas corrélée au pouvoir pathogène des souches. Ceci pourrait s'expliquer par la redondance des systèmes de transport (transporteurs ABC ou perméases à substrats multiples) et des systèmes de biosynthèse des acides aminés mis en évidence lors du séquençage complet des génomes de *S. agalactiae* (Glaser *et al.*, 2002, Tettelin *et al.*, 2002). De même, il apparaît que *S. agalactiae* contient dans son génome de nombreux transporteurs de type PTS (phosphotransferase system) permettant le transport actif de substrats via l'hydrolyse du phosphoénolpyruvate (Glaser *et al.*, 2002). Mais, contrairement aux transporteurs ABC, ces transporteurs PTS possèdent une étroite spécificité de substrat. Comme le montre notre travail, l'IS*1548* possède un site d'insertion dans un des composants d'un transporteur PTS, l'enzyme transmembranaire IIBC (*sag0667*) impliquée spécifiquement dans le transport du lactose (Postma *et al.*, 1993). Cette insertion pourrait engendrer l'inactivation du transporteur et expliquer l'inadaptation des souches portant l'IS*1548* à ce *locus*. En effet, elles deviendraient incapables d'occuper les niches écologiques riches en lactose telles que les tissus épithéliaux de la mamelle des bovins. Il a été rapporté que les souches bovines fermentaient le lactose dans leur quasi totalité alors que les souches humaines en étaient incapables (Finch & Martin, 1984, Duarte *et al.*, 2004, Mosabi *et al.*, 1997). Or, nos résultats montrent que l'IS*1548* est, précisément, un marqueur exclusif de l'écosystème humain. L'insertion de l'IS*1548* dans le gène codant l'enzyme IIBC du transporteur du lactose pourrait donc avoir un impact direct dans le tropisme d'hôte de

S. agalactiae.

Outre l'identification de transporteurs impliqués dans l'adaptation de *S. agalactiae*, notre travail a identifié un système de transport impliqué dans la virulence de la bactérie. Il s'agit de l'opéron zinc pour lequel la LM-PCR a révélé une insertion de l'IS*1548,* située entre le gène codant le transporteur ABC du zinc (*sag0155*) et son répresseur *adcR* (*sag0154*). Chez *S. pyogenes*, la concentration extracellulaire du zinc influe sur la synthèse d'une protéine connue chez *S. agalactiae* pour médier l'attachement à la laminine humaine, la protéine Lmb (Spellerberg *et al.*, 1999b, Elsner *et al.*, 2002). De faibles concentrations en zinc (comme celles présentes dans les muqueuses et les tissus) entraînent une faible sécrétion d'AdcR, qui possède un site d'attachement en amont du gène *lmb*, ce qui favorise l'expression des gènes des transporteurs du zinc et des protéines d'adhésion comme Lmb (Panina *et al.*, 2003). Cette phase correspondrait à l'étape d'adhésion de *S. agalactiae* intervenant au moment de la colonisation des muqueuses ou au moment du franchissement de la barrière hémato-encéphalique. De plus fortes concentrations en zinc (comme celles rencontrées dans le sang) entraînent une sécrétion accrue d'AdcR qui inhibe alors la synthèse des transporteurs du zinc et des protéines d'adhésion (Panina *et al.*, 2003). Cette phase correspondrait à l'étape d'envahissement sanguin par *S. agalactiae*. Des éléments objectifs ont étayé ultérieurement le scénario proposé par Panina *et al.* en 2003 (Panina *et al.*, 2003). Chez *S. suis,* l'inactivation du gène *adcR* entraîne une dérépression des transporteurs du zinc et des facteurs de virulence (Aranda *et al.*, 2009). Chez *S. agalactiae*, la transcription du gène *AdcC* (transporteur du zinc) est diminuée en cas d'incubation de la bactérie avec du sang humain (Mereghetti *et al.*, 2008). Nos résultats montrent que l'insertion de l'IS*1548* dans l'opéron zinc est observée uniquement chez des souches responsables de méningites. L'IS*1548* pourrait donc avoir inactivé le répresseur *adcR* et entraîné ainsi la surexpression de certains gènes de virulence tels que celui de la protéine Lmb qui favorise l'invasion de la barrière hémato-encéphalique (Tenenbaum *et al.*, 2007). L'étude de variants isogéniques dans des modèles d'infection à *S. agalactiae in vitro*, *ex vivo* ou *in vivo,* serait nécessaire pour valider cette hypothèse.

Le gène *lmb* est également concerné par l'insertion des EGM à un degré plus direct. En effet, nos résultats montrent que le fragment intergénique *scpB-lmb* est un

site à haute fréquence d'insertion, ce qui confirme la qualité de haute plasticité génomique affectée à cette région chromosomique (Broker & Spellerberg, 2004, Tettelin *et al.*, 2005, Brochet *et al.*, 2008b). Nous avons noté une insertion systématique mais jamais concomitante de l'un des deux EGM à ce site, et pour toutes les souches possédant IS*1548* et/ou GBSi1. Un travail réalisé au sein de notre équipe a montré que l'insertion de l'IS*1548* (mais pas celle de GBSi1) entraînait une augmentation significative de la transcription du gène *lmb* (Al Safadi *et al.*, 2009). La production de protéine Lmb était significativement plus élevée chez les souches possédant l'IS*1548* par rapport aux souches possédant GBSi1 ou ne possédant aucun EGM au niveau du locus *scpB-lmb* (Al Safadi *et al.*, 2009). De même, l'adhésion à la laminine était significativement augmentée chez les souches possédant l'IS*1548*. Ce site (site X) a été suspecté d'être le premier site d'insertion de GBSi1 par Luan *et al.* (Luan *et al.*, 2003). Ces auteurs retrouvaient par ailleurs systématiquement ce site associé à deux autres sites d'insertion, les sites Y (*lplA1-sag0883*) et Z (*ftsY-sag0728*) (Luan *et al.*, 2003). Nous avons montré que les sites Y et Z étaient effectivement systématiquement associés au site X, mais pour deux souches (L23 et H9) le site X était l'unique site d'insertion, prouvant ainsi qu'il s'agit très probablement du site d'insertion initial de GBSi1. Le site X, encadré par deux ISSag2, constituerait un transposon composite acquis suite à un transfert horizontal (Franken *et al.*, 2004, Franken *et al.*, 2001). L'hypothèse d'une transposition réplicative de celui-ci pourrait expliquer plusieurs phénomènes : i) la présence en double du site X chez certaines souches possédant GBSi1 (L13, L17 et G64), ii) la différence de taille du fragment de restriction du site d'insertion X observée chez L23 et H9, et iii) la similitude de séquences observée entre les différents fragments GBSi1 de la souche H14.

L'analyse du locus *cps* a montré l'existence d'une forte corrélation entre la présence de l'IS*1548* dans le gène *cpsD* et le caractère non invasif des souches : toutes les souches vaginales et gastriques possèdent une copie de l'IS*1548* au niveau de ce site contre une seule des dix souches de méningites. Le locus *cps* est un opéron impliqué dans la synthèse de la capsule polysaccharidique, facteur de virulence majeur de *S. agalactiae* car il bloque l'opsonisation de la bactérie (Maisey 2009). Paradoxalement, ce locus semble être un « hot-spot » d'insertion des EGM car l'équipe de Sellin *et al.* avait précédemment observé l'insertion de l'IS*861* et de l'IS*1548* dans le gène *cpsD* (Sellin *et al.*, 1995, Sellin *et al.*, 2000). L'insertion de l'IS*1548*

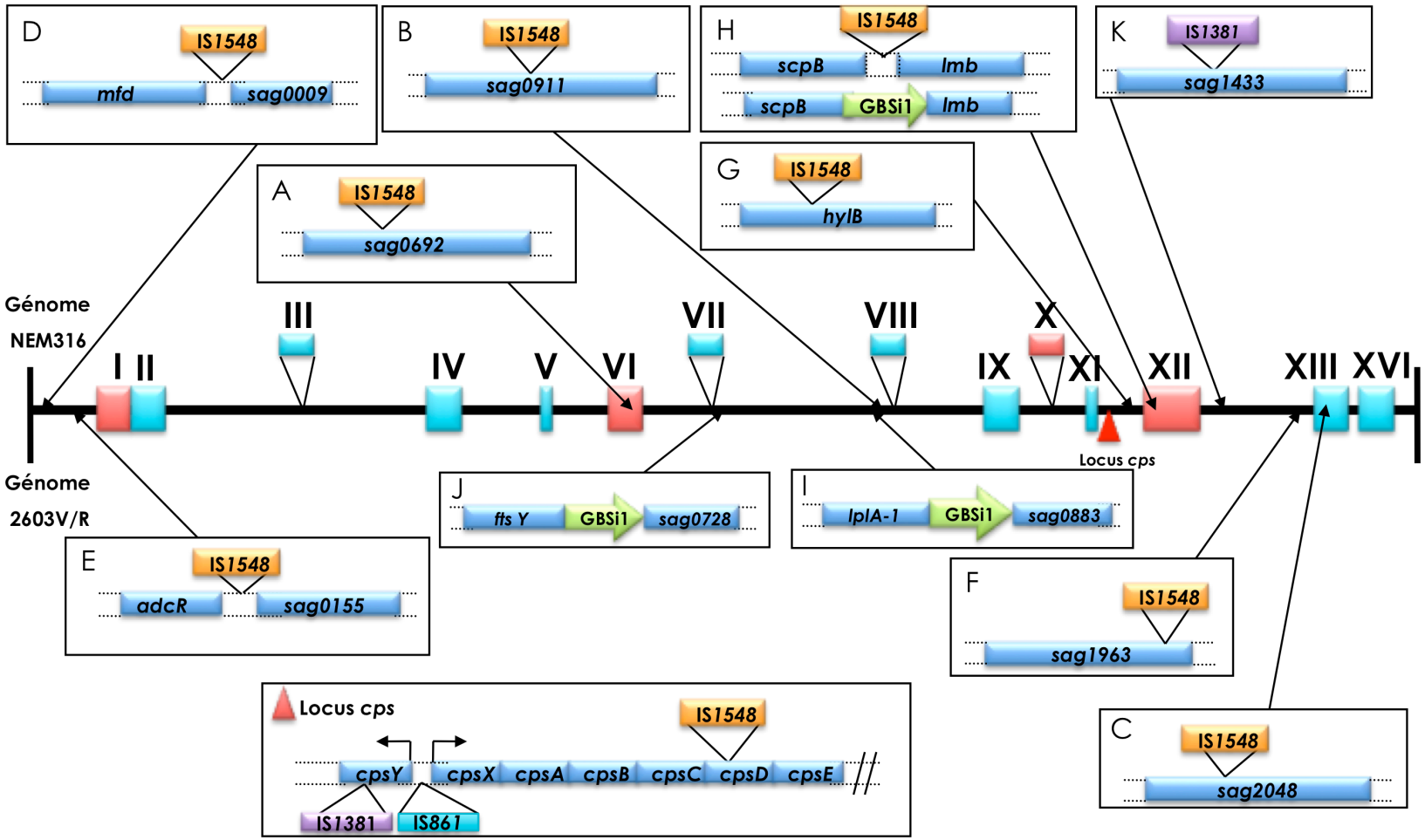


Figure 35. Représentation des différents sites d'insertion d'éléments génétiques mobiles dans le génome de *S. agalactiae* de sérotype III (NEM316) et de sérotype V (2603V/R) (adapté de Herbert *et al.*, 2005). Les similitudes entre les génomes sont indiquées par la ligne horizontale. Les îlots génomiques (I à XVI) sont représentés par des rectangles bleus. Les îlots de pathogénicité sont représentés par des rectangles rouges. L'IS*1548* s'insère dans le gène *sag0692* (A), *sag0911* (B), *sag2048* (C), en aval du gène *mfd* (D), entre le gène *adcR* et *sag0155* (E), dans le gène *sag1963* (F), dans le gène *hylB* (G), et entre les gènes *scpB* et *lmb* (H). GBSi1 s'insère entre les gènes *scpB* et *lmb* (H), entre les gènes *lplA-1* et *sag0883* (I), et entre les gènes *ftsY* et *sag0728* (J). L'IS*1381* s'insère dans le gène *sag1433* (K). Une partie du locus *cps*, représenté par un triangle rouge, comporte l'insertion de l'IS*1548* dans le gène *cpsD*, de l'IS*1381*dans le gène *cpsY* et de l'IS*861* en aval du gène *cpsX*.

dans *cpsD* affecterait directement la virulence en entraînant l'inactivation de la galactosyl-transférase, l'enzyme indispensable pour la synthèse de la capsule (Sellin *et al.*, 2000, Rubens *et al.*, 1993). L'identification de nouveaux sites d'insertion pour l'IS*861* et l'IS*1548* situés à l'intérieur ou à proximité des gènes *cpsX* et *cpsY*, codant des régulateurs transcriptionnels du locus *cps*, renforce par ailleurs l'hypothèse d'une influence des EGM dans la modulation des polysaccharides de surface.

L'analyse comparative des souches NEM316 et 2603V/R, qui montre une architecture conservée entre les deux génomes, montre d'ailleurs que les différences sont principalement observées au niveau du locus associé à la capsule et dans la distribution des EGM ; ceci constituerait la preuve du rôle des EGM en tant que déterminants génétiques importants dans le pouvoir pathogène des souches (Herbert *et al.*, 2005). Trois des sites d'insertion mis en évidence dans notre étude sont associés aux îlots de pathogénicité définis par Herbert *et al.* (**Figure 35**) (Herbert *et al.*, 2005): *gbs0667* à l'îlot VI, *sag2048* à l'îlot XIII et le site X (*scpB-lmb*) à l'îlot XII. Par ailleurs, l'analyse de la localisation des sites d'insertion des EGM sur les génomes annotés montre la présence fréquente d'ADN phagique, plasmidique ou de transposon à proximité. C'est par exemple le cas de *gbs0667* situé à côté des gènes du transposon TN5252 chez la souche NEM316, ou de *sag2048* situé à proximité des gènes phagiques chez la souche 2603V/R. Ces observations sont en faveur d'une acquisition des EGM par transfert horizontal, comme pour GBSi1 qui aurait été transmis *via* un plasmide (Luan *et al.*, 2003). Par leur activité de transposition, les EGM participent aux évènements de recombinaison et de réarrangements génomiques voire d'élaboration des îlots de pathogénicité (Chain *et al.*, 2004). Cependant, chez *S. agalactiae*, aucun des îlots ne serait par lui-même suffisant pour fixer le pouvoir pathogène d'une souche (Herbert *et al.*, 2005), ce qui pourrait suggérer que la virulence de certaines souches proviendrait plus de la modulation de gènes par de petits EGM comme les séquences d'insertion que de l'acquisition de grosses structures comme les îlots de pathogénicité.

IV-CONCLUSION ET PERSPECTIVES

Responsable de mammites chez les bovins, simple commensal des épithéliums de l'homme, première étiologie des infections néonatales chez l'homme, pathogène émergent chez les sujets immunodéprimés, *S. agalactiae* est une bactérie au pouvoir pathogène complexe, capable de s'adapter à diverses niches écologiques. Cette adaptabilité, liée à l'évolution de l'espèce, est due en partie à la présence d'EGM qui, en participant aux évènements de recombinaison homologue et aux transferts horizontaux, représentent une source majeure de diversité génétique. C'est la raison pour laquelle nous avons choisi de nous intéresser au rôle des EGM dans l'évolution et la virulence de *S. agalactiae*.

Notre travail par MLST a montré l'existence d'un niveau important d'événements de recombinaison dans la population de *S. agalactiae*, ceux-ci brouillant les signaux phylogénétiques, particulièrement pour CC19. Notre analyse phylogénétique approfondie a démontré que le fond génétique de CC19 était partagé par une partie de la population de *S. agalactiae*. Ce résultat est concordant avec le modèle d'évolution proposé par Brochet *et al.* dans lequel un seul clone, génétiquement proche de CC19, aurait disséminé dans la population et les différents complexes clonaux en seraient dérivés après échanges de grands fragments d'ADN (Brochet *et al.*, 2008b). Par ailleurs, l'analyse par MLST a permis d'identifier l'existence de la lignée clonale CC23, alors que cette dernière n'était pas différenciée par MLEE. La différenciation de ce clone est importante car il présente des spécificités épidémiologiques et biologiques intéressantes, notamment une plus grande propension à déclencher des infections néonatales précoces (Martins *et al.*, 2007).

La poursuite de l'analyse phylogénétique réalisée dans la première partie de notre travail, nous a conduit à déterminer la prévalence des EGM présents chez *S. agalactiae*. Une très bonne corrélation entre la distribution de certains EGM et les divisions (et sous-divisions) phylogénétiques a été démontré, ce qui nous a permis de formuler une hypothèse expliquant la chronologie d'acquisition des EGM au cours de l'évolution de *S. agalactiae*. Pour cette raison, les EGM peuvent être retenus comme des marqueurs pertinents de l'évolution, hypothèse qui n'est pas contradictoire avec leur capacité de mobilité. Certains EGM apparaissent même

comme des marqueurs de l'origine humaine ou bovine des souches. L'observation la plus significative est que la lignée humaine hyperinvasive CC17 possède l'intron GBSi1, marqueur spécifique de l'écosystème bovin. Dans la mesure où le clone humain CC17 est phylogénétiquement plus proche du clone bovin CC67 que des autres lignées humaines, l'hypothèse de l'émergence d'un clone responsable de la majorité des infections néonatales invasives à partir d'un ancêtre bovin s'en trouve renforcée (Bisharat *et al.*, 2004). Du reste, l'émergence des principales lignées phylogénétiques semble fréquemment associée à un événement génétique lié aux EGM. Ainsi, l'acquisition de l'IS*Sag1* et de l'IS*Sag9*, observée chez toutes les lignées humaines, pourrait offrir un avantage sélectif favorisant l'adaptation à l'homme. L'excision de SAG0448, observée notamment chez le clone hyperinvasif CC17, pourrait permettre la récupération de la fonctionnalité d'un transporteur de magnésium impliqué dans la virulence de la bactérie.

Les corrélations observées entre les EGM, les lignées phylogénétiques et le potentiel de virulence, nous ont conduit à rechercher si les EGM, plus que de simples marqueurs, pouvaient être des acteurs directs de la virulence de *S. agalactiae*. La première étape a été de déterminer le nombre de copies des EGM et de cartographier les sites d'insertion de ces EGM. Nous avons confirmé l'existence de sites déjà connus dans la littérature, mais surtout mis en évidence huit sites d'insertion non publiés à ce jour. Deux de ces sites d'insertion sont plus particulièrement intéressants car ils sont significativement associés à l'origine écologique des souches. L'un, retrouvé uniquement chez les souches responsables de méningites, correspond à l'insertion de l'IS*1548* entre le gène codant un transporteur ABC du zinc et son répresseur *adcR*. L'autre, associé préférentiellement aux souches de colonisation, correspond à l'insertion de l'IS*1548* dans le gène *cpsD,* gène de l'opéron contrôlant la synthèse de la capsule, principal facteur de virulence de *S. agalactiae*.

La stratégie expérimentale envisagée pour l'approfondissement de notre travail à court terme prévoit l'identification par LM-PCR des sites d'insertion de l'IS*Sag9* et SAG0448, deux EGM utiles à la compréhension de l'évolution de *S. agalactiae*. L'identification des sites d'insertion de l'IS*Sag9* peut mettre en évidence, soit la potentialisation de gènes indispensables à l'adaptation à la niche écologique humaine, soit l'inactivation de gènes métaboliques inutiles dans les

écosystèmes humains dont le « silencing » aurait engendré une augmentation du fitness de la bactérie. L'identification des sites d'insertion de SAG0448 permettra de mettre en évidence un gène ayant un rôle dans la physiopathologie des méningites à *S. agalactiae* dont l'insertion de SAG0448 aurait engendré l'inactivation chez les souches de colonisation. Le rôle potentiel de ces EGM dans l'adaptation et/ou la virulence de *S. agalactiae* sera interprété en fonction de la distribution de ces sites d'insertion au sein de la population de souches. Le travail inclura les souches bovines du CC67, car l'excision de SAGG448 est corrélée à l'émergence de CC17 à partir de CC67 selon notre hypothèse de chronologie d'acquisition des EGM. L'étude des souches bovines sera également contributive pour l'analyse de GBSi1, EGM pour lequel la LM-PCR n'a pas identifié de nouveaux sites d'insertion, alors qu'il semble être un EGM important dans la virulence de CC17. Ce travail pourrait identifier un ou plusieurs sites d'insertion additionnels, à partir desquels l'excision de GBSi1 aurait engendré l'émergence d'un clone virulent pour l'homme.

A moyen terme, il est nécessaire d'évaluer l'impact des EGM aux sites d'insertion/excision identifiés. Le dosage par RT-PCR quantitative des transcrits des gènes impliqués est la première étape de la validation des hypothèses moléculaires élaborées pour chaque EGM et chaque site d'insertion d'intérêt. Il serait intéressant de débuter l'analyse par l'IS*1548* en raison des *loci cpsD* et *sag0155*, et d'étudier leur niveau de transcription en fonction de la présence ou non de l'EGM et du potentiel infectieux des souches. On s'attend à ce que les transcrits *cpsD* soient diminués voire absents en cas d'insertion de l'IS*1548,* et qu'*a contrario*, les transcrits *sag0155* soient augmentés en cas d'insertion de l'IS*1548* dans le répresseur *adcR*.

A plus long terme, et en fonction des résultats obtenus lors des étapes précédentes, le rôle des EGM dans la virulence de *S. agalactiae* sera envisagé par l'étude de mutants isogéniques dans des modèles d'infection *in vitro*, sur cultures cellulaires telles que celles développées dans notre équipe de recherche (Al Safadi *et al.*, 2009, Rosenau *et al.*, 2007), *ex vivo*, par culture dans différents liquides biologiques d'origine humaine (liquide gastrique, liquide amniotique, sang, LCR, …) (Mereghetti *et al.*, 2008, Sitkiewicz *et al.*, 2009), voire *in vivo,* sur des modèles animaux.

V-BIBLIOGRAPHIE

Achard, A. & R. Leclercq, (2007) Characterization of a small mobilizable transposon, MTn*Sag1*, in *Streptococcus agalactiae*. *J Bacteriol* **189**: 4328-4331.

Achard, A., C. Villers, V. Pichereau & R. Leclercq, (2005) New *lnu*(C) gene conferring resistance to lincomycin by nucleotidylation in *Streptococcus agalactiae* UCN36. *Antimicrob Agents Chemother* **49**: 2716-2719.

Al Safadi, R., A. Amor, G. Héry-Arnaud, B. Spellerberg, P. Lanotte, F. Gannier, R. Quentin & A. Rosenau, (2009) The insertion element IS*1548* induces the expression of the laminin-binding gene *lmb* in *Streptococcus agalactiae*. *Infect Immun* **soumis**.

Amin, A., Y. M. Abdulrazzaq & S. Uduman, (2002) Group B streptococcal serotype distribution of isolates from colonized pregnant women at the time of delivery in United Arab Emirates. *J Infect* **45**: 42-46.

Anonymous, (1996) Prevention of perinatal group B streptococcal disease: a public health perspective. Centers for Disease Control and Prevention. *MMWR Recomm Rep* **45**: 1-24.

Anonymous, (2001) Prévention anténatale du risque infectieux bactérien néonatal précoce. Recommandations pour la pratique clinique. *Agence Nationale d'Accréditation et d'Evaluation en Santé* **Paris**.

Anonymous, (2002) From the Centers for Disease Control. Revision of guidelines for the prevention of perinatal group B streptococcal disease. *JAMA* **287**: 1106-1107.

Anonymous, (2009) Trends in perinatal group B streptococcal disease - United States, 2000-2006. *MMWR Morb Mortal Wkly Rep* **58**: 109-112.

Aranda, J., M. E. Garrido, N. Fittipaldi, P. Cortes, M. Llagostera, M. Gottschalk & J. Barbe, (2009) Protective capacities of cell surface-associated proteins of *Streptococcus suis* mutants deficient in divalent cation-uptake regulators. *Microbiology* **155**: 1580-1587.

Badri, M. S., S. Zawaneh, A. C. Cruz, G. Mantilla, H. Baer, W. N. Spellacy & E. M. Ayoub, (1977) Rectal colonization with group B streptococcus: relation to vaginal colonization of pregnant women. *J Infect Dis* **135**: 308-312.

Baker, C. J., (1997) Group B streptococcal infections. *Clin Perinatol* **24**: 59-70.

Band, J. D., H. W. Clegg, 2nd, P. S. Hayes, R. R. Facklam, J. Stringer & R. E. Dixon, (1981) Transmission of group B streptococci. Traced by use of multiple epidemiologic markers. *Am J Dis Child* **135**: 355-358.

Barcaite, E., A. Bartusevicius, R. Tameliene, M. Kliucinskas, L. Maleckiene & R. Nadisauskiene, (2008) Prevalence of maternal group B streptococcal colonisation in European countries. *Acta Obstet Gynecol Scand* **87**: 260-271.

Bergeron, M. G., D. Ke, C. Menard, F. J. Picard, M. Gagnon, M. Bernier, M. Ouellette, P. H. Roy, S. Marcoux & W. D. Fraser, (2000) Rapid detection of group B streptococci in pregnant women at delivery. *N Engl J Med* **343**: 175-179.

Bergseng, H., J. E. Afset, A. Radtke, K. Loeseth, R. V. Lyng, M. Rygg & K. Bergh, (2009) Molecular and phenotypic characterization of invasive group B streptococcus strains from infants in Norway 2006-2007. *Clin Microbiol Infect*.

Bidet, P., N. Brahimi, C. Chalas, Y. Aujard & E. Bingen, (2003) Molecular characterization of serotype III group B-streptococcus isolates causing neonatal meningitis. *J Infect Dis* **188**: 1132-1137.

Bingen, E., E. Denamur, N. Lambert-Zechovsky, Y. Aujard, N. Brahimi, P. Geslin & J. Elion, (1992) Analysis of DNA restriction fragment length polymorphism extends the evidence for breast milk transmission in *Streptococcus agalactiae* late-onset neonatal infection. *J Infect Dis* **165**: 569-573.

Bisharat, N., D. W. Crook, J. Leigh, R. M. Harding, P. N. Ward, T. J. Coffey, M. C. Maiden, T. Peto & N. Jones, (2004) Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor. *J Clin Microbiol* **42**: 2161-2167.

Bisharat, N., N. Jones, D. Marchaim, C. Block, R. M. Harding, P. Yagupsky, T. Peto & D. W. Crook, (2005) Population structure of group B streptococcus from a low-incidence region for invasive neonatal disease. *Microbiology* **151**: 1875-1881.

Bishop, E. J., C. Shilton, S. Benedict, F. Kong, G. L. Gilbert, D. Gal, D. Godoy, B. G. Spratt & B. J. Currie, (2007) Necrotizing fasciitis in captive juvenile *Crocodylus porosus* caused by *Streptococcus agalactiae*: an outbreak and review of the animal and human literature. *Epidemiol Infect* **135**: 1248-1255.

Blancas, D., M. Santin, M. Olmo, F. Alcaide, J. Carratala & F. Gudiol, (2004) Group B streptococcal disease in nonpregnant adults: incidence, clinical characteristics, and outcome. *Eur J Clin Microbiol Infect Dis* **23**: 168-173.

Bliss, S. J., S. D. Manning, P. Tallman, C. J. Baker, M. D. Pearlman, C. F. Marrs & B. Foxman, (2002) Group B Streptococcus colonization in male and nonpregnant female university students: a cross-sectional prevalence study. *Clin Infect Dis* **34**: 184-190.

Blumberg, H. M., D. S. Stephens, M. Modansky, M. Erwin, J. Elliot, R. R. Facklam, A. Schuchat, W. Baughman & M. M. Farley, (1996) Invasive group B streptococcal disease: the emergence of serotype V. *J Infect Dis* **173**: 365-373.

Bohnsack, J. F., A. Whiting, M. Gottschalk, D. M. Dunn, R. Weiss, P. H. Azimi, J. B. Philips, 3rd, L. E. Weisman, G. G. Rhoads & F. Y. Lin, (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.

Bohnsack, J. F., A. A. Whiting, R. D. Bradford, B. K. Van Frank, S. Takahashi & E. E. Adderson, (2002) Long-range mapping of the *Streptococcus agalactiae* phylogenetic lineage restriction digest pattern type III-3 reveals clustering of virulence genes. *Infect Immun* **70**: 134-139.

Brimil, N., E. Barthell, U. Heindrichs, M. Kuhn, R. Lutticken & B. Spellerberg, (2006) Epidemiology of *Streptococcus agalactiae* colonization in Germany. *Int J Med Microbiol* **296**: 39-44.

Brochet, M., E. Couve, R. Bercion, J. M. Sire & P. Glaser, (2009a) Population structure of human isolates of *Streptococcus agalactiae* from Dakar and Bangui. *J Clin Microbiol* **47**: 800-803.

Brochet, M., E. Couve, P. Glaser, G. Guedon & S. Payot, (2008a) Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. *J Bacteriol* **190**: 6913-6917.

Brochet, M., E. Couve, M. Zouine, T. Vallaeys, C. Rusniok, M. C. Lamy, C. Buchrieser, P. Trieu-Cuot, F. Kunst, C. Poyart & P. Glaser, (2006) Genomic diversity and evolution within the species *Streptococcus agalactiae*. *Microbes Infect* **8**: 1227-1243.

Brochet, M., V. Da Cunha, E. Couve, C. Rusniok, P. Trieu-Cuot & P. Glaser, (2009b) Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Mol Microbiol* **71**: 948-959.

Brochet, M., C. Rusniok, E. Couve, S. Dramsi, C. Poyart, P. Trieu-Cuot, F. Kunst & P. Glaser, (2008b) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* **105**: 15961-15966.

Broker, G. & B. Spellerberg, (2004) Surface proteins of *Streptococcus agalactiae* and horizontal gene transfer. *Int J Med Microbiol* **294**: 169-175.

Bruen, T. C., H. Philippe & D. Bryant, (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665-2681.

Burrus, V., G. Pavlovic, B. Decaris & G. Guedon, (2002a) Conjugative transposons: the tip of the iceberg. *Mol Microbiol* **46**: 601-610.

Burrus, V., G. Pavlovic, B. Decaris & G. Guedon, (2002b) The ICE*St1* element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* **48**: 77-97.

Caetano-Anolles, G., B. J. Bassam & P. M. Gresshoff, (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology (N Y)* **9**: 553-557.

Campbell, A., D. E. Berg, D. Botstein, E. M. Lederberg, R. P. Novick, P. Starlinger & W. Szybalski, (1979) Nomenclature of transposable elements in prokaryotes. *Gene* **5**: 197-206.

Castor, M. L., C. G. Whitney, K. Como-Sabetti, R. R. Facklam, P. Ferrieri, J. M. Bartkus, B. A. Juni, P. R. Cieslak, M. M. Farley, N. B. Dumas, S. J. Schrag & R. Lynfield, (2008) Antibiotic resistance patterns in invasive group B streptococcal isolates. *Infect Dis Obstet Gynecol* **2008**: 727505.

Chain, P. S., E. Carniel, F. W. Larimer, J. Lamerdin, P. O. Stoutland, W. M. Regala, A. M. Georgescu, L. M. Vergez, M. L. Land, V. L. Motin, R. R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V. Chenal-Francisque, B. Souza, D. Dacheux, J. M. Elliott, A. Derbise, L. J. Hauser & E. Garcia, (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **101**: 13826-13831.

Chatellier, S., H. Huet, S. Kenzi, A. Rosenau, P. Geslin & R. Quentin, (1996) Genetic diversity of rRNA operons of unrelated *Streptococcus agalactiae* strains isolated from cerebrospinal fluid of neonates suffering from meningitis. *J Clin Microbiol* **34**: 2741-2747.

Chatellier, S., C. Ramanantsoa, P. Harriau, K. Rolland, A. Rosenau & R. Quentin, (1997) Characterization of *Streptococcus agalactiae* strains by randomly amplified polymorphic DNA analysis. *J Clin Microbiol* **35**: 2573-2579.

Chen, C. C. & P. P. Cleary, (1990) Complete nucleotide sequence of the streptococcal C5a peptidase gene of *Streptococcus pyogenes*. *J Biol Chem* **265**: 3161-3167.

Cheng, Q., D. Stafslien, S. S. Purushothaman & P. Cleary, (2002) The group B streptococcal C5a peptidase is both a specific protease and an invasin. *Infect Immun* **70**: 2408-2413.

Chhuy, T., G. Mansour, A. Zejli, C. Bouquigny, S. Bock & P. Abboud, (2005) [Group B streptococcus screening: a retrospective study in 1,674 pregnancies]. *J Gynecol Obstet Biol Reprod (Paris)* **34**: 328-333.

Christensen, K. K. & P. Christensen, (1978) Typing of group B streptococci from the throat and urogenital tract of females. *Scand J Infect Dis* **10**: 209-212.

Cieslewicz, M. J., D. Chaffin, G. Glusman, D. Kasper, A. Madan, S. Rodrigues, J. Fahey, M. R. Wessels & C. E. Rubens, (2005) Structural and genetic diversity of group B streptococcus capsular polysaccharides. *Infect Immun* **73**: 3096-3103.

Cousineau, B., D. Smith, S. Lawrence-Cavanagh, J. E. Mueller, J. Yang, D. Mills, D. Manias, G. Dunny, A. M. Lambowitz & M. Belfort, (1998) Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**: 451-462.

Culebras, E., I. Rodriguez-Avial, C. Betriu & J. J. Picazo, (2005) Differences in the DNA sequence of the translational attenuator of several constitutively expressed *erm*(A) genes from clinical isolates of *Streptococcus agalactiae*. *J Antimicrob Chemother* **56**: 836-840.

Dai, L., N. Toor, R. Olson, A. Keeping & S. Zimmerly, (2003) Database for mobile group II introns. *Nucleic Acids Res* **31**: 424-426.

Dai, L. & S. Zimmerly, (2002) Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res* **30**: 1091-1102.

Dai, L. & S. Zimmerly, (2003) ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA* **9**: 14-19.

Daniels, J., J. Gray, H. Pattison, T. Roberts, E. Edwards, P. Milner, L. Spicer, E. King, R. K. Hills, R. Gray, L. Buckley, L. Magill, N. Elliman, B. Kaambwa, S. Bryan, R. Howard, P. Thompson & K. S. Khan, (2009) Rapid testing for group B streptococcus during labour: a test accuracy study with evaluation of acceptability and cost-effectiveness. *Health Technol Assess* **13**: 1-154, iii-iv.

Darling, C. L., (1975) Standardization and evaluation of the CAMP reaction for the prompt, presumptive identification of *Streptococcus agalactiae* (Lancefield group B) in clinical material. *J Clin Microbiol* **1**: 171-174.

Davies, H. D., M. A. Miller, S. Faro, D. Gregson, S. C. Kehl & J. A. Jordan, (2004) Multicenter study of a rapid molecular-based assay for the diagnosis of group B streptococcus colonization in pregnant women. *Clin Infect Dis* **39**: 1129-1135.

Denning, D. W., C. J. Baker, N. J. Troup & L. S. Tompkins, (1989) Restriction endonuclease analysis of human and bovine group B streptococci for epidemiologic study. *J Clin Microbiol* **27**: 1352-1356.

Devriese, L. A., (1991) Streptococcal ecovars associated with different animal species: epidemiological significance of serogroups and biotypes. *J Appl Bacteriol* **71**: 478-483.

Dmitriev, A., A. Shen, X. Shen & Y. Yang, (2004) IS*Sa4*-based differentiation of *Streptococcus agalactiae* strains and identification of multiple target sites for IS*Sa4* insertions. *J Bacteriol* **186**: 1106-1109.

Dobbins, C. N., Jr., (1977) Mastitis losses. *J Am Vet Med Assoc* **170**: 1129-1132.

Dogan, B., Y. H. Schukken, C. Santisteban & K. J. Boor, (2005) Distribution of serotypes and antimicrobial resistance genes among *Streptococcus agalactiae* isolates from bovine and human hosts. *J Clin Microbiol* **43**: 5899-5906.

Domelier, A. S., N. van der Mee-Marquet, P. Y. Sizaret, G. Héry-Arnaud, M. F. Lartigue, L. Mereghetti & R. Quentin, (2009) Molecular characterization and lytic activities of *Streptococcus agalactiae* bacteriophages and determination of lysogenic-strain features. *J Bacteriol* **191**: 4776-4785.

Duarte, R. S., O. P. Miranda, B. C. Bellei, M. A. Brito & L. M. Teixeira, (2004) Phenotypic and molecular characteristics of *Streptococcus agalactiae* isolates recovered from milk of dairy cows in Brazil. *J Clin Microbiol* **42**: 4214-4222.

Easmon, C. S., M. J. Hastings, A. J. Clare, B. Bloxham, R. Marwood, R. P. Rivers & J. Stringer, (1981) Nosocomial transmission of group B streptococci. *Br Med J (Clin Res Ed)* **283**: 459-461.

Edwards, M. S. & C. J. Baker, (2005) Group B streptococcal infections in elderly adults. *Clin Infect Dis* **41**: 839-847.

Eickel, V., B. Kahl, B. Reinisch, A. Dubbers, P. Kuster, C. Brandt & B. Spellerberg, (2009) Emergence of respiratory *Streptococcus agalactiae* isolates in cystic fibrosis patients. *PLoS One* **4**: e4650.

Eickhoff, T. C., J. O. Klein, A. K. Daly, D. Ingall & M. Finland, (1964) Neonatal Sepsis and Other Infections Due to Group B Beta-Hemolytic Streptococci. *N Engl J Med* **271**: 1221-1228.

Ekelund, K. & H. B. Konradsen, (2004) Invasive group B streptococcal disease in infants: a 19-year nationwide study. Serotype distribution, incidence and recurrent infection. *Epidemiol Infect* **132**: 1083-1090.

Ekelund, K., P. Skinhoj, J. Madsen & H. B. Konradsen, (2005) Invasive group A, B, C and G streptococcal infections in Denmark 1999-2002: epidemiological and clinical aspects. *Clin Microbiol Infect* **11**: 569-576.

El Helali, N., J. C. Nguyen, A. Ly, Y. Giovangrandi & L. Trinquart, (2009) Diagnostic accuracy of a rapid real-time polymerase chain reaction assay for universal *intrapartum* group B streptococcus screening. *Clin Infect Dis* **49**: 417-423.

Elliott, J. A., R. R. Facklam & C. B. Richter, (1990) Whole-cell protein patterns of nonhemolytic group B, type Ib, streptococci isolated from humans, mice, cattle, frogs, and fish. *J Clin Microbiol* **28**: 628-630.

Elliott, J. A., K. D. Farmer & 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. *J Clin Microbiol* **36**: 2115-2116.

Elsner, A., B. Kreikemeyer, A. Braun-Kiewnick, B. Spellerberg, B. A. Buttaro & A. Podbielski, (2002) Involvement of Lsp, a member of the Lral-lipoprotein family in *Streptococcus pyogenes*, in eukaryotic cell adhesion and internalization. *Infect Immun* **70**: 4859-4869.

Enright, M. C. & B. G. Spratt, (1999) Multilocus sequence typing. *Trends Microbiol* **7**: 482-487.

Evans, J. J., P. H. Klesius, D. J. Pasnik & J. F. Bohnsack, (2009) Human *Streptococcus agalactiae* isolate in Nile tilapia (*Oreochromis niloticus*). *Emerg Infect Dis* **15**: 774-776.

Evans, J. J., D. J. Pasnik, P. H. Klesius & S. Al-Ablani, (2006) First report of *Streptococcus agalactiae* and *Lactococcus garvieae* from a wild bottlenose dolphin (*Tursiops truncatus*). *J Wildl Dis* **42**: 561-569.

Farley, M. M., (2001) Group B streptococcal disease in nonpregnant adults. *Clin Infect Dis* **33**: 556-561.

Fayet, O., P. Ramond, P. Polard, M. F. Prere & M. Chandler, (1990) Functional similarities between retroviruses and the IS*3* family of bacterial insertion sequences? *Mol Microbiol* **4**: 1771-1777.

Feikin, D. R., P. Thorsen, S. Zywicki, M. Arpi, J. G. Westergaard & A. Schuchat, (2001) Association between colonization with group B streptococci during pregnancy and preterm delivery among Danish women. *Am J Obstet Gynecol* **184**: 427-433.

Feil, E. J., E. C. Holmes, D. E. Bessen, M. S. Chan, N. P. Day, M. C. Enright, R. Goldstein, D. W. Hood, A. Kalia, C. E. Moore, J. Zhou & B. G. Spratt, (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci U S A* **98**: 182-187.

Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage & B. G. Spratt, (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186**: 1518-1530.

Feil, E. J., J. M. Smith, M. C. Enright & B. G. Spratt, (2000) Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* **154**: 1439-1450.

Feil, E. J. & B. G. Spratt, (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55**: 561-590.

Ferat, J. L., M. Le Gouar & F. Michel, (1994) Multiple group II self-splicing introns in mobile DNA from *Escherichia coli*. *C R Acad Sci III* **317**: 141-148.

Ferat, J. L. & F. Michel, (1993) Group II self-splicing introns in bacteria. *Nature* **364**: 358-361.

Ferrieri, P., P. P. Cleary & A. E. Seeds, (1977) Epidemiology of group-B streptococcal carriage in pregnant women and newborn infants. *J Med Microbiol* **10**: 103-114.

Finch, L. A. & D. R. Martin, (1984) Human and bovine group B streptococci: two distinct populations. *J Appl Bacteriol* **57**: 273-278.

Franken, C., C. Brandt, G. Broker & B. Spellerberg, (2004) IS*Sag1* in streptococcal strains of human and animal origin. *Int J Med Microbiol* **294**: 247-254.

Franken, C., G. Haase, C. Brandt, J. Weber-Heynemann, S. Martin, C. Lammler, A. Podbielski, R. Lutticken & B. Spellerberg, (2001) Horizontal gene transfer and host specificity of beta-haemolytic streptococci: the role of a putative composite transposon containing *scpB* and *lmb*. *Mol Microbiol* **41**: 925-935.

Frost, L. S., R. Leplae, A. O. Summers & A. Toussaint, (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**: 722-732.

Gagneur, A., G. Héry-Arnaud, S. Croly-Labourdette, G. Gremmo-Feger, S. Vallet, J. Sizun, R. Quentin & D. Tandé, (2009) Infected breast milk associated with late-onset and recurrent group B streptococcal infection in neonatal twins: a genetic analysis. *Eur J Pediatr* **168**: 1155-1158.

Galask, R. P., M. W. Varner, C. R. Petzold & S. L. Wilbur, (1984) Bacterial attachment to the chorioamniotic membranes. *Am J Obstet Gynecol* **148**: 915-928.

Garland, S. M., N. Kelly & A. M. Ugoni, (2000) Is antenatal group B streptococcal carriage a predictor of adverse obstetric outcome? *Infect Dis Obstet Gynecol* **8**: 138-142.

Gherardi, G., M. Imperi, L. Baldassarri, M. Pataracchia, G. Alfarone, S. Recchia, G. Orefici, G. Dicuonzo & R. Creti, (2007) Molecular epidemiology and distribution of serotypes, surface proteins, and antibiotic resistance among group B streptococci in Italy. *J Clin Microbiol* **45**: 2909-2916.

Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couve, L. Lalioui, C. Poyart, P. Trieu-Cuot & F. Kunst, (2002) Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol Microbiol* **45**: 1499-1513.

Gleich-Theurer, U., S. Aymanns, G. Haas, S. Mauerer, J. Vogt & B. Spellerberg, (2009) Human serum induces streptococcal c5a peptidase expression. *Infect Immun* **77**: 3817-3825.

Gordillo, M. E., K. V. Singh, C. J. Baker & B. E. Murray, (1993) Typing of group B streptococci: comparison of pulsed-field gel electrophoresis and conventional electrophoresis. *J Clin Microbiol* **31**: 1430-1434.

Granlund, M., F. Michel & M. Norgren, (2001) Mutually exclusive distribution of IS*1548* and GBSi1, an active group II intron identified in human isolates of group B streptococci. *J Bacteriol* **183**: 2560-2569.

Granlund, M., L. Oberg, M. Sellin & M. Norgren, (1998) Identification of a novel insertion element, IS*1548*, in group B streptococci, predominantly in strains causing endocarditis. *J Infect Dis* **177**: 967-976.

Hacker, J., L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre & W. Goebel, (1990) Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb Pathog* **8**: 213-225.

Hacker, J., G. Blum-Oehler, I. Muhldorfer & H. Tschape, (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23**: 1089-1097.

Hacker, J. & J. B. Kaper, (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**: 641-679.

Hakansson, S., P. Axemo, K. Bremme, A. L. Bryngelsson, M. C. Wallin, C. M. Ekstrom, M. Granlund, B. Jacobsson, K. Kallen, E. Spetz & I. Tessin, (2008) Group B streptococcal carriage in Sweden: a national study on risk factors for mother and infant colonisation. *Acta Obstet Gynecol Scand* **87**: 50-58.

Hammerschmidt, S., R. Hilse, J. P. van Putten, R. Gerardy-Schahn, A. Unkmeir & M. Frosch, (1996) Modulation of cell surface sialic acid expression in *Neisseria meningitidis* via a transposable genetic element. *EMBO J* **15**: 192-198.

Harrison, L. H., J. A. Elliott, D. M. Dwyer, J. P. Libonati, P. Ferrieri, L. Billmann & A. Schuchat, (1998) Serotype distribution of invasive group B streptococcal isolates in Maryland: implications for vaccine formulation. Maryland Emerging Infections Program. *J Infect Dis* **177**: 998-1002.

Hartl, D. L. & S. A. Sawyer, (1988) Why do unrelated insertion sequences occur together in the genome of *Escherichia coli*? *Genetics* **118**: 537-541.

Haug, R. H., R. Gudding & G. Bakken, (1981) Serotyping and bacteriophage typing of human and bovine group-B streptococci. *J Med Microbiol* **14**: 479-482.

Hauge, M., C. Jespersgaard, K. Poulsen & M. Kilian, (1996) Population structure of *Streptococcus agalactiae* reveals an association between specific evolutionary lineages and putative virulence factors but not disease. *Infect Immun* **64**: 919-925.

Hedegaard, J., M. Hauge, J. Fage-Larsen, K. K. Mortensen, M. Kilian, H. U. Sperling-Petersen & K. Poulsen, (2000) Investigation of the translation-initiation factor IF2 gene, *infB*, as a tool to study the population structure of *Streptococcus agalactiae*. *Microbiology* **146 (Pt 7)**: 1661-1670.

Helmig, R., N. Uldbjerg, J. Boris & M. Kilian, (1993) Clonal analysis of *Streptococcus agalactiae* isolated from infants with neonatal sepsis or meningitis and their mothers and from healthy pregnant women. *J Infect Dis* **168**: 904-909.

Hennig, S. & W. Ziebuhr, (2008) A transposase-independent mechanism gives rise to precise excision of IS*256* from insertion sites in *Staphylococcus epidermidis*. *J Bacteriol* **190**: 1488-1490.

Henning, K. J., E. L. Hall, D. M. Dwyer, L. Billmann, A. Schuchat, J. A. Johnson & L. H. Harrison, (2001) Invasive group B streptococcal disease in Maryland nursing home residents. *J Infect Dis* **183**: 1138-1142.

Herbert, M. A., C. J. Beveridge, D. McCormick, E. Aten, N. Jones, L. A. Snyder & N. J. Saunders, (2005) Genetic islands of *Streptococcus agalactiae* strains NEM316 and 2603VR and their presence in other Group B streptococcal strains. *BMC Microbiol* **5**: 31.

Horaud, T., G. de Cespedes & P. Trieu-Cuot, (1996) Chromosomal gentamicin resistance transposon Tn*3706* in *Streptococcus agalactiae* B128. *Antimicrob Agents Chemother* **40**: 1085-1090.

Hunter, P. R. & M. A. Gaston, (1988) Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity. *J Clin Microbiol* **26**: 2465-2466.

Huson, D. H. & D. Bryant, (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254-267.

Jackson, L. A., R. Hilsdon, M. M. Farley, L. H. Harrison, A. L. Reingold, B. D. Plikaytis, J. D. Wenger & A. Schuchat, (1995) Risk factors for group B streptococcal disease in adults. *Ann Intern Med* **123**: 415-420.

Jain, N. C., (1979) Common mammary pathogens and factors in infection and mastitis. *J Dairy Sci* **62**: 128-134.

Janulczyk, R., J. Pallon & L. Bjorck, (1999) Identification and characterization of a *Streptococcus pyogenes* ABC transporter with multiple specificity for metal cations. *Mol Microbiol* **34**: 596-606.

Jaureguy, F., M. Carton, J. Teboul, M. J. Butel, P. Panel, J. C. Ghnassia & F. Doucet-Populaire, (2003) [Risk factors and screening strategy for group B streptococcal colonization in pregnant women: results of a prospective study]. *J Gynecol Obstet Biol Reprod (Paris)* **32**: 132-138.

Jelinkova, J. & J. Motlova, (1985) Worldwide distribution of two new serotypes of group B streptococci: type IV and provisional type V. *J Clin Microbiol* **21**: 361-362.

Jensen, N. E., (1979) Production and evaluation of antisera for serological type determination of group-B streptococci by double diffusion in agarose gel. *Acta Pathol Microbiol Scand B* **87B**: 77-83.

Jeong, E. L. & J. N. Timmis, (2000) Novel insertion sequence elements associated with genetic heterogeneity and phenotype conversion in *Ralstonia solanacearum*. *J Bacteriol* **182**: 4673-4676.

Jolley, K. A., E. J. Feil, M. S. Chan & M. C. Maiden, (2001) Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**: 1230-1231.

Jones, N., J. F. Bohnsack, S. Takahashi, K. A. Oliver, M. S. Chan, F. Kunst, P. Glaser, C. Rusniok, D. W. Crook, R. M. Harding, N. Bisharat & B. G. Spratt, (2003) Multilocus sequence typing system for group B streptococcus. *J Clin Microbiol* **41**: 2530-2536.

Jones, N., K. Oliver, Y. Jones, A. Haines & D. Crook, (2006a) Carriage of group B streptococcus in pregnant women from Oxford, UK. *J Clin Pathol* **59**: 363-366.

Jones, N., K. A. Oliver, J. Barry, R. M. Harding, N. Bisharat, B. G. Spratt, T. Peto & D. W. Crook, (2006b) Enhanced invasiveness of bovine-derived neonatal sequence type 17 group B streptococcus is independent of capsular serotype. *Clin Infect Dis* **42**: 915-924.

Jordan, H. T., M. M. Farley, A. Craig, J. Mohle-Boetani, L. H. Harrison, S. Petit, R. Lynfield, A. Thomas, S. Zansky, K. Gershman, B. A. Albanese, W. Schaffner & S. J. Schrag, (2008) Revisiting the need for vaccine prevention of late-onset neonatal group B streptococcal disease: a multistate, population-based analysis. *Pediatr Infect Dis J* **27**: 1057-1064.

Jourdan-Da Silva, N., D. Antona, C. Six, S. Georges, V. Goulet, P. Judlin & D. Lévy-Bruhl, (2008) Infections néonatales à streptocoque B en France : données d'incidence de 1997 à 2006 et pratiques de prévention en maternité *Bull Epidemiol Hebd* **14**: 110-113.

Kazazian, H. H., Jr., (2004) Mobile elements: drivers of genome evolution. *Science* **303**: 1626-1632.

Keefe, G. P., I. R. Dohoo & E. Spangler, (1997) Herd prevalence and incidence of *Streptococcus agalactiae* in the dairy industry of Prince Edward Island. *J Dairy Sci* **80**: 464-470.

Kleckner, N., (1981) Transposable elements in prokaryotes. *Annu Rev Genet* **15**: 341-404.

Klein, J. R. & G. M. Dunny, (2002) Bacterial group II introns and their association with mobile genetic elements. *Front Biosci* **7**: d1843-1856.

Kong, F., S. Gowan, D. Martin, G. James & G. L. Gilbert, (2002) Serotype identification of group B streptococci by PCR and sequencing. *J Clin Microbiol* **40**: 216-226.

Koonin, E. V. & Y. I. Wolf, (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**: 6688-6719.

Koskiniemi, S., M. Sellin & M. Norgren, (1998) Identification of two genes, *cpsX* and *cpsY*, with putative regulatory function on capsule expression in group B streptococci. *FEMS Immunol Med Microbiol* **21**: 159-168.

Krause, J. C., P. Ghandil, M. Chrabieh, J. L. Casanova, C. Picard, A. Puel & C. B. Creech, (2009) Very Late-Onset Group B Streptococcus Meningitis, Sepsis, and Systemic Shigellosis due to Interleukin-1 Receptor-Associated Kinase-4 Deficiency. *Clin Infect Dis*.

Krohn, M. A., S. L. Hillier & C. J. Baker, (1999) Maternal peripartum complications associated with vaginal group B streptococci colonization. *J Infect Dis* **179**: 1410-1415.

Kumar, S., M. Nei, J. Dudley & K. Tamura, (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* **9**: 299-306.

Kumar, S., K. Tamura & M. Nei, (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**: 150-163.

Kusumoto, M., Y. Nishiya & Y. Kawamura, (2000) Reactivation of insertionally inactivated Shiga toxin 2 genes of *Escherichia coli* O157:H7 caused by nonreplicative transposition of the insertion sequence. *Appl Environ Microbiol* **66**: 1133-1138.

Lachenauer, C. S., R. Creti, J. L. Michel & L. C. Madoff, (2000) Mosaicism in the alpha-like protein genes of group B streptococci. *Proc Natl Acad Sci U S A* **97**: 9630-9635.

Lachenauer, C. S., D. L. Kasper, J. Shimada, Y. Ichiman, H. Ohtsuka, M. Kaku, L. C. Paoletti, P. Ferrieri & L. C. Madoff, (1999) Serotypes VI and VIII predominate among group B streptococci isolated from pregnant Japanese women. *J Infect Dis* **179**: 1030-1033.

Lambowitz, A. M. & S. Zimmerly, (2004) Mobile group II introns. *Annu Rev Genet* **38**: 1-35.

Lammler, C., A. Abdulmawjood & R. Weiss, (1998) Properties of serological group B streptococci of dog, cat and monkey origin. *Zentralbl Veterinarmed B* **45**: 561-566.

Lancefield, R. C. & E. H. Freimer, (1966) Type-specific polysaccharide antigens of group B streptococci. *J Hyg (Lond)* **64**: 191-203.

Lartigue, M. F., G. Héry-Arnaud, E. Haguenoer, A. S. Domelier, P. O. Schmit, N. van der Mee-Marquet, P. Lanotte, L. Mereghetti, M. Kostrzewa & R. Quentin, (2009) Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* **47**: 2284-2287.

Le Thomas-Bories, I., F. Fitoussi, P. Mariani-Kurkdjian, J. Raymond, N. Brahimi, P. Bidet, V. Lefranc & E. Bingen, (2001) Clonal relationship between U.S. and French serotype V group B streptococcus isolates. *J Clin Microbiol* **39**: 4526-4528.

Lederberg, E. M., (1981) Plasmid reference center registry of transposon (Tn) allocations through July 1981. *Gene* **16**: 59-61.

Lee, N. Y., J. J. Yan, J. J. Wu, H. C. Lee, K. H. Liu & W. C. Ko, (2005) Group B streptococcal soft tissue infections in non-pregnant adults. *Clin Microbiol Infect* **11**: 577-579.

Lefebure, T. & M. J. Stanhope, (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol* **8**: R71.

Lefebvre, N., E. Forestier, M. Mohseni-Zadeh, V. Remy, O. Lesens, C. Kuhnert, V. Poindron, P. Riegel, Y. Piemont, D. Christmann & Y. Hansmann, (2007) [Invasive *Streptococcus agalactiae* infections in non-pregnant adults]. *Med Mal Infect* **37**: 796-801.

Lehmann, K. B. & R. Neumann, (1896) *Atlas and Grund-riss der Bakteriologie und Lehrbuch der speciellen bakteriologischen Diagnostik I and II*: 126.

Lin, F. Y., A. Whiting, E. Adderson, S. Takahashi, D. M. Dunn, R. Weiss, P. H. Azimi, J. B. Philips, 3rd, L. E. Weisman, J. Regan, P. Clark, G. G. Rhoads, C. E. Frasch, J. Troendle, P. Moyer & J. F. Bohnsack, (2006) Phylogenetic lineages of invasive and colonizing strains of serotype III group B Streptococci from neonates: a multicenter prospective study. *J Clin Microbiol* **44**: 1257-1261.

Loulergue, J., C. Couhé, C. Grasmick, P. Laudat & R. Quentin, (2004) Sensibilité aux antibiotiques des souches de streptocoque du groupe B de portage vaginal isolées en France, 2003. *Bull Epidemiol Hebd* **18**: 69-70.

Luan, S. L., M. Granlund & M. Norgren, (2003) An inserted DNA fragment with plasmid features is uniquely associated with the presence of the GBSi1 group II intron in *Streptococcus agalactiae*. *Gene* **312**: 305-312.

Luan, S. L., M. Granlund, M. Sellin, T. Lagergard, B. G. Spratt & M. Norgren, (2005) Multilocus sequence typing of Swedish invasive group B streptococcus isolates indicates a neonatally associated genetic lineage and capsule switching. *J Clin Microbiol* **43**: 3727-3733.

Lyon, B. R., J. W. May & R. A. Skurray, (1984) Tn*4001*: a gentamicin and kanamycin resistance transposon in *Staphylococcus aureus*. *Mol Gen Genet* **193**: 554-556.

Lyras, D., V. Adams, S. A. Ballard, W. L. Teng, P. M. Howarth, P. K. Crellin, T. L. Bannam, J. G. Songer & J. I. Rood, (2009) tIS*Cpe8*, an IS*1595*-family lincomycin resistance element located on a conjugative plasmid in *Clostridium perfringens*. *J Bacteriol*.

Mahillon, J. & M. Chandler, (1998) Insertion sequences. *Microbiol Mol Biol Rev* **62**: 725-774.

Mahillon, J., C. Leonard & M. Chandler, (1999) IS elements as constituents of bacterial genomes. *Res Microbiol* **150**: 675-687.

Maiden, M. C., (2006) Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**: 561-588.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman & B. G. Spratt, (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**: 3140-3145.

Maisey, H. C., K. S. Doran & V. Nizet, (2008) Recent advances in understanding the molecular basis of group B Streptococcus virulence. *Expert Rev Mol Med* **10**: e27.

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

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

Marchaim, D., S. Efrati, R. Melamed, L. Gortzak-Uzan, K. Riesenberg, R. Zaidenstein & F. Schlaeffer, (2006) Clonal variability of group B Streptococcus among different groups of carriers in southern Israel. *Eur J Clin Microbiol Infect Dis* **25**: 443-448.

Martinez, G., J. Harel, R. Higgins, S. Lacouture, D. Daignault & M. Gottschalk, (2000) Characterization of *Streptococcus agalactiae* isolates of bovine and human origin by randomly amplified polymorphic DNA analysis. *J Clin Microbiol* **38**: 71-78.

Martinez-Abarca, F. & N. Toro, (2000) Group II introns in the bacterial world. *Mol Microbiol* **38**: 917-926.

Martins, E. R., M. A. Pessanha, M. Ramirez & J. Melo-Cristino, (2007) Analysis of group B streptococcal isolates from infants and pregnant women in Portugal revealing two lineages with enhanced invasiveness. *J Clin Microbiol* **45**: 3224-3229.

Matsubara, K., H. Mikamo, M. Numa, G. Yamamoto, H. Kusano & Y. Takamine, (2006) Three fatal cases of invasive serotype VI group B streptococcal infection. *J Infect* **53**: e139-142.

Mc Clintock, B., (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* **36**: 344-355.

Mereghetti, L., I. Sitkiewicz, N. M. Green & J. M. Musser, (2008) Extensive adaptive changes occur in the transcriptome of *Streptococcus agalactiae* (group B streptococcus) in response to incubation with human blood. *PLoS One* **3**: e3143.

Michaelis, G., H. Saedler, P. Venkov & P. Starlinger, (1969) Two insertions in the galactose operon having different sizes but homologous DNA sequences. *Mol Gen Genet* **104**: 371-377.

Michel, F. & J. L. Ferat, (1995) Structure and activities of group II introns. *Annu Rev Biochem* **64**: 435-461.

Mizuuchi, K., (1992) Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu Rev Biochem* **61**: 1011-1051.

Mosabi, J. M., S. M. Arimi & E. K. Kang'ethe, (1997) Isolation and characterization of group B streptococci from human and bovine sources within and around Nairobi. *Epidemiol Infect* **118**: 215-220.

Motlova, J., M. Wagner & J. Jelinkova, (1986) A search for new group-B streptococcal serotypes. *J Med Microbiol* **22**: 101-105.

Moyo, S. R., J. A. Maeland & K. Bergh, (2002) Typing of human isolates of *Streptococcus agalactiae* (group B streptococcus, GBS) strains from Zimbabwe. *J Med Microbiol* **51**: 595-600.

Murayama, S. Y., C. Seki, H. Sakata, K. Sunaoshi, E. Nakayama, S. Iwata, K. Sunakawa & K. Ubukata, (2009) Capsular type and antibiotic resistance in *Streptococcus agalactiae* isolates from patients, ranging from newborns to the elderly, with invasive infections. *Antimicrob Agents Chemother* **53**: 2650-2653.

Musser, J. M., S. J. Mattingly, R. Quentin, A. Goudeau & R. K. Selander, (1989) Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B streptococcus) causing invasive neonatal disease. *Proc Natl Acad Sci USA* **86**: 4731-4735.

Nagano, N., Y. Nagano, R. Nakano, R. Okamoto & M. Inoue, (2006) Genetic diversity of the C protein beta-antigen gene and its upstream regions within clonally related groups of type Ia and Ib group B streptococci. *Microbiology* **152**: 771-778.

Nagano, Y., N. Nagano, S. Takahashi, K. Murono, K. Fujita, F. Taguchi & Y. Okuwaki, (1991) Restriction endonuclease digest patterns of chromosomal DNA from group B beta-haemolytic streptococci. *J Med Microbiol* **35**: 297-303.

Nocard, E. & R. Mollereau, (1887) Sur une mammite contagieuse des vaches laitières. *Ann Inst Pasteur* **I**: 109-126.

Nolla, J. M., C. Gomez-Vaquero, X. Corbella, S. Ordonez, C. Garcia-Gomez, A. Perez, J. Cabo, J. Valverde & J. Ariza, (2003) Group B streptococcus (*Streptococcus agalactiae*) pyogenic arthritis in nonpregnant adults. *Medicine (Baltimore)* **82**: 119-128.

Norcross, N. L. & N. Oliver, (1976) The distribution and characterization of group B streptococci in New York State. *Cornell Vet* **66**: 242-248.

Novak, R., A. Cauwels, E. Charpentier & E. Tuomanen, (1999) Identification of a *Streptococcus pneumoniae* gene locus encoding proteins of an ABC phosphate transporter and a two-component regulatory system. *J Bacteriol* **181**: 1126-1133.

Ochman, H. & R. K. Selander, (1984) Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* **157**: 690-693.

Oliveira, I. C., M. C. de Mattos, T. A. Pinto, B. T. Ferreira-Carvalho, L. C. Benchetrit, A. A. Whiting, J. F. Bohnsack & A. M. Figueiredo, (2006) Genetic relatedness between group B streptococci originating from bovine mastitis and a human group B Streptococcus type V cluster displaying an identical pulsed-field gel electrophoresis pattern. *Clin Microbiol Infect* **12**: 887-893.

Orle, K. A. & N. L. Craig, (1991) Identification of transposition proteins encoded by the bacterial transposon Tn*7*. *Gene* **104**: 125-131.

Osborn, A. M. & D. Boltner, (2002) When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* **48**: 202-212.

Panina, E. M., A. A. Mironov & M. S. Gelfand, (2003) Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proc Natl Acad Sci U S A* **100**: 9912-9917.

Papp-Wallace, K. M., M. Nartea, D. G. Kehres, S. Porwollik, M. McClelland, S. J. Libby, F. C. Fang & M. E. Maguire, (2008) The CorA Mg2+ channel is required for the virulence of *Salmonella enterica* serovar typhimurium. *J Bacteriol* **190**: 6517-6523.

Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell & D. J. Maskell, (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* **35**: 32-40.

Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead & B. G. Barrell, (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**: 523-527.

Pavlovic, G., V. Burrus, B. Gintz, B. Decaris & G. Guedon, (2004) Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICE*St1*-related elements from *Streptococcus thermophilus*. *Microbiology* **150**: 759-774.

Phares, C. R., R. Lynfield, M. M. Farley, J. Mohle-Boetani, L. H. Harrison, S. Petit, A. S. Craig, W. Schaffner, S. M. Zansky, K. Gershman, K. R. Stefonek, B. A. Albanese, E. R. Zell, A. Schuchat & S. J. Schrag, (2008) Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005. *JAMA* **299**: 2056-2065.

Picard, B., J. S. Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion & E. Denamur, (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* **67**: 546-553.

Posada, D., (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucleic Acids Res* **34**: W700-703.

Posada, D. & K. A. Crandall, (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Postma, P. W., J. W. Lengeler & G. R. Jacobson, (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* **57**: 543-594.

Poyart, C., H. Reglier-Poupet, A. Tazi, A. Billoet, N. Dmytruk, P. Bidet, E. Bingen, J. Raymond & P. Trieu-Cuot, (2008) Invasive group B streptococcal infections in infants, France. *Emerg Infect Dis* **14**: 1647-1649.

Poyart, C., A. Tazi, H. Reglier-Poupet, A. Billoet, N. Tavares, J. Raymond & P. Trieu-Cuot, (2007) Multiplex PCR assay for rapid and accurate capsular typing of group B streptococci. *J Clin Microbiol* **45**: 1985-1988.

Puopolo, K. M., D. C. Klinzing, M. P. Lin, D. L. Yesucevitz & M. J. Cieslewicz, (2007) A composite transposon associated with erythromycin and clindamycin resistance in group B Streptococcus. *J Med Microbiol* **56**: 947-955.

Puopolo, K. M., L. C. Madoff & E. C. Eichenwald, (2005) Early-onset group B streptococcal disease in the era of maternal screening. *Pediatrics* **115**: 1240-1246.

Quentin, R., H. Huet, F. S. Wang, P. Geslin, A. Goudeau & R. 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. *J Clin Microbiol* **33**: 2576-2581.

Quentin, R., V. Morange-Saussier & S. Watt, (2002) [Obstetrical management of *Streptococcus agalactiae*]. *J Gynecol Obstet Biol Reprod (Paris)* **31**: 4S65-64S73.

Ramaswamy, S. V., P. Ferrieri, A. E. Flores & L. C. Paoletti, (2006) Molecular characterization of nontypeable group B streptococcus. *J Clin Microbiol* **44**: 2398-2403.

Regan, J. A., M. A. Klebanoff & R. P. Nugent, (1991) The epidemiology of group B streptococcal colonization in pregnancy. Vaginal Infections and Prematurity Study Group. *Obstet Gynecol* **77**: 604-610.

Regan, J. A., M. A. Klebanoff, R. P. Nugent, D. A. Eschenbach, W. C. Blackwelder, Y. Lou, R. S. Gibbs, P. J. Rettig, D. H. Martin & R. Edelman, (1996) Colonization with group B streptococci in pregnancy and adverse outcome. VIP Study Group. *Am J Obstet Gynecol* **174**: 1354-1360.

Reimmann, C. & D. Haas, (1990) The *istA* gene of insertion sequence IS*21* is essential for cleavage at the inner 3' ends of tandemly repeated IS*21* elements in vitro. *EMBO J* **9**: 4055-4063.

Roberts, A. P., M. Chandler, P. Courvalin, G. Guedon, P. Mullany, T. Pembroke, J. I. Rood, C. J. Smith, A. O. Summers, M. Tsuda & D. E. Berg, (2008) Revised nomenclature for transposable genetic elements. *Plasmid* **60**: 167-173.

Robinson, D. A., S. K. Hollingshead, J. M. Musser, A. J. Parkinson, D. E. Briles & M. J. Crain, (1998) The IS*1167* insertion sequence is a phylogenetically informative marker among isolates of serotype 6B *Streptococcus pneumoniae*. *J Mol Evol* **47**: 222-229.

Rolland, K., C. Marois, V. Siquier, B. Cattier & R. Quentin, (1999) Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *hylB* gene analysis. *J Clin Microbiol* **37**: 1892-1898.

Rolland, K., L. Mereghetti, S. Watt, S. Chatellier & R. 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**: 1493-1499.

Rosenau, A., K. Martins, S. Amor, F. Gannier, P. Lanotte, N. van der Mee-Marquet, L. Mereghetti & 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 *fbs*A and *fbs*B genes. *Infect Immun* **75**: 1310-1317.

Rosini, R., C. D. Rinaudo, M. Soriani, P. Lauer, M. Mora, D. Maione, A. Taddei, I. Santi, C. Ghezzo, C. Brettoni, S. Buccato, I. Margarit, G. Grandi & J. L. Telford, (2006) Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. *Mol Microbiol* **61**: 126-141.

Rubens, C. E., L. M. Heggen, R. F. Haft & M. R. Wessels, (1993) Identification of *cpsD*, a gene essential for type III capsule expression in group B streptococci. *Mol Microbiol* **8**: 843-855.

Rubens, C. E., L. M. Heggen & J. M. Kuypers, (1989) IS*861*, a group B streptococcal insertion sequence related to IS*150* and IS*3* of *Escherichia coli*. *J Bacteriol* **171**: 5531-5535.

Rubens, C. E., M. R. Wessels, L. M. Heggen & D. L. Kasper, (1987) Transposon mutagenesis of type III group B Streptococcus: correlation of capsule expression with virulence. *Proc Natl Acad Sci U S A* **84**: 7208-7212.

Ryc, M., J. Jelinkova, J. Motlova & M. Wagner, (1988) Immuno-electronmicroscopic demonstration of capsules on group-B streptococci of new serotypes and type candidates. *J Med Microbiol* **25**: 147-149.

Salyers, A. A. & N. B. Shoemaker, (1995) Conjugative transposons: the force behind the spread of antibiotic resistance genes among *Bacteroides* clinical isolates. *Anaerobe* **1**: 143-150.

Salyers, A. A., N. B. Shoemaker & L. Y. Li, (1995a) In the driver's seat: the *Bacteroides* conjugative transposons and the elements they mobilize. *J Bacteriol* **177**: 5727-5731.

Salyers, A. A., N. B. Shoemaker, A. M. Stevens & L. Y. Li, (1995b) Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* **59**: 579-590.

Sambola, A., J. M. Miro, M. P. Tornos, B. Almirante, A. Moreno-Torrico, M. Gurgui, E. Martinez, A. Del Rio, M. Azqueta, F. Marco & J. M. Gatell, (2002) *Streptococcus agalactiae* infective endocarditis: analysis of 30 cases and review of the literature, 1962-1998. *Clin Infect Dis* **34**: 1576-1584.

Sanchez-Beato, A. R., E. Garcia, R. Lopez & J. L. Garcia, (1997) Identification and characterization of IS*1381*, a new insertion sequence in *Streptococcus pneumoniae*. *J Bacteriol* **179**: 2459-2463.

Schenkman, D. I., R. J. Rahija, K. L. Klingenberger, J. A. Elliott & C. B. Richter, (1994) Outbreak of group B streptococcal meningoencephalitis in athymic mice. *Lab Anim Sci* **44**: 639-641.

Schrag, S., R. Gorwitz, K. Fultz-Butts & A. Schuchat, (2002) Prevention of perinatal group B streptococcal disease. Revised guidelines from CDC. *MMWR Recomm Rep* **51**: 1-22.

Schrag, S. J., S. Zywicki, M. M. Farley, A. L. Reingold, L. H. Harrison, L. B. Lefkowitz, J. L. Hadler, R. Danila, P. R. Cieslak & A. Schuchat, (2000) Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis. *N Engl J Med* **342**: 15-20.

Schuchat, A., (1995) Group B streptococcal disease in newborns: a global perspective on prevention. *Biomed Pharmacother* **49**: 19-25.

Schuchat, A., (1998) Epidemiology of group B streptococcal disease in the United States: shifting paradigms. *Clin Microbiol Rev* **11**: 497-513.

Schuchat, A., (1999) Group B streptococcus. *Lancet* **353**: 51-56.

Schwartz, B., A. Schuchat, M. J. Oxtoby, S. L. Cochi, A. Hightower & C. V. Broome, (1991) Invasive group B streptococcal disease in adults. A population-based study in metropolitan Atlanta. *JAMA* **266**: 1112-1114.

Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour & T. S. Whittam, (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* **51**: 873-884.

Sellin, M., S. Hakansson & M. Norgren, (1995) Phase-shift of polysaccharide capsule expression in group B streptococci, type III. *Microb Pathog* **18**: 401-415.

Sellin, M., C. Olofsson, S. Hakansson & M. Norgren, (2000) Genotyping of the capsule gene cluster (*cps*) in nontypeable group B streptococci reveals two major *cps* allelic variants of serotypes III and VII. *J Clin Microbiol* **38**: 3420-3428.

Shapiro, J. A., (1969) Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *J Mol Biol* **40**: 93-105.

Shet, A. & P. Ferrieri, (2004) Neonatal & maternal group B streptococcal infections: a comprehensive review. *Indian J Med Res* **120**: 141-150.

Siefert, J. L., (2009) Defining the mobilome. *Methods Mol Biol* **532**: 13-27.

Siguier, P., L. Gagnevin & M. Chandler, (2009) The new IS*1595* family, its relation to IS*1* and the frontier between insertion sequences and transposons. *Res Microbiol* **160**: 232-241.

Siguier, P., J. Perochon, L. Lestrade, J. Mahillon & M. Chandler, (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32-36.

Simon, D. M., N. A. Clarke, B. A. McNeil, I. Johnson, D. Pantuso, L. Dai, D. Chai & S. Zimmerly, (2008) Group II introns in *eubacteria* and *archaea*: ORF-less introns and new varieties. *RNA* **14**: 1704-1713.

Sitkiewicz, I., N. M. Green, N. Guo, A. M. Bongiovanni, S. S. Witkin & J. M. Musser, (2009) Transcriptome adaptation of group B Streptococcus to growth in human amniotic fluid. *PLoS One* **4**: e6114.

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

Slotved, H. C., F. Kong, L. Lambertsen, S. Sauer & G. L. Gilbert, (2007) Serotype IX, a Proposed New *Streptococcus agalactiae* Serotype. *J Clin Microbiol* **45**: 2929-2936.

Smith, J. M., N. H. Smith, M. O'Rourke & B. G. Spratt, (1993) How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**: 4384-4388.

Soto, C. Y., M. C. Menendez, E. Perez, S. Samper, A. B. Gomez, M. J. Garcia & C. Martin, (2004) IS*6110* mediates increased transcription of the *phoP* virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. *J Clin Microbiol* **42**: 212-219.

Southern, E. M., (1974) An improved method for transferring nucleotides from electrophoresis strips to thin layers of ion-exchange cellulose. *Anal Biochem* **62**: 317-318.

Spellerberg, B., S. Martin, C. Franken, R. Berner & R. Lutticken, (2000) Identification of a novel insertion sequence element in *Streptococcus agalactiae Gene* **241**: 51-56.

Spellerberg, B., B. Pohl, G. Haase, S. Martin, J. Weber-Heynemann & R. Lutticken, (1999a) Identification of genetic determinants for the hemolytic activity of *Streptococcus agalactiae* by IS*1* transposition. *J Bacteriol* **181**: 3212-3219.

Spellerberg, B., E. Rozdzinski, S. Martin, J. Weber-Heynemann, N. Schnitzler, R. Lutticken & A. Podbielski, (1999b) Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human laminin. *Infect Immun* **67**: 871-878.

Spratt, B. G., W. P. Hanage, B. Li, D. M. Aanensen & E. J. Feil, (2004) Displaying the relatedness among isolates of bacterial species -- the eBURST approach. *FEMS Microbiol Lett* **241**: 129-134.

Springman, A. C., D. W. Lacher, G. Wu, N. Milton, T. S. Whittam, H. D. Davies & S. D. Manning, (2009) Selection, recombination, and virulence gene diversity among group B streptococcal genotypes. *J Bacteriol* **191**: 5419-5427.

Stoll, B. J. & A. Schuchat, (1998) Maternal carriage of group B streptococci in developing countries. *Pediatr Infect Dis J* **17**: 499-503.

Stringer, J. & W. R. Maxted, (1979) Phage typing of group B streptococci. *Lancet* **1**: 328.

Suara, R. O., R. A. Adegbola, C. J. Baker, O. Secka, E. K. Mulholland & B. M. Greenwood, (1994) Carriage of group B Streptococci in pregnant Gambian mothers and their infants. *J Infect Dis* **170**: 1316-1319.

Sukhnanand, S., B. Dogan, M. O. Ayodele, R. N. Zadoks, M. P. Craver, N. B. Dumas, Y. H. Schukken, K. J. Boor & M. Wiedmann, (2005) Molecular subtyping and characterization of bovine and human *Streptococcus agalactiae* isolates. *J Clin Microbiol* **43**: 1177-1186.

Sun, Y., F. Kong, Z. Zhao & G. L. Gilbert, (2005) Comparison of a 3-set genotyping system with multilocus sequence typing for *Streptococcus agalactiae* (Group B Streptococcus). *J Clin Microbiol* **43**: 4704-4707.

Swofford, D. L., (2003) PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, Mass.

Takahashi, S., E. E. Adderson, Y. Nagano, N. Nagano, M. R. Briesacher & J. F. Bohnsack, (1998) Identification of a highly encapsulated, genetically related group of invasive type III group B streptococci. *J Infect Dis* **177**: 1116-1119.

Takahashi, S., S. Detrick, A. A. Whiting, A. J. Blaschke-Bonkowksy, Y. Aoyagi, E. E. Adderson & J. 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. *J Infect Dis* **186**: 1034-1038.

Tamura, G. S., M. Herndon, J. Przekwas, C. E. Rubens, P. Ferrieri & S. L. Hillier, (2000) Analysis of restriction fragment length polymorphisms of the insertion sequence IS*1381* in group B Streptococci. *J Infect Dis* **181**: 364-368.

Taylor, A. L., (1963) Bacteriophage-Induced Mutation in *Escherichia coli*. *Proc Natl Acad Sci U S A* **50**: 1043-1051.

Tenenbaum, T., B. Spellerberg, R. Adam, M. Vogel, K. S. Kim & H. Schroten, (2007) *Streptococcus agalactiae* invasion of human brain microvascular endothelial cells is promoted by the laminin-binding protein Lmb. *Microbes Infect* **9**: 714-720.

Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. Deboy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli & C. M. Fraser, (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci USA* **102**: 13950-13955.

Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F.

Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi & C. M. Fraser, (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci USA* **99**: 12391-12396.

Thong, K. L., G. Y. Ling, L. W. Kong, L. C. Theam & Y. F. Ngeow, (2004) Macrorestriction analysis of *Streptococcus agalactiae* (group B Streptococcus) isolates from Malaysia. *J Med Microbiol* **53**: 991-997.

Tompkins, L. S., (1992) The use of molecular methods in infectious diseases. *N Engl J Med* **327**: 1290-1297.

Toor, N., G. Hausner & S. Zimmerly, (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**: 1142-1152.

Toussaint, A. & C. Merlin, (2002) Mobile elements as a combination of functional modules. *Plasmid* **47**: 26-35.

Trivalle, C., E. Martin, P. Martel, B. Jacque, J. F. Menard & J. F. Lemeland, (1998) Group B streptococcal bacteraemia in the elderly. *J Med Microbiol* **47**: 649-652.

Turner, K. M. & E. J. Feil, (2007) The secret life of the multilocus sequence type. *Int J Antimicrob Agents* **29**: 129-135.

Turner, K. M., W. P. Hanage, C. Fraser, T. R. Connor & B. G. Spratt, (2007) Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC Microbiol* **7**: 30.

Valkenburg-van den Berg, A. W., A. J. Sprij, F. W. Dekker, P. J. Dorr & H. H. Kanhai, (2009) Association between colonization with Group B Streptococcus and preterm delivery: a systematic review. *Acta Obstet Gynecol Scand* **88**: 958-967.

van der Mee-Marquet, N., A. S. Domelier, L. Mereghetti, P. Lanotte, A. Rosenau, W. van Leeuwen & R. Quentin, (2006) Prophagic DNA fragments in *Streptococcus agalactiae* strains and association with neonatal meningitis. *J Clin Microbiol* **44**: 1049-1058.

van der Mee-Marquet, N., L. Fourny, L. Arnault, A. S. Domelier, M. Salloum, M. F. Lartigue & R. Quentin, (2008) Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites. *J Clin Microbiol* **46**: 2906-2911.

van der Mee-Marquet, N., C. Jouannet, A. S. Domelier, L. Arnault, M. F. Lartigue & R. Quentin, (2009) Genetic diversity of *Streptococcus agalactiae* strains and density of vaginal carriage. *J Med Microbiol* **58**: 169-173.

Van Dyke, M. K., C. R. Phares, R. Lynfield, A. R. Thomas, K. E. Arnold, A. S. Craig, J. Mohle-Boetani, K. Gershman, W. Schaffner, S. Petit, S. M. Zansky, C. A. Morin, N. L. Spina, K. Wymore, L. H. Harrison, K. A. Shutt, J. Bareta, S. N. Bulens, E. R. Zell, A. Schuchat & S. J. Schrag, (2009) Evaluation of universal antenatal screening for group B streptococcus. *N Engl J Med* **360**: 2626-2636.

Varon, E., (2009) [Epidemiology of acute bacterial meningitis in adult patients in France]. *Med Mal Infect* **39**: 432-444.

Vokes, S. A., S. A. Reeves, A. G. Torres & S. M. Payne, (1999) The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. *Mol Microbiol* **33**: 63-73.

von Both, U., A. John, K. Fluegge, A. Siedler & R. Berner, (2008) Molecular epidemiology of invasive neonatal *Streptococcus agalactiae* isolates in Germany. *Pediatr Infect Dis J* **27**: 903-906.

Ward, P. N., M. T. Holden, J. A. Leigh, N. Lennard, A. Bignell, A. Barron, L. Clark, M. A. Quail, J. Woodward, B. G. Barrell, S. A. Egan, T. R. Field, D. Maskell, M. Kehoe, C. G. Dowson, N. Chanter, A. M. Whatmore, S. D. Bentley & J. Parkhill, (2009) Evidence for niche adaptation in the genome of the bovine pathogen *Streptococcus uberis*. *BMC Genomics* **10**: 54.

Watanabe, T. & T. Fukasawa, (1962) Episome-mediated transfer of drug resistance in *Enterobacteriaceae* IV. Interactions between resistance transfer factor and F-factor in *Escherichia coli* K-12. *J Bacteriol* **83**: 727-735.

Wibawan, I. W., C. Lammler & J. Smola, (1993) Properties and type antigen patterns of group B streptococcal isolates from pigs and nutrias. *J Clin Microbiol* **31**: 762-764.

Wilkinson, H. W., (1978) Analysis of group B streptococcal types associated with disease in human infants and adults. *J Clin Microbiol* **7**: 176-179.

Wilson, J. M. & Y. G. Jungner, (1968) [Principles and practice of mass screening for disease]. *Bol Oficina Sanit Panam* **65**: 281-393.

Winn, H. N., (2007) Group B streptococcus infection in pregnancy. *Clin Perinatol* **34**: 387-392.

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

Yow, M. D., L. J. Leeds, P. K. Thompson, E. O. Mason, Jr., D. J. Clark & C. W. Beachler, (1980) The natural history of group B streptococcal colonization in the pregnant woman and her offspring. I. Colonization studies. *Am J Obstet Gynecol* **137**: 34-38.

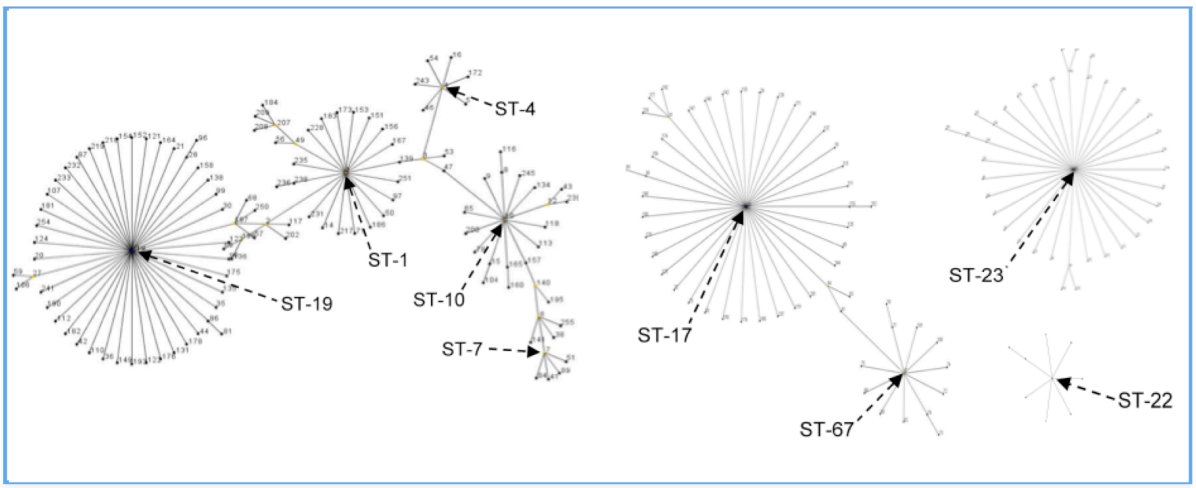
Zhong, S., A. Khodursky, D. E. Dykhuizen & A. M. Dean, (2004) Evolutionary genomics of ecological specialization. *Proc Natl Acad Sci U S A* **101**: 11719-11724.

Zimmerly, S., G. Hausner & X. Wu, (2001) Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res* **29**: 1238-1250.

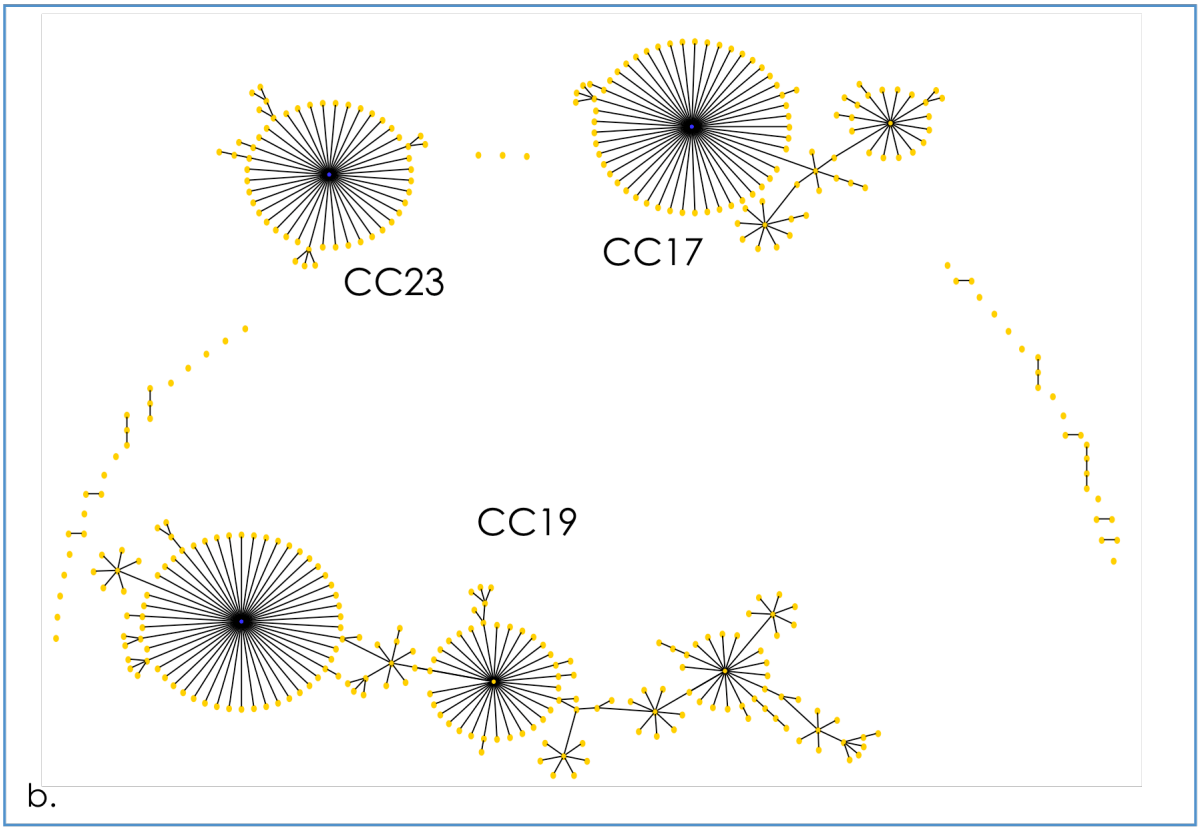
Annexes

Annexe A. Représentation graphique de la totalité de la base de données multilocus sequence typing (MLST) de *S. agalactiae* obtenue avec l'algorithme eBURST v4 qui détermine les lignées clonales appelées complexes clonaux (CC) en MLST. Pour chaque CC, le sequence type (ST) fondateur est indiqué.

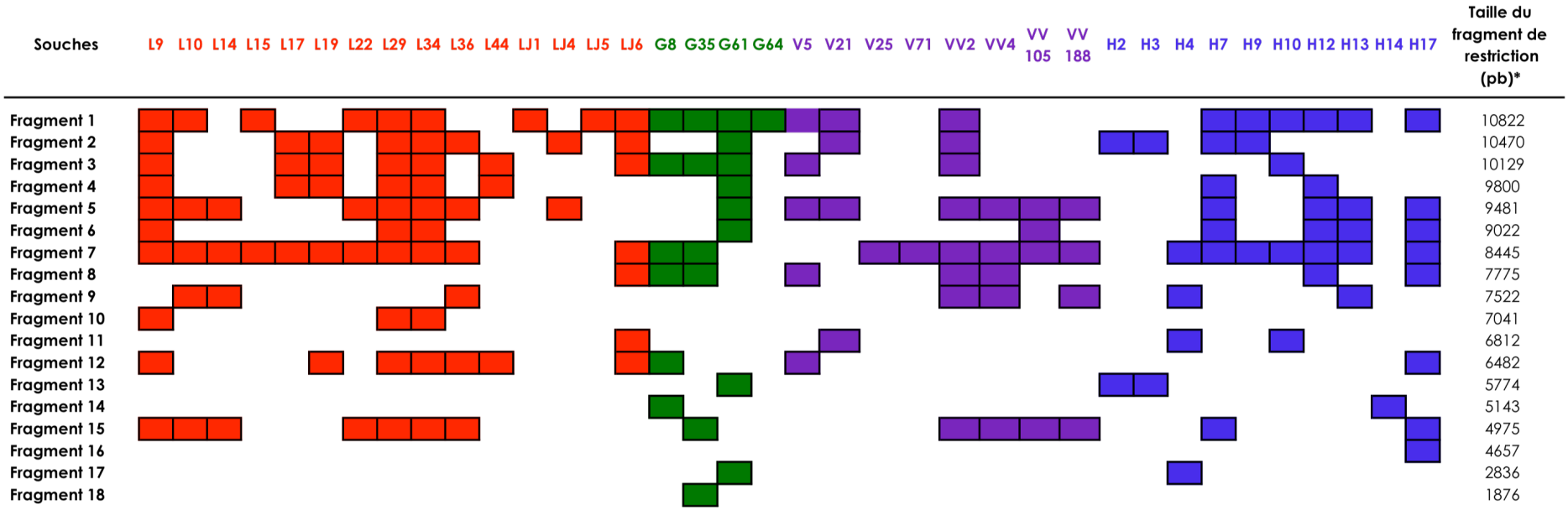
a. En février 2006, les 259 ST référencés se répartissent sur neuf CC : CC19, CC10, CC1, CC4, CC17, CC67, CC23 et CC22.



b. En août 2009, les 449 ST référencés se répartissent dans trois grands groupes principaux centrés autour de CC19, CC17 et CC23.

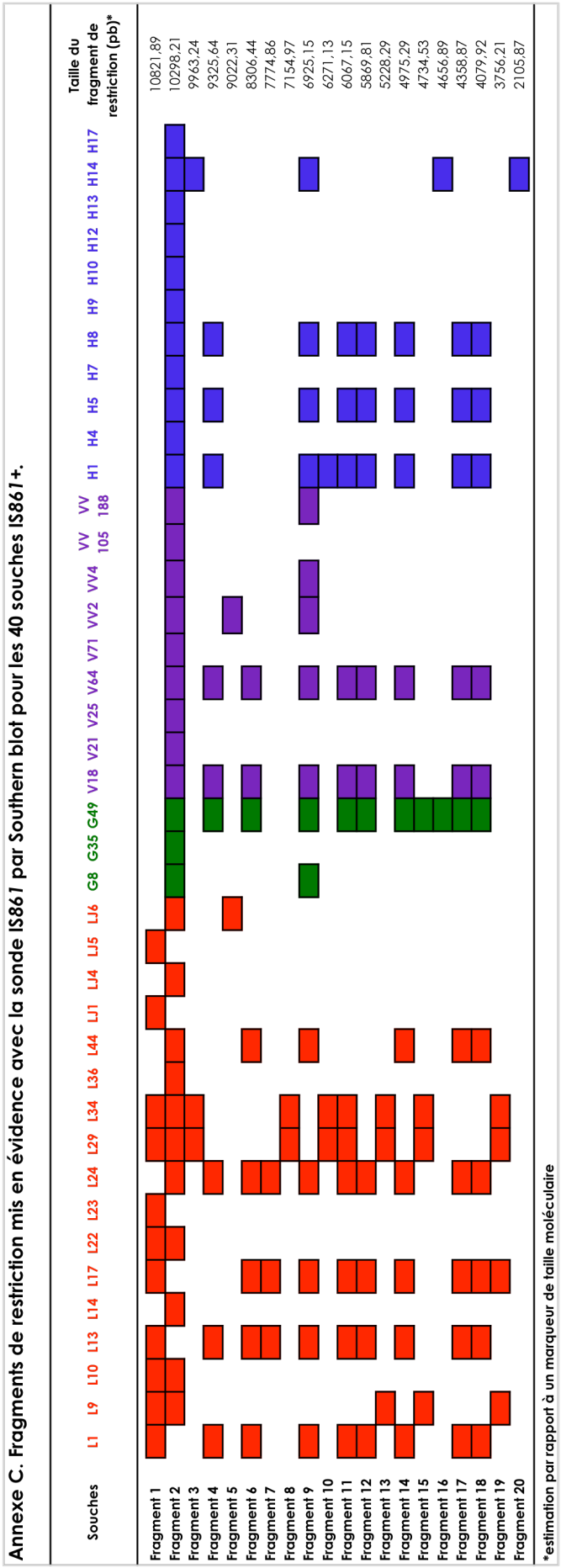


Annexe B. Fragments de restriction mis en évidence avec la sonde IS*1381* par Southern blot pour les 37 souches IS*1381*+.

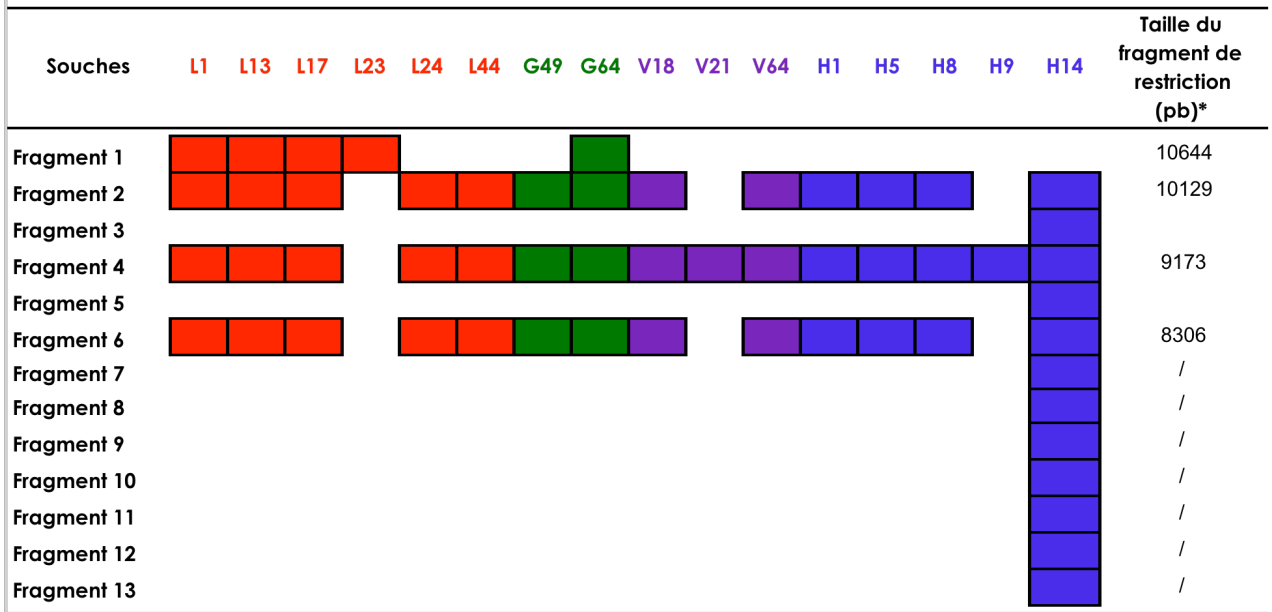


*estimation par rapport à un marqueur de taille moléculaire

Annexe C. Fragments de restriction mis en évidence avec la sonde IS861 par Southern blot pour les 40 souches IS861+.

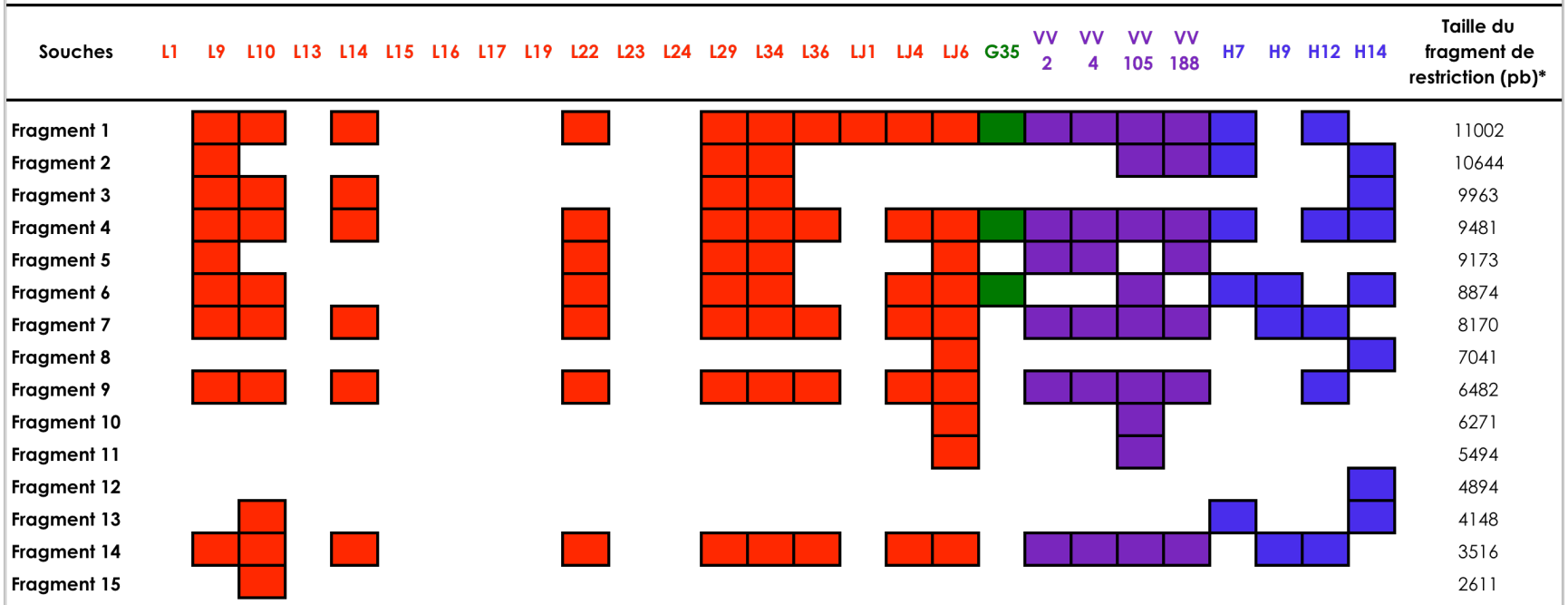


Annexe D. Fragments de restriction mis en évidence avec la sonde GBSi1 par Southern blot pour les 16 souches GBSi1+.



*estimation par rapport à un marqueur de taille moléculaire

Annexe E. Fragments de restriction mis en évidence avec la sonde IS*1548* par Southern blot pour les 19 souches IS*1548*+.



*estimation par rapport à un marqueur de taille moléculaire

Annexe F. Fragments de restriction mis en évidence avec la sonde IS*Sa4* par Southern blot pour les 4 souches IS*Sa4*+.

Souches	G64	H6	H15	H16	Taille du fragment de restriction (pb)*
Fragment 1					/
Fragment 2					/
Fragment 3					9326
Fragment 4					/
Fragment 5					/
Fragment 6					/
Fragment 7					/
Fragment 8					/
Fragment 9					7158
Fragment 10					/
Fragment 11					/
Fragment 12					/
Fragment 13					/
Fragment 14					/
Fragment 15					5143
Fragment 16					/
Fragment 17					4505
Fragment 18					4287
Fragment 19					3947
Fragment 20					/

*estimation par rapport à un marqueur de taille moléculaire

Publications et communications scientifiques

-ARTICLE 1-

Acquisition of Insertion Sequences and the GBSi1 Intron by *Streptococcus agalactiae* Isolates Correlates with the Evolution of the Species

Geneviève Héry-Arnaud,[1] Guillaume Bruant,[1†] Philippe Lanotte,[1] Stella Brun,[1] Agnès Rosenau,[1] Nathalie van der Mee-Marquet,[1,2] Roland Quentin,[1,2] and Laurent Mereghetti[1*]

EA 3854 "Bactéries et risque materno-fœtal",[1] Département de Microbiologie Médicale et Moléculaire, Université François Rabelais,1 and Centre National de Référence des Streptocoques—Laboratoire associé, Service de Bactériologie et Hygiène Hospitalière, Centre Hospitalier Universitaire,[2] Tours, France

*Corresponding author

JOURNAL OF BACTERIOLOGY, SEPT. 2005, P. 6248–6252 VOL. 187, NO. 17

JOURNAL OF BACTERIOLOGY, Sept. 2005, p. 6248–6252 Vol. 187, No. 17
0021-9193/05/$08.00+0 doi:10.1128/JB.187.17.6248–6252.2005
Copyright © 2005, American Society for Microbiology. All Rights Reserved.

Acquisition of Insertion Sequences and the GBSi1 Intron by *Streptococcus agalactiae* Isolates Correlates with the Evolution of the Species

Geneviève Héry-Arnaud,[1] Guillaume Bruant,[1]† Philippe Lanotte,[1] Stella Brun,[1] Agnès Rosenau,[1] Nathalie van der Mee-Marquet,[1,2] Roland Quentin,[1,2] and Laurent Mereghetti[1]*

EA 3854 "Bactéries et risque materno-fœtal," Département de Microbiologie Médicale et Moléculaire, Université François Rabelais,[1] and Centre National de Référence des Streptocoques—Laboratoire associé, Service de Bactériologie et Hygiène Hospitalière, Centre Hospitalier Universitaire,[2] Tours, France

Received 4 May 2005/Accepted 16 June 2005

The prevalence of insertion sequences IS*1548*, IS*861*, IS*1381*, and ISSa4 and of the group II intron GBSi1 within *Streptococcus agalactiae* human isolates strongly correlates with the genetic lineages obtained by multilocus sequence typing. Our results yielded an evolutionary scheme for the acquisition of these genetic elements linked to the ecosystems from which the isolates were obtained.

Mobile genetic elements drive bacterial evolution and adaptation via recombination and horizontal transfer events (12), and they can influence the virulence of bacteria by modulating gene expression (9). These mobile genetic elements include bacteriophages, transposons, superintegrons, insertion sequences (ISs), and group II introns. ISs are small genetic elements, usually less than 2.5 kb in size, encoding only their own mobility (15), whereas group II introns have a more complicated mechanism and transpose via an RNA intermediate that acts like a ribozyme (17). *Streptococcus agalactiae* has emerged, during the last 35 years, as the most common pathogen implicated in neonatal infectious diseases (24). It harbors several mobile genetic elements. Five are well characterized: four ISs, IS*1548* (8), IS*861* (18), IS*1381* (26), and ISSa4 (24), and one group II intron, GBSi1 (7). Some of these elements have been reported to be associated with virulence genes: IS*1548* with the hyaluronate lyase gene (8), IS*1548* and GBSi1 with the C5a-peptidase gene and the gene encoding the laminin-binding protein Lmb (7), IS*861* and IS*1548* with the capsule cluster *cps* gene (21, 22), and ISSa4 with the *cylB* gene, which encodes the membrane-spanning domain of a putative hemolysin transporter (25). Given the influence of these repeat elements on genome plasticity, the aim of this study was to correlate the presence of IS*1548*, IS*861*, IS*1381*, ISSa4, and GBSi1 with the evolution of the *S. agalactiae* species analyzed by multilocus sequence typing (MLST).

Our collection was composed of 52 epidemiologically unrelated *S. agalactiae* isolates, representative of the various human sites where *S. agalactiae* is found. They were collected in France from 1986 to 2003: 20 were isolated from the cerebrospinal fluid of neonates, 17 from the blood cultures of patients suffering from endocarditis, 10 from the vaginas of asymptomatic women, and 5 from the gastric fluids of neonates. Serotyping was performed with a Pastorex rapid latex agglutination test (Bio-Rad, Hercules, Calif.) and by a molecular serotyping process as described by Kong et al. (13). Thirty isolates were from serotype III (57.7% of isolates), seven from serotype Ib (13.5%), four from serotype II (7.7%), three from serotype Ia (5.8%), two from serotype V (3.8%), and one from serotype IV (1.9%). Five isolates were nontypeable (9.6%).

We evaluated the prevalence of IS*1548*, IS*861*, IS*1381*, ISSa4, and GBSi1 within the 52 isolates by PCR using previously described primers and amplification conditions (1, 4). PCR products corresponding to each IS and GBSi1 were checked by sequencing. Forty of the 52 isolates possessed IS*861* (76.9%), 37 possessed IS*1381* (71.2%), 19 possessed IS*1548* (36.5%), 16 possessed GBSi1 (30.8%), and 4 possessed ISSa4 (7.7%). Based on the presence or absence of each of the five elements, nine genetic variants, designated V1 to V9, were defined among the 52 isolates in our collection (Table 1).

The genomic positions of the copies of each IS and GBSi1 were determined by Southern blot analysis for each of the 52 isolates. Briefly, DNA was digested with EcoRI, for which there were no cleavage sites in the four ISs or GBSi1. Hybridizations were carried out as previously described (6), with probes specific for each IS and GBSi1. Vector NTI v 9.0 software was employed to estimate the number and the size of EcoRI fragments by using the whole genome sequence of *S. agalactiae* NEM316 (GenBank accession number AL732656). The Smart Ladder molecular-size standard (Eurogentec), which was run in parallel, allowed us to locate the copies of the ISs and of GBSi1 on the restriction fragments generated in silico.

Phylogenetic analysis was performed by the MLST technique with seven housekeeping genes (*adhP*, *pheS*, *atr*, *glnA*, *sdhA*, *glcK*, and *tkt*) as described by Jones et al. (11). Nucleotide sequences were compared with those available on the MLST database (http://pubmlst.org/sagalactiae/). After assigning an allele number to each locus, the sequence type (ST), which takes into account the allele combinations of the seven

* Corresponding author. Mailing address: EA 3854 "Bactéries et risque materno-fœtal," Département de Microbiologie Médicale et Moléculaire, Faculté de Médecine de Tours, 37032 Tours, France. Phone: 33-247478056. Fax: 33-247473812. E-mail: laurent.mereghetti@med.univ-tours.fr.

† Present address: Groupe de recherche sur les maladies infectieuses du porc, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, Canada.

TABLE 1. Nine genetic variants defined by combination of the presence or absence of each of the four ISs and of the group II intron GBSi1[a]

Genetic variant	No. of different ISs and/or group II intron	Result for:					No. (%) of isolates	Serotype (no. of isolates)[b]
		IS*1548*	GBSi1	IS*861*	IS*1381*	ISSa4		
V1	0	−	−	−	−	−	2 (4)	III (1), NT (1)
V2	1	−	−	−	+	−	6 (11)	NT (2), IV (1), V (1), Ia (1), Ib (1)
V3	1	−	−	−	−	+	3 (6)	III (3)
V4	2	−	+	+	−	−	10 (9)	III (10)
V5	2	−	−	+	+	−	8 (15)	Ib (6), Ia (1), NT (1)
V6	3	+	−	+	+	−	17 (33)	III (15), II (1), NT (1)
V7	3	−	+	+	+	−	3 (6)	Ia (1), II (1), III (1)
V8	3	−	+	−	+	+	1 (2)	II (1)
V9	4	+	+	+	+	−	2 (4)	II (1), V (1)

[a] Serotypes of the isolates are indicated in the last column.
[b] NT, nontypeable.

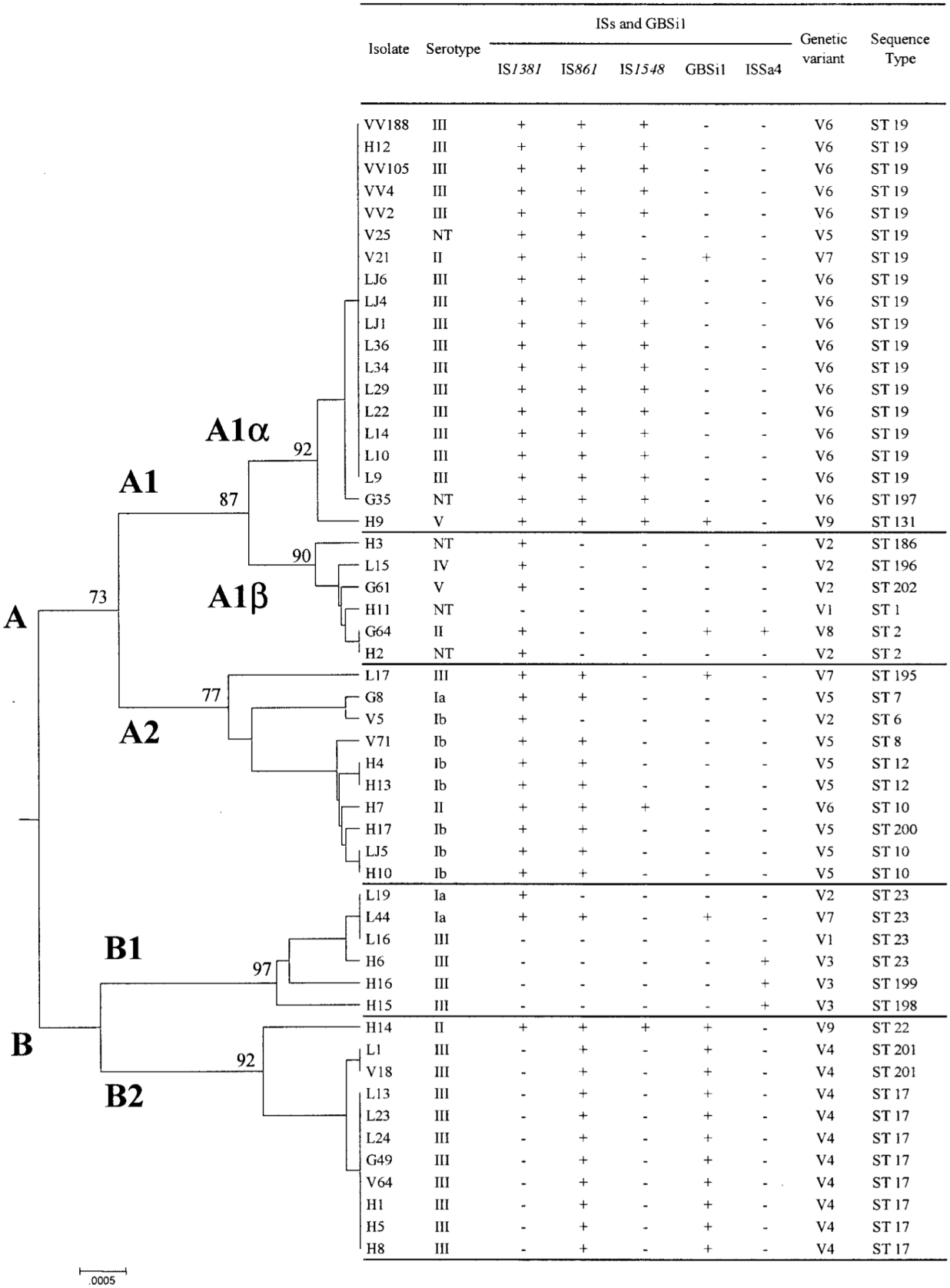
loci, was determined for each isolate. Between three (genes *pheS* and *glnA*) and eight (gene *adhP*) alleles were present for each locus (Table 2). Based on the combinations of the alleles for the seven loci, 21 STs were identified, of which 14 corresponded to a single isolate; three STs, ST 19 ($n = 17$), ST 17 ($n = 8$), and ST 23 ($n = 4$), accounted for 55.8% of the whole collection. These STs corresponded to three of the four major STs previously defined by Jones et al., confirming the worldwide spread of ST 17, ST 19, and ST 23. These results also confirm the close correlation of ST 19 and ST 17 with serotype III (11). Moreover, nucleotide sequences of the seven genes were concatenated into a single "supergene" (3,456 bp), which allows the construction of a phylogenetic tree by use of the MEGA software version 2.1 (http://www.megasoftware.net) with the UPGMA algorithm and the Kimura two-parameter mutation model of genetic distance. The level of statistical support for the nodes on the tree was evaluated by examining their percentage of recovery in 1,000 resampled trees by use of the bootstrap test (23). The MEGA software grouped the isolates into two main genomic divisions: division A, which could be subdivided into A1α, A1β, and A2, and division B, which could be subdivided into B1 and B2 (Fig. 1).

A strong correlation was observed between the distribution of the ISs and GBSi1 and the divisions obtained by MLST (Fig. 1). IS*1381* was significantly associated with division A (34 of the 37 isolates [92%] harboring IS*1381* were distributed in

TABLE 2. STs assigned on the basis of allelic profiles[a]

ST	Allelic profile	No. (%) of:			Total no. (%) of isolates ($n = 52$)	Serotype (no. of isolates)[d]
		CSF[b] isolates ($n = 20$)	Endocarditis isolates ($n = 17$)	Colonizing isolates[c] ($n = 15$)		
19	1, 1, 3, 2, 2, 2, 2	10 (50)	1 (5.9)	6 (40)	17 (32.7)	III (15), NT (1), II (1)
17	2, 1, 1, 2, 1, 1, 1	3 (15)	3 (17.7)	2 (13.3)	8 (15.4)	III (8)
23	5, 4, 6, 3, 2, 1, 3	3 (15)	1 (5.9)	0 (0)	4 (7.7)	Ia (2), III (2)
10	9, 1, 4, 1, 3, 3, 2	1 (5)	2 (11.8)	0 (0)	3 (5.81)	Ib (2), II (1)
12	10, 1, 4, 1, 3, 3, 2	0 (0)	2 (11.8)	0 (0)	2 (3.85)	Ib (2)
2	1, 1, 3, 1, 1, 2, 2	0 (0)	1 (5.9)	1 (6.6)	2 (3.85)	II (1), NT (1)
201	2, 1, 25, 2, 1, 1, 1	1 (5)	0 (0)	1 (6.6)	2 (3.85)	III (2)
1	1, 1, 2, 1, 1, 2, 2	0 (0)	1 (5.9)	0 (0)	1 (1.92)	NT (1)
6	9, 1, 2, 1, 3, 2, 2	0 (0)	0 (0)	1 (6.6)	1 (1.92)	Ib (1)
7	10, 1, 2, 1, 3, 2, 2	0 (0)	0 (0)	1 (6.6)	1 (1.92)	Ia (1)
8	4, 1, 4, 1, 3, 3, 2	0 (0)	0 (0)	1 (6.6)	1 (1.92)	Ib (1)
22	13, 3, 1, 3, 1, 1, 1	0 (0)	1 (5.9)	0 (0)	1 (1.92)	II (1)
131	1, 1, 3, 2, 2, 3, 2	0 (0)	1 (5.9)	0 (0)	1 (1.92)	V (1)
186	1, 1, 2, 1, 1, 2, 5	0 (0)	1 (5.9)	0 (0)	1 (1.92)	NT (1)
195	9, 1, 1, 1, 3, 2, 1	1 (5)	0 (0)	0 (0)	1 (1.92)	III (1)
196	1, 1, 3, 1, 1, 12, 2	1 (5)	0 (0)	0 (0)	1 (1.92)	IV (1)
197	1, 1, 3, 1, 2, 2, 2	0 (0)	0 (0)	1 (6.6)	1 (1.92)	NT (1)
198	13, 4, 6, 3, 2, 1, 3	0 (0)	1 (5.9)	0 (0)	1 (1.92)	III (1)
199	5, 4, 6, 3, 2, 1, 2	0 (0)	1 (5.9)	0 (0)	1 (1.92)	III (1)
200	9, 1, 4, 1, 3, 3, 5	0 (0)	1 (5.9)	0 (0)	1 (1.92)	Ib (1)
202	40, 1, 3, 1, 1, 2, 2	0 (0)	0 (0)	1 (6.6)	1 (1.92)	V (1)

[a] The allelic profiles, available on the MLST database website (http://pubmlst.org/sagalactiae/), were obtained by combination of each of the seven alleles, in the following order: *adhP*, *pheS*, *atr*, *glnA*, *sdhA*, *glcK*, and *tkt*. The ecological origins and serotypes of the *S. agalactiae* isolates are also indicated for each ST.
[b] CSF, cerebrospinal fluid.
[c] Isolated from vaginas of asymptomatic women and from the gastric fluids of neonates.
[d] NT, nontypeable.



Isolate	Serotype	ISs and GBSi1					Genetic variant	Sequence Type
		IS*1381*	IS*861*	IS*1548*	GBSi1	ISSa4		
VV188	III	+	+	+	-	-	V6	ST 19
H12	III	+	+	+	-	-	V6	ST 19
VV105	III	+	+	+	-	-	V6	ST 19
VV4	III	+	+	+	-	-	V6	ST 19
VV2	III	+	+	+	-	-	V6	ST 19
V25	NT	+	+	-	-	-	V5	ST 19
V21	II	+	+	-	+	-	V7	ST 19
LJ6	III	+	+	+	-	-	V6	ST 19
LJ4	III	+	+	+	-	-	V6	ST 19
LJ1	III	+	+	+	-	-	V6	ST 19
L36	III	+	+	+	-	-	V6	ST 19
L34	III	+	+	+	-	-	V6	ST 19
L29	III	+	+	+	-	-	V6	ST 19
L22	III	+	+	+	-	-	V6	ST 19
L14	III	+	+	+	-	-	V6	ST 19
L10	III	+	+	+	-	-	V6	ST 19
L9	III	+	+	+	-	-	V6	ST 19
G35	NT	+	+	+	-	-	V6	ST 197
H9	V	+	+	+	+	-	V9	ST 131
H3	NT	+	-	-	-	-	V2	ST 186
L15	IV	+	-	-	-	-	V2	ST 196
G61	V	+	-	-	-	-	V2	ST 202
H11	NT	-	-	-	-	-	V1	ST 1
G64	II	+	-	-	+	+	V8	ST 2
H2	NT	+	-	-	-	-	V2	ST 2
L17	III	+	+	-	+	-	V7	ST 195
G8	Ia	+	+	-	-	-	V5	ST 7
V5	Ib	+	-	-	-	-	V2	ST 6
V71	Ib	+	+	-	-	-	V5	ST 8
H4	Ib	+	+	-	-	-	V5	ST 12
H13	Ib	+	+	-	-	-	V5	ST 12
H7	II	+	+	+	-	-	V6	ST 10
H17	Ib	+	+	-	-	-	V5	ST 200
LJ5	Ib	+	+	-	-	-	V5	ST 10
H10	Ib	+	+	-	-	-	V5	ST 10
L19	Ia	+	-	-	-	-	V2	ST 23
L44	Ia	+	+	-	+	-	V7	ST 23
L16	III	-	-	-	-	-	V1	ST 23
H6	III	-	-	-	-	+	V3	ST 23
H16	III	-	-	-	-	+	V3	ST 199
H15	III	-	-	-	-	+	V3	ST 198
H14	II	+	+	+	+	-	V9	ST 22
L1	III	-	+	-	+	-	V4	ST 201
V18	III	-	+	-	+	-	V4	ST 201
L13	III	-	+	-	+	-	V4	ST 17
L23	III	-	+	-	+	-	V4	ST 17
L24	III	-	+	-	+	-	V4	ST 17
G49	III	-	+	-	+	-	V4	ST 17
V64	III	-	+	-	+	-	V4	ST 17
H1	III	-	+	-	+	-	V4	ST 17
H5	III	-	+	-	+	-	V4	ST 17
H8	III	-	+	-	+	-	V4	ST 17

FIG. 1. UPGMA dendrogram showing the genetic relationships between the 52 *S. agalactiae* isolates. The dendrogram was constructed with the MEGA 2.1 software by using the nucleotide sequences of the seven housekeeping genes (*pheS*, *atr*, *tkt*, *glcK*, *sdhA*, *glnA*, and *adhP*) concatenated into a single "supergene." Bootstrap values are shown at the major nodes. Isolates can be divided into two divisions (A and B) and five subdivisions (A1α, A1β, A2, B1, and B2). For each isolate, the serotype, ST, presence or absence of the four ISs and of the group II intron GBSi1, and genetic variants (deduced from the combination of the four ISs and GBSi1 and named V1 to V9) are indicated. L and LJ refer to meningitis isolates, H to endocarditis isolates, G to gastric fluid isolates, and V and VV to vaginal isolates.

division A), whereas this IS was almost completely absent from division B ($P < 0.0001$, chi-square test) (Fig. 1). IS*861* was significantly associated with subdivisions A1α, A2, and B2 (39 of the 40 isolates [98%] harboring IS*861* were distributed in subdivisions A1α, A2, and B2) ($P < 0.0001$) (Fig. 1). IS*1548* was significantly associated with subdivision A1α (17 of 19 [89%] isolates harboring IS*1548* were distributed in subdivision A1α) ($P < 0.0001$) (Fig. 1). Similarly, GBSi1 was significantly associated with subdivision B2 (11 of 16 [69%] isolates harboring GBSi1 were distributed in subdivision B2) ($P < 0.0001$); conversely, 100% of the isolates from subdivision B2 were GBSi1 positive (Fig. 1). Finally, ISSa4 was significantly associated with subdivision B1 (3 of 4 isolates harboring ISSa4 were distributed in subdivision B1) ($P < 0.001$), although the small number of isolates in B1 and the small number of ISSa4-positive isolates must be taken into account (Fig. 1). The rare exceptions we noticed in the correlation of IS distribution within MLST lineages were probably due to horizontal gene transfers.

Therefore, based on the prevalence of each IS and GBSi1 in each MLST group, an evolutionary scheme corresponding to the acquisition of these mobile genetic elements by *S. agalactiae* during its evolution can be proposed. Indeed, IS*1381* may have been acquired during the differentiation of the common ancestor (Fig. 2). IS*1548*, GBSi1, and ISSa4 seem to have been acquired more recently by subdivisions A1α, B2, and B1, respectively. The hypothesis that ISSa4 was acquired recently is consistent with the empirical observation of Spellerberg et al., in that all ISSa4-positive isolates in their study were isolated after 1996 (25). IS*861*, which is present in both divisions A and B but not in every subdivision, may have been acquired recently by subdivisions A1α, A2, and B2 (Fig. 2a). Alternatively, the earliest common ancestor may have harbored IS*861*, which was subsequently lost by subdivisions A1β and B1 (Fig. 2b).

One striking finding, already pointed out by Granlund et al. (8) and Bohnsack et al. (3), is the rarity of IS*1548*$^+$ GBSi1$^+$ isolates (only two isolates in our collection harbored both). To date, three insertion sites on the *S. agalactiae* chromosome are known for IS*1548*, named site X, site Y, and site Z (7, 14, 22). Site X corresponds to the intergenic fragment located between the *scpB* and *lmb* genes, encoding the C5a-peptidase and the laminin-binding protein, respectively. As IS*1548* and GBSi1 may share this insertion site, the simultaneous insertion of IS*1548* and GBSi1 at the same insertion site is unlikely. Indeed, in our two IS*1548*$^+$ GBSi1$^+$ isolates, IS*1548* was not found at site X or within *hylB*. The IS*1548*-free site X could therefore be explained by the presence of GBSi1 at site X, as confirmed by Southern blot analysis.

Interestingly, the evolutionary scheme presented here suggests another explanation for the mutually exclusive distribution of IS*1548* and GBSi1. In fact, subdivisions A1α (carrying IS*1548*) and B2 (carrying GBSi1) separated early during the evolutionary process (when the ancestors A and B separated) (Fig. 1 and Fig. 2). Furthermore, GBSi1 was associated with IS*1381* only once within subdivision B2 isolates (Fig. 1); similarly, ISSa4 was associated with other ISs only once (Fig. 1), in agreement with Dmitriev et al. (4). Conversely, GBSi1 and IS*1548* were always associated with IS*861*, except for one isolate (Fig. 1). Such a preferential association has already been observed in the *Escherichia coli* reference (ECOR) collection

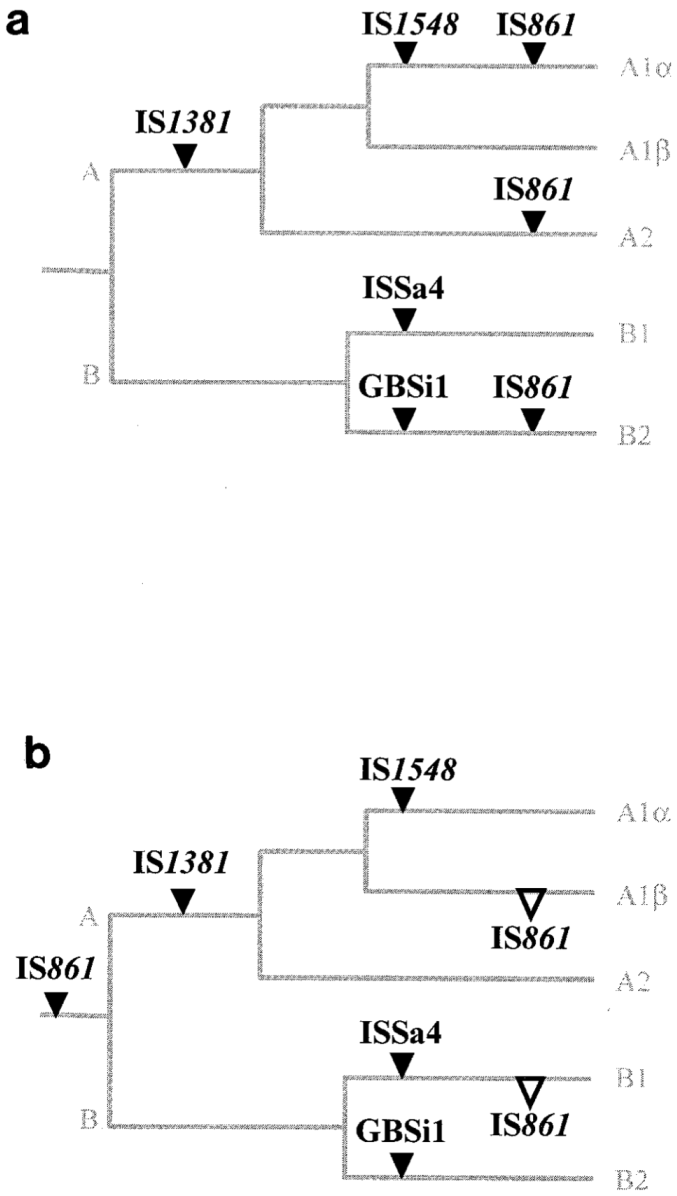


FIG. 2. Evolutionary scheme showing the chronology of acquisition of the four ISs and of the group II intron GBSi1 during *S. agalactiae* evolution. This scheme was deduced from the prevalence of each mobile genetic element within the divisions and subdivisions obtained by MLST as indicated in Fig. 1. IS*1381* may have been acquired (acquisition of ISs or GBSi1, ▼) by division A, IS*1548* by subdivision A1α, ISSa4 by subdivision B2, and GBSi1 by subdivision B2. IS*861* may have been acquired by subdivisions A1α, A2, and B2 (Fig. 2a), or it may have been acquired by the common ancestor and then lost (loss of ISs or GBSi1, ▽) by subdivisions A1β and B1 (Fig. 2b). Branch values are not representative of genetic distances.

(19). The mutually exclusive distribution of ISs, as well as the preferential association of ISs, is consistent with the delivery of ISs by a vector carrying multiple ISs. Although transmission of ISs by phages has not yet been reported for *S. agalactiae*, it has been observed for *E. coli*, *Pseudomonas aeruginosa*, and *Bacillus thuringiensis* (16). Despite the low prevalence of plasmids within the species *S. agalactiae* (10), many of the genomic islands flanked by ISs in *S. agalactiae* are associated with genes related to phages, integrative plasmids, and conjugative trans-

posons (5, 27). Therefore, the peculiar distribution of ISs (mutually exclusive mode and preferential association mode) could be explained by the link between the isolates and the ecosystems from which they were isolated. It is tempting to consider that ISs and GBSi1 reflect the original ecosystem of the isolates and that many of the IS-mediated mutations contribute to the adaptation of bacterial populations to their environment (20). This hypothesis is in accordance with a recent study reporting that the ST 17 complex, that is, our subdivision B2 characterized by GBSi (Fig. 1), was closer to bovine allelic profiles, whereas the ST 19 complex, i.e., our subdivision A1α characterized by IS*1548* (Fig. 1), was closer to allelic profiles initially described for *S. agalactiae* human isolates (2).

We thank Sandra Bourdon, Anne Bouvet, Christophe Burucoa, Julien Loubinoux, Christian Martin, Didier Tandé, Anne Tristan, and François Vandenesch for providing some of the isolates used in this study. We thank Nicola Jones for providing us with new allele numbers and new sequence type numbers. We are indebted to Alain Goudeau for critical reading of the manuscript. We also thank Alain Moreau and Pascal Vaudin for technical advice.

REFERENCES

1. **Bidet, P., N. Brahimi, C. Chalas, Y. Aujard, and E. Bingen.** 2003. Molecular characterization of serotype III group B-streptococcus isolates causing neonatal meningitis. J. Infect. Dis. **188:**1132–1137.
2. **Bisharat, N., D. W. Crook, J. Leigh, R. M. Harding, P. N. Ward, T. J. Coffey, M. C. Maiden, T. Peto, and N. Jones.** 2004. Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor. J. Clin. Microbiol. **42:**2161–2167.
3. **Bohnsack, J. F., A. A. Whiting, G. Martinez, N. Jones, E. E. Adderson, S. Detrick, A. J. Blaschke-Bonkowsky, N. Bisharat, and M. Gottschalk.** 2004. Serotype III *Streptococcus agalactiae* from bovine milk and human neonatal infections. Emerg. Infect. Dis. **10:**1412–1419.
4. **Dmitriev, A., A. Shen, X. Shen, and Y. Yang.** 2004. ISSa4-based differentiation of *Streptococcus agalactiae* strains and identification of multiple target sites for ISSa4 insertions. J. Bacteriol. **186:**1106–1109.
5. **Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couve, L. Lalioui, C. Poyart, P. Trieu-Cuot, and F. Kunst.** 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. Mol. Microbiol. **45:**1499–1513.
6. **Gousset, N., A. Rosenau, P.-Y. Sizaret, and R. Quentin.** 1999. Nucleotide sequences of genes coding for fimbrial proteins in a cryptic genospecies of *Haemophilus* spp. isolated from neonatal and genital tract infections. Infect. Immun. **67:**8–15.
7. **Granlund, M., F. Michel, and M. Norgren.** 2001. Mutually exclusive distribution of IS*1548* and GBSi1, an active group II intron identified in human isolates of group B streptococci. J. Bacteriol. **183:**2560–2569.
8. **Granlund, M., L. Oberg, M. Sellin, and M. Norgren.** 1998. Identification of a novel insertion element, IS*1548*, in group B streptococci, predominantly in strains causing endocarditis. J. Infect. Dis. **177:**967–976.
9. **Hacker, J., and J. B. Kaper.** 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. **54:**641–679.
10. **Horodniceanu, T., D. H. Bouanchaud, G. Bieth, and Y. A. Chabbert.** 1976. R plasmids in *Streptococcus agalactiae* (group B). Antimicrob. Agents Chemother. **10:**795–801.
11. **Jones, N., J. F. Bohnsack, S. Takahashi, K. A. Oliver, M. S. Chan, F. Kunst, P. Glaser, C. Rusniok, D. W. Crook, R. M. Harding, N. Bisharat, and B. G. Spratt.** 2003. Multilocus sequence typing system for group B streptococcus. J. Clin. Microbiol. **41:**2530–2536.
12. **Kazazian, H. H., Jr.** 2004. Mobile elements: drivers of genome evolution. Science **303:**1626–1632.
13. **Kong, F., S. Gowan, D. Martin, G. James, and G. L. Gilbert.** 2002. Serotype identification of group B streptococci by PCR and sequencing. J. Clin. Microbiol. **40:**216–226.
14. **Luan, S. L., M. Granlund, and M. Norgren.** 2003. An inserted DNA fragment with plasmid features is uniquely associated with the presence of the GBSi1 group II intron in *Streptococcus agalactiae*. Gene **312:**305–312.
15. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
16. **Mahillon, J., C. Leonard, and M. Chandler.** 1999. IS elements as constituents of bacterial genomes. Res. Microbiol. **150:**675–687.
17. **Michel, F., and J. L. Ferat.** 1995. Structure and activities of group II introns. Annu. Rev. Biochem. **64:**435–461.
18. **Rubens, C. E., L. M. Heggen, and J. M. Kuypers.** 1989. IS*861*, a group B streptococcal insertion sequence related to IS*150* and IS*3* of *Escherichia coli*. J. Bacteriol. **171:**5531–5535.
19. **Sawyer, S. A., D. E. Dykhuizen, R. F. DuBose, L. Green, T. Mutangadura-Mhlanga, D. F. Wolczyk, and D. L. Hartl.** 1987. Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. Genetics **115:**51–63.
20. **Schneider, D., and R. E. Lenski.** 2004. Dynamics of insertion sequence elements during experimental evolution of bacteria. Res. Microbiol. **155:**319–327.
21. **Sellin, M., S. Hakansson, and M. Norgren.** 1995. Phase-shift of polysaccharide capsule expression in group B streptococci, type III. Microb. Pathog. **18:**401–415.
22. **Sellin, M., C. Olofsson, S. Hakansson, and M. Norgren.** 2000. Genotyping of the capsule gene cluster (*cps*) in nontypeable group B streptococci reveals two major *cps* allelic variants of serotypes III and VII. J. Clin. Microbiol. **38:**3420–3428.
23. **Sitnikova, T., A. Rzhetsky, and M. Nei.** 1995. Interior-branch and bootstrap tests of phylogenetic trees. Mol. Biol. Evol. **12:**319–333.
24. **Spellerberg, B.** 2000. Pathogenesis of neonatal *Streptococcus agalactiae* infections. Microbes Infect. **2:**1733–1742.
25. **Spellerberg, B., S. Martin, C. Franken, R. Berner, and R. Lutticken.** 2000. Identification of a novel insertion sequence element in *Streptococcus agalactiae*. Gene **241:**51–56.
26. **Tamura, G. S., M. Herndon, J. Przekwas, C. E. Rubens, P. Ferrieri, and S. L. Hillier.** 2000. Analysis of restriction fragment length polymorphisms of the insertion sequence IS*1381* in group B streptococci. J. Infect. Dis. **181:**364–368.
27. **Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi, and C. M. Fraser.** 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. USA **99:**12391–12396.

-ARTICLE 2-

Mobile Genetic Elements Provide Evidence for a Bovine Origin of Clonal Complex 17 of *Streptococcus agalactiae*

Geneviève Héry-Arnaud,[1] Guillaume Bruant,[1†] Philippe Lanotte,[1] Stella Brun,[1] Bertrand Picard,[3] Agnès Rosenau,[1] Nathalie van der Mee-Marquet,[1,2] Pascal Rainard,[4] Roland Quentin,[1,2] and Laurent Mereghetti[1*]

EA 3854 "Bactéries et risque materno-fœtal",[1] Département de Microbiologie Médicale et Moléculaire, Université François Rabelais,1 and Centre National de Référence des Streptocoques—Laboratoire associé, Service de Bactériologie et Hygiène Hospitalière, Centre Hospitalier Universitaire,[2] Tours, France, Service de Microbiologie, Hôpital Avicenne, Assistance Publique-Hôpitaux de Paris, Bobigny, France,[3] and Institut National de la Recherche Agronomique, UR1282, Infectiologie Animale et Santé Publique, Nouzilly, France[4]

*Corresponding author

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, JULY 2007, P. 4668–4672 VOL. 73, NO. 14

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, July 2007, p. 4668–4672 Vol. 73, No. 14
0099-2240/07/$08.00+0 doi:10.1128/AEM.02604-06
Copyright © 2007, American Society for Microbiology. All Rights Reserved.

Mobile Genetic Elements Provide Evidence for a Bovine Origin of Clonal Complex 17 of *Streptococcus agalactiae*▽

Geneviève Héry-Arnaud,[1] Guillaume Bruant,[1]† Philippe Lanotte,[1,2] Stella Brun,[1] Bertrand Picard,[3] Agnès Rosenau,[1] Nathalie van der Mee-Marquet,[1,2] Pascal Rainard,[4] Roland Quentin,[1,2] and Laurent Mereghetti[1,2]*

Université François-Rabelais, IFR 136, Faculté de Médecine, EA 3854 Bactéries et Risque Materno-Foetal, Tours, France[1]; Centre Hospitalier Universitaire, Tours, France[2]; Service de Microbiologie, Hôpital Avicenne, Assistance Publique-Hôpitaux de Paris, Bobigny, France[3]; and Institut National de la Recherche Agronomique, UR1282, Infectiologie Animale et Santé Publique, Nouzilly, France[4]

Received 8 November 2006/Accepted 12 May 2007

We sought an explanation for epidemiological changes in *Streptococcus agalactiae* infections by investigating the link between ecological niches of the bacterium by determining the prevalence of 11 mobile genetic elements. The prevalence of nine of these elements differed significantly according to the human or bovine origin of the isolate. Correlating this distribution with the phylogeny obtained by multilocus sequence analysis, we observed that human isolates harboring GBSi1, a clear marker of the bovine niche, clustered in clonal complex 17. Our results are thus consistent with the emergence of this virulent human clone from a bovine ancestor.

Epidemiological changes in the pattern of *Streptococcus agalactiae* infections over the last six decades from predominantly bovine infections (13) to a leading cause of human neonatal infections (25, 26) and, more recently, to a cause of invasive disease in elderly adults (6) require explanation. Genomic lineages grouping the most invasive isolates have been described by multilocus enzyme electrophoresis (17, 23), by pulsed-field gel electrophoresis (1, 24), and more recently by the analysis of the polymorphism in the allelic profiles of housekeeping genes (11) but are insufficient to account for the emergence of *S. agalactiae* from such diverse ecosystems. A link between the bovine and human niches has been proposed (8, 19) but remains a matter of debate. Indeed, it was recently suggested that human and bovine *S. agalactiae* isolates are largely unrelated (3) and represent distinct populations or subtypes (5, 27). Conversely, another study provided evidence to suggest that the major hyperinvasive sequence type 17 (ST-17) clone implicated in neonatal invasive diseases had arisen from bovine lineage ST-61 (2), whereas others suggested that the genome of the common ancestor was closer to that of human ST-17 strains than to that of the ST-61 strains of bovine origin (4). We hypothesized that mobile genetic elements (MGEs), such as insertion sequences (ISs) or group II introns, which drive bacterial evolution (12, 30), might provide insight into changes in the epidemiology of *S. agalactiae* infections.

(This work was presented in part at the XVIth Lancefield International Symposium on Streptococci and Streptococcal Diseases, Cairns, Australia, 25 to 29 September 2005.)

We tested this hypothesis by analyzing a collection of 98 epidemiologically unrelated isolates, mostly collected in France (3 isolates from Germany and 1 isolate from Malaysia) between 1966 and 2003. Of the 98 isolates, 63 human isolates were from colonized individuals, adults with invasive diseases, and neonates with meningitis (52 of which were previously typed by multilocus sequence typing [MLST] and studied for the prevalence of IS*1548*, IS*861*, IS*1381*, ISSa4, and GBSi1) (10) and 35 bovine isolates were obtained from cows with clinical evidence of udder infection.

The prevalence of 11 MGEs (one group II intron [GBSi1] and 10 IS elements corresponding to six ISs [IS*1548*, IS*861*, IS*1381*, ISSa4, ISSag1, and ISSag2] and four transposase genes corresponding to remnants of IS elements [ISSag9, SAG0448, SAG1241, and SAG0610, relating to the identification of the open reading frames in strains A909 and 2605V/R] [28, 29]) was evaluated by PCR. IS*1548*, IS*861*, IS*1381*, ISSa4, ISSag1, ISSag2, and GBSi1 were detected as previously described (10). Other transposase genes were detected with the primers listed in Table 1. The prevalence of all the MGEs except IS*861* and ISSag2 differed significantly as a function of the original reservoir (Fig. 1). This was particularly true for IS*1381*, IS*1548*, ISSag1, and ISSag9, which were mostly present in human isolates, and for GBSi1 and ISSa4, which were mostly present in bovine isolates (Fig. 1).

Multilocus sequence analysis, carried out with the standard MLST scheme of Jones et al. (11), resolved the 98 isolates into 47 sequence types (STs). A phylogenetic tree was generated,

* Corresponding author. Mailing address: EA 3854 "Bactéries et Risque Materno-Foetal," Laboratoire de Bactériologie, Faculté de Médecine de Tours, Université François-Rabelais, 2 bd. Tonnellé, 37032 Tours, France. Phone: 33 247478056. Fax: 33 247473812. E-mail: laurent.mereghetti@med.univ-tours.fr.

† Present address: Groupe de Recherche sur les Maladies Infectieuses du Porc, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, Canada.

▽ Published ahead of print on 25 May 2007.

TABLE 1. Primers used for the amplification of transposase genes from the *S. agalactiae* isolates

Target	GenBank accession no.[a]	Sequence of primer (5′–3′)		T_m[b] (°C)	Amplicon size (bp)
		Forward	Reverse		
ISSag9	NC_004116	GACCGTAATGACATCCAATCTGG	GGCTTGAAATGGACACGGTT	50	600
SAG0448	NC_004116	CCACAGAATTACTTAACTTCCTAGC	CGACCAACCTTATTGAATCACT	50	994
SAG1241 (*orfA*)	NC_004116	CTCTTGTTAACCTTTATCAAACTGG	CCTAAAAAAAGCAAGTGCCA	50	218
SAG0610	NC_004116	TACTCCACATTAGCTAAAGAACAGG	GGCCTGATTTAACCGAGATT	50	394

[a] NC_004116 refers to strain 2603 V/R (29).
[b] T_m, melting temperature.

based on the nucleotide sequences of the supergene obtained by concatenation of the seven loci for the 98 isolates, using MEGA software version 3.1 (www.megasoftware.net/) with the neighbor-joining (NJ) method (14). Clonal complexes (CCs) (isolates sharing six or seven identical alleles) were defined by eBURST analysis (http://eburst.mlst.net/) (7). Seven main NJ phylogenetic divisions were identified, closely matching the seven main CCs identified by eBURST analysis (Fig. 2): subclonal complex 19 (subCC19) (n = 21), subCC2 (n = 11), CC10 (n = 12), CC7 (n = 6), CC17 (n = 13), CC67 (n = 15), and CC23 (n = 10). Three of these CCs were "host specific": CC17 and subCC19 consisted almost exclusively of human isolates (100% and 95% of the isolates, respectively), whereas CC67 was almost exclusively composed of bovine isolates (93.8% of the isolates). Moreover, the NJ dendrogram shows that human lineage CC17 arose from bovine lineage CC67 (Fig. 2), consistent with the assertions of Bisharat et al. (2).

A factorial analysis of correspondence was performed on the whole data set, with the 11 MGEs as active variables and the three main clonal complexes (CC19, CC17, and CC67) and the bovine or human origins of the strains as illustrative variables. Computations were performed with SPAD.N software (Centre International de Statistiques et Informatique Appliquées, St. Mandé, France) as previously described (22). The projections of the active and illustrative variables and of the strains on the F1/F2 plane accounting for 40.97% of the total variance (Fig. 3) separated two groups of variables and of strains. GBSi1 and ISSa4 (and to a lesser degree SAG0448 and SAG1241), bovine origin, CC17 and CC67, and 94.3% of bovine isolates were projected onto the negative values of F1. IS*1548*, IS*1381*, ISSag9, and ISSag1 (and to a lesser degree SAG0610), human origin, CC19, and 79.4% of human isolates were projected onto the positive values of F1. These findings confirm those of Luan et al. in that the two major human *S. agalactiae* lineages are each marked by one MGE: subCC19 (defined as CC19 by Luan et al. [16]) is characterized by IS*1548*, whereas CC17 is characterized by GBSi1. All but one of the human isolates projected with the bovine isolates and variable CC67 onto the negative values of F1 belonged to CC17, highlighting the stronger linkage of CC17 with the bovine lineage CC67 than with any of the other human lineages (Fig. 3). Moreover, our data also suggest that the presence of GBSi1 in the bovine reservoir predates its presence in the human reservoir because of the following. (i) All our bovine isolates harbored GBSi1, whereas this element was present in less than one-third of human isolates (almost all of which belonged to CC17). (ii) The bovine isolates had higher copy numbers and more-complex patterns for GBSi1 than the human isolates did (data not shown), consistent with an increase in IS pattern complexity over time (18). MGEs therefore provide evidence for a bovine ancestor of the human clonal lineage CC17, all the isolates of which harbor GBSi1. This statement is consistent with the pro-

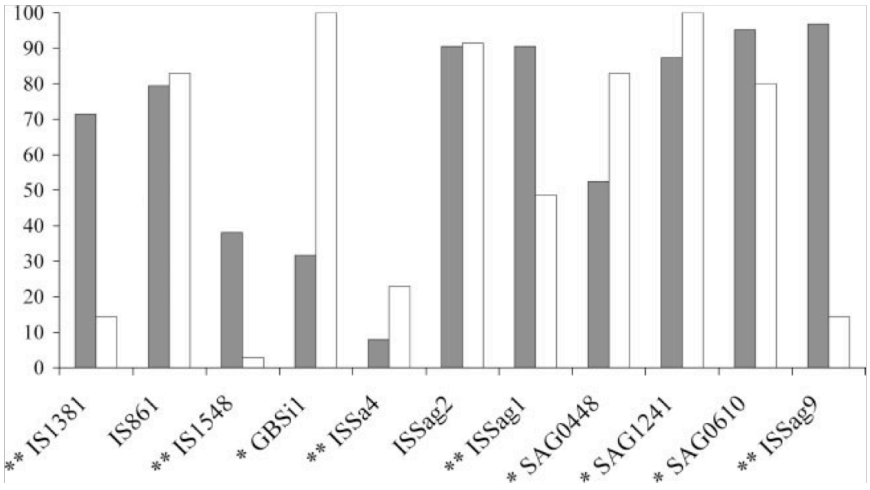
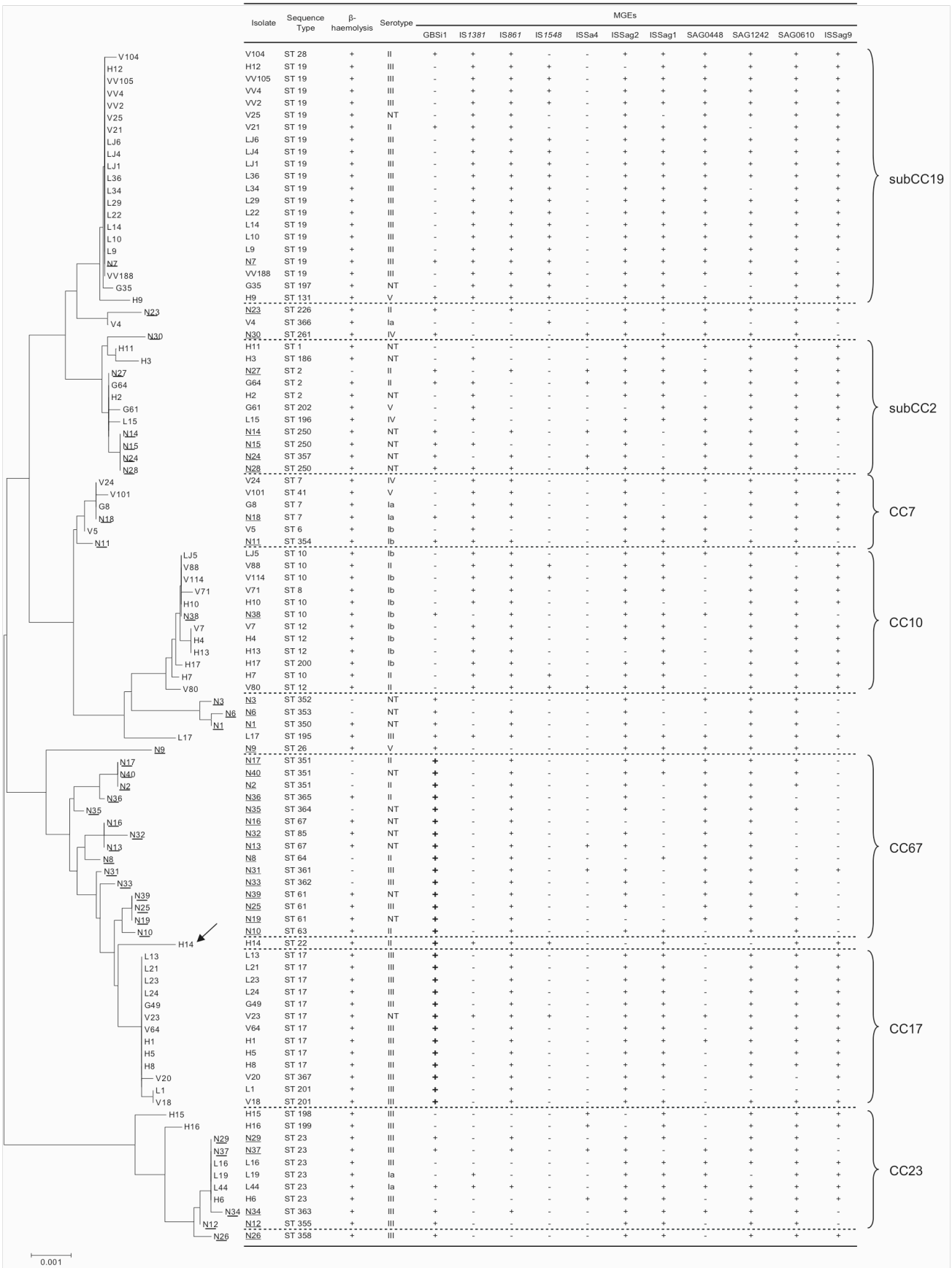


FIG. 1. Prevalence of the MGEs within the human (filled bars) and bovine (open bars) *S. agalactiae* isolates. Significant differences (without correction for bias due to multiple comparisons) are indicated by one asterisk for P values below 0.05 and two asterisks for P values below 0.001.



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	+	+	+	-	-	+	+	+	+	+	+
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	+	-	+	-	-	+	+	+	+	+	+
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	-	+	+	-	-	+	+	+	+	+	+
V101	ST 41	+	V	-	+	+	-	-	+	-	-	+	+	+
G8	ST 7	+	Ia	-	+	+	-	-	+	-	-	+	+	+
N18	ST 7	+	Ia	+	+	+	-	-	+	+	+	+	+	+
V5	ST 6	+	Ib	-	+	-	-	-	+	+	+	-	+	+
N11	ST 354	+	Ib	+	+	+	-	-	+	+	+	+	+	-
LJ5	ST 10	+	Ib	-	+	+	-	-	+	+	+	+	+	+
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	+	-	-	-	-	+	-	+	+	+	-
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	+	-	+	-	-	+	+	-	+	+	+
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	-	-	-	-	+	-	+	-	+	+	+
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	+	-	-	-	-	+	+	-	+	+	+

subCC19
subCC2
CC7
CC10
CC67
CC17
CC23

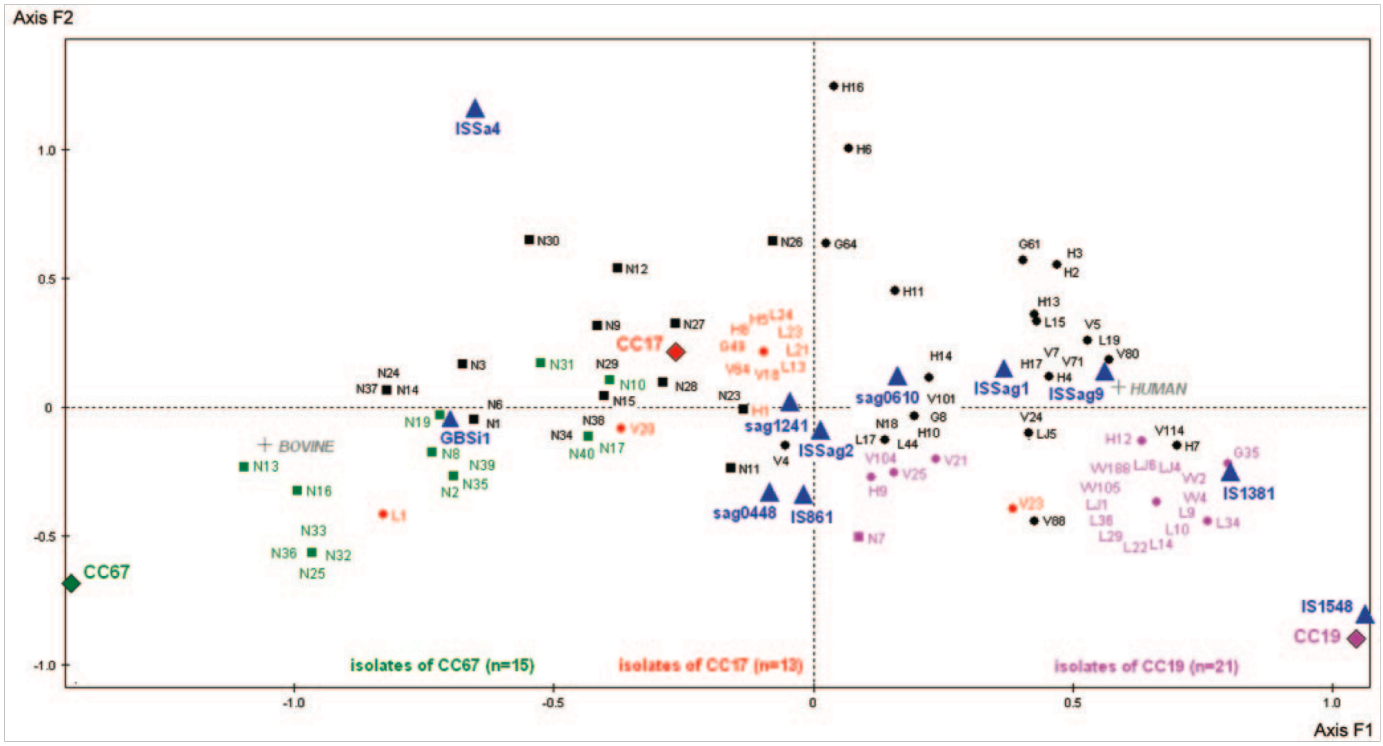


FIG. 3. Factorial analysis of correspondence of 98 *S. agalactiae* isolates based on the MGE data. Eleven active variables (the MGEs [▲] IS*1381*, IS*861*, IS*1548*, GBSi1, ISSa4, ISSag2, ISSag1, SAG0448, SAG1241, SAG0610, and ISSag9), five illustrative variables (the main clonal complexes CC19 [red diamond], CC17 [pink diamond], and CC67 [green diamond] and the human [circles] and bovine [squares] origins of the isolates), and the 98 isolates are projected onto the F1/F2 plane. This plane, obtained by computation, is defined by the two principal axes of the analysis; the first axis, F1, explains most of the variance, and the second axis, F2 (orthogonal to F1), explains most of the variance not explained by F1. Together, F1 and F2 accounted for 40.97% of the total variance. Isolates of CC17 (n = 13) are shown in red, isolates of CC67 (n = 15) are shown in green, and isolates of subCC19 (n = 21) are shown in pink.

posed evolutionary scheme of MGE acquisition, in which GBSi1 is thought to have been acquired recently by human isolates (10).

The particular distribution of MGEs among bovine and human isolates (Fig. 3) is also consistent with a physical barrier reducing the exchange of genetic material between the bovine and human reservoirs. The human isolate H14, which occupies a prominent position on the NJ dendrogram between the bovine CC67 and the human CC17 (Fig. 2), displayed two particular features. First, the copy number of GBSi1 in this isolate was much higher than that in other isolates (human isolates had a median of 3 copies of GBSi1, whereas H14 harbored 12 copies [data not shown]). Parkhill et al. suggested that IS expansion occurs during evolutionary bottlenecks (20, 21). They argued that in *Yersinia pestis*, IS expansion coincided with a change in the ecological niche of the species from a commensal organism of the gut to a systemic pathogen (21). Second, isolate H14 harbored both IS*1548* and GBSi1. In our study, as in previous studies (9, 15), very few isolates harbored both GBSi1 and IS*1548* (Fig. 2). Granlund et al. explained this mutually exclusive distribution by the presence of a closed insertion site for the two MGEs on the bacterial chromosome (9), but this distribution could also be explained by IS*1548* and GBSi1 originating from two different ecological niches. Thus, human clone CC17 seems to have evolved from the bovine lineage containing GBSi1 and has retained the GBSi1 genetic marker from its bovine donor.

In conclusion, our results provide at least a partial explanation of the emergence of *S. agalactiae* as a major cause of neonatal meningitis, despite its original identification as the causal agent of bovine mastitis. Assuming that CC17 evolved from a bovine ancestor, as shown both by MLST analysis and by the presence of the bovine marker GBSi1, the isolates of this CC may have genetic features different from those of other *S. agalactiae* isolates that have adapted to their human hosts

FIG. 2. Relatedness of the 98 isolates of *S. agalactiae* analyzed by the NJ method, using MEGA v3.1 software and the Kimura two-parameter mutation model of genetic distance. The nucleotide differences between the supergene sequences obtained by multilocus sequence analysis method of Jones et al. (11) were used to infer phylogenetic relationships between isolates. We used eBURST v3 software to group the 98 isolates into CCs. For each isolate, the ST, beta-hemolysis, serotype, and presence (+) or absence (−) of the 11 MGEs are indicated. CCs are also indicated on the right of the figure (several isolates were not attributed a CC number by the Start software and are considered "singletons"). The prominent position of isolate H14 is indicated by a black arrow. The human isolates are named L or LJ (meningocerebral fluid), H (adult), V or VV (vaginal carriage), or G (gastric fluid); the bovine isolates are named N (bovine isolates are underlined). NT, nontypeable.

over a longer period of time. These features probably concern genes not present in the core genome, as recently defined by Tettelin et al., but that have been acquired by horizontal exchange due to the environmental gene pool available in other ecosystems (28). Indeed, we recently reported a link between prophage DNA fragments and *S. agalactiae* isolates from neonatal cases of meningitis, suggesting the presence of additional genetic material in these strains (10).

We thank Nicolas Jones, who manages the MLST website database, for assigning new allele numbers and new sequence type numbers. We also thank Stéphanie Gouriou for her help in using the SPAD.N software.

REFERENCES

1. **Bidet, P., N. Brahimi, C. Chalas, Y. Aujard, and E. Bingen.** 2003. Molecular characterization of serotype III group B-streptococcus isolates causing neonatal meningitis. J. Infect. Dis. **188:**1132–1137.
2. **Bisharat, N., D. W. Crook, J. Leigh, R. M. Harding, P. N. Ward, T. J. Coffey, M. C. Maiden, T. Peto, and N. Jones.** 2004. Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor. J. Clin. Microbiol. **42:**2161–2167.
3. **Bohnsack, J. F., A. A. Whiting, G. Martinez, N. Jones, E. E. Adderson, S. Detrick, A. J. Blaschke-Bonkowsky, N. Bisharat, and M. Gottschalk.** 2004. Serotype III *Streptococcus agalactiae* from bovine milk and human neonatal infections. Emerg. Infect. Dis. **10:**1412–1419.
4. **Brochet, M., E. Couve, M. Zouine, T. Vallaeys, C. Rusniok, M. C. Lamy, C. Buchrieser, P. Trieu-Cuot, F. Kunst, C. Poyart, and P. Glaser.** 2006. Genomic diversity and evolution within the species *Streptococcus agalactiae*. Microbes Infect. **8:**1227–1243.
5. **Dogan, B., Y. H. Schukken, C. Santisteban, and K. J. Boor.** 2005. Distribution of serotypes and antimicrobial resistance genes among *Streptococcus agalactiae* isolates from bovine and human hosts. J. Clin. Microbiol. **43:**5899–5906.
6. **Edwards, M. S., M. A. Rench, D. L. Palazzi, and C. J. Baker.** 2005. Group B streptococcal colonization and serotype-specific immunity in healthy elderly persons. Clin. Infect. Dis. **40:**352–357.
7. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186:**1518–1530.
8. **Finch, L. A., and D. R. Martin.** 1984. Human and bovine group B streptococci: two distinct populations. J. Appl. Bacteriol. **57:**273–278.
9. **Granlund, M., F. Michel, and M. Norgren.** 2001. Mutually exclusive distribution of IS*1548* and GBSi1, an active group II intron identified in human isolates of group B streptococci. J. Bacteriol. **183:**2560–2569.
10. **Héry-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. J. Bacteriol. **187:**6248–6252.
11. **Jones, N., J. F. Bohnsack, S. Takahashi, K. A. Oliver, M. S. Chan, F. Kunst, P. Glaser, C. Rusniok, D. W. Crook, R. M. Harding, N. Bisharat, and B. G. Spratt.** 2003. Multilocus sequence typing system for group B streptococcus. J. Clin. Microbiol. **41:**2530–2536.
12. **Kazazian, H. H., Jr.** 2004. Mobile elements: drivers of genome evolution. Science **303:**1626–1632.
13. **Keefe, G. P., I. R. Dohoo, and E. Spangler.** 1997. Herd prevalence and incidence of *Streptococcus agalactiae* in the dairy industry of Prince Edward Island. J. Dairy Sci. **80:**464–470.
14. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. **5:**150–163.
15. **Luan, S. L., M. Granlund, and M. Norgren.** 2003. An inserted DNA fragment with plasmid features is uniquely associated with the presence of the GBSi1 group II intron in *Streptococcus agalactiae*. Gene **312:**305–312.
16. **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. J. Clin. Microbiol. **43:**3727–3733.
17. **Musser, J. M., S. J. Mattingly, R. Quentin, A. Goudeau, and R. K. Selander.** 1989. Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B streptococcus) causing invasive neonatal disease. Proc. Natl. Acad. Sci. USA **86:**4731–4735.
18. **Naas, T., M. Blot, W. M. Fitch, and W. Arber.** 1995. Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. Mol. Biol. Evol. **12:**198–207.
19. **Norcross, N. L., and N. Oliver.** 1976. The distribution and characterization of group B streptococci in New York State. Cornell Vet. **66:**242–248.
20. **Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell.** 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat. Genet. **35:**32–40.
21. **Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature **413:**523–527.
22. **Picard, B., J. S. Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion, and E. Denamur.** 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infect. Immun. **67:**546–553.
23. **Quentin, R., H. Huet, F. S. Wang, P. Geslin, A. Goudeau, and R. 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. J. Clin. Microbiol. **33:**2576–2581.
24. **Rolland, K., C. Marois, V. Siquier, B. Cattier, and R. Quentin.** 1999. Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *hylB* gene analysis. J. Clin. Microbiol. **37:**1892–1898.
25. **Schrag, S. J.** 2004. The past and future of perinatal group B streptococcal disease prevention. Clin. Infect. Dis. **39:**1136–1138.
26. **Schuchat, A.** 2001. Group B streptococcal disease: from trials and tribulations to triumph and trepidation. Clin. Infect. Dis. **33:**751–756.
27. **Sukhnanand, S., B. Dogan, M. O. Ayodele, R. N. Zadoks, M. P. Craver, N. B. Dumas, Y. H. Schukken, K. J. Boor, and M. Wiedmann.** 2005. Molecular subtyping and characterization of bovine and human *Streptococcus agalactiae* isolates. J. Clin. Microbiol. **43:**1177–1186.
28. **Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. Deboy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, and C. M. Fraser.** 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. USA **102:**13950–13955.
29. **Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi, and C. M. Fraser.** 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. USA **99:**12391–12396.
30. **Toro, N.** 2003. Bacteria and Archaea group II introns: additional mobile genetic elements in the environment. Environ. Microbiol. **5:**143–151.

APPORT DU TRAVAIL DE THESE A D'AUTRES PUBLICATIONS SCIENTIFIQUES

1. Lartigue MF, **Héry-Arnaud G**, Haguenauer E, Domelier AS, Schmit PO, van der Mee-Marquet N, Lanotte P, Laurent Mereghetti, Markus Kostrzewa and Roland Quentin**. 2009.** "Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by Matrix-Assisted Laser Desorption/Ionization -Time of Flight Mass Spectrometry"**. J Clin Microbiol.** 2009 Jul;47(7):2284-7. Epub 2009 Apr 29.

2. Domelier AS, van der Mee-Marquet N, Sizaret PY, **Héry-Arnaud G,** Lartigue MF, Mereghetti L, Quentin R. **2009**. "Molecular characterization and lytic activities of *Streptococcus agalactiae* bacteriophages and determination of lysogenic strain features". **J Bacteriol.** 2009 Aug;191(15):4776-85. Epub 2009 May 22.

PUBLICATIONS SCIENTIFIQUES SOUMISES / EN PREPARATION

<u>*Soumis*</u> *:*

Al Safadi, R., Amor, A., **Héry-Arnaud, G**., Spellerberg, B., Lanotte, P., Gannier, F., Quentin, R., Rosenau, A.

"The insertion element IS*1548* induces the expression of the laminin-binding gene *lmb* in *Streptococcus agalactiae*".

Infection and Immunity.

<u>*En préparation :*</u>

Héry-Arnaud, G., Jany, JL, Lanotte, P., Gouriou, S., van der Mee-Marquet, N., Lartigue, M. F., Rosenau, A., Quentin, R., Mereghetti, L.

"New insights in *Streptococcus agalactiae* evolution provided by multilocus sequence typing".

Applied Environmental Microbiology.

COMMUNICATIONS SCIENTIFIQUES

- **Communications orales:**

Héry-Arnaud G, 2006. « Rôle des éléments génétiques mobiles dans l'évolution de *Streptococcus agalactiae* ». Présentation, dans le cadre du doctorat, au **forum de l'école doctorale « Santé, Sciences, Technologies », Université François Rabelais, Tours.**

Héry-Arnaud G, Bruant G, Lanotte P, Brun S, Rosenau A, van der Mee-Marquet N, Rainard P, Quentin R, Mereghetti L. 2005 « The prevalence of insertion sequence and GBSi1 intron within *Streptococcus agalactiae* is linked to the animal or human origin of isolates. ». **XVIth Lancefield International Symposium on Streptococci and Streptococcal Diseases, Cairns, Australia.**

Héry-Arnaud G, 2004. « Rôle des éléments génétiques mobiles dans l'évolution et la virulence de *Streptococcus agalactiae* ». Présentation à la **journée de l'IFR 136, Agents transmissibles et infectiologie, Université François Rabelais, Tours (01/10/04).**

- **Communications affichées:**

Mereghetti L, **Héry-Arnaud G,** Legeais M, Loulergue J, Giraudeau B, Lanotte P, Brun S, Rosenau A, van der Mee-Marquet N, Quentin R. 2006 « Different mutator phenotypes of animal and human isolates of *Streptococcus agalactiae* ». **7th conference on Streptococcal Genetics, ASM conferences, Saint-Malo, France.**

Héry-Arnaud G, Bruant G, Lanotte P, Brun S, Rosenau A, van der Mee-Marquet N, Rainard P, Quentin R, Mereghetti L. 2005 « Les éléments génétiques mobiles, marqueurs de l'origine humaine ou bovine de *Streptococcus agalactiae* ». **25e Réunion Interdisciplinaire de Chimiothérapie Anti Infectieuse (RICAI), Paris.**

Héry-Arnaud G, Bruant G, Lanotte P, Brun S, Rosenau A, van der Mee-Marquet N, Quentin R, Mereghetti L. 2004 « Chronology of the acquisition of mobile genetic elements within Streptococcus agalactiae ». **24e Réunion Interdisciplinaire de Chimiothérapie Anti Infectieuse (RICAI), Paris.**

Geneviève HERY-ARNAUD

RÔLE DES ÉLÉMENTS GÉNÉTIQUES MOBILES DANS L'ÉVOLUTION ET LA VIRULENCE DE *STREPTOCOCCUS AGALACTIAE*

Résumé

Nous avons étudié le rôle des éléments génétiques mobiles (EGM) dans l'évolution et la virulence de *Streptococcus agalactiae*, bactérie pathogène opportuniste responsable d'infections chez l'homme. La structure génétique de la population a été analysée par multilocus sequence typing à partir d'un souchier représentatif de la diversité de l'espèce. La distribution de 11 EGM (sept séquences d'insertion, trois transposases, un intron) a été confrontée à la structure de la population. Cette confrontation a permis i) de démontrer que l'acquisition des EGM corrélait avec l'évolution de l'espèce, ii) de proposer une hypothèse de la chronologie d'acquisition des EGM, iii) d'identifier certains EGM comme marqueurs de l'écosystème d'origine des souches, et iv) de conforter l'hypothèse de l'émergence du clone humain invasif CC17 à partir d'un ancêtre bovin. Nous avons également cartographié les copies des EGM présentes sur le génome de *S. agalactiae*, puis nous avons identifié huit nouveaux sites d'insertion, dont deux sont significativement associés à l'origine écologique de la souche.

Mots-clés : *Streptococcus agalactiae*, multilocus sequence typing, éléments génétiques mobiles, évolution, virulence.

Abstract

We studied the role of mobile genetic elements (MGE) on the evolution and the virulence of *Streptococcus agalactiae*, an opportunistic bacterium responsible for human infections. The genetic population structure was analyzed by multilocus sequence typing using a collection representative of the species diversity. The distribution of 11 MGE (seven insertion sequences, three transposases, one intron) was compared to the population structure. We demonstrated that the MGE prevalence strongly correlates with the genetic lineages. We proposed an evolutionary scheme for the acquisition of the MGE. Several MGE appeared to be markers of the origin of the strains. MGE analysis brought evidence for a bovine origin of the human virulent clone CC17. We also identified the position of the MGE copies on the *S. agalactiae* genome and we identified eight new MGE insertion sites, from which two were significantly associated with the ecological origin of the isolates.

Keywords : *Streptococcus agalactiae*, multilocus sequence typing, mobile genetic elements, evolution, virulence.